MARKOV PROCESSES

Theorems and Problems

Evgenii B. Dynkin and Aleksandr A. Yushkevich
Moscow University

Translated from Russian by James S. Wood

PLENUM PRESS • NEW YORK • 1969

Professor *Evgenii Borisovich Dynkin,* who teaches at Moscow University, has written two earlier textbooks on Markov processes and is also the author of a series of popular books on mathematics.

Aleksandr Adol'fovich Yushkevich holds an associate professorship at Moscow University. His many publications include a series of articles on Markov processes.

The original Russian text, published by Nauka Press in Moscow in 1967 as part of a series on Probability Theory and Mathematical Statistics, has been corrected by the authors for the present edition.

Е. Б. Дынкин, А. А. Юшкевич

ТЕОРЕМЫ И ЗАДАЧИ О ПРОЦЕССАХ МАРКОВА

TEOREMY I ZADACHI O PROTSESSAKH MARKOVA

Library of Congress Catalog Card Number 69-12529

A Division of Plenum Publishing Corporation
227 West 17 Street, New York, New York 10011

Printed in the United States of America

Foreword

The concepts and methods of probability theory are finding a growing number of applications in the natural sciences and engineering, and they are making deeper and deeper inroads into the various domains of mathematics itself. Access to these methods is equally advantageous to mathematicians in diverse specializations, to physicists, and to engineers, and yet the elementary textbooks are only able to give a limited notion as to the present development of the subject, while the monographs devoted to the latest trends in the field are normally written for specialists and employ elaborate set-theoretic and analytic tools.

In order to grasp new mathematical concepts, one must be aware of their power and visualize how they operate. This is best initiated not with general theorems, but with specific problems. The problems must be realistic and the situation typical, but not cluttered with the kind of incidental technical difficulties that arise in a more pedantic systematic framework.

The goal of the present book is to introduce, following the indicated modus operandi, the reader to the latest findings in the theory of Markov processes.

Markov processes comprise a class of stochastic processes that has been thoroughly investigated and has enjoyed numerous applications. Those branches of the theory of Markov processes which have already achieved classical status are presented in a few beautifully written books (e.g., [1] and [2]). In recent years, however, we find a growing number of new and significant trends, as well as the discovery of new relationships between Markov processes and mathematical analysis. These problems are comprehensively discussed in several monographs ([3]-[11]) which, however, are scarcely geared for a beginning familiarization with the subject. Yet the whole problem area is based mainly on transparent and graphic ideas, the study of which provides a rich body of material for conditioning one to the probabilistic way of thinking.

The book is divided into four chapters, each of which introduces the reader to a definite problem area: potentials, harmonic and excessive functions, and the limiting behavior of the paths taken by a process (Chapt. I); the probabilistic solution of differential equations (Chapt. II); certain optimal control problems (Chapt. III), and the probabilistic aspect of boundary problems in analysis (Chapt. IV).

In the first chapter we consider the simplest Markov chain, viz., a symmetric random walk on a lattice. It is explained that the familiar concepts of a harmonic function, potential, capacitance, and others from classical analysis have their analogs in this discrete model and may be used for the solution of purely probabilistic problems, such as the problem of the number of visits to a given set. The foundation for this chapter is the work of Ito and McKean [12].

It is shown in Chapt. II how probabilistic notions are used to obtain analytical results. In particular, this approach is used to prove the existence of a solution to the Dirichlet problem for the Laplace equation in a broad class of domains.*

The connections between Markov processes and potentials find an unexpected application in Chapt. III for the investigation of the problem of the optimal stopping of a Markov process. The source for this chapter is [13].

Many researchers lately have focused their attention on the problem of the broadest classes of boundary conditions for differential and other equations. These problems are treated in Chapt. IV from the probabilistic point of view. The analysis of an ulti-

---

*The relationship between probability theory and the Dirichlet problem was brought to light long before the birth of the general theory of Markov processes (by H. B. Phillips and N. Wiener in 1923; R. Courant, K. Friedrichs, and H. Lèvy in 1928). This idea was exhaustively developed in the work of A. Ya. Khinchin (1933) and I. G. Petrovskii (1934). A formula expressing the solution of the Dirichlet problem in terms of the trajectories of a Wiener process was derived by J. Doob (1954). However, Doob applied it in a direction converse to our own, namely, for the derivation of the properties of paths from theorems of analysis.

mately simple model (birth and death processes) makes it possible to work strictly with completely elementary devices. The pioneer in the application of the probabilistic approach to boundary problems is Feller. He discusses birth and death processes in [14]. However, even though Feller is guided by probabilistic intuition, he still introduces all his constructions in purely analytical form. Our approach is based on a consideration of the properties of paths and rests on the concept of the characteristic operator (a brief outline of Chapt. IV is contained in [15]).

At the end of each chapter is a set of problems which go beyond the simple function of providing material for exercises in that they supplement the text proper and present certain new information. Thus, the problems in Chapt. III include a discussion of the Martin boundary for a denumerable Markov chain.

So as not to impede the main flow of the probabilistic discussions, some of the ancillary analytical problems are treated in the Appendix.

Besides the primary literature sources mentioned above, we will have occasion throughout the book to refer to other works (usually in the problems and examples).

All that is expected of the reader is an acquaintance with the basic tenets of probability theory and classical analysis. Some of the problems, however, require greater erudition. We have consciously avoided references in the main text to measure theory and measurability. The reader who has these concepts at his command will have no difficulty in grasping the presentation at a more rigorous set-theoretic level.

The book is basically the outgrowth of lectures given by the first author at Moscow University in 1962-63 (the lectures were recorded by M. B. Malyutov, S. A. Molchanov, and M. I. Freidlin). This material was subsequently augmented and radically revised, with the addition of problems to round out the content of the book.

The authors wish to thank I. L. Genis, who performed a tremendous service in preparing the manuscript for publication.

January 22, 1966

E. B. Dynkin
A. A. Yushkevich

Contents

Chapter I

A Criterion of Recurrence

§1. Symmetric Random Walk

Consider a particle moving along the integer-valued points $0, \pm 1, \pm 2, \ldots$ on the x axis, executing unit jumps to the left or the right at equal intervals. If in each such instance the probabilities of jumping to the right or to the left are the same and equal to 1/2, we say that the particle executes a s y m m e t r i c r a n d o m w a l k o n a l i n e.

The points $0, +1, -1, \ldots$ that the particle can hit are called s t a t e s.

We propose to show that a particle with arbitrary initial position will with probability one sooner or later enter any possible state. Inasmuch as all states are clearly equally probable, it is sufficient to show that a particle leaving any state will at some time hit 0. We denote by $\pi(x)$ the probability of hitting 0 from a point x. Then $\pi(0) = 1$, and, according to the total-probability formula,

$$\pi(x) = \frac{1}{2}\pi(x-1) + \frac{1}{2}\pi(x+1) \qquad (1)$$

for $x \neq 0$. Consider the graph of the function $\pi(x)$, $x = 0, 1, 2, \ldots, k, \ldots$. Equation (1) means that any three neighboring points of this graph lie on a single line. Consequently, all points of the graph of the function $\pi(x)$ for $x \geq 0$ lie on one line. Inasmuch as $\pi(0) = 1$, this line emanates from the point (0, 1). If $\pi(x)$ were smaller than one for some positive x, this line would have to intersect the x axis, and $\pi(x)$ would be negative for sufficiently large x. This is impossible, hence $\pi(x)=1$ for all $x \geq 0$. Due to the symmetry of the random walk, $\pi(x)=1$ for $x<0$ as well. Thus, for any initial state the probability of attaining zero is equal to one.

A logical generalization of the random walk on a line is a random walk on an l-dimensional integer-valued lattice H^l. This lattice consists of points (vectors) of the type

$$x = x_1 e_1 + \dots + x_l e_l,$$

where $\mathbf{e}_1, \dots, \mathbf{e}_l$ comprises the orthonormal basis of an l-dimensional space, and the coordinates $x_1, \dots, x_l$ are arbitrary integers. Increasing or decreasing one of the coordinates by one and leaving the other coordinates unchanged, we obtain the $2l$ neighboring lattice points to x (thus, in the two-dimensional case every point of the lattice has four neighbors: one to the right, one to the left, one above, and one below). At each step the particle has equal probabilities $1/2l$ of intersecting one of the neighboring states, regardless of its position at the preceding instant.

It turns out in the two-dimensional, as in the one-dimensional, case that the particle, on leaving any point of the lattice, has a probability one of hitting any other point (see the problems at the end of the chapter). For lattices of three or more dimensions, on the other hand, the probability of reaching one state from another, as we shall see, is less than one. The probability of reaching a certain set B rather than a single point can be either equal to or less than one. We designate this probability $\pi_B(x)$, where x is the initial state of the particle. We call the set B recurrent if $\pi_B(x) = 1$ for points x of the lattice, nonrecurrent (transient) if $\pi_B(x) < 1$ for at least one point x. In the present chapter we shall deduce a criterion whereby it is possible to distinguish between recurrent and nonrecurrent sets.

§2 The Transition Function

We let x(0) represent the initial position of a particle executing a random walk and let x(n) $(n=1, 2, 3, \dots)$ represent its position after n steps.

The probability of some event A connected with a random walk, of course, depends on the point x from which the walk was initiated.

We designate this probability $\mathbf{P}_x\{A\}$ and represent the mathematical expectation of the random variable ξ corresponding to the distribution $\mathbf{P}_x$ by the symbol $\mathbf{M}_x\xi$.

Next we denote by p(n, x, y) the probability that a particle leaving the point x will hit the point y after n steps:

$$p(n, x, y) = \mathbf{P}_x\{x(n) = y\}.$$

The function p(n, x, y) is an important characteristic of the random walk and is called its *transition function*. Clearly, p(0, x, x) = 1, p(0, x, y) = 0 for $x \neq y$. It is also clear that $\sum_y p(n, x, y) = 1$.* The quantity

$$\sum_{y \in B} p(n, x, y) = \mathbf{P}_x\{x(n) \in B\},$$

where B is some set in l-dimensional space, is called the *transition probability* from x to B in n steps.

An essential property of the random walk and one that contributes to its analysis is the mutual independence of the jumps $\xi_k = x(k) - x(k-1)$ (k = 1, 2 ...). The vectors ξ_k are also independent of the initial state of the particle, and they all have the same distribution. Specifically, any of the vectors ξ_k assumes with equal probability every one of the values $\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_l$. Utilizing this fact, we derive an appropriate integral representation for the transition function p(n, x, y).

We denote by $\theta(x)$ a linear form assuming the value θ_k on the vector $\mathbf{e}_k$. This means that if $x = x_1\mathbf{e}_1 + \ldots + x_l\mathbf{e}_l$, then $\theta(x) = \theta_1 x_1 + \ldots + \theta_l x_l$. Consider the function

$$F(\theta) = \sum_y p(n, x, y)\, e^{i\theta(y)} = \mathbf{M}_x e^{i\theta(x(n))}, \tag{2}$$

i.e., the characteristic function of the random vector x(n). [In fact, the series in Eq. (2) contains only a finite number of nonzero terms,

---

* Here and elsewhere the symbol $\sum_y$ indicates summation over all points of the lattice H^l.

because after n steps the particle can visit not more than $(2l)^n$ different states.] The transition function p(n, x, y) is easily expressed in terms of the function $F(\theta)$. Thus, we let Q be the set of all linear forms $\theta(z) = \theta_1 z_1 + \ldots + \theta_l z_l$ with coefficients $\theta_1, \ldots, \theta_l$ whose absolute values do not exceed π. We multiply Eq. (2) by $e^{-i\theta(z)}$ (z is a point of the lattice H^l) and integrate over Q Since y and z are vectors with integer-valued coordinates,

$$\int e^{i\theta(y)-i\theta(z)}\,d\theta = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{i[\theta_1(y_1-z_1)+\ldots+\theta_l(y_l-z_l)]}\,d\theta_1 \ldots d\theta_l$$

$$= \prod_{k=1}^{l} \int_{-\pi}^{\pi} e^{i\theta_k(y_k-z_k)}\,d\theta_k = \begin{cases} (2\pi)^l & \text{for} \quad y = z, \\ 0 & \text{for} \quad y \neq z, \end{cases}$$

so that, consequently,

$$p(n, x, z) = \frac{1}{(2\pi)^l} \int_Q F(\theta)\, e^{-i\theta(z)}\,d\theta. \tag{3}$$

Let us find the function $F(\theta)$. Inasmuch as

$$x(n) = x(0) + \sum_{k=1}^{n} \xi_k,$$

where ξ_k is the jump at the kth step, we have

$$F(\theta) = M_x e^{i\theta(x(n))} = M_x e^{i\theta(x(0))} \prod_{k=1} e^{i\theta(\xi_k)}.$$

Since $x(0) = x$ here with probability one, whereas the random vectors ξ_k are independent and distributed identically, we have

$$F(\theta) = e^{i\theta(x)}\Phi^n(\theta), \tag{4}$$

where $\Phi(\theta) = M_x e^{i\theta(\xi_1)}$. The vector ξ_1 assumes any of the values $\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_l$ with probability $1/2l$, hence

$$\Phi(\theta) = \frac{1}{2l} \sum_{m=1}^{l} \left(e^{i\theta_m} + e^{-i\theta_m}\right) = \frac{1}{l} \sum_{m=1}^{l} \cos\theta_m. \tag{5}$$

Substituting the ensuing expressions into Eq. (3) and replacing z by y, we obtain

$$p(n, x, y) = \frac{1}{(2\pi)^l} \int_Q e^{i\theta(x-y)} \Phi^n(\theta)\, d\theta. \tag{6}$$

§ 3. Behavior of the Path of the Particle as n → ∞

We now assume that $l \geq 3$, and set out to show that the length of the vector x(n) tends to infinity with probability one as $n \to \infty$ We will see that this leads to nonrecurrence of any bounded set.

If we make a sequence of trials, where the probability of success in the nth trial is equal to p_n, the sum $p_1 + p_2 + \ldots + p_n + \ldots$ expresses the expectation of the numbers of successes (in fact, the number of successes η is equal to the sum $\eta_1 + \eta_2 + \ldots + \eta_n + \ldots$, where $\eta_n = 1$ if the nth trial results in success, otherwise $\eta_n = 0$).

We now consider a random walk initiated from the point x and assume that the nth trial yields success if $x(n) = y$. Then $p_n = p(n, x, y)$, and the sum

$$g(x, y) = \sum_{n=0}^{\infty} p(n, x, y) \tag{7}$$

represents the *mathematical expectation of the number of hits at the point y.*

We will prove that

$$g(x, y) < \infty. \tag{8}$$

[It can be shown that $g(x, y) = \infty$ for all x, y in the one- and two-dimensional cases (see the problems).]

The function $\Phi(\theta)$ defined by Eq. (5) is continuous, and $|\Phi(\theta)| < 1$ on all of Q except the point $(0, \ldots, 0)$ and 2^l points of the type

$(\pm\pi, \ldots, \pm\pi)$, at which $|\Phi(\theta)| = 1$. Therefore, according to (6),

$$(2\pi)^l g(x, y) \leqslant \sum_{n=0}^{\infty} \int_Q |\Phi^n(\theta)|\, d\theta = \int_Q \frac{d\theta}{1-|\Phi(\theta)|}. \tag{9}$$

Inasmuch as $\cos\alpha \sim 1 - \alpha^2/2$ for $\alpha \to 0$, there exists a neighborhood U of the point $\theta = (0, \ldots, 0)$ in which

$$0 < \cos\theta_m \leqslant 1 - \frac{\theta_m^2}{4} \qquad (m = 1, \ldots, l),$$

and, according to Eq. (5),

$$|\Phi(\theta)| = \Phi(\theta) < 1 - \frac{1}{4l}(\theta_1^2 + \ldots + \theta_l^2).$$

Consequently,

$$\int_U \frac{d\theta}{1-|\Phi(\theta)|} < \int_U \frac{4l\, d\theta}{\theta_1^2 + \ldots + \theta_l^2} < \infty$$

for $l \geq 3$. The convergence of the integral (9) is tested analogously in the neighborhoods of the points $\theta = (\pm\pi, \ldots, \pm\pi)$. Thus,

$$\int_Q \frac{d\theta}{1-|\Phi(\theta)|} < \infty, \tag{10}$$

and the inequality (8) stands proved.

It follows from this inequality that the number of hits of the particle at the point y is finite with probability one. Consequently, the particle has a probability one of occupying any given point of the lattice only a finite number of times. Since the intersection of a denumerable number of certain events is itself a certain event, the particle has a probability one of not occupying a single point infinitely often. The probability is one, therefore, that a time will come for any bounded set of lattice points after which the particle is never present in that set.

It is now a simple matter to prove the nonrecurrence of any bounded set B. First we assume B to be recurrent. Then the probability of the event $A_n = \{$The particle visits B after the nth step$\}$ is equal to the following for any initial state x and any n, according to the total-probability formula:

$$\sum_y p(n, x, y)\pi_B(y) = \sum_y p(n, x, y) = 1.$$

Consequently, all events A_n are realized with probability one, i.e., the particle visits B at arbitrarily remote times. This contradicts the fact that the particle at some time will leave B with probability one.

It follows from the relations (9) and (10) that the series

$$e^{i\theta(x-y)} \sum_{n=0}^{\infty} \Phi^n(\theta)$$

can be integrated term-by-term over Q. Therefore,

$$g(x, y) = \sum_{n=0}^{\infty} \frac{1}{(2\pi)^l} \int_Q e^{i\theta(x-y)} \Phi^n(\theta)\, d\theta = \frac{1}{(2\pi)^l} \int_Q \frac{e^{i\theta(x-y)}\, d\theta}{1-\Phi(\theta)}. \qquad (11)$$

The latter equation makes it possible to advance the following asymptotic estimate for the function g(x, y) for $l \geq 3$:

$$g(x, y) \sim \frac{c_l}{\|x-y\|^{l-2}} \quad \text{for} \quad \|x-y\| \to \infty, \qquad (12)$$

where $\|x\|$ denotes the length of the vector x, and the c_l are certain positive constants (see the Appendix, §1). We will have occasion to use this estimate in the derivation of our recurrence criterion.

§4. Harmonic Functions

Let $f(x)$ be a function at the points of a lattice H^l. We set

$$Pf(x) = M_x f(x(1)) = \sum_y p(1, x, y) f(y). \qquad (13)$$

It is logical to call P the (one-step) shift operator of the function.

Inasmuch as $p(1, x, x+\mathbf{e}_k) = 1/2l$, P is also the *averaging operator*

$$Pf(x) = \frac{1}{2l} \sum f(x + e_k)$$

(k spans the values ±1, ..., $\pm l$, and $\mathbf{e}_{-k} = -\mathbf{e}_k$). It was pointed out long ago that the linear operator

$$A = P - E,$$

where P is the averaging operator and E is the unit operator, is the discrete analog of the operator $\frac{1}{2}\Delta$, where

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_l^2}$$

is the Laplace operator.

It is well known that for sufficiently smooth functions $f(x)$, specified over all space,

$$\Delta f(x) = \lim_{h \to 0} \frac{\sum f(x + he_k) - 2lf(x)}{h^2},$$

so that the Laplace operator is obtained by passing to the limit from the operator P − E as the lattice is infinitely partitioned.

The similarity between the operators $\frac{1}{2}\Delta$ and A has far-reaching implications. Guided by this similarity, we will apply the terms for their analogs in the theory of differential equations to a number of the concepts associated with random walks.

We call function $f(x)$ on the lattice H^l *harmonic* if $Af(x) = 0$ and *superharmonic* if $Af(x) \leq 0$ (for all x). In other words, the function f is harmonic if $Pf = f$, and superharmonic if $Pf \leq f$.

Any constant, clearly, is a harmonic function. We will show that *any bounded harmonic function f is a constant.*

This is very easy to prove if the function f reaches its maximum value at some point y_0. Thus, if $y_1, y_2, \ldots, y_{2l}$ are the neighbors of the point y_0, the arithmetic mean of the numbers $f(y_0) - f(y_k)$ is equal to zero [since $Pf(y_0) = f(y_0)$]. Inasmuch as these numbers are nonnegative, they are equal to zero. Therefore, the set where the function f reaches its maximum value contains, in addition to each point thereof, all of its neighbors. It is clear that this function is a constant.

For any bounded function φ there exists a least upper bound M. Generally speaking, this bound is not attained anywhere, but for every $\varepsilon > 0$ there exists a point y at which $\varphi(y) \geq M - \varepsilon$. Reiterating the arguments of the preceding paragraph, we readily show that if φ is harmonic, then at any point y' neighboring upon y the estimate $\varphi(y') \geq M - 2l\varepsilon$. Hence, if $M > 0$, then it is possible to pick a chain of points y_0, $y_1 = y_0 + \mathbf{e}_1$, $y_2 = y_1 + \mathbf{e}_1$, ..., $y_n = y_{n-1} + \mathbf{e}_1$ for which the sum

$$s = \varphi(y_0) + \varphi(y_1) + \dots + \varphi(y_n)$$

is greater than any prespecified number N.

If now f is an arbitrary bounded harmonic function, the function $\varphi(x) = f(x + \mathbf{e}_1) - f(x)$ is also harmonic and bounded. For it the sum

$$s = f(y_n + e_1) - f(y_0)$$

also does not exceed twice the upper bound of f. Therefore, the exact upper bound of the function φ cannot be positive. This implies that for any x

$$\varphi(x) = f(x + e_1) - f(x) \leqslant 0.$$

In the foregoing discussion it is admissible to replace the vector $\mathbf{e}_1$ with the vector $-\mathbf{e}_1$. Consequently,

$$f(x + e_1) = f(x).$$

It is proved analogously that $f(x + \mathbf{e}_k) = f(x)$ for any k.

An example of a harmonic function is the function $\bar{\pi}_B(x)$ expressing the probability, starting from x, of visiting the set B infinitely often. Consequently,

$$P\bar{\pi}_B(x) = \sum_y p(1, x, y)\bar{\pi}_B(y)$$

is the probability, starting from x, of visiting B infinitely often after the first step. Clearly, this probability is equal to $\bar{\pi}_B(x)$.

Inasmuch as the function $\bar{\pi}_B(x)$ is bounded, it is constant according to the above proof. We will show that it is equal to one or zero depending on whether the set B is recurrent or nonrecurrent. Let the set B first be nonrecurrent. We denote by q(n, y) the probability, starting from x, of visiting B for the first time in the nth step and of occupying the state y at that time, and we denote by $\pi_B(x)$, as before, the probability, starting from x, of visiting B for the time. Clearly,

$$\pi_B(x) = \sum_{n=0}^{\infty} \sum_{y \in B} q(n, y).$$

In order to visit B infinitely often, the particle must visit B for the first time in some step, then visit B infinitely often. Computing the probability of this event according to the total probability formula, we obtain

$$\bar{\pi}_B = \bar{\pi}_B(x) = \sum_{n=0}^{\infty} \sum_{y \in B} q(n, y)\, \bar{\pi}_B(y) = \bar{\pi}_B \cdot \pi_B(x), \tag{14}$$

where x is any point of the lattice. Inasmuch as B is nonrecurrent, there exists an x at which $\pi_B(x) < 1$, so that $\bar{\pi}_B = 0$.

If the set B is recurrent, then, clearly, the probability of the event $C_n = \{$The particle never visits B after the nth step$\}$ is equal to zero for any $n \geq 0$ and any initial state x. Therefore,

$$\begin{aligned} 1 - \bar{\pi}_B(x) &= \mathbf{P}_x\{\text{The particle will visit B only a finite number of times}\} = \mathbf{P}_x\{C_0 \cup C_1 \cup C_2 \cup \ldots\} \\ &\leq \mathbf{P}_x\{C_0\} + \mathbf{P}_x\{C_1\} + \mathbf{P}_x\{C_2\} + \ldots = 0; \end{aligned}$$

consequently, $\bar{\pi}_B = 1$.

Consequently, there is another possible definition for recurrence. *A set B is recurrent if a particle starting from any point of the lattice visits B infinitely often with probability one. If, however, the probability of this event is less than one for some x, it is equal to zero for all x, and B is nonrecurrent.*

In concluding this section, we note that not only bounded harmonic functions, but also harmonic functions bounded below (or above), are constant on the entire lattice H^l (see the problems). The class of harmonic functions unbounded above and below is considerably broader. For example, any linear function of the coordinates $x_1, \dots, x_l$ of a vector x satisfies the equation $Pf = f$, hence it is harmonic.

§5. The Potential

The Laplace operator Δ ties in closely with the concept of the Newtonian potential. Let a mass be distributed in a three-dimensional space R^3 with a density $\varphi(y)$. According to the universal gravitation law of Newton, this mass acts on a unit mass situated at a point x with a force proportional to the gradient of the function

$$f(x) = \frac{1}{2\pi} \int_{R^3} \frac{\varphi(y)\, dy}{\|x - y\|}, \tag{15}$$

where $\|x - y\|$ indicates the distance between the points x and y. The function $f(x)$ is called the potential of the distribution $\varphi(y)$. It may also be interpreted as the potential of an electrostatic field created by a charge distribution φ.

It turns out, given very mild constraints on the function φ, that the potential f is the solution of the Poisson equation

$$\frac{1}{2} \Delta f(x) = -\varphi(x). \tag{16}$$

In complete analogy, the solution of Eq. (16) in an l-dimensional space R^l ($l \geq 3$) is the integral

$$f(x) = b_l \int_{R^l} \frac{\varphi(y)\, dy}{\|x - y\|^{l-2}}, \tag{17}$$

where b_l is some positive constant. This integral is called the potential of the distribution φ in the l-dimensional space.

In the discrete case Eq. (16) goes over to the equation

$$Af(x) = -\varphi(x), \tag{18}$$

where f and φ are functions on the lattice H^l. Let us consider the operator

$$G\varphi = \varphi + P\varphi + P^2\varphi + \dots + P^n\varphi + \dots, \tag{19}$$

where $\varphi \geq 0$. Let

$$f = G\varphi. \tag{20}$$

According to (19),

$$PG\varphi = G\varphi - \varphi ,$$

hence

$$Af = (P - E)f = (P - E)G\varphi = G\varphi - \varphi - G\varphi = -\varphi.$$

Consequently, the operator G is analogous to the integral operator specified by Eq. (17). We therefore refer to the function $G\varphi$ as the potential of the function φ ($\varphi \geq 0$).

The discrete potential has a straightforward probabilitistic interpretation. As a matter of fact,

$$P^n\varphi(x) = \sum_y p(n, x, y)\varphi(y) = M_x\varphi(x(n)). \tag{21}$$

For $n=0$ Eq. (21) reduces to the equation $\varphi(x) = \varphi(x)$, while for $n=1$ it reduces to the definition of the operator P [see Eq. (13)]. For all other values of n Eq. (21) is proved by induction.

According to the total probability formula,

$$p(n+1, x, y) = \sum_z p(1, x, z)\, p(n, z, y).$$

Assuming that Eq. (21) has already been proved for n, we deduce that

$$M_x\varphi(x(n+1)) = \sum_y p(n+1, x, y)\varphi(y) = \sum_z p(1, x, z)\left[\sum_y p(n, z, y)\varphi(y)\right]$$

$$= \sum_z p(1, x, z)[P^n\varphi(z)] = P^{n-1}\varphi(x),$$

i.e., that (21) is true for n - 1 as well.

It follows from (21) that

$$G\varphi(x) = \sum_{n=0}^{\infty} M_x\varphi(x(n)) = M_x \sum_{n=0}^{\infty} \varphi(x(n)). \tag{22}$$

This equation leads to the following important interpretation of the potential. Let every hit at the point y bring a payoff $\varphi(y)$. Then $G\varphi(x)$ is the mean value of the payoff obtained during a random walk of a particle with initial point x.

Using the notation

$$g(x, y) = \sum_{n=0}^{\infty} p(n, x, y),$$

which was introduced in §3, we rewrite the expression for the potential in the form

$$G\varphi(x) = \sum_y g(x, y)\varphi(y). \tag{23}$$

As noted at the end of §3, for large $||x-y||$

$$g(x, y) \sim \frac{c_l}{\|x-y\|^{l-2}}.$$

Hence,

$$G\varphi(x) = \sum_y g(x, y)\varphi(y) \sim c_l \sum \frac{\varphi(y)}{\|x-y\|^{l-2}}.$$

for $||x|| \to \infty$ in every case when $\varphi(y)$ is different from zero only at a finite number of points. Consequently, for large $||x||$ the discrete potential behaves like the Newtonian potential (17).

We now show that if $f = G\varphi$ and τ is the time of first visit of the particle to the set B, then

$$f(x) - M_x f(x(\tau)) = M_x \sum_{k=0}^{\tau-1} \varphi(x(k)) \tag{24}$$

[if the particle never visits B, we say that $\tau = \infty$, $f(x(\tau)) = 0$].

Let

$$f(x) = G\varphi(x) = M_x[\varphi(x(0)) + \varphi(x(1)) + \dots + \varphi(x(n)) + \dots]. \tag{25}$$

Dividing the path of the particle into two parts, viz., the part before the time τ and the part after the time τ, we write that

$$\begin{aligned} f(x) = M_x[\varphi(x(0)) + \dots + \varphi(x(\tau - 1))] \\ + M_x[\varphi(x(\tau)) + \varphi(x(\tau + 1)) + \dots]. \end{aligned} \tag{26}$$

The first term in (26) clearly represents the average payoff during the random walk prior to visiting B, while the second term represents the average payoff after the first visit to B. In order to obtain (24) from (26), we need only verify that

$$M_x[\varphi(x(\tau) + \varphi(x(\tau + 1)) + \dots] = M_x f(x(\tau)).$$

Making use of the probability $q(n, y) = P_x\{\tau = n, x(n) = y\}$, we write that

$$M_x\varphi(x(\tau + k)) = \sum_{n, y} q(n, y) M_y \varphi(x(k)),$$

$$\sum_n q(n, y) = P_x\{x(\tau) = y\},$$

where n varies from 0 to ∞, and y spans the values from B. Consequently,

$$M_x \sum_{k=0}^{\infty} \varphi(x(\tau + k)) = \sum_{k=0}^{\infty} \sum_{n, y} q(n, y) M_y \varphi(x(k))$$

$$= \sum q(n, y) \sum_{k=0}^{\infty} \mathbf{M}_y \varphi(x(k)) = \sum_{n, y} q(n, y) f(y)$$
$$= \sum_{y} f(y) \mathbf{P}_x \{x(\tau) = y\} = \mathbf{M}_x f(x(\tau)).$$

§6. Excessive Functions

We recall that a function $f(x)$ $(x \in H^I)$ is called superharmonic if $Pf \leq f$. Nonnegative superharmonic functions play an important part in the theory of Markov processes and are commonly called excessive functions.

Inasmuch as $Pf = f$ for a harmonic function, a harmonic function is excessive if it is nonnegative. Moreover, if $f = G\varphi$ $(\varphi \geq 0)$, then

$$f - Pf = (\varphi + P\varphi + P^2\varphi + \ldots)$$
$$- P(\varphi + P\varphi + P^2\varphi + \ldots) = \varphi \geqslant 0.$$

Therefore, the potential of any nonnegative function is excessive.

We will show that any excessive function is equal to the sum of a nonnegative harmonic function and the potential of a nonnegative function (this result is the discrete analog of the well-known Riesz theorem in the theory of differential equations).

Let f be an excessive function. We set $f - Pf = \varphi$, noting that $\varphi \geq 0$, and writing the obvious identity

$$f = \varphi + P\varphi + \ldots + P^{n-1}\varphi + P^n f. \tag{27}$$

It follows from the estimate

$$\varphi + P\varphi + \ldots + P^{n-1}\varphi = f - P^n f \leqslant f$$

that

$$G\varphi = \varphi + P\varphi + \ldots + P^n\varphi + \ldots < \infty.$$

Equation (27) implies, therefore, that $h = \lim_{n \to \infty} P^n f$ exists and that

$$f = G\varphi + h. \tag{28}$$

Clearly, $Ph = h$, so that h is a harmonic function.

An example of an excessive function is the probability $\pi_B(x)$ of visiting a set B. Thus, we consider the sequence of events

$A_n = \{$The particle visits the set B after the nth step$\}$.

It is plain that $A_0 \supseteq A_1 \supseteq \ldots \supseteq A_n \supseteq \ldots$. We observe that $\mathbf{P}_x\{A_0\} = \pi_B(x)$. According to Eq. (21),

$$\mathbf{P}_x\{A_n\} = \sum_y p(n, x, y)\pi_B(y) = P^n\pi_B(x). \tag{29}$$

In particular, $P\pi_B(x) = \mathbf{P}_x\{A_1\} \leq \mathbf{P}_x\{A_0\} = \pi_B(x)$, hence the function $\pi_B(x)$ is excessive.

Let us write the expansion (27) for the function $\pi_B(x)$:

$$\pi_B(x) = G\varphi_B(x) + \bar{\pi}_B(x), \tag{30}$$

where $\bar{\pi}_B(x) = \lim_{n\to\infty} P^n\pi_B(x)$, and $\varphi_B(x) = \pi_B(x) - P\pi_B(x)$. According to Eq. (29), $\bar{\pi}_B(x) = \lim_{n\to\infty} \mathbf{P}_x\{A_n\} = \mathbf{P}_x\left\{\bigcap_n A_n\right\}$. Consequently, $\bar{\pi}_B(x)$ is the probability that the particle will visit B at arbitrarily remote times, in other words, that it will visit B infinitely often. We have already encountered this probability in §4, where we showed that it is identically equal to zero if B is nonrecurrent, and identically equal to one if B is recurrent. Consequently, *for a nonrecurrent set*

$$\pi_B(x) = G\varphi_B(x),$$

i.e., *the probability* $\pi_B(x)$ *is the potential of a nonnegative function* φ_B. Here, according to Eq. (29), $\varphi_B(x) = \pi_B(x) - P\pi_B(x) = \mathbf{P}_x\{A_0\} - \mathbf{P}_x\{A_1\} = \mathbf{P}_x\{A_0 \setminus A_1\}$ is the probability, on leaving x, of being situated at the initial instant in the set B and of always exiting from B in the first step. It is clear that this probability can only have a nonzero value for $x \in B$, hence, *outside the set B the function* φ_B *is equal to zero*.

The first term in Eq. (30) is the probability of visiting B a positive finite number of times. Reiterating the arguments of the preceding paragraph, we readily show that $P^n\varphi_B(x) = \mathbf{P}_x\{A_n \setminus A_{n+1}\}$ so that the expansion

$$G\varphi_B = \sum_{n=0}^{\infty} P^n\varphi_B$$

corresponds to the expansion of this probability into the sum of the probabilities of visiting B the last time in the nth step.

It is apparent from the relation (24) derived at the end of the preceding section that if $f = G\varphi$ $(\varphi \geq 0)$ and τ is the time of first visit to the set B, then

$$\mathbf{M}_x f(x(\tau)) \leqslant f(x). \tag{31}$$

It follows from the expansion (28) that this inequality is valid for any bounded excessive function (because, according to the results of §4, a bounded harmonic function h is constant). We will see later that the assumption of boundedness on the part of the function f is superfluous and that the inequality (31) is satisfied for a broader class of times τ (see Chapt. III, §3). Equation (31) is reminiscent of the inequality $\mathbf{M}_x f(x(1)) = Pf(x) \leq f(x)$ that occurs in the definition of the excessive function, the only difference being that now τ is a random time.

§7. Capacity

The Newtonian potential is closely allied with the notion of capacity. The capacity of a body B is defined as follows in electrostatics. Let us consider all the distributions φ of positive charges on B whose potentials at any point of space do not exceed one. It has been proved that there is a maximum among these potentials. This is called the equilibrium potential, and the corresponding charge distribution is known as the equilibrium distribution φ. The total charge

$$C(B) = \int_B \varphi(y)\,dy$$

for the equilibrium distribution φ is called the capacity of the body B. Proceeding from Eq. (17), we generate a definition of capacity in an l-dimensional space for $l > 3$. Capacity is one of the focal concepts in the theory of the Laplace equation.

Starting with discrete potentials $f = G\varphi$, we seek to develop analogous formulations for functions defined on a lattice H^l. We fix a subset B of such a lattice and investigate the class K_B of all functions $\varphi \geq 0$ equal to zero outside B and such that $G\varphi \leq 1$.

For the function $f = G\varphi$, where $\varphi \in K_B$, Eq. (24) of §5 takes the form

$$f(x) = M_x f(x(\tau)), \tag{32}$$

where τ is the time of first visit of the particle to the set B. It follows from the inequality $f \leq 1$ that $\mathbf{M}_x f(x(\tau)) \leq \mathbf{P}_x\{\tau < \infty\} = \pi_B(x)$. We deduce from Eq. (32), therefore, that

$$f(x) \leqslant \pi_B(x). \tag{33}$$

If the set B is nonrecurrent, then, as we saw in the preceding section, $\pi_B = G\varphi_B$, where $\varphi_B = \pi_B - P\pi_B \in K_B$. Consequently, π_B is reasonably called the equilibrium potential, φ_B is known as the equilibrium distribution, and the capacity of the set B is defined by the formula

$$C(B) = \sum_y \varphi_B(y). \tag{34}$$

For recurrent sets the concept of capacity cannot be met. We recall that all finite sets are nonrecurrent.

We now set forth an extremal property of the equilibrium distribution φ_B, forming the discrete analog of the Gauss theorem in Newtonian potential theory. Let a set B be nonrecurrent. We will show that for any function $\varphi \in K_B$ in this case

$$\sum_y \varphi(y) \leqslant \sum_y \varphi_B(y) = C(B). \tag{35}$$

The quantity $\sum_y \varphi(y)$ is logically called the total charge corresponding to the distribution φ. The inequality tells us that the capacity of a nonrecurrent set B is definable as the maximum total charge concentrated on B whose potential does not exceed one.

For the proof of the relation (35) we introduce the abbreviated notation

$$(f_1, f_2) = \sum_{y \in B} f_1(y) f_2(y).$$

By virtue of the symmetry of the random walk, p(n, x, y) = p(n, y, x), and, therefore, g(x, y) = g(y, x). Consequently,

$$(Gf_1, f_2) = (f_1, Gf_2).$$

Utilizing the fact that $\pi_B(x) = 1$ for $x \in B$ and that $G\varphi \leq 1$ for $\varphi \in K_B$, we deduce that

$$\sum_y \varphi(y) = (\varphi, \pi_B) = (\varphi, G\varphi_B) = (G\varphi, \varphi_B) \leqslant (1, \varphi_B) = C(B).$$

§ 8. The Recurrence Criterion

We now establish a necessary and sufficient condition for the recurrence of a subset B of a three-dimensional lattice. This condition is formulated in terms of capacities and is very similar in its substance and in its formulation to the Wiener criterion for the regularity of boundary points in the theory of differential equations (see, e.g., [16], Chapt. IV, §4). The reader will have little difficulty extending the discussion to the case $l > 3$.

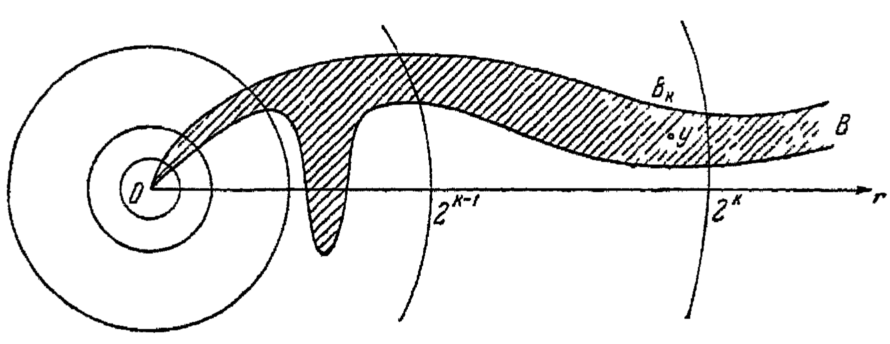

Fig. 1

Inasmuch as any bounded set is nonrecurrent, the recurrence of the set B does not depend on the structure of B inside any predetermined sphere. As it turns out, the recurrence of B depends on how rapidly the number of points of B falling within a sphere of radius r grows as $r \to \infty$.

Let us consider an expanding sequence of spheres with centers at zero and radii $r = 1, 2, 2^2, \ldots, 2^k, \ldots$, which are growing in a geometric progression. We denote by B_k that part of the set B which falls between the kth and the (k+1)st spheres (more precisely, the set of those x from B for which $2^{k-1} < \|x\| \leq 2^k$; see Fig.1). The set B_k is finite, hence for it the capacity $C(B_k)$ is defined. The following criterion holds.

In order for the set B to be recurrent, it is necessary and sufficient that the following series diverge:

$$\sum_k \frac{C(B_k)}{2^k}. \tag{36}$$

We first prove the necessity of this condition, i.e., we show that convergence of the series (36) implies nonrecurrence of the set B.

We first verify that, along with the series (36), the following series also converges:

$$\sum_k \pi_{B_k}(0). \tag{37}$$

We do this with the aid of the asymptotic estimate given at the end of §3:

$$g(x, y) \sim \frac{Q}{\|x-y\|} \qquad (\|x-y\| \to \infty), \tag{38}$$

where $Q = c_3$ [see Eq. (12)]. By virtue of (38), there exists an $N > 0$ such that the following inequality is fulfilled for $y \in B_k$, $k > N$:

$$g(0, y) \leqslant \frac{2Q}{\|y\|}. \tag{39}$$

Since $\pi_{B_k}(x)$ is the equilibrium potential for the set B_k, we have

$$\pi_{B_k}(0) = G\varphi_{B_k}(0) = \sum_y g(0, y)\varphi_{B_k}(y),$$

where φ_{B_k} is the equilibrium distribution on the set B_k. Making use of the estimate (39), the zero value of φ_{B_k} outside B_k, and the fact that $\|y\| > 2^{k-1}$ for $y \in B_k$, we obtain

$$\pi_{B_k}(0) \leqslant \sum_y \frac{2Q\varphi_{B_k}(y)}{\|y\|} \leqslant \frac{Q}{2^{k-2}} \sum_y \varphi_{B_k}(y) = 4Q\frac{C(B_k)}{2^k}.$$

Hence, the series (37) is majorized by the series (36), correct to a constant factor, and also converges.

We note that since the event $\left\{\text{The particle visits the set } \bar{B}_n = \bigcup_{k=n}^{\infty} B_k\right\}$ comprises the union of the events {The particle visits B_k}, $k = n, n+1, \ldots$, then

$$\pi_{\bar{B}_n}(0) \leqslant \sum_{k=n}^{\infty} \pi_{B_k}(0).$$

Consequently, for sufficiently large n we have $\pi_{\bar{B}_n}(0) < 1$, so that the set $\bar{B} = \bar{B}_n$ is nonrecurrent. As the finite set $\bar{\bar{B}} = B \setminus \bar{B}$ is also nonrecurrent, we have only to prove that the union of two nonrecurrent sets is nonrecurrent.

We recall for this purpose the second definition of recurrence, according to which a set is nonrecurrent if a particle visits that set a finite number of times with probability one (see §4). Inasmuch as the intersection of two events of probability one is itself an event of probability one, in our case the particle has probability one of visiting both $\bar{B}$ and $\bar{\bar{B}}$ a finite number of times, hence the same is true of their union $B = \bar{B} \cup \bar{\bar{B}}$. Consequently, the set B is nonrecurrent.

The sufficiency of the recurrence condition is more difficult to prove. Let the series (36) diverge. We decompose it into four

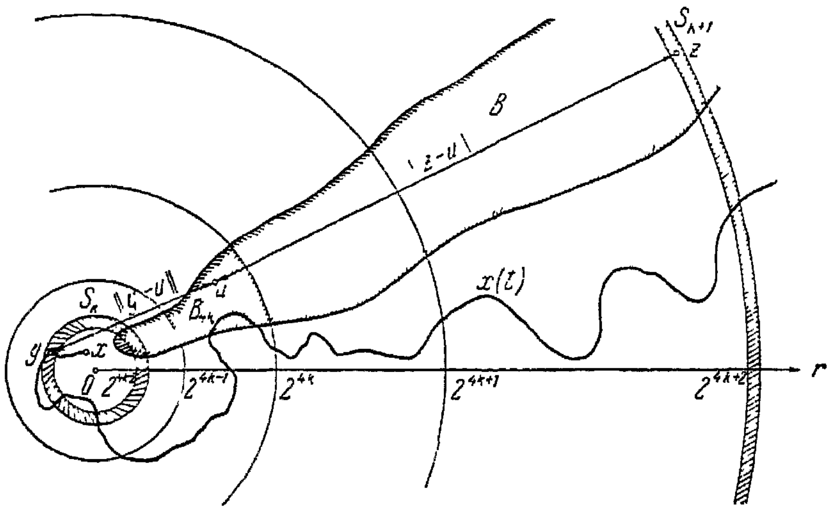

Fig. 2

series, each of which contains terms of the series (36) with indices yielding identical remainders when divided by four. At least one of these four series diverges. Let us suppose for definiteness that the following does so:

$$\sum_k \frac{C(B_{4k})}{2^{4k}}. \tag{40}$$

We denote by S_k the set of lattice points lying in the spherical layer bounded by the spheres of radius 2^{4k-2} and $2^{4k-2}+1$ (the set of all y from H^3 for which $2^{4k-2} \leq \|y\| \leq 2^{4k-2}+1$) (Fig. 2). The set B_{4k} is then contained between the layers S_k and S_{k+1} and considerably nearer to S_k than to S_{k+1}.

Inasmuch as the distance of the particle from the coordinate origin does not vary by more than unity during the first step, the particle cannot jump through the layer S_k without intersecting it. The particle goes to infinity with probability one, so it passes with probability one through all layers S_k enclosing the initial state x. We investigate *the event* A_k *wherein the particle visits the set* B_{4k} *on the path from the layer* S_k *to the layer* S_{k+1} (more precisely, betwen the times of first visit to S_k and to S_{k+1}). We show that for all sufficiently large k

$$\mathbf{P}_y\{A_k\} \geqslant Q_1 \frac{C(B_{4k})}{2^{4k}} \quad \text{for} \quad y \in S_k, \tag{41}$$

where $Q_1 > 0$ does not depend on y or k.

If a particle, on leaving $y \in S_k$, visits B_{4k}, then either the event A_k or the event $D_k = \{$The particle visits B_{4k} after hitting the layer $S_{k+1}\}$ occurs. Therefore,

$$\pi_{B_{4k}}(y) \leqslant \mathbf{P}_y\{A_k\} + \mathbf{P}_y\{D_k\}.$$

It is clear that

$$\mathbf{P}_y\{D_k\} \leqslant \max_{z \in S_{k+1}} \pi_{B_{4k}}(z)$$

[this inequality is formally proved by the introduction of the probabilities q(n, z) of first hitting the layer S_{k+1} at the time n at the point z; see the discussion at the end of §5]. Consequently,

$$\mathbf{P}_y\{A_k\} \geqslant \pi_{B_{4k}}(y) - \max_{z \in S_{k+1}} \pi_{B_{4k}}(z). \tag{42}$$

It remains for us to evaluate the function $\pi_{B_{4k}}$. Since this function is the potential of the equilibrium distribution $\varphi_{B_{4k}}$, while the function $\varphi_{B_{4k}}$ is equal to zero outside B_{4k}, and the total charge of the equilibrium distribution is equal to the capacity $C(B_{4k})$, it follows from the inequality (42) that for $y \in S_k$

$$\mathbf{P}_y\{A_k\} \geqslant \sum_u g(y, u)\varphi_{B_{4k}}(u) - \max_{z \in S_{k+1}} \sum_u g(z, u)\varphi_{B_{4k}}(u)$$

$$\geqslant C(B_{4k})\left[\min_{\substack{y \in S_k \\ u \in B_{4k}}} g(y, u) - \max_{\substack{z \in S_{k+1} \\ u \in B_{4k}}} g(z, u)\right].$$

Applying the asymptotic function (38) here, we see that for sufficiently large k and $y \in S_k$

$$\mathbf{P}_y\{A_k\} \geqslant C(B_{4k})\left(\frac{5Q}{6r_k} - \frac{7Q}{6R_k}\right),$$

where r_k is the largest distance between the points $y \in S_k$ and $u \in B_{4k}$, and R_r is the least distance between $z \in S_{k+1}$ and $u \in B_{4k}$. It follows from the relative position of the sets S_k, B_{4k}, and S_{k+1}

that

$$r_k \leqslant 2^{4k-2} + 1 + 2^{4k} \leqslant 2 \cdot 2^{4k},$$
$$R_k \geqslant 2^{4k+2} - 2^{4k} = 3 \cdot 2^{4k}.$$

Consequently, for sufficiently large k

$$\mathbf{P}_y\{A_k\} \geqslant \frac{Q}{36} \frac{C(B_{4k})}{2^{4k}} \qquad (y \in S_k),$$

and the inequality (41) is thus proved.

Now that we have the inequality (41), it is no longer a difficult matter to prove the recurrence of the set B. We pick a number m such that the initial state x lies inside the layer S_m and the inequality (41) is satisfied for all $k \geq m$. We denote by τ_k the time of first visit to the layer S_k. The opposite of the event A_k is the event $\bar{A}_k = \{$During the period $[\tau_k, \tau_{k+1}]$ the particle did not visit $B_{4k}\}$. It follows from the estimate (41) that, irrespective of the values of τ_k and $x(\tau_k)$ or the nature of the motion prior to the time τ_k, the probability of $\bar{A}_k$ is not greater than

$$1 - Q_1 \frac{C(B_{4k})}{2^{4k}}.$$

For any s, therefore,

$$\mathbf{P}_x\{\bar{A}_m \cap \bar{A}_{m+1} \cap \ldots \cap \bar{A}_{m+s}\} \leqslant \prod_{k=m}^{m+s} \left(1 - Q_1 \frac{C(B_{4k})}{2^{4k}}\right).$$

Indeed, let

$$q_k(n, y) = \mathbf{P}_x\{\tau_k = n,\ x(\tau_k) = y,\ \bar{A}_m \cap \ldots \cap \bar{A}_{k-1}\}.$$

Then

$$\mathbf{P}_x\{\bar{A}_m \cap \ldots \cap \bar{A}_{k-1} \cap \bar{A}_k\} = \sum_{n,\,y} q_k(n, y)\, \mathbf{P}_y\{\bar{A}_k\}$$
$$\leqslant \left(1 - Q_1 \frac{C(B_{4k})}{2^{4k}}\right) \mathbf{P}_x\{\bar{A}_m \cap \ldots \cap \bar{A}_{k-1}\}.$$

Passing to the limit as $s \to \infty$ and bearing in mind that the series $\sum Q_1 \frac{C(B_{4k})}{2^{4k}}$ diverges, we infer that

$$P_x\{A_m \cup A_{m+1} \cup \cdots \cup A_{m+n} \cup \cdots\}$$

$$= 1 - P_x\{\bar{A}_m \cap \bar{A}_{m+1} \cap \cdots \cap \bar{A}_{m+n} \cap \cdots\} = 1$$

and hence, that the particle with probability one visits one of the sets B_{4k} belonging to B. Thus, the set B must be recurrent.

§9. Recurrence of a Set Situated on the Axis

Making use of the recurrence criterion developed in the preceding section, we now try to imagine what recurrent and nonrecurrent sets of a three-dimensional lattice look like.

It is clear that any subset of a nonrecurrent set is also nonrecurrent, and that if a set contains a recurrent subset, the set itself is recurrent. Moreover, we know that any bounded set is nonrecurrent.

We denote the coordinates of the point x(n) by $x_1(n)$, $x_2(n)$, and $x_3(n)$. We will show that the coordinate plane $x_3 = 0$ is a recurrent set. Clearly, the value of $x_3(n)$ varies according to the following law: During unit time it increases or decreases by unity with probabilities 1/6 and keeps the same value with probability 2/3. The probability that the value of $x_3(n)$ will retain the same value k times in succession is equal to $(2/3)^k$. It tends to zero as $k \to \infty$, hence the value of $x_3(n)$ must change sooner or later. It is evident from symmetry considerations that the first increment of the quantity $x_3(n)$ will be equal to either -1 or 1 with probability 1/2. Therefore, the random variation law governing $x_3(n)$ differs from the one-dimensional symmetric random walk described at the beginning of §1 only in the possibility of the particle becoming trapped for some finite period of time in every state in which it happens to be situated. It is clear that such waiting times do not alter the probabilities of attaining the value 0, but only affect the speed of motion, not the configuration of the path. Inasmuch as the point O is accessible from any other point with probability one in a simple random walk, $x_3(n)$ reaches zero at some time. Thus, the coordinate plane $x_3 = 0$ is a recurrent set.

Proceeding from the fact that the point O is also accessible from any other point with probability one in a two-dimensional symmetric random walk (see the problems), it could be proved analogously that the coordinate axis $x_2 = x_3 = 0$ forms a recurrent set. Invoking the Wiener criterion, it is possible not only to prove this fairly simple fact, but also to obtain the following test of recurrence of the set B consisting of the points $\{b_n, 0, 0\}$, where $0 < b_1 < b_2 < \ldots$ (the b_n, of course, are integers):

If the series $\sum \frac{1}{b_n}$ converges, the set B is nonrecurrent, and if the series $\sum \frac{1}{b_n}$ diverges and if for large n

$$b_{n+1} - b_n \geqslant c \log_2 b_n \qquad (c = \text{const} > 0), \tag{43}$$

then the set B is recurrent.

The relationship of the recurrence of the set B to the divergence of the series $\sum \frac{1}{b_n}$ is entirely reasonable; divergence of the series tells us that the points $\{b_n, 0, 0\}$ are situated close together, whereas convergence tells us that b_n is rapidly increasing. The condition (43) is related to the method used in the proof for estimating capacities.* It is satisfied only by very slowly diverging series.

---

*The condition (43) can be weakened by demanding the existence of a subsequence b_{n_k} such that $\sum\limits_k \frac{1}{b_{n_k}} = \infty$ and that for all k

$$b_{n_{k+1}} - b_{n_k} \geqslant c \log_2 b_{n_k}.$$

In this case the set B contains a recurrent subset, hence it is recurrent itself. It is only natural to ask whether it is impossible to pick from a sequence b_n for which $\sum\limits_n \frac{1}{b_n} = \infty$ a sequence b_{n_k} for which $\sum\limits_k \frac{1}{b_{n_k}} = \infty$ and the condition (43) is satisfied. Examples formulated by S. M. Gusein-zade and L. A. Ivanov show that this cannot always be done.

Is divergence of the series $\sum\limits_n \frac{1}{b_n}$ a sufficient condition for recurrence of the set B? R. S. Bucy [17] gives an example refuting this hypothesis. On the other hand, it is shown in the same paper that the auxiliary condition (43) can be replaced by the requirement that the difference $b_{n+1} - b_n$ be increasing.

For example, if $b_n = [n \log_2 n]$ ([x] denotes the integer part of x), then for large n

$$b_{n+1} - b_n \geqslant (n+1)\log_2(n+1) - 1 - n\log_2 n \geqslant \log_2 n - 1$$
$$= \log_2 \frac{n}{2} \geqslant \log_2 \sqrt{n \log_2 n} \geqslant \frac{1}{2} \log_2 b_n$$

and the inequality (43) is satisfied. Therefore, for $b_n = [n \log_2 n]$ the set B is recurrent. But if $b_n = [n \log_2^{\alpha} n]$, where $\alpha > 1$, the series $\sum \frac{1}{b_n}$ converges, and B is nonrecurrent.

Thus, let the series $\sum \frac{1}{b_n}$ converge. We note that the capacity of a finite set does not exceed the number of elements of that set. This is apparent from the definition of capacity:

$$C(B) = \sum_{x \in B} \varphi_B(x),$$

where φ_B, like the probability, cannot be greater than unity. Let us estimate the number of elements of the set B_k involved in the recurrence criterion. We call this number $|B_k|$. If $b_n \in B_k$,* then $2^{k-1} < b_n \leq 2^k$, whence $1/2^k \leq 1/b_n$. Summing these inequalities over all points belonging to a given B_k, we obtain

$$\frac{|B_k|}{2^k} \leqslant \sum_{b_n \in B_k} \frac{1}{b_n}$$

and, therefore,

$$\frac{C(B_k)}{2^k} \leqslant \sum_{b_n \in B_k} \frac{1}{b_n}.$$

Consequently, the series (36) majorizes the series $\sum \frac{1}{b_n}$ and also converges. According to the recurrence criterion, the set B is nonrecurrent.

*Here and elsewhere for brevity we simply write b_n instead of $\{b_n, 0, 0\}$. The reader will have no trouble distinguishing when b_n means a number and when it means a point of the lattice H^3.

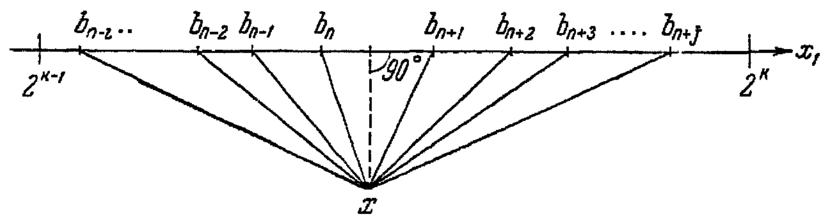

Fig. 3

Let us now assume that the series $\sum \frac{1}{b_n}$ diverges. If the inequalities (43) are fulfilled, then, as we will show presently, for any x

$$\sum_{y \in B_k} g(x, y) \leqslant M, \tag{44}$$

where M is some fixed number. Using this inequality, let us find a lower estimate for $C(B_k)$. We consider the function $\varphi(y)$, which is equal to $1/M$ for $y \in B_k$ and equal to zero at all other points. The potential of this function is

$$f(x) = \sum_y g(x, y)\,\varphi(y) = \frac{1}{M} \sum_{y \in B_k} g(x, y)$$

and, by virtue of (44), does not exceed unity. Recalling the definition given at the end of §7 for the capacity as the maximum total charge whose potential is not greater than one, we deduce that

$$C(B_k) \geqslant \sum_y \varphi(y) = \frac{|B_k|}{M}.$$

If $b_n \in B_k$, then $2^{k-1} < b_n \leq 2^k$, whence $1/b_n < 1/2^{k-1}$ Summing these inequalities over $b_n \in B_k$, we obtain

$$\sum_{b_n \in B_k} \frac{1}{b_n} \leqslant \frac{|B_k|}{2^{k-1}} \leqslant 2M \frac{C(B_k)}{2^k}.$$

Consequently, the series $\sum \frac{1}{b_n}$ is majorized, correct to a factor 2M, by the series (36). The divergence of the series $\sum \frac{1}{b_n}$ implies the divergence of the series (36) and recurrence of the set B.

It is now left for us to prove the inequality (44). Clearly, we may assume that $k \geq 2$. We note that the quantities $g(x, y)$ and $\|x - y\| g(x, y)$ are bounded by some number Q, by virtue of the asymptotic estimate (12). Let b_n and b_{n+1} be the two points nearest x from the set B_k, and let $b_{n-1}, b_{n-2}, \ldots, b_{n-i}$ and $b_{n+2}, b_{n+3}, \ldots, b_{n+j}$ be all the remaining points of this set (Fig. 3).

According to the foregoing remark,

$$\sum_{y \in B_k} g(x, y) \leqslant 2Q + Q\left(\frac{1}{\|x - b_{n-1}\|} + \frac{1}{\|x - b_{n-2}\|} + \ldots + \frac{1}{\|x - b_{n-i}\|}\right)$$
$$+ Q\left(\frac{1}{\|x - b_{n+2}\|} + \frac{1}{\|x - b_{n+3}\|} + \ldots + \frac{1}{\|x - b_{n+j}\|}\right).$$

It follows from the condition (43) that within the limits of the set B_k

$$b_{l+1} - b_l \geqslant c \cdot \log_2 b_l \geqslant c \log_2 2^{k-1} = c(k-1).$$

Therefore (see Fig. 3),

$$\begin{aligned}
&\|x - b_{n-1}\| \geqslant b_n - b_{n-1} \geqslant c(k-1),\\
&\|x - b_{n-2}\| \geqslant b_n - b_{n-2} \geqslant 2c(k-1),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&\|x - b_{n-i}\| \geqslant b_n - b_{n-i} \geqslant ic(k-1),\\
&\|x - b_{n+2}\| \geqslant b_{n+2} - b_{n+1} \geqslant c(k-1),\\
&\|x - b_{n+3}\| \geqslant b_{n+3} - b_{n+1} \geqslant 2c(k-1),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&\|x - b_{n+j}\| \geqslant b_{n+j} - b_{n+1} \geqslant (j-1)c(k-1).
\end{aligned}$$

Consequently,

$$\sum_{y \in B_k} g(x, y) \leqslant 2Q + \frac{Q}{c(k-1)}\left(1 + \frac{1}{2} + \ldots + \frac{1}{i} + 1 + \frac{1}{2} + \ldots + \frac{1}{j-1}\right).$$

Since i and j do not exceed 2^{k-1}, i.e., the number of points included on the abscissa axis between the spheres of radius 2^{k-1} and 2^k, we have

$$\sum_{y \in B_k} g(x, y) \leqslant 2Q + \frac{2Q}{c(k-1)}\left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{2^{k-1}}\right).$$

Inasmuch as

$$\frac{1}{n} \leqslant \int_{n-1}^{n} \frac{dx}{x},$$

we then have

$$\sum_{y \subset B_k} g(x, y) \leqslant 4Q + \frac{2Q}{c(k-1)} \int_{1}^{2^{k-1}} \frac{dx}{x} = 4Q + \frac{2Q}{c(k-1)} \ln 2^{k-1} = 2Q\left(2 + \frac{\ln 2}{c}\right).$$

PROBLEMS

Two-Dimensional Lattice

1. For a symmetric random walk on a two-dimensional lattice $g(x, x) = \infty$.

Hint Inasmuch as $p(2k+1, x, x) = 0$, it follows that

$$g(x, x) = \sum_{k=0}^{\infty} p(2k, x, x).$$

In the representation (6) for $p(2k, x, x)$ the integrand is positive. Term-by-term integration is therefore admissible, and

$$g(x, x) = \frac{1}{\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{d\theta_1\, d\theta_2}{4 - (\cos\theta_1 + \cos\theta_2)^2}.$$

The resulting integral is evaluated with the aid of the inequality $\cos\alpha \geq 1 - (\alpha^2/2)$ $(|\alpha| < \alpha_0)$.

2. We denote by $r(x)$ the probability that a particle leaving the state x will return to that state again at some time. Then

$$g(x, x) = \sum_{n=0}^{\infty} r(x)^n.$$

Hint. Consider the random variable ξ_k, which is equal to one if the particle returns to the point x at least k times and is equal to zero otherwise (k = 1, 2, ...). Then

$$g(x, x) = 1 + \mathbf{M}_x(\xi_1 + \xi_2 + \dots).$$

3. In a symmetric random walk on a two-dimensional lattice the one-point set x is recurrent.

Hint. It follows from Problems 1 and 2 that $r(x) = 1$. On the other hand, $1 - r(x) \geq s(x, y)\,[1 - \pi_x(y)]$, where $s(x, y)$ is the probability of hitting y from x before the first return to x.

Extremal Points of a Convex Set

Let H be a finite or denumerable space, and let E be a certain set of functions given on H. We say that a sequence of functions $\{f_n\}$ converges to the function f if $f_n(x) \to f(x)$ for every $x \in H$. The set E is called closed if it follows from $f_n \to f$, $f_n \in E$ that $f \in E$, and is called compact if it is closed and if it is possible from any sequence $\{f_n\}$, $f_n \in E$, to pick a convergent subsequence. If it follows from $f_1 \in E$, $f_2 \in E$ that $pf_1 + qf_2 \in E$ for any $p > 0$, $q > 0$, $p + q = 1$, then we say that the set E is convex.

4. A closed set E is compact when and only when there exists a function $c(x)$ such that $|f(x)| \leq c(x)$ for $f \in E$ and $x \in H$.

Hint. For the proof of sufficiency we index all points of the space H and use the diagonal process.

A compact subset A of a set E is called extremal if it follows from $f \in A$, $f = pf_1 + qf_2$, and $f_1, f_2 \in E$, $p > 0$, $q > 0$, $p + q = 1$ that f_1 and $f_2 \in A$.

5. If A is an extremal set, $f \in A$ and

$$f = \alpha_1 f_1 + \dots + \alpha_n f_n,$$

where $f_1, \dots, f_n \in E$; $\alpha_1, \dots, \alpha_n > 0$, $\alpha_1 + \dots + \alpha_n = 1$, then $f_1, \dots, f_n \in A$.

We say that a functional $l(f)$ is linear if $l(af_1 + bf_2) = al(f_1) + bl(f_2)$ for any numbers a and b and functions f_1 and f_2 and if it follows from $f_n \to f$ that $l(f_n) \to l(f)$. An example of a linear functional is $l(f) = f(x_0)$, where x_0 is a fixed point of the space H.

6. Let A be an extremal set, let l be a linear functional, and let

$$M = \max_{f \in A} l(f).$$

Then the set A_1 of functions $f \in A$ for which $l(f) = M$ is also an extremal set.

7. Let $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$ be a sequence of compact extremal sets. The set $A_\infty = \bigcap_n A_n$ is also extremal.

An extremal set consisting of one point is called an *extremal point*.

8. Any compact convex set E has an extremal point.

Hint. According to Problem 4, the functional $l_x(f) = f(x)$, where x is a fixed point of H, reaches its maximum on any compact and can be used to "shrink" the already-existing extremal set (see Problem 6). Beginning with the entire set E and choosing all the functions l_x in some definite order, we obtain in the limit an extremal set A_∞ (see Problem 7). This set comprises a single function.

It has been proved that any compact convex set consists of all functions of the form $\alpha_1 f_1 + \ldots + \alpha_n f_n$, where $\alpha_1, \ldots, \alpha_n > 0$, $\alpha_1 + \ldots + \alpha_n = 1$, $f_1, \ldots, f_n$ are extremal points, and all limits of such functions (Krein-Milman theorem; see, e.g., [18], § 3. The special case of this theorem when E has only one extremal point is sufficient for our purposes.

9. If a compact convex set E has only one extremal point g, then E comprises a single function g.

Hint. Suppose that $h \in E$, $h \neq g$. Then for some $x \in H$ the inequality $l(h) > l(g)$ is fulfilled for one of the linear functions $l(f) = \pm f(x)$. According to Problem 6, there exists a compact convex extremal set A not containing g. The set A has an extremal point $g_1 \neq g$, which is also an extremal point for all of E.

Positive Harmonic Functions

10. If a nonnegative harmonic function has a zero at some point, it is equal to zero everywhere.

11. The limit of a sequence of harmonic functions is a harmonic function.

12. If f is a positive harmonic function, then

$$f(x \pm e_k) \leqslant 2l f(x),$$

where l is the dimensionality of the lattice.

We denote by E the class of positive harmonic functions equal to one at the point $x=0$.

13. The set E is convex and compact.

Hint. Use Problems 12, 11, and 4.

14. If $f \in E$, then for any integer-valued vector a the function $g(x) = [f(x+a)]/f(a)$ belongs to E.

15. If g is an extremal point of the set E, then

$$g(x \pm e_k) = g(e_k)^{\pm 1} g(x).$$

Hint. Use the equation $g = Pg$ and Problems 14 and 5.

16. Under the conditions of the preceding problem $g(x) = 1$ for all $x \in H^l$.

Hint. Since $g(\mathbf{e}_k) + g(\mathbf{e}_k)^{-1} \geq 2$ and the equality is only possible for $g(\mathbf{e}_k) = 1$, it follows from the equation $g = Pg$ that $g(\mathbf{e}_k) = 1$.

It is inferable from Problems 9, 13, and 16 that the set E comprises the single function 1, and, hence, that any positive harmonic function is constant.

17. If a harmonic function is bounded below (above), it is constant.

The Dirichlet Problem

Let B be a subset of points of a lattice H^l. We call a point $x \bar{\in} B$ a *boundary point* for the set B if at least one point of the type $x \pm \mathbf{e}_k$ belongs to B. We call the collection of boundary points of the set B the *boundary* of B, designating it ∂B. We say that a function $f(x)$, $x \in B \cup \partial B$, is *harmonic (superharmonic)* on B if the equation $f(x) = Pf(x)$ [the inequality $f(x) > Pf(x)$]

is fulfilled for all $x \in B$. We call the set B *connected* if for any two points $x, y \in B$ there exists a chain of points $x_1 = x, x_2, x_3, \ldots, x_n = y$ from B such that each of the differences $x_i - x_{i-1}$ coincides with one of the vectors $\pm \mathbf{e}_k$.

18. If the set B is connected, f is a superharmonic function on B, and f reaches its minimum value on $B \cup \partial B$ at a point $x \in B$, then f is constant on $B \cup \partial B$.

19. If the set B is finite and the functions f_1 and f_2 harmonic on B are identical on ∂B, then they are also identical on B.

Hint. Apply Problem 18 to the functions $f_1 - f_2$ and $f_2 - f_1$.

In Problems 20-24 below, τ denotes the time of first visit to the set ∂B, and φ is an arbitrary function given on ∂B.

20. If the set B is connected, then the expectation $\mathbf{M}_x \varphi(x(\tau))$ exists, or does not exist, simultaneously for all $x \in B$.

21. The function $f(x) = \mathbf{M}_x \varphi(x(\tau))$ is a harmonic function on B and coincides with φ on ∂B (assuming that this expectation exists for all $x \in B$).

It follows from Problems 19 and 21 that for any set B distinct from the entire lattice and any bounded function φ on ∂B, there exists a function f harmonic on B and equal to φ on ∂B (the Dirichlet problem has a solution), and that this solution is unique in the case of a finite set B. A sufficient condition for uniqueness of a bounded solution of the Dirichlet problem for an infinite set B is given in Problem 22.

22. If the boundary ∂B of the set B is recurrent and the function φ is bounded, then the unique function that is bounded, harmonic on B, and equal to φ on ∂B is the function $f(x) = \mathbf{M}_x f(x(\tau))$.

Hint. Let g(x) be a bounded function, harmonic on B, and equal to φ on ∂B, let K be an l-dimensional cube with center at zero and side a, and let τ_1 be the time of first visit to $\partial(B \cap K)$. Then $g(x) = \mathbf{M}_x g(x(\tau_1))$ for $x \in B \cap K$, and as $a \to \infty$, this equation goes over to $g(x) = \mathbf{M}_x \varphi(x(\tau)) = f(x)$ (due to the recurrence of ∂B, the probability of the equality $\tau_1 = \tau$ tends to one as $a \to \infty$).

23. If the set ∂B is nonrecurrent, the statement of Problem 22 is in general untrue.

Hint. Consider the functions $f = 1$ and $g = \pi_{\partial B}(x)$.

24. Functions harmonic on the entire lattice can be determined as follows: f is harmonic if for any x and any finite set B containing x the equation $f(x) = M_x f(x(\tau))$ is satisfied.

Properties of Potentials

In Problems 25-31 the term "potential" (symbol $G\varphi$) means, unless special mention is made otherwise, a finite potential of a nonnegative function.

25. There exist points at which the potential assumes values arbitrarily close to zero.

Hint. Apply the operator P^n to the inequality $G\varphi \geq h$ ($h =$ const > 0), and let n tend to infinity.

26. If $\varphi(x) \geq \varepsilon$ on some recurrent set B (where ε is a positive number), the potential $G\varphi$ is infinite.

27. If an excessive function does not exceed some potential, it is a potential itself.

Hint. Use the formula $f = G\varphi + h$ and Problem 25.

28. (Principle of the envelope.) For any family of potentials $\{f_\alpha\}$ the function $f(x) = \inf_\alpha f_\alpha(x)$ is also a potential.

We define the carrier of the potential $G\varphi$ as the set of points x at which $\varphi(x) > 0$.

29. (Domination principle.) If $G\varphi_1 \geq G\varphi_2$ on the carrier of $G\varphi_2$, then $G\varphi_1 \geq G\varphi_2$ everywhere.

Hint. Use Eq. (24), taking the time of first visit to the carrier of $G\varphi_2$ as the value of τ.

30. (Principle of sweeping out.) The function $f(x) = M_x G\varphi(x(\tau))$, where τ is the time of first visit to the set B, has the following properties: 1) f is a potential; 2) f is equal to $G\varphi$ on B; 3) f does not exceed $G\varphi$; 4) the carrier of f does not exceed the limits of B. The function f is uniquely defined by the conditions 1-4.

31. Is the following statement true: If $G\varphi_1 \geq G\varphi_2$, then $\varphi_1 \geq \varphi_2$?

Hint. Set $\varphi_1(0) = \varphi_1(\mathbf{e}_1) = 1$, $\varphi_1(x) = 0$ for all other x and $\varphi_2(0) = 1 + \varepsilon$, $\varphi_2(x) = 0$ for $x \neq 0$. For large values of $\|x\|$ the inequality $G\varphi_1 \geq G\varphi_2$ is fulfilled simultaneously for all ε from the interval (0, 1/2) by virtue of the asymptotic estimate (12). For all the other x the required inequality is attained by decreasing ε, because for $\varepsilon = 0$ we have $G\varphi_1 > G\varphi_2$, and $G\varphi_2$ depends continuously on ε.

Excessive Functions

32. The representation of an excessive function in the form $f = G\varphi + h$, where φ, $h \geq 0$, and h is a harmonic function, is unique.

33. For $l = 1$ or 2 all excessive functions are constants.

34. Excessive functions can be determined as follows: f is excessive if for any state x and any (including an empty) set B the inequality $f(x) \geq \mathbf{M}_x f(x(\tau))$ is fulfilled, where τ is the time of first visit to B.

Properties of the Capacity

In Problems 35-41 all the sets discussed are assumed to be nonrecurrent, and ∂B denotes the boundary of the set B (see the text leading up to Problem 18).

35. If $A \subset B$, then $C(A) \leq C(B)$.

36. $C(A \cup B) \leqslant C(A) + C(B)$.

Hint. Consider nonintersecting sets first.

37. The equilibrium distribution for the set $B \cup \partial B$ as a whole is concentrated on ∂B.

38. The equilibrium distributions on the sets $B \cup \partial B$ and ∂B are the same, and, in particular, $C(B \cup \partial B) = C(\partial B)$.

Hint. On entering B, a particle has a probability one of leaving by way of ∂B, because the set $H^l \setminus B$ is recurrent.

39. The capacity of a point x is equal to $1/[g(x, x)]$.

40. The capacity of a set consisting of n points tends to $n/[g(0, 0)]$ as the paired distances between the points of this set grow without limit.

Hint. Use the asymptotic estimate (12) and Problems 35 and 39.

41. The capacity of a nonrecurrent infinite set is infinite.

Asymmetric Random Walk

In a random walk on points of a lattice H^2 let the probabilities of jumping to the right, to the left, upward, and downward be equal to p, q, r, and s, respectively (p, q, r, s, >0, p+q+r+s=1), irrespective of the character of the previous motion. We set

$$Pf(x) = pf(x+e_1) + qf(x-e_1) + rf(x+e_2) + sf(x-e_2)$$

and call the function f harmonic if $Pf = f$. As in the symmetric case, it is easily established that the class E of positive harmonic functions equal to one at zero is a convex compact (cf. Problems 11-13). We wish to find the extremal points of the set E.

42. If $\Lambda(x)$ is an extremal point of the set E, then

$$\Lambda(x_1e_1 + x_2e_2) = \lambda^{x_1}\mu^{x_2}, \tag{45}$$

where λ and μ are positive numbers satisfying the relation

$$p\lambda + \frac{q}{\lambda} + r\mu + \frac{s}{\mu} = 1. \tag{46}$$

Hint. Compare Problem 15.

In Problems 43-47 it is demonstrated that the function $\Lambda(x)$ specified by Eq. (45) is in fact an extremal point (cf., the discussion in §4, pp. 8-9).

43. For a harmonic function φ let

$$\sup_x \frac{\varphi(x)}{\Lambda(x)} = M < \infty.$$

If at a point y

$$\frac{\varphi(y)}{\Lambda(y)} \geqslant M - \varepsilon,$$

then

$$\frac{\varphi(y+e_1)}{\Lambda(y+e_1)} \geqslant M - c\varepsilon,$$

where

$$c = 1 + \frac{1}{p}\left(\frac{q}{\lambda^2} + \frac{r\mu}{\lambda} + \frac{s}{\lambda\mu}\right).$$

44. If in the preceding problem $M > 0$, then for any number N there exists a chain of states y_0, $y_1 = y_0 + \mathbf{e}_1$, ..., $y_n = y_{n-1} + \mathbf{e}_1$ such that

$$\varphi(y_0) + \frac{\varphi(y_1)}{\lambda} + \ldots + \frac{\varphi(y_n)}{\lambda^n} \geqslant N\Lambda(y_0).$$

45. If $f \in E$ and

$$\sup_x \frac{f(x)}{\Lambda(x)} < \infty,$$

then for all x

$$f(x + e_1) \leqslant \lambda f(x).$$

Hint. In Problem 44 set

$$\varphi(x) = f(x + e_1) - \lambda f(x).$$

46. Under the conditions of the preceding problem $f = \Lambda$.

Hint. Apply the same arguments to the vectors $-\mathbf{e}_1$, $\mathbf{e}_2$, and $-\mathbf{e}_2$.

47. The function Λ is an extremal point of the set E.

It follows from Problems 42 and 47 that the extremal points of the set E are in one-to-one correspondence with the positive solutions (λ, μ) of Eq. (46). It is easily verified that for $p = q$ and $r = s$ this equation has a unique positive solution $\lambda = \mu = 1$, while in all other cases it defines a certain oval in the quadrant $\lambda > 0$, $\mu > 0$.

Chapter II

Probabilistic Solution of Certain Equations

§1. Definition of the Wiener Process

In the preceding chapter we investigated a random walk on an integral-valued l-dimensional lattice. Let us now imagine that the length of the step between adjacent lattice points is not equal to one, but to some number δ (which we choose to call the lattice parameter). Logically in this case the distance over which the particle moves in n steps becomes proportional to δ. We therefore vary the transition frequency as a function of δ in such fashion that for any δ the particle will move, on the average, the same distance in the same period of time. It is to be expected that in the limit as $\delta \to 0$ a continuous random walk will result, its properties resembling those of a random walk on a lattice.

In order to find the proper relation between the shrinking parameter δ and the growing transition frequency and to obtain the limiting distribution of the particle displacement in the time t, we invoke the central limit theorem for the sums of independent random vectors. In the special case when the vectors ξ_i $(i = 1, 2, \ldots)$ are mutually independent, are identically distributed, and have zero expectations and finite second moments, this theorem confirms the fact that the normed-sum distribution

$$\frac{\xi_1 + \cdots + \xi_n}{\sqrt{n}} \tag{1}$$

converges as $n \to \infty$ to a normal distribution with zero expectation and the same matrix of second moments as that of the random vector ξ_1.

We denote by ξ_i the displacement in the ith step of a particle executing a walk on the unit lattice. By virtue of the symmetry of the random walk, $M\xi_i = 0$. Let us find the second moments of the random vector ξ_i. Let $x_1, \ldots, x_l$ be the coordinates of this vector. Since only one of the coordinates $x_1, \ldots, x_l$ can have a nonzero value, all the product moments Mx_jx_k $(j \neq k)$ are equal to zero. Considering that the probability of x_j assuming values of ± 1 is $1/2l$, and of assuming a value of zero is $1 - 1/l$, it follows that $Mx_1^2 = \ldots = Mx_l^2 = 1/l$. In our situation, therefore, the vector (1) in the limit has a spherically symmetric normal distribution with a variance equal to $1/l$ in any direction.

We now observe that the displacement of the particle in n steps during a random walk on a lattice with parameter δ is equal to

$$\delta(\xi_1 + \ldots + \xi_n). \tag{2}$$

Comparing (1) and (2), we see that in order to obtain a reasonable limiting distribution, the parameter δ must be of the order $1/\sqrt{n}$ or, what amounts to the same thing, the number of steps n must be of the order $1/\delta^2$. We assume therefore that the time interval between successive jumps of the particle is equal to δ^2/l (the coefficient $1/l$ is introduced to simplify the final relation). The position of the particle at time t in this random walk is denoted by x(t) [realizing, of course, that x(t) is not yet defined for all t, but only for those t which are multiples of δ^2/l]. At the instant t the particle has executed $n = lt/\delta^2$ jumps. This means that

$$x(t) - x(0) = \sqrt{\frac{lt}{n}}(\xi_1 + \ldots + \xi_n), \text{ where } n = \frac{lt}{\delta^2}.$$

Consequently, the vector $x(t) - x(0)$ is derived from the vector (1) by multiplying by a constant coefficient $\sqrt{lt}$. Hence, in the limit as $\delta \to 0$ the increment $x(t) - x(0)$ has a symmetric normal distribution with a variance equal to $(1/l)(\sqrt{tl})^2 = t$ in any direction. The density of this distribution is

$$p(t, y) = p(t, y_1, \ldots, y_l) = \frac{1}{(2\pi t)^{l/2}} e^{-\frac{y_1^2 + \ldots + y_l^2}{2t}} = \frac{1}{(2\pi t)^{l/2}} e^{-y^2/2t} \tag{3}$$

(if $y = y_1\mathbf{e}_1 + \ldots + y_l\mathbf{e}_l$, we set $y^2 = y_1^2 + \ldots + y_l^2$). In the limit, t can now be any positive number, and x(t) and x(0) can be any points in the l-dimensional space.

As apparent from Eq. (3), the coordinates of the increment $x(t) - x(0)$ are mutually independent in the limit. We note that the coordinates of the terms ξ_i depend on one another; if one of them is different from zero, the others are equal to zero. The limiting distribution of each coordinate $x(t) - x(0)$, regardless of the dimensionality of the space, is a normal distribution with zero expectation and a variance equal to t.

Thus, it is reasonable to believe that a random walk on a lattice goes over in the limit to a continuous process, for which a random displacement of the particle during a time t has the density (3). The mathematical theory of the chaotic motion discovered by the botanist Brown in 1828 in the activity of fine particles suspended in a liquid also reduces to this process. A theory of Brownian motion was formulated in 1906 by Einstein and Smoluchowski. The mathematically correct formulation of the corresponding stochastic process was first enunciated by Wiener in 1923. We now set out to define this process, which has come to be known as the *Wiener process*.

Consider a space X consisting of functions x(t), $t \geq 0$, which assume values in an l-dimensional vector space R. These functions are interpreted as all the possible paths of Brownian motion. Let a family of distributions (i.e., probability measures) $\mathbf{P}_x$ be given on X, where x is any point of R. The measure $\mathbf{P}_x$ is to be interpreted as the distribution of random paths of a particle initating its movement from the point x at time $t = 0$. The expectation corresponding to the measure $\mathbf{P}_x$ is designated $\mathbf{M}_x$. (In cases when $\mathbf{P}_x$ or $\mathbf{M}_x$ does not depend on x, we write simply $\mathbf{P}$ or $\mathbf{M}$.*)

*Occasionally one encounters random variables ξ which are not defined for all paths. We define the expectation of these variables in terms of the usual expression, but with the integration or summation extending over, rather than the total space of elementary events Ω, only in the domain Ω_ξ in which the variable ξ is specified. An equivalent definition of the expectation $M_x\xi$ may be obtained by setting $\xi = 0$ in those cases when ξ is not defined.

We say that the set of probability measures $\mathbf{P}_x$ on X defines a Wiener process x(t) if the following conditions are met:

a) The space X contains only continuous functions.

b) $\mathbf{P}_x\{x(0)=x\}=1$.

c) The random increment $x(t+s)-x(s)$ has a symmetric normal distribution with the density (3) for $s\geq 0$, $t>0$, and this increment does not depend on any events or any random variables defined in terms of the path x(t) up to the moment s.*

In particular, for any domain $\Gamma \in R$

$$\mathbf{P}_x\{x(t)\in\Gamma\}=\mathbf{P}_x\{x(t)-x(0)\in\Gamma-x\}$$
$$=\int_{\Gamma-x} p(t, y)\,dy=\int_{\Gamma} p(t, z-x)\,dz \tag{4}$$
$$(t>0,\ x\in R).$$

We designate this probability $P(t, x, \Gamma)$ and refer to it as the transition probability of the Wiener process.

It sometimes proves useful to consider, instead of a path that begins at a fixed point x, one that begins from a random point having a distribution μ. The probability of any event A in this case is computed according to the relation

$$\mathbf{P}_\mu\{A\}=\int_R \mathbf{P}_x\{A\}\,\mu(dx).$$

We call this process a Wiener process with an initial distribution μ. We observe that for any random variable ξ

$$\mathbf{M}_\mu\xi=\int_R \mathbf{M}_x\xi\,\mu(dx). \tag{5}$$

*Events and random variables of this type refer to sets and functions measurable relative to the minimum σ-algebra in the space X containing all sets of the type $\{x(u)\in\Gamma\}$, where Γ is a domain of R, $u\leq s$. We will not devote any attention below to the problems of measurability. The reader who wishes to pursue this aspect of the problem is advised to refer, for example, to Chapt. 3 of [4].

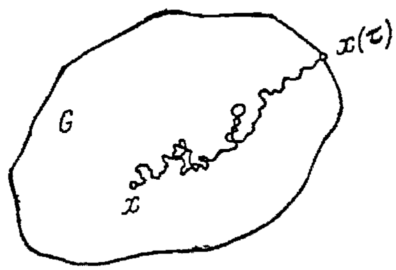

Fig. 4

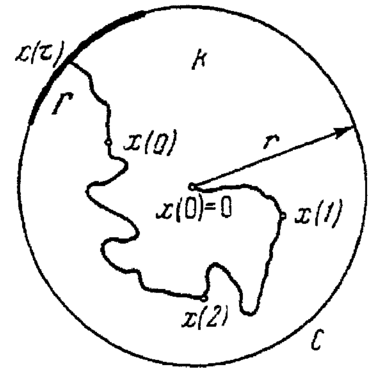

Fig. 5

§2. Distribution at the Time of Exit from a Circle and Mean Exit Time

In Chapt. I we found out which sets are reached with probability 1 in a random walk on a lattice H^l. Inasmuch as the attainment of a set A implies exit from the complementary set $H^l \setminus A$, we therefore sought to find out from which sets a particle would escape somewhere with probability one. It is reasonable to pose the same problem in the case of a Wiener process. Let G be an arbitrary domain, and let τ be the time of first exit of the path x(t) from this domain (Fig. 4). We are concerned not only with the escape probability $\mathbf{P}_x\{\tau < \infty\}$, $x \in G$, but also with the mean value $\mathbf{M}_x\tau$, as well as the distribution of the position $x(\tau)$ of the particle at the time τ.

It might also be interesting to explore the analogous problems for a random walk on a lattice, but in the continuous case the process has considerable symmetry, so that the solutions take on a simple analytical form. As we shall see, the stated problems reduce to boundary-value problems for the Laplace equation $\Delta u = 0$ and the Poisson equation $\Delta u = -2$. We are confronted with the possibility, on the one hand, of using analytical tools for the analysis of a random walk and, on the other, of invoking probabilistic considerations for the solution of analytical problems.

We will consider in the ensuing discussion a Wiener process on a plane. All the results are easily translated to a space having any number of dimensions.*

*Particularly simple relations are obtained in the one-dimensional case (see the problems at the end of the chapter).

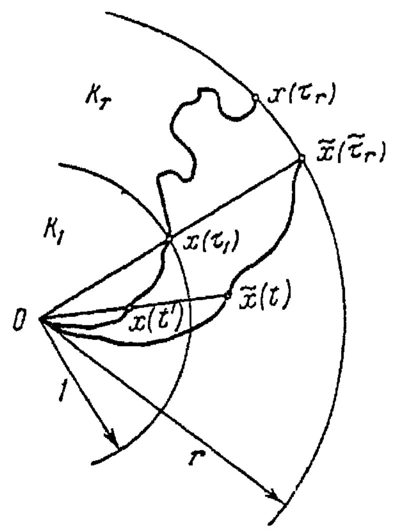

Fig. 6

As a first step, we investigate the time τ of first exit from a circular domain K of radius r on the assumption that the particle initiates its motion from the center of the circle. Since the increments $x(t) - x(0)$ do not depend on the starting point $x(0)$, the position of the circle on the plane is of no consequence, and we will assume that its center coincides with the origin (Fig. 5).

We first of all verify the fact that *the particle will indeed exit from the circle K with probability one.* Clearly, if the path $x(t)$ remains inside K until a time $t = n$, all the increments $x(1) - x(0)$, $x(2) - x(1)$, ..., $x(n) - x(n-1)$ will have absolute values smaller than 2r. These increments are independent and have the same normal distribution. Consequently,

$$\mathbf{P}_0\{\tau \geqslant n\} \leqslant [\mathbf{P}\{|x(1) - x(0)| < 2r\}]^n = \alpha^n, \tag{6}$$

where α is definitely less than 1. Hence $\mathbf{P}_0\{\tau = \infty\} \leq \alpha^n$ for any n, so that $\mathbf{P}_0\{\tau = \infty\} = 0$.

We now find the distribution of the point $x(\tau)$. Clearly, $x(\tau)$ is situated on the circumference C of the circle K. Since the distribution density of any increment $x(t) - x(s)$ depends only on $t - s$ and the length of the vector $x(t) - x(s)$, the distribution of the Wiener path does not vary on rotation of the plane through any angle about the point 0. Consequently, the distribution of the random point $x(\tau)$ is invariant relative to all rotations of the circumference C. The only distribution that has this property is a uniform distribution for which the probability of hitting any arc Γ is proportional to the length of that arc. Thus, *$x(\tau)$ is uniformly distributed over the circumference C.*

Finally, we show that *the mean time $\mathbf{M}_0\tau$ is proportional to the square of the radius of the circle K.* To do this, we make use of the following property of a Wiener process (*self-similarity*): if the plane R is given an r-fold dilation and the time axis is given an r^2-fold elongation ($r > 0$), the new process obtained from the original Wiener process is a Wiener process. In fact, given this type of transformation, clearly, the continuity of the paths and independence of the increments are preserved, and the normal distribution with the density (3) is left unchanged.

Let us consider two circles with centers at the origin: a circle K_r of radius r and a circle K_1 of radius 1 (Fig. 6). We initiate a Wiener path x(t) from the origin and denote by τ_r and τ_1 the times of first exit from K_r and K_1. Given the curve x(t), we construct the path $\tilde{x}(t) = rx(t')$, where $t = r^2t'$, and we denote by $\tilde{\tau}_r$ the time of first exit of $\tilde{x}(t)$ from the circle K_r. It is clear now that $\tilde{\tau}_r = r^2\tau_1$, hence $\mathbf{M}_0\tilde{\tau}_r = r^2\mathbf{M}_0\tau_1$. On the other hand, according to the self-similarity property, the process $\tilde{x}(t)$ is also a Wiener process, so that, consequently,

$$\mathrm{M}_0\tau_r = \mathrm{M}_0\tilde{\tau}_r = cr^2, \tag{7}$$

where $c = \mathbf{M}_0\tau_1$ is a certain constant.

Of course, we do not yet know whether c is equal to infinity or not. The finiteness of $\mathbf{M}_0\tau_1$ is readily deduced from the estimate (6). Thus, if F(t) is the distribution function of τ, then

$$\mathrm{M}_0\tau = \int_0^\infty t\,dF(t) \leqslant \sum_{n=1}^\infty \int_{n-1}^n t\,dF(t) \leqslant \sum_{n=1}^\infty n \int_{n-1}^n dF(t)$$

$$\leqslant \sum_{n=1}^\infty n\mathrm{P}_0\{\tau \geqslant n-1\} \leqslant \sum_{n=1}^\infty n\alpha^{n-1} < \infty.$$

The results thus obtained are valid for a Wiener process in a space having any number of dimensions. Notice that for any l Eq. (7) operates with the exponent 2 (not l !). The constant c in Eq. (7), of course, depends on the dimensionality l of the space. Closer examination shows that $c = 1/l$ (see the problems). Consequently, in the two-dimensional case

$$\mathrm{M}_0\tau_r = \frac{1}{2}r^2. \tag{8}$$

The uniformity of the distribution $x(\tau)$ in the one-dimensional case means that a particle beginning to move from the midpoint of a segment hits each end of that segment (before hitting the other end) with a probability of 1/2.

§3. The Markov and Strong Markov Properties

We now require one additional property of a Wiener process x(t) initiated from an arbitrary point x. We note that the sum of any random vector with distribution μ and, independent of that vector, a Wiener process initiated from zero represents a Wiener process with initial distribution μ. On the other hand, the difference $x(s+t) - x(s)$, where s is fixed and t varies from 0 to $+\infty$, is, by definition, a Wiener process initiated from zero and independent of the behavior of x(t) prior to the time s. The value of x(s) is determined by the behavior of x(t) prior to the time s. Therefore, representing the random path $x(s+t)$, $0 \leq t < +\infty$, as the sum of $x(s+t) - x(s)$ and x(s), we find that, given any conditions A on the behavior of the Wiener process prior to the time $s > 0$, the process $y(t) \equiv x(s+t)$ is a Wiener process with initial distribution $\mu(\Gamma)$ $\mathbf{P}_x\{A, x(s) \in \Gamma\}$ [Markov property of the process x(t)]. In the wide sense the Markov property means independence of the future process $x(s+t)$ from the past $x(s-t)$ when the present x(s) is known. Random sequences x(0), x(1), x(2), ... having this property were first investigated by Markov in 1907. We observe that if there are no conditions A imposed, $\mu(\Gamma)$ reverts to the transition probability $P(s, x, \Gamma) = \mathbf{P}_x\{x(s) \in \Gamma\}$

For a broad class of probabilistic processes, including the Wiener process, the Markov property is preserved if the current process is regarded not only as x(s) with fixed s, but also as $x(\tau)$ for some random τ. For example, a Wiener process along a line beginning at a point $x > 0$ behaves exactly the same after first hitting

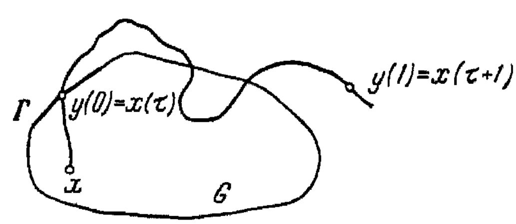

Fig. 7

zero as a process initiated from zero. This assertion, despite its apparent obviousness, is in need of proof, and it has indeed been proved. It is shown that in general if $x(t)$ is a Wiener process in a space of any dimensionality and τ is the time of first exit of $x(t)$ from an arbitrary domain (Fig. 7), then, given any conditions A on the behavior of $x(t)$ prior to the time τ, the process $y(t) \equiv x(\tau + t)$, $0 \leq t < +\infty$, is a Wiener process with initial distribution $\mu(\Gamma) = \mathbf{P}_x\{A, x(\tau) \in \Gamma\}$ (strong Markov property).* If there are no conditions A, the measure μ is simply the distribution of the vector $x(\tau)$.

The strong Markov property is not only fulfilled for the time of first exit, but for any random time τ in general, the occurrence or nonrecurrence of which is determined by the behavior of the process on the time interval $[0, \tau]$ (for example, for $\tau = \sigma + 1$, where σ is the time of first exit). We call all such random variables *Markov times*.

§4. Harmonicity of the Escape Probabilities

In §2 we investigated the time τ of first exit of the Wiener process from a circle, assuming that the motion of the particle is initiated at the center of the circle. If the initial point x does not coincide with the center of the circle, the symmetry of the random walk prior to the time τ is violated, and the circle problem is not much simpler than that of an arbitrary domain. We now turn our attention to this general case.

Let G be some domain on the plane and let $x(0) = x \in G$ (Fig. 8). At the time τ of first exit from G the particle is situated on the boundary L of the domain G. The probability that $x(\tau)$ will be in a definite set Γ of the boundary L is a function of the starting point x. We call this function $f(x)$.

We note that

$$f(x) = \mathbf{M}_x \varphi(x(\tau)), \tag{9}$$

*The systematic investigation of the strong Markov property was begun by E. B. Dynkin and A. A. Yushkevich (1956) and R. Blumenthal (1957). For the proof of this characteristic see, for example, [3] (Chapt. 5, §6).

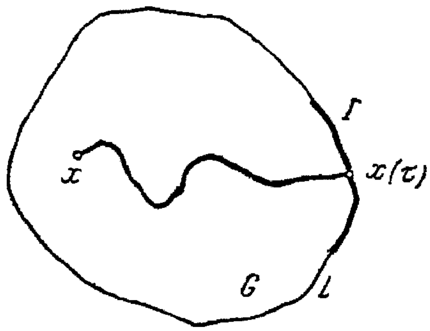

Fig. 8

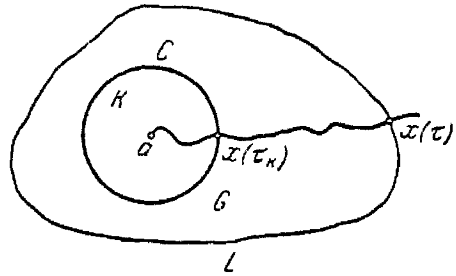

Fig. 9

where $\varphi(y)$ is a function defined on L, equal to one on Γ and equal to zero outside Γ. For the investigation of the function $f(x)$ the special character of the function φ is of no particular significance. It is sufficient to assume that the function φ is bounded (and measurable).

We will show in this section that the function $f(x)$ is harmonic in the domain G, i.e., that it satisfies the Laplace equation $\Delta f = 0$. We see below that, given rather broad assumptions, the function $f(x)$ becomes $\varphi(y)$ on the boundary of the domain, so that Eq. (9) defines the harmonic function inside the domain by means of its values on the boundary thereof (gives a solution to the Dirichlet problem).

We will show that for any circle K contained within the domain G the value of the function f in the center a of that circle is equal to the mean value of f along the circumference C bounding K:

$$f(a) = \int_C f(x)\mu(dx), \tag{10}$$

where μ is a unit measure, uniformly distributed over C (Fig. 9).

Thus, we let the motion of the particle commence at the point a. Then, before hitting L, the path must exit from the circle K. At the time τ_k of first exit from K, as we know, the particle is uniformly distributed over the circumference C. This implies, according to the strong Markov property, that the process x(t) beginning with the time τ_k may be regarded as a Wiener process with uniform initial distribution μ, "forgetting" about the behavior of x(t) prior to the time τ_k. Therefore, according to (5),

$$M_a\varphi(x(\tau)) = M_\mu\varphi(x(\tau)) = \int_C M_x\varphi(x(\tau))\mu(dx) ;$$

but this is none other than Eq. (10).

We now verify that every function having the property (10) (for any circle K) satisfies the Laplace equation.

Our task is to verify that the function $f(x)$ is differentiable in G any number of times.* We complete the definition of $f(x)$ by setting it equal to zero outside G We "smooth" the function $f(x)$, averaging it by means of a function g(x), which is infinitely differentiable, positive at the point 0, equal to zero outside an ε-neighborhood of zero, and invariant with respect to rotations:

$$\tilde{f}(x) = \int_R f(x+y)\, g(y)\, dy. \tag{11}$$

Making the change of variable $y = z - x$, we obtain

$$\tilde{f}(x) = \int_R f(z)\, g(z-x)\, dz.$$

The latter integral may be differentiated with respect to x any number of times; hence $\tilde{f}(x)$ is an infinitely differentiable function.

If the point $x \in G$ is located at a distance greater than ε from the boundary L of the domain G, the integral (11) is readily computed by going over to polar coordinates with pole at the point x. Integrating first over a circle of radius $\rho \leq \varepsilon$, one can take outside the integral the Jacobian of the transformation, ρ, and the weighting function g(y), which are constant on this circle; the remaining integral of $f(x+y)$ over the circle yields the mean value of the function f on this circle, correct to constants depending on ρ, and this value is equal to the value of f at the center of the circle, i.e., it is equal

*The measurability of $f(x)$ follows from the general properties of Markov processes (see the footnote on p. 42).

to $f(x)$. Then, taking $f(x)$ outside the outer integral over ρ, we find that $\mathcal{F}(x)$, correct to a constant factor, is equal to $f(x)$ for all points x which, together with their ε-neighborhoods, belong to G. Therefore, at these points the function $f(x)$ is differentiable any number of times. In view of the arbitrariness of ε, this fact is true for all points of the domain G.

We expand the function $f(x)$ in the neighborhood of an arbitrary point $a \in G$ into a Taylor series:

$$f(x) = f(a) + \frac{\partial f}{\partial x_1}(x_1 - a_1) + \frac{\partial f}{\partial x_2}(x_2 - a_2)$$

$$+ \frac{1}{2}\left[\frac{\partial^2 f}{\partial x_1^2}(x_1 - a_1)^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1 - a_1)(x_2 - a_2) + \frac{\partial^2 f}{\partial x_2^2}(x_2 - a_2)^2\right] + \alpha(x), \tag{12}$$

where the derivatives are taken at the point a with coordinates (a_1, a_2), and $|\alpha(x)| \leq k\rho^3$ $(\rho = |x - a|)$ for sufficiently small ρ. We integrate both sides of Eq. (12) along the circle C with center a over the unit uniform measure μ. If the radius ρ of the circle C is chosen such that it is entirely inside the domain G, then on the left side, according to (10), we obtain $f(a)$. On the right side the integrals of the terms $x_1 - a_1$, $x_2 - a_2$, and $(x_1 - a_1) \cdot (x_2 - a_2)$ revert to zero, insofar as symmetry causes the integrals over the upper and lower (or over the right and left) semicircles to cancel one another. Due to the invariance under rotation,

$$\int_C (x_1 - a_1)^2 \mu(dx) = \int_C (x_2 - a_2)^2 \mu(dx)$$

$$= \frac{1}{2}\int_C [(x_1 - a_1)^2 + (x_2 - a_2)^2]\,\mu(dx) = \frac{1}{2}\rho^2.$$

Thus, we find that for sufficiently small ρ

$$f(a) = f(a) + \frac{1}{4}\rho^2\left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2}\right) + \int_C \alpha(x)\,\mu\,dx,$$

whence

$$\left|\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2}\right| = \frac{4}{\rho^2}\left|\int_C \alpha(x)\,\mu(dx)\right| \leqslant \frac{4}{\rho^2}\cdot k\rho^3 = 4k\rho.$$

Letting ρ tend to zero and denoting the Laplace operator $\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}$ by the symbol Δ, we thus find that

$$\Delta f = 0$$

at the point a. This proves the harmonicity of the function $f(x)$.

§5. Regular and Irregular Boundary Points

We now analyze the behavior of the harmonic function

$$f(x) = \mathbf{M}_x \varphi(x(\tau)),$$

when x tends to some point a belonging to the boundary L of the domain G. We will assume here that the boundary function φ is continuous at the point a and bounded.

We note that if $x = a$, i.e., if the motion is initiated from the point a itself, then $\tau = 0$, $x(\tau) = a$, and $f(x) = \varphi(a)$.

We will demonstrate the fact that if the boundary L is "sufficiently nice" in the neighborhood of the point a, then

$$\lim_{\substack{x \in G \\ x \to a}} f(x) = \varphi(a). \tag{13}$$

This relation is derived from the obvious fact that a path beginning from a point x close to a has a high probability of meeting the boundary soon. During this time the particle does not succeed in advancing very far from the initial point x, hence it reaches the boundary L near the point a. However, this obvious picture does not always conform to reality. For example, if G represents a unit circular domain with deleted center a (Fig. 10), then, no matter how near a the path begins, it will exit with probability one from G at a boundary point different from a (this will be shown in §7). Consequently, here τ does not always tend to zero as $x \to a$. These considerations lead to the following definition. A point a belonging to the boundary is called r e g u l a r if for any $h > 0$

$$\mathbf{P}_x\{\tau > h\} \to 0 \quad \text{as} \quad x \to a \tag{14}$$

(τ converges in probability to zero as $x \to a$).

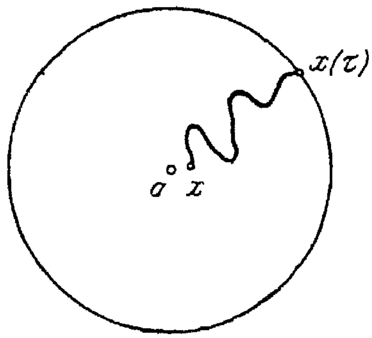

Fig. 10

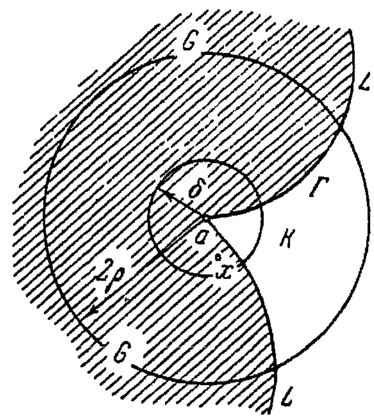

Fig. 11

We will prove that the boundary condition (13) is satisfied at any regular point.

Let Γ be the part of the boundary L contained inside a circle K of radius 2ρ with center at a regular boundary point a (Fig. 11). We will show that for any $\rho > 0$

$$\lim_{\substack{x \to a \\ x \in G}} \mathbf{P}_x \{x(\tau) \in \Gamma\} = 1 \qquad (15)$$

[the point $x(\tau)$ converges in probability to a].

We point out that the maximum displacement of a Brownian particle in time h

$$z(h) = \max_{0 \leqslant t \leqslant h} |x(t) - x(0)|$$

does not depend on the initial point $x = x(0)$. For every fixed path z(h) becomes smaller than the number ρ when $h \downarrow 0$. This means that the event $\{z(h) < \rho\}$ gives a sure event in the limit as $h \downarrow 0$. Therefore,

$$\lim_{h \downarrow 0} \mathbf{P}\{z(h) < \rho\} = 1 \qquad (16)$$

[z(h) converges in probability to zero].

Specifying an arbitrary number $\varepsilon > 0$, we choose $h > 0$ such that

$$\mathbf{P}\{z(h) < \rho\} > 1 - \varepsilon. \tag{17}$$

For this h, according to (14), there exists a $\delta > 0$ such that for $|x - a| < \delta$

$$\mathbf{P}_x\{\tau < h\} > 1 - \varepsilon. \tag{18}$$

Clearly, it may be assumed that $\delta < \rho$.

If simultaneously $z(h) < \rho$ and $\tau < h$ for $|x - a| < \delta$, $x \in G$, the path manages to hit L in the period h without deviating from x by more than ρ, i.e., it hits L before actually exiting from the circle K. Consequently, $x(\tau) \in \Gamma$ and therefore

$$\mathbf{P}_x\{x(\tau) \in \Gamma\} \geqslant \mathbf{P}_x\{z(h) < \rho,\ \tau < h\}.$$

From this inequality and the estimates (17) and (18) we obtain

$$\mathbf{P}_x\{x(\tau) \in \Gamma\} > 1 - 2\varepsilon$$

for $|x - a| < \delta$, $x \in G$. This proves Eq. (15).

It is now an altogether simple task to verify that the condition (13) is satisfied at a regular boundary point a for any bounded function $\varphi(y)$ continuous in a. Let μ_x be the distribution of a random point $x(\tau)$ under the condition $x(0) = x \in G$. Then

$$f(x) - \varphi(a) = \int_L \varphi(y)\mu_x(dy) - \varphi(a) = \int_\Gamma [\varphi(y) - \varphi(a)]\,\mu_x(dy)$$

$$+ \int_{L \setminus \Gamma} \varphi(y)\mu_x(dy) - \varphi(a)[1 - \mu_x(\Gamma)],$$

where, as before, Γ is the part of the boundary L in a circle K with center at the point a. Inasmuch as the function $\varphi(y)$ is continuous at the point a, the circle K may be chosen such that $\varphi(y)$ differs

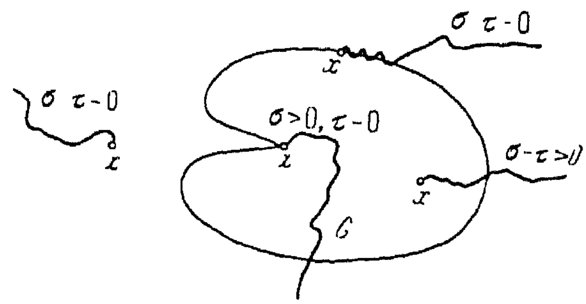

Fig. 12

from $\varphi(a)$ therein by less than an arbitrary number $\varepsilon > 0$. We then obtain

$$|f(x)-\varphi(a)| < \varepsilon\mu_x(\Gamma) + \lambda\mu_x(L \setminus \Gamma) + \lambda[1-\mu_x(\Gamma)],$$

where λ is a number bounding $|\varphi(y)|$. Since $\mu_x(\Gamma) \to 1$ as $x \to a$ and $\mu_x(L \setminus \Gamma) \to 0$, it follows that for x sufficiently close to a the right side of the inequality is less than 2ε. This means that $f(x)$ tends to $\varphi(a)$ as $x \to a$, $x \in G$.

Paths beginning from an infinite number of distinct x contribute to the condition (14) defining the regularity of the boundary point a, and the immediate verification of this condition is difficult. The recognition of regular boundary points is facilitated by a regularity criterion that can be formulated in terms of paths beginning from the point a itself. For this, in place of the time τ of first exit from G, which is identically equal to zero for $x=a$, we need to consider the first of the *positive definite* times at which the path is situated outside G (the time σ of *first exit from G after zero*). If none of the times $t>0$ for which $x(t) \notin G$ is the first (as will be the case when the path falls outside G for arbitrarily small positive values of t), then σ is considered to be equal to zero. (The relations between the times τ and σ for various paths are shown in Fig. 12.)

It turns out that *the boundary point a is regular if*

$$\mathbf{P}_a\{\sigma = 0\} = 1. \tag{19}$$

To prove this criterion, we consider the event A_u: "In the time interval [u, h] the trajectory lies inside G" $(0 < u < h)$. Clearly, the events A_u contract with diminishing u, and their intersection is congruent with the event $\{\sigma > h\}$. Consequently, for any x of R

$$\mathbf{P}_x\{\sigma > h\} = \lim_{u \downarrow 0} \mathbf{P}_x\{A_u\},$$

where the function $\mathbf{P}_x\{A_u\}$ decreases monotonically for $u \downarrow 0$.

We will show that for fixed $u > 0$ the probability $\mathbf{P}_x\{A_u\}$ is continuous with respect to x. In fact, the event A_u depends only on the values of $x(t+u)$, $t \geq 0$, and reduces to the event $\{\tau > h - u\}$ for the process $y(t) \equiv x(u+t)$. By virtue of the Markov property, y(t) is a Wiener process with initial distribution $\mu(\Gamma) = P(u, x, \Gamma)$. Consequently,

$$\mathbf{P}_x\{A_u\} = \int_R \mathbf{P}_y\{\tau > h - u\}\, P(u, x, dy) = \int_R \mathbf{P}_y\{\tau > h - u\}\, p(u, x - y)\, dy$$

[see Eq. (4)]. Since the density $p(u, x - y)$ is continuous over x and the integral converges uniformly in any finite domain, the integral is also continuous over x.

Inasmuch as the probability $\mathbf{P}_x\{\sigma > h\}$, which depends on x, is the limit of a monotonically decreasing sequence of continuous functions $\mathbf{P}_x\{A_u\}$, for any point a and any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbf{P}_x\{\sigma > h\} \leqslant \mathbf{P}_a\{\sigma > h\} + \varepsilon \tag{20}$$

for $|x - a| < \delta$.

Thus, by virtue of the monotonic convergence of $\mathbf{P}_a\{A_u\}$ to $\mathbf{P}_a\{\sigma > h\}$, it is possible to choose a $u > 0$ such that

$$\mathbf{P}_a\{A_u\} \leqslant \mathbf{P}_a\{\sigma > h\} + \frac{\varepsilon}{2}.$$

Having decided upon this u, we find a $\sigma > 0$ for the continuous function $\mathbf{P}_x\{A_u\}$ such that

$$\mathbf{P}_x\{A_u\} \leqslant \mathbf{P}_a\{A_u\} + \frac{\varepsilon}{2}$$

for $|x-a|<\delta$. Since $\mathbf{P}_x\{\sigma>h\}\le\mathbf{P}_x\{A_u\}$, it follows that the inequality (20) is valid for $|x-a|<\delta$.

If the condition (19) is satisfied at the point a, then $\mathbf{P}_a\{\sigma>h\}=0$, and it follows from the inequality (20) that

$$\lim_{x\to a}\mathbf{P}_x\{\sigma>h\}=0.$$

Inasmuch as $\tau\le\sigma$,

$$\mathbf{P}_x\{\tau>h\}\leqslant\mathbf{P}_x\{\sigma>h\},$$

and, consequently, $\mathbf{P}_x\{\tau>h\}$ also tends to zero. This proves the regularity of the point a.

It can be shown that the converse is also true: If for any bounded function $\varphi(y)$ continuous at the point a the boundary condition (13) is fulfilled, the point a is regular, and $\sigma=0$ at this point with probability one (see the problems).

Based on the criterion (19), in the next section we derive a simple geometric criterion of the regularity of a boundary point.

§6. The Zero-One Law; a Sufficient Criterion of Regularity

We now show that the probability $\mathbf{P}_a\{\sigma>0\}$ cannot assume values other than zero or one (zero-one law).

This requires us to prove first of all that for any fixed $h>0$ the random point $x(h)$ does not depend on the occurrence or nonoccurence of the event $\sigma>0$. This is attributable to the fact that the occurrence of the event $\sigma>0$ is determined by the behavior of the path on an arbitrarily small time interval $[0, s]$. Since the paths of the process are continuous, for small s the point $x(s)$, even though its distribution depends on the event $\{\sigma>0\}$, will have a high probability of being near the initial point a. The displacement of the particle in the remaining time interval $[s, h]$, on the other hand, does not depend on the behavior of the process prior to the time s or, in particular, on which of the events $\{\sigma>0\}$ or $\{\sigma=0\}$ took place. The random vector $x(h)$, as the sum of the "almost constant" vector x(s) and the vector x(h) − x(s), which is independent of the event $\{\sigma > 0\}$, is also "almost independent" of the event $\{\sigma > 0\}$. Passing to the limit as $s \to 0$, we deduce the "total" independence of the point x(h) from the event $\{\sigma > 0\}$.

In order to translate this intuitive line of reasoning into precise mathematical terms, we observe that for an arbitrary domain Γ, by virtue of the Markov property,

$$\mathbf{P}_a\{\sigma > 0,\ x(h) \in \Gamma\} = \int_R P(h-s,\ y,\ \Gamma)\,\mu(dy), \tag{21}$$

where μ is the distribution of the particle at the time s, given the additional requirement $\sigma > 0$:

$$\mu(\Gamma') = \mathbf{P}_a\{\sigma > 0,\ x(s) \in \Gamma'\}.$$

It is required in the integral (21) to pass to the limit as $s \downarrow 0$. The integrand of (21) is bounded, and it is apparent from Eq. (4) that it is continuous over the set of variables y and s for $h > s$. The measure $\mu(\Gamma')$ does not exceed $P(s, a, \Gamma')$ and therefore tends to zero as $s \downarrow 0$ for the exterior of any circle K centered at a. This means that the same measure tends to the complete measure $\mu(R)$ for the circle K itself, i.e., to the number $\mathbf{P}_a\{\sigma > 0\}$. Choosing a sufficiently small circle K, and then the number $s > 0$, we partition the integral (21) into two integrals over the circle K and its complement $\overline{K}$. In the integral over K the integrand differs only by an arbitrarily small amount from the number $P(h, a, \Gamma)$, while the measure $\mu(K)$ is almost equal to $\mathbf{P}_a\{\sigma > 0\}$. The integral over $\overline{K}$ does not exceed $\mu(\overline{K})$ and is therefore close to zero. Consequently, the entire integral (21) differs by an arbitrarily small amount from the product $\mathbf{P}_a\{\sigma > 0\} \cdot P(h, a, \Gamma)$ and is therefore simply equal to this product. Thus,

$$\mathbf{P}_a\{\sigma > 0,\ x(h) \in \Gamma\} = \mathbf{P}_a\{\sigma > 0\} \cdot \mathbf{P}_a\{x(h) \in \Gamma\}, \tag{22}$$

i.e., the vector x(h) and the event $\{\sigma > 0\}$ are independent.

Moreover, it follows from the Markov principle of the independence of future and past events given the present that inasmuch as the "current" event x(h) does not depend on the "past event" $\{\sigma > 0\}$, neither does a "future" event depend on that event, i.e., any event determined during the process after the time h is independent of that event.

Formally, according to the Markov property, any event A for the process $y(t) = x(h+t)$, $t \geq 0$, has a probability

$$\int_R \mathbf{P}_y\{A\}\, P(h, a, dy),$$

while in combination with the event $\{\sigma > 0\}$, by virtue of (22), the probability is

$$\int_R \mathbf{P}_y\{A\}\, \mathbf{P}_a\{\sigma > 0\}\, P(h, a, dy) = \mathbf{P}_a\{\sigma > 0\} \cdot \int_R \mathbf{P}_y\{A\}\, P(h, a, dy).$$

Hence, the probabilities of the event $\{\sigma > 0\}$ and the event A for the process y(t) are multiplicative, and these events are independent.

Inasmuch as h may be chosen arbitrarily small, it is possible to construct the event $\{\sigma > 0\}$ itself from among the "future" events, which are independent of $\{\sigma > 0\}$, and to deduce that this event is independent of itself, i.e., to arrive at the zero-one law.

Specifically, we consider the event A(h, t) = {The path is situated in the domain G during the time interval [h, t]}. By what has been demonstrated so far, for $h > 0$ the events A(h, t) and $\{\sigma > 0\}$ are independent, i.e.,

$$\mathbf{P}_a\{\sigma > 0,\ A(h, t)\} = \mathbf{P}_a\{\sigma > 0\} \cdot \mathbf{P}_a\{A(h, t)\}. \tag{23}$$

Clearly, as h is diminished, the events A(h, t) contract, and their limit for $h \downarrow 0$ becomes the event {The path is situated in the domain G during any time interval [h, t], where $0 < h < t$}, which is equivalent to the event $\{\sigma > t\}$. Hence, letting h tend to zero in Eq. (23) and recognizing the fact that $\{\sigma > 0, \sigma > t\} = \{\sigma > t\}$, we obtain

$$\mathbf{P}_a\{\sigma > t\} = \mathbf{P}_a\{\sigma > 0\} \cdot \mathbf{P}_a\{\sigma > t\}.$$

Now, letting t tend to zero, we find

$$\mathbf{P}_a\{\sigma > 0\} = [\mathbf{P}_a\{\sigma > 0\}]^2.$$

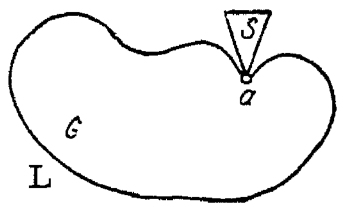

Fig. 13

Hence it is evident that $\mathbf{P}_a\{\sigma > 0\}$ can only be equated either to zero or to one.

We observe that the zero-one law just proved for the event $\{\sigma > 0\}$ is a special case of the following result, set forth by Blumenthal*: If the occurrence or nonoccurrence of the event A is determined in the course of a Wiener process on an arbitrarily small time interval $[0, t]$ $(t > 0)$, then the probability of A is either zero or one.

The proof of this law in its general formulation follows the same plan of attack as for the event $\{\sigma > 0\}$. It is first established that any event A_h depending on the behavior of the Wiener particle only after the time h does not depend on A. Then, approximating A by events of the type A_h (now it is mandatory to consider the measurability aspects of the events), it is deduced that $\mathbf{P}_x\{A\} = [\mathbf{P}_x\{A\}]^2$ and therefore that $\mathbf{P}_x\{A\} = 0$ or 1.

Let a be an irregular point. It follows from the regularity criterion set forth in the preceding section that the probability $\mathbf{P}_a\{\sigma > 0\} > 0$. By virtue of the zero-one law, in fact, at this kind of point

$$\mathbf{P}_a\{\sigma > 0\} = 1.$$

This allows us to deduce the following sufficient criterion of regularity: *A point a of the boundary L of a domain G is regular if it is the vertex of some triangle S which lies outside the domain* (Fig. 13). Thus, should the point a be irregular, the probability would be one that the particle, on leaving the point a, would lie during some positive time interval $(0, \sigma)$ inside G, and, therefore, certainly outside the triangle S. Due to the invariance of a Wiener process with respect to rotations, this result would be equally true for any triangle obtained by the rotation of S about the point a. It is implied by our assumption, therefore, that a particle at a will with probability one appear instantaneously outside the neighborhood of a. This contradicts the continuity of the Wiener trajectory. Hence, the point a is in fact regular.

* See, e.g., [3], Chapt. 5, §6.

The criterion just demonstrated implies regularity of the boundary points for a very broad class of domains, in particular, for any domains bounded by smooth curves.

For the regularity of a boundary point in a three-dimensional space it is sufficient if the point is the vertex of some tetrahedron outside the domain or, in the one-dimensional case, is the end point of a line segment.

§7. The Dirichlet Problem

We are now in a position to draw certain conclusions. We have shown that *if G is a domain with a regular boundary and if φ is any continuous bounded function on the boundary, then the formula*

$$f(x) = M_x \varphi(x(\tau)) \tag{24}$$

defines in the domain G a harmonic function, which assumes the values of φ on the boundary.

Consequently, not only have we proved the existence of a solution to the Dirichlet problem for a very broad class of domains, we have also obtained an explicit relation giving that solution. This relation can be used either for the qualitative analysis of the solution or for numerical calculations. The Wiener process in this case is modeled by means of random-number tables, then the mean value of the function φ at a random point of exit on the boundary is calculated. One of the techniques for the modeling of Brownian motion consists in simulating it by a random walk on a lattice. Of course, other models are equally possible. The advantage of the method developed here is the fact that it permits us to treat the solution to the Dirichlet problem for discontinuous boundary functions and domains with irregular boundaries.

On the other hand, Eq. (24) may also be used, of course, for the analytical investigation of escape probabilities. In a number of cases the solution to the Dirichlet problem is found directly, thus making it possible to obtain valuable information regarding the behavior of the paths of a Wiener process.

As we know, in a bounded domain G the Dirichlet problem cannot have two distinct solutions (this ensues from the fact that a

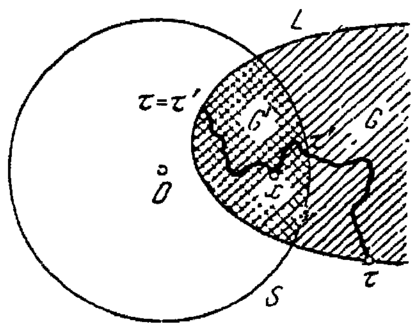

Fig. 14

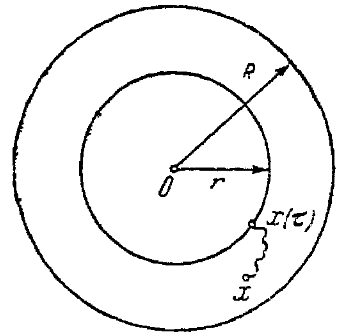

Fig. 15

harmonic function attains its maximum and minimum on the boundary of the domain). For an unbounded domain G the Dirichlet problem does not in general have a unique solution (for $l \geq 3$ not even a unique bounded solution). It is not clear at the outset which of these solutions coincides with $f(x)$. It can be shown that if the continuous boundary function $\varphi(y)$ is nonnegative and the boundary is regular, Eq. (24) gives the minimum nonnegative solution of the Dirichlet problem.

Thus, let S be a circle with center at the origin, and let G' be the part of the domain G lying inside S (Fig. 14).* Inasmuch as any point of the circle S can be brought into contact outside G' with the vertex of a triangle, the boundary of the domain G' is also regular. We denote by τ' the time of first exit of the Wiener path from G' and examine on the boundary of G' the function φ', which is equal to φ wherever the boundaries of the domains G' and G coincide, and is equal to zero at all other boundary points of the domain G'. We set

$$f'(x) = M_x \varphi'(x(\tau')) \quad (x \in G'). \tag{25}$$

Let g(x) be a nonnegative harmonic function in the domain G, which becomes $\varphi(y)$ on the boundary. By virtue of the uniqueness of the solution to the Dirichlet problem for the bounded domain G',

$$g(x) = M_x g(x(\tau')) \qquad (x \in G').$$

* Generally speaking, G' can consist of several unconnected components. This does not affect our discussion in any way.

Since $\varphi'(y) \leq g(y)$ everywhere on the boundary of the domain G', it follows from the above equations that

$$f'(x) \leqslant g(x) \qquad (x \in G'). \tag{26}$$

We now note that if the path x(t) exits from the domain G before exiting from the circle S, then for it $\tau' = \tau < \infty$, and the values of the functions φ' and φ at the point $x(\tau')$ are equal. For all other points of the path $\varphi'(x(\tau'))$ is either equal to zero (if $\tau' < \infty$, $\tau' \neq \tau$) or is undefined (if $\tau' = \infty$). Therefore, in place of Eq. (25) we write

$$f'(x) = M'_x \varphi(x(\tau)),$$

where the prime on **M** indicates that the integration is not taken over the entire path for which $\tau < \infty$, but only over the set of those paths for which $\tau' = \tau < \infty$. Clearly, the event $\{\tau' = \tau < \infty\}$, given an infinite dilation of the circle S, expands and goes over to the event $\{\tau < \infty\}$. Hence it follows that $M'_x \varphi(x(\tau))$ tends for any $x \in G$ to $\mathbf{M}_x \varphi(x(\tau))$ This means that on passing to the limit in the inequality (26), we arrive at $f(x) \leq g(x)$ over the entire domain G

We now consider some examples in which Eq. (24) is used either for the investigation of a Wiener process by means of differential equations or in order to obtain a solution to the Dirichlet problem from probabilistic considerations.

We begin with the observation that the function $\ln|x|$ satisfies the Laplace equation everywhere except the point $x = 0$. This function is constant on any circle with center at the point 0. It is a simple matter, therefore, to pick constants c_1 and c_2 such that the function

$$f(x) = c_1 \ln|x| + c_2$$

is equal to zero on the outer boundary of the ring $G = \{r < |x| < R\}$ and is equal to one on its inner boundary (Fig. 15). Denoting by τ the time of first exit of the path from the ring G, we obtain according to Eq. (24)

$$f(x) = \mathbf{M}_x f(x(\tau)) = 1 \cdot \mathbf{P}_x\{|x(\tau)| = r\} + 0 \cdot \mathbf{P}_x\{|x(\tau)| = R\}$$
$$(x \in G),$$

i.e., we find that $f(x)$ represents the escape probability from the ring G through the inner circle. Choosing c_1 and c_2 as required, we find that

$$f(x)=\frac{\ln R-\ln|x|}{\ln R-\ln r}. \tag{27}$$

The situation is analogous with regard to a space of $l\geq 3$ dimensions. Here, instead of the function $\ln|x|$ we require the function $1/|x|^{l-2}$. Then for the probability of escape from the spherical layer $\{r<|x|<R\}$ through its inner boundary we obtain the relation

$$f(x)=\frac{\frac{1}{|x|^{l-2}}-\frac{1}{R^{l-2}}}{\frac{1}{r^{l-2}}-\frac{1}{R^{l-2}}}. \tag{28}$$

A number of interesting consequences may be inferred from Eqs. (27) and (28).

Thus, it turns out that a *Wiener path on a plane or in a space has a probability one of never hitting a fixed point a different from the initial point of the path.**

We discuss the plane case for definiteness. It is clear that for any $R>|x|>r>0$ the probability of hitting 0 from the point x before hitting the outer boundary is less than or equal to the probability $f(x)$ of hitting the inner circle before the outer one. Equation (27) shows that the probability $f(x)$ will be arbitrarily near zero for fixed $R>|x|>0$ and sufficiently small $r>0$. Consequently, for any initial position $x\neq 0$ the probability of the particle hitting the point 0 before exiting from a circle of radius $R>|x|$ with center at 0 is equal to zero.

But if the particle, starting from the point $x\neq 0$, has a probability one of crossing any fixed circle of radius $R>|x|$ before hitting 0, then it has a probability one of crossing all circles of

*It is readily inferred from this result that it also has a probability one of never returning to the starting point.

radius $n|x|$ ($n=2, 3, \dots$) before hitting 0. However, it is impossible for the particle to cross all these circles in a finite period of time, because its path is continuous. Therefore, the particle, on emerging from the point $x \neq 0$, has a probability one of never hitting the point zero. This same result, of course, is also valid for any other fixed point a on the plane.

Using the terminology of Chapt. I, we say that a one-point set is nonrecurrent for a Wiener process on a plane or in a space. It is readily shown that in the one-dimensional case, conversely, every point is recurrent (see the problems). We recall by way of comparison that a one-point set for a random walk on a lattice is recurrent on a line and on a plane but is nonrecurrent in space.

This disparity between a continuous and a discrete random walk on a plane is attributable to the fact that one point on a continuous plane represents a much "sparser" set on a discrete lattice. The analogy between the discrete and continuous cases is restored if, instead of investigating arrival at a point, we consider arrival inside an arbitrary circle; a Wiener particle on a plane has a probability one of entering any circular domain of positive radius.

Thus, we return to the ring shown in Fig. 15. The probability, on starting from x, of sometime entering a circular domain of radius r is greater than or equal to the probability $f(x)$ of entering this circular domain before hitting the circle of radius R. According to Eq. (27), however, $f(x) \to 1$ as $R \to \infty$. Consequently, the probability of entering the inner circle is equal to one.

Drawing a denumerable number of circles on the plane such that any point of the plane falls inside a circle of arbitrarily small radius, we find that the path x(t) has a probability one of being everywhere dense on the plane.

In a space of $l \geq 3$ dimensions, according to Eq. (28),

$$\lim_{R \to +\infty} f(x) = \frac{r^{l-2}}{|x|^{l-2}}.$$

This limit is equal to the probability, on emerging from x, of sometime entering a sphere U of radius r. In fact, the event {The path enters the sphere U before its first exit from the sphere of radius

R} converges monotonically as $R \to \infty$ to the event {The path enters the sphere U}.

Knowing that the probability of the path entering any sphere is less than one, we may infer that $|x_t| \to \infty$ with probability one as $t \to \infty$ (the analogous result for the discrete case was obtained in Chapt. I, §3).

As another example, we proceed on probabilistic considerations to find the distribution of the position of a Brownian particle executing a random walk on a plane at the time of its first exit on a given line. This enables us to derive a relation giving a solution to the Dirichlet problem for a half-plane.

We note first of all that if the particle initiates its motion at a point situated on the bisector of a given angle (Fig. 16), then, by virtue of symmetry, it has an equal probability of exiting the first time from the angle through either side. Since the particle has a probability one of entering any circular domain K located outside the angle, the probability of escape from the angle is equal to one, hence the probability of first exit on either side of the angle is one half (the probability of hitting the vertex of the angle, as we know, is equal to zero).

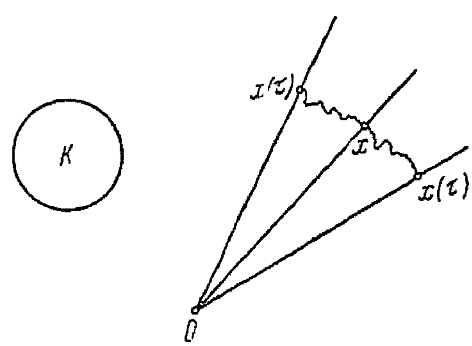

Fig. 16

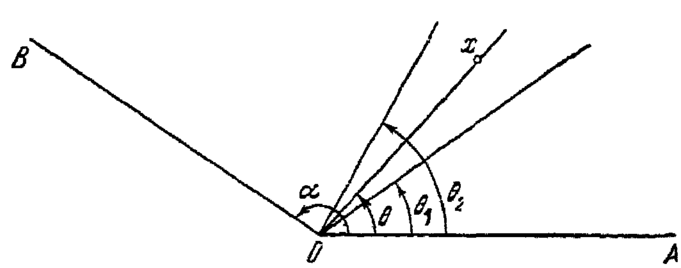

Fig. 17

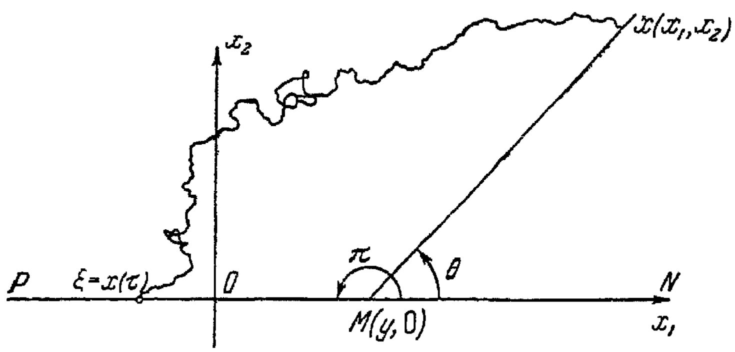

Fig. 18

Making use of this simple observation, we readily find the probability of the particle escaping on a particular side of the given angle before the other side, given any initial position of the particle. Let us consider some angle AOB equal to α $(0 < \alpha < 2\pi)$. Let x be an arbitrary point belonging to this angle, and let p(x) be the probability, on coming from the point x, of exiting from the angle AOB through the side OB (Fig. 17). We will prove that the probability p(x) is constant on every ray emanating from the point O and is equal to the ratio θ/α, where θ is the angle AOx. This assertion is obvious for θ equal to 0 or α. It follows from the above observation that if the formula $p(x) = \theta/\alpha$ is true for $\theta = \theta_1$ and $\theta = \theta_2$, then it is true also for $\theta = (\theta_1 + \theta_2)/2$; in fact, according to the total probability equation for points x with $\theta = (\theta_1 + \theta_2)/2$, we have $p(x) = \frac{1}{2}\frac{\theta_1}{\alpha} + \frac{1}{2}\frac{\theta_2}{\alpha} = \frac{\frac{1}{2}(\theta_1 + \theta_2)}{\alpha}$. Consequently, we establish in succession that the formula $p(x) = \theta/\alpha$ is valid for $\theta = \alpha/2$, then for $\theta = \alpha/4$, $\theta = 3\alpha/4$, ..., i.e., it is valid for any $\theta = k\alpha/2^n$ $(k = 0, 1, 2, \ldots, 2^n; n = 1, 2, 3, \ldots)$.

An arbitrary point x of the angle AOB can for any n include $\theta_1 = k\alpha/2^n$ and $\theta_2 = [(k+1)/2^n]\alpha$ between two rays. Designating the escape probabilities from a point x on each of these rays by q_1 and q_2 $(q_1 + q_2 = 1)$, we write

$$p(x) = q_1 \frac{k}{2^n} + q_2 \frac{k+1}{2^n} = \frac{k}{2^n} + \frac{q_2}{2^n} = \frac{\theta_1}{\alpha} + \frac{q_2}{2^n}.$$

For $n \to \infty$ we then deduce that

$$p(x) = \lim_{n\to\infty}\left(\frac{\theta_1}{\alpha} + \frac{q_2}{2^n}\right) = \lim_{n\to\infty}\frac{\theta_1}{\alpha} = \frac{\theta}{\alpha}.$$

We now consider the time τ of first exit of a Wiener path initiated in the upper half-plane onto the x_1 axis (Fig. 18). We denote the abscissa of the point $x(\tau)$ by the symbol ξ. For any fixed y the event $\xi < y$ means that the path exits from the unfolded angle NMP with vertex M(y, 0) on its side MP. Consequently,

$$P_x\{\xi < y\} = \frac{\theta}{\pi}, \tag{29}$$

where θ is the angle NMx. This angle is expressed in terms of the coordinates x_1, x_2 of the point x by the relation

$$\theta = \tan^{-1}\frac{x_2}{x_1 - y};$$

differentiating Eq. (29) with respect to y, we find the distribution density of the point ξ:

$$p(x_1,\ x_2;\ y) = \frac{1}{\pi}\,\frac{x_2}{x_2^2 + (y - x_1)^2}.$$

This is the Cauchy distribution density.

Knowing the density $p(x_1, x_2; y)$, we at once write an expression for a function $f(x)$, which is harmonic in the upper half-plane and assumes the given values of $\varphi(y)$ on the x_1 axis:

$$f(x) = \int_{-\infty}^{+\infty} \varphi(y)\,p(x_1,\ x_2;\ y)\,dy = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x_2\varphi(y)\,dy}{x_2^2 + (y - x_1)^2}. \tag{30}$$

Utilizing the fact that harmonic functions transform into harmonic functions in a conformal mapping, it is possible to obtain from Eq. (30) a solution to the Dirichlet problem for different domains on a plane and, in particular, to derive the Poisson formula, which gives the solution to the Dirichlet problem for a circle.

Assuming that the plane of x_1, x_2 is the plane of the complex variable $z = x_1 + ix_2$, we rewrite Eq. (30) in the form

$$f(z) = \frac{1}{\pi}\int_{-\infty}^{\infty} \operatorname{Im}\left(\frac{1}{y - z}\right)\varphi(y)\,dy \qquad (\operatorname{Im} z > 0).$$

The upper half-plane $x_2 > 0$ is mapped into a unit circle by means of the function $w = (z - i)/(z + i)$. Making this transformation, we deduce after certain computations that a function $u(w)$, which is harmonic in the circle $|w| < 1$ and assumes the values of $\varphi(y)$ on the boundary of this circle, is given by the formula

$$u(w) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - |w|^2}{|e^{i\theta} - w|^2} \varphi(e^{i\theta})\, d\theta$$

(Poisson integral in complex notation).

§8. Probabilistic Solution of the Poisson Equation

We turn our attention, finally, to the calculation of the mean time $m(x) = \mathbf{M}_x \tau$ to exit of the path from the domain G. We show that, given broad conditions, the function $m(x)$ represents in the domain G a solution of the Poisson equation

$$\Delta m = -2, \tag{31}$$

and it goes to zero on the boundary of the domain. In view of the analogy with the problem of the escape probabilities, we only briefly sketch the line of reasoning involved.

We surround the initial point a with a circular domain K of radius ρ, which lies entirely inside G (Fig. 9), and divide the time τ into two components: the time τ_K to first exit from K and the time $\tau - \tau_K$ that the path requires in order to reach the boundary L of the domain G from the circumference C. It is readily inferred from the strong Markov property that

$$\mathbf{M}_a(\tau - \tau_K) = \int_C m(x)\, \mu(dx),$$

where μ is the uniform distribution on the circumference C [cf. the derivation of Eq. (10) in §4]. We saw in §2 that $\mathbf{M}_a \tau_K = \rho^2/2$. Consequently,

$$m(a) = \frac{1}{2}\rho^2 + \int_C m(x)\, \mu(dx). \tag{32}$$

Let the domain G be bounded. Then it follows from the results of §2 that the function m(x) is finite. It is shown by direct verification that Eq. (32) satisfies the function $-x^2/2 = -(x_1^2 + x_2^2)/2$. Hence, the function $n(x) = m(x) - (x^2/2)$ satisfies the equation

$$n(x) = \int_C n(y)\mu(dy),$$

which coincides with Eq. (10) for the function $f(x) = M_x\varphi(x(\tau))$. Consequently, the function n(x) is harmonic in the domain G. Inasmuch as $\Delta(-x^2/2) = -2$, the mean time m(x) satisfies (31) inside G. For boundary points of the domain G, clearly, m = 0. It is to be expected, therefore, that m(x) tends to zero as x approaches the boundary. This statement is true for regular boundary points and is proved by the method developed in §5.

Thus, for a bounded domain G with a regular boundary function, m(x) is a solution of the Poisson equation (31) and goes to zero on the boundary. We know, of course, that this is a unique solution.

In the case of an unbounded domain G with a regular boundary L it can be shown, reiterating the discussion of §7 with an expanding circle, that m(x) is the minimum of the positive solutions of Eq. (31) in the domain G which go to zero on L, if such solutions exist.

§9. Infinitesimal and Characteristic Operators*

We have discovered an intimate relationship between the Wiener process x(t) and the Laplace operator $\Delta = \sum_{i=1}^{l} \frac{\partial^2}{\partial x_i^2}$. We now wish to delve a little deeper into the nature of this relationship.

We begin with some examples. Let a certain effect be described by the function x(t), where t varies in the interval (a, b). There exists an unpredictable set of different courses taken by this process. However, if we are concerned only with the increment of the function x(t) during a small time interval $(t_0, t_0 + \Delta t)$ and are agreeable to disregarding infinitesimally small quantities

* The reader is referred to [4] for a more detailed presentation of the problems discussed in this section.

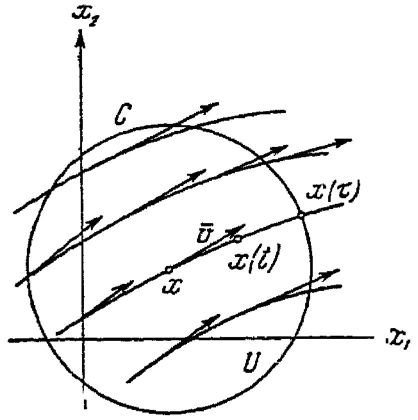

Fig. 19

of higher order (with respect to Δt), we can replace $x(t) - x(t_0)$ by the linear function $x'(t_0)(t - t_0)$, which is described by the single number $x'(t_0)$. This number is completely determined by the behavior of $x(t)$ in an arbitrarily small neighborhood of t_0 and is therefore an *infinitesimal characteristic* of our process at the time t_0. The consideration of fundamental importance here is the fact that the effect is reproducible in the large as long as these infinitesimal characteristics are known for all t.

We next consider a stationary (time-invariant) smooth flow of a fluid in a plane. The plane is filled with the paths of the fluid particles, only one path emanating from each point (Fig. 19). Knowing the position of a particle at a certain instant of time, we can give a good prediction of its future course, as though we knew the entire previous path of the particle (we have already encountered this property in studying the Wiener process, at which time we called it the Markov property). For the infinitesimal characteristic of our flow we have the flow velocity field (the vector field that results when the velocity vector at the given point is placed at each point of the plane). Sometimes, of course, it is more convenient, instead of the velocity field, to investigate its corresponding *operator* A, which is obtained as follows. Let $x(t)$ be the path of a particle situated at time $t = 0$ at the given point x. We trace the variation along the path not only of the coordinates $x_1(t)$ and $x_2(t)$, but also of an arbitrary (smooth) function $f(x) = f(x_1, x_2)$. After a time t the particle goes from x to a point $x(t)$, and the function $f(x(t))$ acquires the increment $f(x(t)) - f(x(0)) = f(x(t)) - f(x)$. Therefore, the rate of change $Af(x)$ of the function f along the path at the point x is equal to

$$\lim_{t\to 0}\frac{f(x(t))-f(x)}{t}=\frac{df(x(t))}{dt}=\frac{\partial f}{\partial x_1}\cdot\frac{dx_1}{dt}+\frac{\partial f}{\partial x_2}\cdot\frac{dx_2}{dt}=v_1\frac{\partial f}{\partial x_1}+v_2\frac{\partial f}{\partial x_2}, \tag{33}$$

where v_1 and v_2 are the projections of the velocity vector $\bar{v}$ on the x_1 and x_2 axes at the given point x, and the derivatives are taken at this point. The operator $A=v_1\frac{\partial}{\partial x_1}+v_2\frac{\partial}{\partial x_2}$ is called the infinitesimal operator of the process. Clearly, the assignment of A is equivalent to assignment of the velocity $\overline{v}$. Solving the equation

$$\frac{df}{dt}=Af$$

for various initial conditions, the time variation of any (smooth) function f, and, hence, of the overall flow in the large can be reproduced according to A.

Let us attempt to approach the Wiener process from this point of view. For a Wiener process the displacement $y(t) = x(t)-x(0) = x(t)-x$ of a particle during the time t is random and has (on the plane) a probability density

$$p(t, y)=\frac{1}{2\pi t}e^{-\frac{y_1^2+y_2^2}{2t}}$$

[see Eq. (3)]. Consequently, the ratio $[f(x(t))-f(x)]/t$ is also a random variable, and its limit, in general, has a probability one that it does not exit (see the problems). Consequently, the literal translation of the preceding arguments to a Wiener process, resulting in the concept of a random velocity and random operator A, is not meaningful. The situation is altered if the random variable $[f(x(t))-f(x)]/t$ is replaced by its mathematical expectation. Then we arrive at the infinitesimal operator of a Wiener process:*

$$Af(x_1, x_2)=\lim_{t\to 0}\frac{M_x f(x(t))-f(x)}{t}$$

$$=\lim_{t\to 0}\left[\frac{1}{2\pi t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x_1+y_1, x_2+y_2)e^{-\frac{y_1^2+y_2^2}{2t}}dy_1\,dy_2-f(x_1, x_2)\right]. \quad (34)$$

---

* In order to completely specify the infinitesimal operator A, it is necessary to indicate its domain of definition D_A. Let C denote a set of continuous bounded functions. Letting $f\in D_A$, if $f\in C$, $Af\in C$ and $[M_x f(x(t))-f(x)]/t$ converges to $Af(x)$ uniformly with respect to x, we obtain the so-called strong infinitesimal operator A. If all functions $f\in C$, for which $Af\in C$, are included in D_A, then if a limit exists at every point, and if the quantity $[M_x f(x(t))-f(x)]/t$ is bounded for all x and $t\to 0$, we obtain a weak infinitesimal operator A.

In many instances another limiting transition is more suitable, not with respect to time, but to space. We refer to the example of a stationary plane fluid flow. We assign a neighborhood U to the point x and consider the position of a particle at the time τ of its first exit from U (Fig. 19). This is a definite point $x(\tau)$ on the boundary of U. Consequently, during the time τ the function $f(x(t))$ acquires the increment $f(x(\tau)) - f(x)$. Shrinking the neighborhood U to the point x, we find that the rate of change of the function f at the point x may be defined as

$$\lim_{U \downarrow x} \frac{f(x(\tau)) - f(x)}{\tau},$$

and this limit clearly coincides with the limit found in Eq. (33). In the case of a Wiener process the time of first exit from U is already random. The position of the particle at the time τ is also random. Proceeding as in the formulation of the operator A and taking the mathematical expectation of the random variables, then shrinking U to x, we arrive at the operator

$$\mathfrak{A} f(x) = \lim_{U \downarrow x} \frac{M_x f(x(\tau)) - f(x)}{M_x \tau}, \tag{35}$$

which is called the *characteristic operator* of the process.* If U is a circle of radius ρ with center at the point x, then, as we know,

$$M_x f(x(\tau)) = \int_C f(y) \mu(dy), \tag{36}$$

where μ is the uniform distribution on the boundary C of the circular domain U, and

$$M_x \tau = \frac{1}{2} \rho^2.$$

---

*All functions f for which the limit (35) exists and is finite at every point x belong to the domain of definition $D_{\mathfrak{A}}$ of the characteristic operator $\mathfrak{A}$. Sometimes it is more convenient to shrink $D_{\mathfrak{A}}$, imposing the added requirements of continuity and boundedness on f and $\mathfrak{A} f$.

As is shown at the end of §4, the integral (36) is equal to

$$f(x)+\frac{\rho^2}{4}\left(\frac{\partial^2 f}{\partial x_1^2}+\frac{\partial^2 f}{\partial x_2^2}\right)+\int\limits_C \alpha(y)\mu(dy),$$

where $|\alpha| \leq k\rho^3$ (the function f is assumed to be sufficiently smooth). Substituting these values into Eq. (35), we obtain

$$\mathfrak{A}f(x)=\frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2}+\frac{\partial^2 f}{\partial x_2^2}\right)=\frac{1}{2}\Delta f(x). \tag{37}$$

A closer analysis shows that the limit (35) exists and is equal to $\frac{1}{2}\Delta f$ for twice continuously differentiable functions f and neighborhoods U of arbitrary configuration.

Thus, the characteristic operator describing the behavior of a Wiener process near a given point coincides, correct to a constant factor, with the Laplace operator.*

If we calculate the limit in Eq. (34), it turns out that it is also equal to $\frac{1}{2}\Delta f(x)$. This coincidence is not fortuitous. It has been demonstrated that it holds for a very broad class of Markov processes.

We have seen that the analysis of the probabilistic characteristics of a Wiener process is closely related to the investigation of the Laplace operator Δ. It is reasonable to inquire for which differential operators besides the Laplace operator it is possible to formulate an analogous theory. In other words, which differential operators are characteristic (infinitesimal) operators?

Let $x(t)=\{x_1(t), x_2(t)\}$ be a Wiener process on a plane, and let

$$\left.\begin{aligned} y_1(t)&=x_1(0)+c_{11}[x_1(t)-x_1(0)]+c_{12}[x_2(t)-x_2(0)]+b_1t,\\ y_2(t)&=x_2(0)+c_{21}[x_1(t)-x_1(0)]+c_{22}[x_2(t)-x_2(0)]+b_2t, \end{aligned}\right\} \tag{38}$$

*Here we investigate the operator $\mathfrak{A}$ only for twice continuously differentiable functions.

where c_{ij} and b_i are arbitrary real constants. Then it can be shown that the characteristic (infinitesimal) operator of the process $y(t) = \{y_1(t), y_2(t)\}$ coincides with the differential operator

$$L = \frac{1}{2}\left[a_{11}\frac{\partial^2}{\partial x_1^2} + 2a_{12}\frac{\partial^2}{\partial x_1\,\partial x_2} + a_{22}\frac{\partial^2}{\partial x_2^2}\right] + b_1\frac{\partial}{\partial x_1} + b_2\frac{\partial}{\partial x_2}, \tag{39}$$

where

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}\begin{pmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{pmatrix}. \tag{40}$$

We assume now that c_{ij} and b_i are smooth* functions of x. It is possible in this case also, using a more complex apparatus, to construct a Markov process for which the characteristic operator (on twice continuously differentiable functions) is given by Eq. (39). This type of process is called a *diffusion process* with generator L. The stationary fluid flow discussed at the beginning of this section may also be interpreted as a special case of the diffusion process (for $a_{ij} = 0$).

It is readily seen that the matrix $\{a_{ij}\}$ defined by Eq. (40) is symmetric and satisfies the condition

$$a_{11}\lambda_1^2 + 2a_{12}\lambda_1\lambda_2 + a_{22}\lambda_2^2 \geqslant 0 \quad \text{for any} \quad \lambda_1,\ \lambda_2.$$

If the stronger condition

$$a_{11}\lambda_1^2 + 2a_{12}\lambda_1\lambda_2 + a_{22}\lambda_2^2 > 0 \quad \text{for} \quad \lambda_1^2 + \lambda_2^2 > 0,$$

is satisfied, the operator L is called an elliptic operator. The probabilistic theory formulated in this chapter for the Laplace operator extends to arbitrary elliptic differential operators with sufficiently smooth coefficients.

PROBLEMS

Kolmogorov-Chapman Equation

1. Verify that the density p(t, y) given by Eq. (3) satisfies the equation

$$p(t+s,\ y) = \int_R p(t,\ x)\, p(s,\ y-x)\, dx$$

* Actually it is sufficient for the coefficients to be continuous in the Hölder sense.

for any $s, t > 0$, $y \in R$. Derive this equation on the basis of probabilistic considerations.

Hint. Represent the increment $x(t+s) - x(0)$ in the form $[x(s) - x(0)] + [x(t+s) - x(s)]$.

Escape Probabilities and Mean Exit Time in the One-Dimensional Case

In Problems 2-6 the variables $p(a; x)$ and $q(a, x)$ represent the probabilities that a Wiener particle commencing its motion at a point $x \in [0, a]$ will be situated at the time of first exit from the interval $(0, a)$ at its right or left end point, respectively, and $m(a; x)$ is the mean time required for this particle to escape from the interval $(0, a)$. The solutions of these problems do not rest on the general results of §§4-8.

2. On the graph of the function $p(a; x)$ points corresponding to the equidistant values $0 = x_1 < x_2 < \ldots < x_n = a$ lie on a single straight line.

Hint. See the discussion of Chapt. I, §1.

3. The function $p(a; x)$ is monotonic in x.

Hint. For $0 \leq x < y \leq a$ we have $p(a; x) = p(y; x) p(a; y)$.

4. $p(a; x) = x/a$, $q(a; x) = (a - x)/a$.

5. Knowing that $m(a; a/2) = c_1 (a/2)^2$ (see §2), derive the relation $m(a; x) = c_1 x(a - x)$.

Hint. Assuming that $x < a/2$, draw the path from the point $a/2$, and from the mean time to exit from $(0, a)$ deduct the mean time to exit from $(x, a - x)$.

6. The probability of the particle arriving from a point $x \neq 0$ at the point zero is equal to one, and the mean time to arrival at zero is infinite.

Hint. In Problems 4 and 5 pass to the limit as $a \to \infty$.

One-Dimensional Wiener Processes with Reflection and Absorption

If those parts of the paths of the Wiener process $x(t)$ on the line on which $x(t) < a$ are reflected symmetrically in the point a, leaving those parts of the paths with $x(t) \geq a$ unchanged, a Wiener process $y(t)$ with left reflection at the point a is obtained on the half-line $[a, +\infty)$. Right reflection at the point a is defined analogously. For $a < b$ a path initiated at a point $x \in [a, b]$ can be reflected first from the left at the point a, the resulting path reflected from the right at the point b, the path resulting therefrom reflected again from the left at the point a, then the new path again from the right at the point b, and so on ad infinitum. The end result is a *Wiener process* y(t) *on the interval* $[a, b]$ *with reflection at the points* a *and* b. If after the first visit of the Wiener path to the point a (at one of the points a or b) the particle always remains at that point, the result is a *Wiener process* z(t) *with absorption at the point* a (at the points a and b).

If a particle starting at x arrives after a time t > 0 in any interval Γ not containing absorbing boundary point with a probability P(t, x, Γ) equal to

$$P(t, x, \Gamma) = \int_{\Gamma} p(t, x, y)\, dy,$$

then the function p(t, x, y) is called the *transition density* of the process. For a Wiener process on the entire line the transition density is given by the formula

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} \tag{41}$$

(see §1).

7. The transition density of a Wiener process with left reflection at zero is equal to

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \left[e^{-\frac{(y-x)^2}{2t}} + e^{-\frac{(y+x)^2}{2t}} \right]$$

$$(x, y \geqslant 0).$$

Hint. For any interval $\Gamma \in [0, +\infty)$ we have $\{y(t) \in \Gamma\} = \{x(t) \in \Gamma \cup \Gamma'\}$, where Γ' is obtained from Γ by reflection at zero.

8. The transition density of a Wiener process with left re-

flection at zero and right reflection at a point a is equal to

$$p(t, x, y)=\frac{1}{\sqrt{2\pi t}}\sum_{n=-\infty}^{\infty}\left[e^{-\frac{(y-x-2na)^2}{2t}}+e^{-\frac{(y-x+2na)^2}{2t}}\right]$$

$$(x, y\in[0, a]).$$

Hint. Arguing as in the preceding problem, we obtain a representation for $\mathbf{P}_x\{y(t)\in\Gamma\}$ in the form of a series of integrals of the transition density (41). A change of variables permits the entire integration to be reduced to a single interval Γ, after which, utilizing the positiveness of all the integrands, the order of summation and integration may be reversed.

The following intuitively obvious statement is used in calculating the transition density for a Wiener process with absorption: If τ is the Markov time for which $x(\tau)=a$, and μ is the distribution of the random time τ, then for any $t>0$ and any interval Γ

$$\mathbf{P}_x\{\tau\leqslant t, x(t)\in\Gamma\}=\int_0^t P(t-s, a, \Gamma)\mu(ds). \tag{42}$$

This statement is inherent in the strong Markov property and can be accurately proved.

9. If a point x and an interval Γ are situated on one side of zero and τ is the time of first arrival at zero, then

$$\mathbf{P}_x\{\tau\leqslant t, x(t)\in\Gamma\}=P(t, -x, \Gamma).$$

Hint. For an initial state $-x$ the events $\{\tau\leqslant t, x(t)\in\Gamma\}$ and $\{x(t)\in\Gamma\}$ coincide Then make use of Eq. (42).

10. The transition density of the Wiener process $z(t)$ on the half-line $[0, +\infty)$ with absorption at zero is equal to

$$p(t, x, y)=\frac{1}{\sqrt{2\pi t}}\left[e^{-\frac{(y-x)^2}{2t}}-e^{-\frac{(y+x)^2}{2t}}\right]$$

$$(x, y>0).$$

Hint. In the notation of the preceding problem

$$P_x\{z(t)\in\Gamma\} = P_x\{\tau > t,\ x(t)\in\Gamma\}$$

(the interval Γ does not contain zero).

11. For an initial state $x > 0$ the time τ of first arrival at zero is distributed with a density

$$p(x,\ t) = \frac{x}{\sqrt{2\pi t^3}}\, e^{-x^2/2t} \qquad (t > 0). \tag{43}$$

Hint. In the notation of Problem 9

$$P_x\{\tau \leqslant t\} = P_x\{z(t) = 0\} = 1 - \int_0^\infty p(t,\ x,\ y)\,dy.$$

In the resulting interval make the substitution $y \pm x = \sqrt{2t}\, u$ and differentiate with respect to t.

The results of Problem 6 can be obtained by integration, using Eq. (43).

12. Let τ_0 be the time of first arrival of $x(t)$ at zero, σ_1 the time of first arrival after τ_0 at $a > 0$, τ_1 the time of first arrival after σ_1 at zero, σ_2 the time of first arrival after τ_1 at a, etc. For an initial state $x > 0$ the time τ_n is distributed the same as the time of first arrival from the point $-2na - x$ at zero, and the time σ_n is distributed the same as the time of first arrival from $2na + x$ at a.

Hint. All the differences $\sigma_{i+1} - \tau_i$ and $\tau_i - \sigma_i$ are mutually independent, and each of them has the same distribution as the time required for the particle to move one unit to the right (or left) at a. The random variable τ_0 does not depend on these differences and is distributed the same as the time required for the particle to move one unit to the right (or left) at x. It may be assumed, therefore, without upsetting the distribution of the sum $\tau_n = \tau_0 + (\sigma_1 - \tau_0) + (\tau_1 - \sigma_1) + \ldots + (\tau_n - \sigma_n)$ or the analogous sum for σ_n, that all the displacements are in the same direction.

13. If τ_0, σ_1, τ_1, σ_2, τ_2, ... is the sequence of random times from Problem 12, and ρ_0, π_1, ρ_1, π_2, ρ_2, ... is the analogous sequence of arrival times at the points a and 0, beginning with the

time of first arrival at a, then

$$\{\tau_n \leqslant t, \quad \rho_n \leqslant t\} = \{\sigma_{n+1} \leqslant t \ \text{ or } \ \pi_{n+1} \leqslant t\},$$
$$\{\sigma_n \leqslant t, \quad \pi_n \leqslant t\} = \{\tau_n \leqslant t \ \text{ or } \ \rho_n \leqslant t\}.$$

14. In the notation of Problems 12 and 13, for any event A

$$\mathbf{P}_x\{A, (\tau_0 \leqslant t \ \text{ or } \ \rho_0 \leqslant t)\} = \sum_{n=0}^{\infty} [\mathbf{P}_x\{A, \tau_n \leqslant t\} + \mathbf{P}_x\{A, \rho_n \leqslant t\}]$$
$$- \sum_{n=1}^{\infty} [\mathbf{P}_x\{A, \sigma_n \leqslant t\} + \mathbf{P}_x\{A, \pi_n \leqslant t\}].$$

Hint. Making use of the result of the preceding problem, apply the formula $\mathbf{P}\{B \cup C\} = \mathbf{P}\{B\} + \mathbf{P}\{C\} - \mathbf{P}\{B \cap C\}$ n times and let n tend to infinity. Passage to the limit is allowed because the continuity of the path permits only a finite number of transitions from 0 to a in a finite period of time, hence the probabilities $P_x\{\tau_n \le t\}$ and $P_x\{\sigma_n \le t\}$ tend to zero.

15. The transition density of the Wiener process z(t) with absorption at the points 0 and a is equal to

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \left[e^{-\frac{(y-x+2na)^2}{2t}} - e^{-\frac{(y+x+2na)^2}{2t}} \right] \tag{44}$$

$$(x, y \in (0, a)).$$

Hint. We have

$$\mathbf{P}_x\{z(t) \in \Gamma\} = \mathbf{P}_x\{x(t) \in \Gamma, \tau_0 > t, \rho_0 > t\}$$

(the interval Γ does not contain the points 0 and a). Now it is required to make successive use of Problem 14, Eq. (42), Problem 12, and the analogous result for the times ρ_n and π_n, as well as the hint to Problem 9 (the result formulated here carries over to the case when the initial point and the set Γ are delimited by the point a rather than by zero). The order of summation and integration may be changed in the expression for $\mathbf{P}_x\{z(t) \in \Gamma\}$ because the series (44) converges uniformly in y for any $t > 0$, $0 < x < a$.

16. For an initial state $x \in (0, a)$ the time τ of first exit from the interval $(0, a)$ is distributed with a density

$$p(x, t)=\frac{1}{\sqrt{2\pi t^3}}\sum_{n=-\infty}^{\infty}\Big[(x+2na)\,e^{-\frac{(x+2na)^2}{2t}}$$

$$+(2na+a-x)\,e^{-\frac{(2na+a-x)^2}{2t}}\Big]. \tag{45}$$

In particular,

$$p\left(\frac{a}{2}, t\right)=\frac{a}{\sqrt{2\pi t^3}}\sum_{k=0}^{\infty}(-1)^k(2k+1)\,e^{-\frac{(2k+1)^2a^2}{8t}} \tag{46}$$

$$(t>0).$$

Hint. Compare Problem 11. Term-by-term differentiation with respect to t is possible, inasmuch as the series (45) converges uniformly for $t \geq t_0 > 0$.

Calculation of the Constants c_l

It is possible with the aid of Eq. (46) to find analytically the mean time to exit of a particle from an interval and thus to determine the value of the constant c_1 (see the end of §2). Knowing c_1, it is a simple matter to calculate c_l for any l. However, it is impossible to calculate

$$\int_0^{\infty} t\,p\left(\frac{a}{2}, t\right)dt$$

directly by termwise integration, because each integral-summand diverges.

17. Using the well-known integral (see [19; Vol. II, p. 460)

$$\int_0^{\infty} e^{-\alpha x^2-\beta/x^2}\,dx=\frac{1}{2}\sqrt{\frac{\pi}{\alpha}}\,e^{-2\sqrt{\alpha\beta}} \qquad (\alpha, \beta>0),$$

calculate

$$\int_0^{\infty} te^{-\lambda t}\,p\left(\frac{a}{2}, t\right)dt \qquad (\lambda>0),$$

where the density p is given by Eq. (46).

<u>Answer</u>·

$$\frac{ae^{-a\sqrt{\lambda/2}}\left(1-e^{-a\sqrt{2\lambda}}\right)}{\sqrt{2\lambda}\left(1+e^{-a\sqrt{2\lambda}}\right)^2}.$$

18. The mean time to arrival of a Wiener path on a line from the point $a/2$ at either of the points 0 or a is equal to $a^2/4$, hence $c_1 = 1$.

19. The mean time to exit of a Wiener particle on a plane from the center of a circle of radius r on the circumference of the circle is equal to $r^2/2$, and hence $c_2 = 1/2$.

<u>Hint.</u> Let τ_1 be the time of first visit of x(t) to the circle $x_1^2 - x_2^2 = r^2$, and let τ_2 be the time of first visit of x(t) to one of the lines $x_1 = \pm r$.

Then

$$\mathbf{M}_0\tau_1 = \mathbf{M}_0\tau_2 - \mathbf{M}_0(\tau_2 - \tau_1) = \mathbf{M}_0\tau_2 - \mathbf{M}_\mu\tau_2,$$

where μ is a uniform distribution on the circle $x_1^2 + x_2^2 = r^2$. Inasmuch as the coordinate $x_1(t)$ represents a one-dimensional Wiener process, Problems 5 and 18 may be used for the calculation of $\mathbf{M}_x\tau_2$.

20. Generalizing the arguments of Problem 19 to the l-dimensional case, verify the fact that $c_l = 1/l$.

<u>Hint.</u> As in Problem 19, the problem reduces to one of averaging the quantity $x_1(2r - x_1)$ over an $(l-1)$-dimensional sphere. This average is readily calculated, for example, by means of the procedure followed at the end of §4.

<u>Nondifferentiability of the Wiener Path</u>

A discussion of the Wiener process on a line is sufficient.

21. Let there correspond to every t from the interval (0, T), $T > 0$, an interval Γ_t on the x axis. If

$$\mathbf{P}_0\{x(t) \in \Gamma_t\} \geqslant \varepsilon > 0 \quad \text{for} \quad 0 < t < T,$$

then for a path x(t) starting from zero the probability is one that there will exist positive times t arbitrarily close to zero such that

$$x(t) \in \Gamma_t.$$

Hint. Use the zero-one law.

22. The ratio

$$\frac{x(t)-x(0)}{t}$$

has a probability one of assuming all the real values in any interval $0<t<\varepsilon$ $(\varepsilon>0)$.

Hint. Apply the preceding problem to the intervals

$$\Gamma_t=(\sqrt{t},+\infty) \text{ and } \Gamma_t=(-\infty,-\sqrt{t}).$$

Necessary and Sufficient Condition for Regularity

We saw in §5 that the following equation was a sufficient condition for regularity of a boundary point a:

$$\mathbf{P}_a\{\sigma=0\}=1 \tag{47}$$

(σ is the time of first exit after zero from the domain G), and that the following provided a necessary condition:

$$\lim_{\substack{x\in G\\ x\to a}} \mathbf{M}_x\varphi(x(\tau))=\varphi(a) \tag{48}$$

for any bounded function φ continuous at the point a and specified on the boundary of the domain (τ is the time of first exit from G). It is proved in Problems 23-26 that these two conditions are necessary and sufficient. This requires the derivation of Eq. (48) from (47). All of the arguments are valid in a space of any dimensionality $l\geq 2$, but the case $l=2$ is considered for greater visualization.

23. A path x(t) starting at a point a has a probability one of never arriving at a for $t>0$.

Hint. Investigate first the time interval $[t_0, +\infty)$, where $t_0>0$, recognizing in the meantime that the probability of arriving from x sometime at a is equal to zero for $x\neq a$.

24. If $\mathbf{P}_a\{\sigma = 0\} = 0$, find a circular domain K of positive radius with center at the point a, such that

$$\mathbf{P}_a\{x(\sigma) \in K\} < \frac{1}{2}.$$

Hint. $\mathbf{P}_a\{x(\sigma) \in K\}$ tends to $\mathbf{P}_a\{x(\sigma) = a\}$ as the circle K is shrunk to the point a.

25. Given the conditions of the preceding problem, on any circumference C lying inside the circular domain K and circumscribing the point a there exists a point x for which

$$\mathbf{P}_x\{x(\tau) \in K\} < \frac{1}{2}.$$

Hint. If μ is the distribution at the time of first arrival from the point a on the circumference C, then

$$\mathbf{P}_a\{x(\sigma) \in K\} \geqslant \int_C \mathbf{P}_x\{x(\tau) \in K\}\, \mu(dx).$$

26. If $\mathbf{P}_a\{\sigma = 0\} < 1$, there exists a continuous bounded function φ for which Eq. (48) is violated.

Hint. By virtue of the zero-one law,

$$\mathbf{P}_a\{\sigma = 0\} = 0,$$

and it is sufficient to pick a function φ equal to zero outside the circle K, not greater than one inside K, and equal to one at the point a.

Strengthening of the Sufficient Criterion of Regularity

27. If a boundary point a of a plane domain G is the endpoint of a line segment outside the domain, then the point a is regular.

Hint. Use the zero-one law. If the probability of intersecting a half-line drawn from the point x(0) after an arbitrarily small positive time was equal to zero, then the probability of the path being situated on either side of the line $x_2 = x_2(0)$ after some positive time would be one.

28. If a boundary point a of a three-dimensional domain G is the vertex of a triangle outside the domain, then the point a is regular.

29. Find the fallacy in the following argument. The probability of the event $A_r = \{$The path x(t) on a plane enters a circular domain K_r of radius r with center at the point 0$\}$ is equal to one for any $r > 0$, and these events are embedded one into another. Passing to the limit as $r \downarrow 0$, we deduce that the path has a probability one of hitting the point 0.

Mean Time to Exit from a Domain

In Problems 30-33 m(x) denotes the time to exit from a point x to the outside of a plane domain G, and a is a regular boundary point of the domain G.

30. If the domain G is bounded, the function m(x) is bounded.

31. If the domain G is bounded, $m(x) \to 0$ as $x \to a$.

Hint. $\mathbf{P}_x\{\tau > \varepsilon\} \to 0$ as $x \to a$, and

$$m(x) \leqslant \varepsilon + \mathbf{P}_x\{\tau > \varepsilon\} \cdot \sup_x m(x)$$

(τ is the time of first exit from the domain G).

32. The mean time to exit from a circular domain K is equal to half the product of the maximum and minimum distances from the initial point $x \in K$ to the circumference of the circle (cf. Problem 5).

33. If $m(x) = \infty$ at one point at least, then $m(x) = \infty$ over the entire domain G.

Hint. It follows from the Poisson integral (see §7) that for any circle $C \subset G$ and any points x and y situated inside C there exists a positive constant c such that

$$\mu_y(\Gamma) = \mathbf{P}_y\{x(\tau) \in \Gamma\} > c\mathbf{P}_x\{x(\tau) \in \Gamma\} = c\mu_x(\Gamma),$$

where τ is the time of first arrival of the path on C, and Γ is any arc of this circle. Therefore,

$$m(y) = \mathbf{M}_y\tau + \int_C m(z)\mu_y(dz) > c\int_C m(z)\mu_x(dz) = c\,[m(x) - \mathbf{M}_x\tau],$$

where $\mathbf{M}_x\tau < \infty$.

Chapter III

The Optimal Stopping Problem

§1. The Problem of Optimal Choice

We start with the following problem. Suppose that we scan n objects in random succession and that from these objects we wish to choose the best one. After an examination of each object in turn, we must either accept or reject that particular one; it is inadmissible to return to an object previously rejected.

The latter condition, of course, is not always a realistic limitation. It is realistic, for example, if we are concerned with an automobile tourist who wishes to stop over in the most comfortable or the least expensive hotel along his route but has no intention of backtracking (assuming that the traveler is apprised of the number of hotels, but knows nothing of their quality). Or consider the astute bride-to-be who wishes to make an unerring choice of the best of all the suitors proposing marriage to her. With this second interpretation our postulate of being unable to return to a previously rejected object is adequately justified On the other hand, the stipulation that the decision-maker has prior knowledge of the total number of objects n appears rather artificial in this case.

We now make a more precise statement of the problem. There exist n objects, ordered in some definite manner according to their quality. We might think of these objects as represented, for example, by points on a line, where points further to the right correspond to "better" objects. We denote by a_1 the first object we come to. Inasmuch as the objects are inspected in random sequence, the probabilities of any of the existing n points turning out to be the point a_1 are identical. From precisely the same point of view the point a_2 has equal probability of being any of the remaining n − 1 points. Continuing to number the objects in the order in which we meet them, we ultimately arrive at a certain set $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$, where any of the n! conceivable permutations appears with equal probability. This permutation becomes gradually known, after the second test we know only the relative position of a_1 and a_2, whereas after the kth test we know the relative position of $a_1, a_2, \ldots, a_k$ (the reader might think of indicator lights flashing one after the other at the points $a_1, a_2, \ldots, a_k, \ldots, a_n$). The problem is to discriminate the rightmost of all the n points at the instant it first appears. It is required to indicate the method by which this result is achieved with maximum probability.

For a better understanding of the problem we consider some elementary selection techniques. We could, for example, decide on the first point a_1. Clearly, the probability of guessing the rightmost point in this case is equal to 1/n (and thus tends to zero as $n \to \infty$). The same result is obtained if we decide on a_2 or on a_3, etc.

It might seem at first glance that the probability of success in any system of selection would tend to zero as $n \to \infty$. But this is not the case. Let us suppose for simplicity that the number of points n is even. Let us assume that we pass over the first n/2 points, then choose the first point that falls to the right of all the preceding ones. Following this strategy, we are certain to achieve our goal if the best object happens to lie in the second half of the sequence $a_1, \ldots, a_n$ and the second best object lies in the first half. The probability of the two best objects being so arranged is equal to $[(n/2)/n] \cdot [(n/2)/(n-1)] > 1/4$. Hence, no matter how large n even is made, there exists a strategy for which the probability of success is greater than 1/4.

Let the allocation of the points $a_1, \ldots, a_k$ on the line be already known (see Fig. 20, where k = 4). We wish to determine the

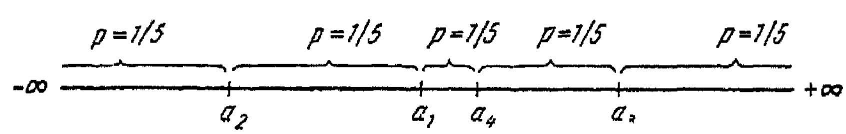

Fig. 20

probability of the next point a_{k+1} falling in each of the k+1 intervals partitioned off on the line by the points $a_1, \ldots, a_k$. The occurrence of a_{k+1} in a fixed interval corresponds to a definite permutation of the k-1 points $a_1, \ldots, a_k, a_{k+1}$ Inasmuch as all the points are equally probable, the probability of any such permutation is the reciprocal of the number of all permutations of k-1 elements and is equal to $1/(k+1)!$. Similarly, the probability of the permutation of the points $a_1, \ldots, a_k$ corresponding to their given position on the line is equal to $1/k!$. Consequently, the conditional probability of the point a_{k-1} falling in any of the k+1 intervals, given the condition that the relative position of the points $a_1, \ldots, a_k$ is known, is equal to $[1/(k+1)!]/(1/k!) = 1/(k+1)$, no matter how the points $a_1, \ldots, a_k$ are arranged. Thus, *the next point observed has equal probability of occurring in any of the intervals into which the line is divided by the existing points, regardless of the order in which these points have appeared.*

If the next point a_k turns out to be to the left of some previously inspected point, it is clearly not the rightmost. Consequently, it is only necessary to make our choice from among the points a_k situated to the right of the previous points $a_1, \ldots, a_{k-1}$. We call these points *maximal points*. It is clear that the point a_1 is always maximal, just as the rightmost of all the points $a_1, \ldots, a_n$ is maximal. This sought-after point is the last maximal point to be counted.

When the next maximal point a_k occurs, it is necessary to make a decision, either to choose that point or to wait until later. At this time the relative position of the points $a_1, \ldots, a_k$, of which a_k is the rightmost, is known. Since it is only possible now to choose from among the points $a_k, a_{k+1}, \ldots, a_n$, the decision rests solely on the prediction regarding the relative position of the points $a_k, a_{k+1}, \ldots, a_n$. With the stipulation that the points $a_1, \ldots, a_k$ are known, nothing affects this decision other than the conditional probabilities of the various permutations of the points $a_k, a_{k+1}, \ldots, a_n$. We will show that the conditional probabilities actually depend only

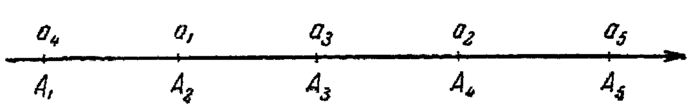

Fig. 21

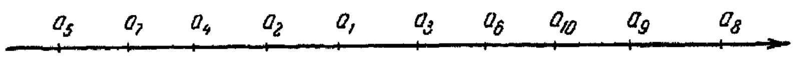

Fig. 22

on the index k and are not influenced in any way by the relative position of the points $a_1, \ldots, a_{k-1}$. In this way we establish the fact that *when a maximal point a_k makes its appearance, a particular decision must be made solely on the basis of the index k of that point* (with due regard, of course, for the number n of all the points).

The points $a_1, \ldots, a_n$ are numbered in the order of their occurrence on the line. We renumber the points $a_1, \ldots, a_k$ that have already occurred in the order of their position on the line from left to right: $A_1, \ldots, A_k$. Since the point a_k is maximal, a_k coincides with A_k (Fig. 21). Assignment of the relative arrangement of the points $a_1, \ldots, a_k$ is equivalent to assignment of the order of occurrence of the points $A_1, \ldots, A_k$. The fact that any permutation of the points $A_k, a_{k+1}, \ldots, a_n$ is independent of the order in which the points $A_1, \ldots, A_{k-1}$ occurred is established with the observation that the individual events, namely the relative position of the points $A_k, a_{k+1}, \ldots, a_n$, as well as their position relative to the points $A_1, \ldots, A_{k-1}$, are independent of the order of accession of the points $A_1, \ldots, A_{k-1}$. This is implied by the fact that each point in turn, as established earlier, has equal probability of falling within any of the intervals into which the line is divided by the preceding points. Specifically, the point a_{k+1} has a probability $1/(k+1)$ of occurring in any of the intervals $(-\infty, A_1)$, (A_1, A_2), ..., $(A_k, +\infty)$, regardless of the accession order of the points $A_1, \ldots, A_{k-1}$, the point a_{k+2} has a probability $1/(k+2)$ of occurring in any of the intervals delimited by the points $A_1, \ldots, A_k$ and a_{k+1}, regardless of the accession order of the points $A_1, \ldots, A_{k-1}$, etc. Multiplying these probabilities, we deduce that the probability of any permutation of the points $A_1, \ldots, A_k, a_{k+1}, \ldots, a_n$ (associated with the natural order of the points $A_1, \ldots, A_k$) is equal to

$$\frac{1}{k+1} \cdot \frac{1}{k+2} \cdots \frac{1}{n}$$

regardless of the accession order of the points $A_1, \ldots, A_{k-1}$. This proves our original assertion.

For example, let n = 10, and let the points $a_1, \ldots, a_{10}$ be arranged as in Fig. 22. Then the maximal points are a_1, a_3, a_6, and a_8. When the point a_1 occurs, it is necessary to make a decision with regard only for the fact that its index is equal to one; when the point a_3 occurs, the decision is based solely on the fact that its index is equal to three (providing, of course, that we have not stopped earlier); etc.

Thus, in order to make the optimal decision,* it is sufficient to analyze only the indices of the maximal points. We designate these indices in increasing order x(0), x(1), x(2), ... As already mentioned, x(0) = 1. The indices x(1), x(2), ... are random, just as the number of these indices is random. None of the indices exceeds n. The last (largest) of the indices x(i) is the index of the rightmost point a_k, and it must be guessed with the highest possible probability. The guessing procedure rests on the fact that when the next random variable x(i) comes up, it is required solely on the basis of its value either to declare this x(i) as the last one or to wait until later. (In particular, it is not required for the optimal choice to know what were the previous indices of the maximal points x(0), ..., x(i − 1) or how many of these indices there were.)

In order to translate the problem completely into the language of the sequence {x(i)}, we look further for the probabilistic law governing this random sequence. We show first of all that the random variables x(0), x(1), ... form a M a r k o v c h a i n . This means that the conditional probability of the event x(i + 1) = l, with the provision that the values of all the preceding random variables x(0), ..., x(i) are known, actually depends only on the value of k acquired by the immediately preceding variable x(i).† Thus, let it be known that x(0) = 1, x(1) = b, ..., x(i) = k. This is equivalent to saying that the maximal of the points $a_1, a_2, \ldots, a_k$ are the points a_1, $a_b, \ldots, a_k$. In other words, it is known that the point a_k is maximal, and something is known also about the relative position of the points $a_1, a_2, \ldots, a_{k-1}$. The event x(i + 1) = l now means that the points $a_{k+1}, \ldots, a_{l-1}$ are situated to the left of a_k and that the point a_l is to the right of it. Consequently, if it is known that x(0) = 1,

*Inasmuch as there only exists a finite number of selection strategies, there is certainly an optimum among them.

†More precisely, this is the definition of a homogeneous Markov chain. In the general inhomogeneous case the indicated conditional probability also depends on the time i. Inhomogeneous Markov chains are not discussed in the present book.

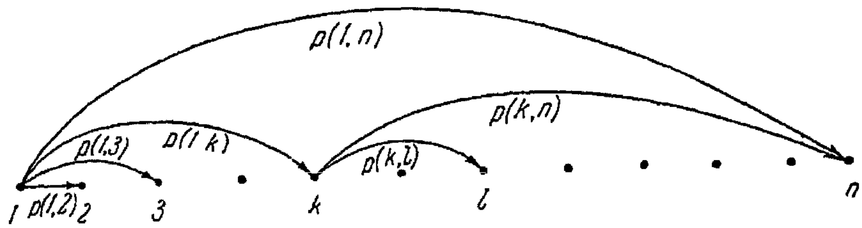

Fig. 23

x(1) = b, ..., x(i) = k, then the event x(i + 1) = l may be described in terms of the relative position of the points a_k, a_{k+1}, ..., a_l, ..., a_n. We found out above, however, that if the point a_k is maximal, the conditional probability of any event referred to the relative position of the points a_k, ..., a_n, given the condition that something is known regarding the relative position of the points a_1, a_2, ..., a_{k-1}, in fact depends only on the index k. Thus the conditional probability

$$P\{x(i+1)=l/x(0)=1,\ x(1)=b,\ \ldots,\ x(i)=k\},$$

apart from l, depends only on k (and possibly on the total number of points n). It is called the t r a n s i t i o n p r o b a b i l i t y of the Markov chain and is designated p(k, l).

The variables x(0), x(1), ... assume the values 1, 2, ..., n. This set of values (called the p h a s e s p a c e) is conveniently represented in the form of points along which a particle executes a random walk (Fig. 23). At the initial instant the particle is located at the point 1, then it jumps to the point j with probability p(1, j). In general, if the particle happens to be situated at the point k at some particular time, then in the next step it has a probability p(k, l) of transferring to the point l, regardless of how it arrived at the point k. In our case $\sum_l p(k, l)$ can turn out to be less than one. It is reasonable to interpret the difference $1 - \sum_l p(k, l)$ as the extinction probability of the particle. The transition of the particle from k to l means that the next maximal point after the maximal point a_k will be the maximal point a_l. Extinction of the particle implies that there are no more maximal points.

Let us calculate the transition probabilities $p(k, l)$. By definition of the conditional probability

$$p(k, l)=\frac{P\{x(i)=k, x(i+1)=l\}}{P\{x(i)=k\}} \quad (k, l=1, \ldots, n).$$

Clearly, $p(k, l)=0$ for $l \leq k$ (only left-to-right jumps are possible in Fig. 23). For $l > k$ the event $\{x(i)=k, x(i+1)=l\}$ implies that the points a_k and a_l (where a_l is to the right of a_k) are the furthest to the right of all the points $a_1, \ldots, a_l$. The probability of this event, considering the equal likelihood of all the points, is equal to $1/l(l-1)$. By complete analogy $\mathbf{P}\{x(i)=k\}=1/k$. Consequently,

$$p(k, l)=\frac{k}{l(l-1)} \quad (1 \leqslant k < l \leqslant n).$$

We proceed now to formulate the optimal choice procedure.

As already mentioned, this method may be obtained by indicating for each index k whether to stop with this number or to look further. It is clearly sufficient to specify the subset Γ of the indices at which it is required to stop. The set of numbers 1, 2, ..., n has 2^n subsets (including the empty subset and the total set). Each of these corresponds to a certain strategy, and it is our problem to decide upon the best of these 2^n strategies.

Of course, there are many other strategies in addition to the procedures listed above. For example, we denote by ξ the first of the values $x(0), x(1), x(2), \ldots$ greater than or equal to k [so that $\xi=x(i)$ for $x(0)<\ldots<x(i-1)<k$, $x(i) \geq k$]. We could conceive of a strategy such that it is prescribed to stop with the number following ξ, i.e., with $x(i+1)$. Strategies of this type are clearly nonoptimal, but we will use them in studying the best selection procedure.

Let us calculate the conditional probability $q(k)$ of payoff by stopping at the point $x(i)=k$:

$$q(k)=1-\sum_{l=k+1}^{n} p(k, l)=1-\sum_{l=k+1}^{n} \frac{k}{l(l-1)}$$

$$= 1 - k \sum_{l=k+1}^{n} \left(\frac{1}{l-1} - \frac{1}{l} \right) = \frac{k}{n} \quad (1 \leqslant k \leqslant n)$$

For comparison we find the conditional payoff probability $q'(k)$ if in the same situation exactly one more step is taken, i.e., if one stops with the number $x(i+1)$. According to the total probability formula

$$q'(k) = \sum_{l=k+1}^{n} p(k, l) q(l) = \sum_{l=k-1}^{n} \frac{k}{l(l-1)} \frac{l}{n} = \frac{k}{n} \left(\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{n-1} \right)$$

$$= q(k) \left(\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{n-1} \right) \qquad (k < n),$$

$$q'(k) = 0 \qquad (k = n).$$

Since the sum $1/k + \ldots + 1/(n-1)$ decreases monotonically with increasing k, the ratio $q'(k)/q(k)$ also decreases monotonically, going to zero for $k=n$. Consequently, the condition $q'(k) \leq q(k)$ is satisfied by some interval k_n, $k_n + 1$, ..., n of the series of numbers 1, 2, ..., n.

We will show that the set $\Gamma = k_n, \ldots, n$ *corresponds to the optimal strategy* (in other words, that *scanning must be continued as long as* $x(i) < k_n$, *and stopped the first time* $x(i) \geq k_n$).

We assume below that the number of objects $n \geq 3$. For $n = 1$ there is in general no choice, and for $n = 2$ there are equal chances of success in stopping with either of two objects. It is at once apparent that in both of these cases the set $\Gamma = \{k_n, \ldots, n\}$ leads to an optimal strategy, but the next argument is inapplicable to these cases, because $k_n = 1$ for $n = 1$ or 2.

For $n \geq 3$

$$q'(1) = q(1) \left(1 + \frac{1}{2} + \ldots + \frac{1}{n-1} \right) > q(1)$$

and, therefore, $k_n > 1$. This means that a strategy requiring us to stop with the number $1 = x(0)$ is nonoptimal; in fact, this type of strategy is successful with probability $q(1)$, whereas a strategy calling for the choice of the number $x(1)$ yields a payoff with a greater probability $q'(1)$.

Thus, we can only seek the optimal selection technique among those strategies which require going past the first index. Since $p(1, k) > 0$ for $2 \leq k \leq n$, we have a positive probability $p_A(k)$, applying such a strategy A, of waiting until any index k. Let us suppose that the strategy A prescribes us to stop with a number $k < k_n$. Then the strategy A', which coincides with A as long as the particle arrives at k but requires exactly one more step after arriving at k, is clearly better than A. In fact, with the strategy A' the probability of success will be greater than with the strategy A by an amount $p_A(k)[q'(k) - q(k)]$. Consequently, the optimal strategy excludes stopping at the points $1, \ldots, k_n - 1$.

We now show by induction from a larger to a smaller value of k that at points of the interval $\{k_n + 1, \ldots, n\}$ the optimal strategy A requires stopping at once. Clearly, at points of this interval we have the strict inequality $q'(k) < q(k)$. If the strategy A required passing over the number n, the strategy A' prescribing stopping at the point n and otherwise coinciding with A would increase the probability of success relative to A by an amount $p_A(n)$, and the strategy A would not be optimal. Hence, for $k = n$ our assertion is true. Suppose now that it has already been proved for the points $k+1, k+2, \ldots, n$ $(k \geq k_n + 1)$. If A had prescribed passing over the number k, then the strategy A' requiring that we stop at the point k and otherwise coinciding with A would have been better than A. Actually, on arriving at the point k, the strategy A' would in fact prescribe stopping at once, while the strategy A, according to the induction hypothesis, would prescribe stopping after precisely one more step. Therefore, the probability of success would be greater for A' than for A, by an amount $p_A(k)[q(k) - q'(k)]$, and the strategy A would not be the optimal one. Consequently, A also requires stopping at the point k.

We have established the fact that the optimal strategy A forbids stopping at the points $1, \ldots, k_n - 1$ and, conversely, requires stopping at the points $k_n + 1, \ldots, n$. If for $k = k_n$ we have the strict inequality $q'(k_n) < q(k_n)$, the induction process can be continued to $k = k_n$ and the fact verified that the strategy A also requires stopping at the point k_n. But if for some n it turns out that $q'(k_n) = q(k_n)$, then, applying the same arguments, it is immaterial how we arrive at the point k_n. For convenience in this case we attach the point k_n to the set Γ.*

Thus, the best method of selection consists in passing over the first $k_n - 1$ objects and then choosing the first object that is better than all the preceding ones.

The number k_n is the smallest integer for which $q'(k) \leq q(k)$, i.e., for which

$$\frac{1}{k}+\frac{1}{k+1}+\ldots+\frac{1}{n-1}\leqslant 1.$$

Therefore, k_n is found from the double inequality

$$\frac{1}{k_n}+\ldots+\frac{1}{n-1}\leqslant 1<\frac{1}{k_n-1}+\frac{1}{k_n}+\ldots+\frac{1}{n-1}. \qquad (1)$$

We now determine the probability of success using the optimal strategy. We first calculate the probability s_m that the first object coming after the first $k_n - 1$ rejected objects and better than all the preceding objects will have the index m. This event means that the rightmost of all the points $a_1, \ldots, a_m$ will be a_m and that the next one to it will be any of the points $a_1, \ldots, a_{k_n-1}$. By virtue of the equal weight of the objects, the probability of this event is $(1/m)\cdot(k_n - 1)/(m - 1)$. Consequently,

$$s_m=\frac{k_n-1}{m\,(m-1)}.$$

The conditional probability of success in this case is equal to $q(m) = m/n$. Hence, in the large, the probability of a correct decision

* Actually, the equation $q'(k) = q(k)$ holds only for $n = 2$ and $k = 1$. Thus, of the numbers $k, k+1, \ldots, n-1$, exactly one is divisible by the highest power of the number two not in excess of $n-1$, so that after reduction of the sum

$$s=\frac{1}{k}+\frac{1}{k+1}+\ldots+\frac{1}{n-1}$$

to a common denominator an odd number is obtained in the numerator. For $n > 2$ the denominator will be even, hence the sum s will be different from one.

is equal to

$$p_n = \sum_{m=k_n}^{n} s_m q(m) = \sum_{m=k_n}^{n} \frac{k_n - 1}{m(m-1)} \cdot \frac{m}{n}$$

$$= \frac{k_n - 1}{n}\left(\frac{1}{k_n - 1} + \frac{1}{k_n} + \ldots + \frac{1}{n-1}\right). \qquad (2)$$

For example, for n = 10 we have the following table:

TABLE 1

k	$\frac{1}{k}$	$\frac{1}{k}+\ldots+\frac{1}{n-1}$	k	$\frac{1}{k}$	$\frac{1}{k}+\ldots+\frac{1}{n-1}$
9	0.111	0.111	4	0.250	0.996
8	0.125	0.236	3	0.333	1.329
7	0.143	0.379	2	0.500	. . .
6	0.167	0.546	1	1.000	. . .
5	0.200	0.746			

It is apparent from this table that $k_n = 4$. Consequently, it is necessary first to reject three objects, then to choose the first object better than all the preceding ones. The probability of success in this case is

$$p_{10} = 0.3 \cdot 1.329 = 0.399.$$

Similar calculations are easily carried out for any n, as long as it is not too large. We now derive relations giving a better approximation for k_n and p_n for large values of n. For any $m \geq 2$ we have

$$\ln(m+1) - \ln m = \int_m^{m+1} \frac{dx}{x} < \frac{1}{m} < \int_{m-1}^{m} \frac{dx}{x} = \ln m - \ln(m-1).$$

Summing these inequalities from m = k to m = n − 1, we deduce that

$$\ln\frac{n}{k} < \frac{1}{k} + \frac{1}{k+1} + \ldots + \frac{1}{n-1} < \ln\frac{n-1}{k-1}.$$

It follows from these estimates and from the inequalities (1) that

$$\ln \frac{n}{k_n} < 1 < \ln \frac{n-1}{k_n - 2},$$

whence

$$\frac{n}{e} < k_n < \frac{n}{e} + \left(2 - \frac{1}{e}\right).$$

Inasmuch as no more than two integers can fall within an interval of length $2 - 1/e$, the above inequalities permit k_n to be found for any n with an error no greater than one. For large n an error of one in the calculation of k_n has little effect on the probability of correct choice.

It is evident from the inequalities (1) that the sum $1/(k_n - 1) + 1/k_n \ldots + 1/(n - 1)$ differs from unity by less than $1/(k_n - 1)$. Since $k_n \to \infty$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \left(\frac{1}{k_n - 1} + \frac{1}{k_n} + \ldots + \frac{1}{n-1}\right) = 1.$$

From Eq. (2), therefore, we find

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} \frac{k_n - 1}{n} = \frac{1}{e} \approx 0\ 368.$$

§2. Optimal Stopping Problem for a Markov Chain

In the preceding section we solved the optimal choice problem by the construction of a special Markov chain. We now investigate the general problem of the optimal stopping of an arbitrary Markov chain.

Let a certain particle (or system) exist at each instant of time in one of the states formed by a finite or denumerable set E (*phase space*). If the particle is found at some instant in the state x, then after a unit of time it is found in the state y with a probability p(x, y) (regardless of when and by what route it arrived at the point x). We say then that we have specified a *Markov chain with transition probabilities* p(x, y).

The probabilities p(x, y) may be any nonnegative numbers obeying the condition

$$\sum_y p(x, y) \leqslant 1 \qquad (x \in E).$$

If $\sum_y p(x, y) < 1$, for some x, then the variable $q(x) = 1 - \sum_y p(x, y)$ represents the extinction probability in the next step for a particle situated at x. An extinct particle cannot be recreated, hence the chain in this case is terminated once and for all.

Examples of Markov chains are the random walk on a lattice studied in Chapt. I and the sequence of indices of the maximal points in the optimal choice problem In the former example the chain sometimes fails to terminate, and in the latter the probability is one that it will terminate no later than the nth step.

We denote by x(n) the position of the particle at the instant n. Let us suppose that we observe the path x(0), x(1), ..., x(n), ..., and can at any instant n stop the migrating particle. If at the time of stopping the particle is situated at the point x, we acquire a payoff f(x), where f is a known function. If we do not stop the process (either because it succeeded in terminating itself or because we wait an infinitely long time), the payoff is zero. We wish to inquire how to optimize the payoff.

Let us refine the statement of the problem. We first of all describe the class of possible stopping times τ. The time τ, generally speaking, is random because it depends on the random path of the particle. However, it is not an arbitrary integer-valued random variable. As a matter of fact, at the time τ we do not know how the process would behave after τ, and we have to solve the problem knowing the process prior to the time τ. We therefore consider only those integer-valued random variables τ for which the occurrence or nonoccurrence of the event $\{\tau = t\}$ is uniquely determined according to the values of x(0), x(1), ..., x(t). These random times are called the M a r k o v t i m e s (the Markov times for a Wiener process have already been discussed in Chapt. II, §4).

The sum $\sum_{t=0}^{\infty} \mathrm{P}_x\{\tau = t\}$ can be less than one (and even equal to zero). Instead of saying that τ is indeterminate for the corresponding paths of the particle, we sometimes write $\tau = \infty$.

A typical Markov time is the time of first visit to some subset Γ of the set E (there are, of course, other Markov times, for example, $\tau = 5$ or $\tau = \tau_1 + 2$, where τ_1 is a Markov time, etc.).

If the time τ is chosen (in other words, if the strategy of the person stopping the process is given), the gain turns out to be a random variable $f(x(\tau))$. It is required to choose τ such that the mean value $\mathbf{M}_x f(x(\tau))$ is as large as possible (as usual, $\mathbf{M}_x$ indicates the expectation for the initial position of the particle at the point x).* In order for the expectation $\mathbf{M}_x f(x(\tau))$ to have meaning for any τ, certain restrictions must be imposed on the function f. It is sufficient to demand that f be bounded.

In summary, the problem is stated as follows· *A Markov chain with transition probabilities* $p(x, y)$ *and a bounded function* $f(x)$ *are given on a finite or denumerable set E. It is required to: 1) calculate the variable* $v(x) = \sup_\tau \mathbf{M}_x f(x(\tau))$, *where* τ *represents all the possible Markov times, 2) find the Markov time* τ_0 *for which* $\mathbf{M}_x f(x(\tau_0)) = v(x)$

By analogy with the theory of games, the variable v(x) is called the *value of a game*, and the Markov time τ_0 is called the *optimal strategy*.

In order to gain better insight into the problem, we consider some special cases and examples.

If $f \leq 0$ over the entire phase space E, the problem has a trivial solution, clearly, $\tau_0 = \infty$ (i.e., never stopping the process) may be adopted as the optimal strategy, and $v(x) = 0$. We exclude this uninteresting case right away and assume that $\sup_x f(x) > 0$.

We next consider a random walk on a one-dimensional lattice. As we know (see Chapt. I, §1), in this type of random walk the particle has a probability one of sooner or later visiting any state x. Consequently, here $v(x) = c$, where $c = \sup_x f(x)$, because it is permissible to wait until the particle attains a state in which $f(x)$ is arbitrarily close to c. If the value of c is attained on a subset Γ

*In calculating the expectation $\mathbf{M}_x f(x(\tau))$, the summation extends only over the elementary events for which τ is finite (see the footnote on page 41).

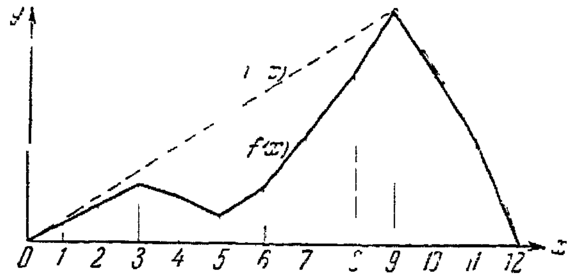

Fig. 24

of points of the lattice, then in order to obtain an optimal strategy it is sufficient to set τ_0 equal to the time of first visit to Γ. If, on the other hand, c is not attained at any point, an optimal strategy does not exist, even though it might be possible to obtain a payoff arbitrarily close to c.

It is clear that the same pattern will be observed in any Markov chain in which the particle has a probability one of occupying all states (such chains are called *recurrent*).

We next examine a homogeneous random walk on a line segment with absorption at the ends (Fig. 24). The particle has a probability 1/2 of jumping from the states 1-11 to the nearest point to the right or left, but on arriving at the state 0 or 12 it always stays there. A graph of the function $f(x)$ is shown in Fig. 24 (the adjacent points of the graph are joined for clarity).

Inasmuch as it is impossible to exit from the points 0 and 12, we have $v(0) = f(0) = 0$, $v(12) = f(12) = 0$. At these points there is nothing to wait for, and the process may be stopped at once. Similarly, it must be stopped immediately in the state 9; in this state $f(x)$ attains an absolute maximum, hence any continuation of the process can only diminish the payoff. Consequently, $v(9) = f(9)$. At the point 5, where $f(x)$ has a relative minimum, conversely, it is unfavorable to stop, even after one step it is possible to obtain a payoff greater than $f(x)$. Therefore, $v(5) > f(5)$. What is the situation in the other states? At the point 3, for example, where $f(x)$ has a relative maximum, a postponement by one or two steps clearly diminishes the average payoff. If one waits longer, there is hope of arriving in the domain of a second or higher peak, where the payoff would be considerably greater than $f(3)$. But then there is the danger of becoming trapped at the point 0 and of gaining nothing.

Moving ahead, we point out that the value of the game v(x) in this example is the least concave function greater than or equal to $f(x)$. In other words, in order to generate the graph of v(x), it is necessary to run a line above the graph of $f(x)$ between the points 0 and 12 [in Fig. 24 the graph of v(x) is indicated by a dashed line]. The optimal strategy is to stop the chain at the time τ_0 of first arrival of the particle at the point where $f(x) = v(x)$.

It will be shown that the problem has an analogous solution in the general case of a chain with a finite number of states. The role of the concave functions in this case is taken by the class of excessive functions associated with the given Markov chain.

The optimal choice problem analyzed in §1 is a special case of our general problem. In fact, in §1 we constructed a Markov chain x(i) with states 1, 2, ..., n, and the problem was one of stopping this chain with maximum probability at the instant immediately prior to termination. If the particle is situated in the state k, the chain terminates in the next instant with a probability q(k) = k/n. Consequently, the probability of success with the strategy τ is equal to

$$\sum_{k=1}^{n} \mathsf{P}_1 \{x(\tau) = k\} \cdot \frac{k}{n} = \mathbf{M}_1 \frac{x(\tau)}{n} = \mathbf{M}_1 q(x(\tau)).$$

[The subscript 1 attached to P and M indicates that the path x(0), x(1), ... is initiated at the point 1.] Therefore, the optimal choice problem reduces to the optimal stopping problem for the payoff function $f(x) = q(x)$ and the initial state x = 1.

§3. Excessive Functions

We begin our investigation of the optimum stopping problem for an arbitrary Markov chain with a study of the payoff functions f for which the optimal strategy consists in stopping the process at once. Clearly, these must be functions f that satisfy the following inequality at any Markov time τ:

$$f(x) \geqslant \mathbf{M}_x f(x(\tau)) \qquad (x \in E). \tag{3}$$

Since the number of Markov times, in general, is infinite, it would be difficult to test the condition (3) directly for every Markov time τ. As we shall see, it is sufficient for (3) to hold for $\tau=\infty$ and $\tau=1$; then this condition will be satisfied for all other Markov times.

For $\tau=\infty$ the condition (3) leads to the inequality

$$f(x)\geqslant 0 \qquad (x\in E). \tag{4}$$

For $\tau=1$ it becomes the condition

$$f(x)\geqslant Pf(x), \tag{5}$$

where P denotes an operator operating according to the formula $Pf(x)=\sum_y p(x, y)f(y)$ (the one-step shift operator).

We are well acquainted with the requirements (4) and (5) from Chapt. I; they comprised the definition of the excessive function for a symmetric random walk on a lattice. It is logical to introduce analogous definitions in the case of an arbitrary Markov chain as well. Nonnegative functions f for which $Pf\leq f$ are called excessive functions.

We will prove that if a function f is excessive, the inequality (3) is satisfied for any Markov time τ.*

This statement has already been proved for a random walk on a lattice in Chapt. I, §6. Of course, τ was interpreted there as the time of first visit to a certain set, but, as is readily observed, the arguments are wholly applicable to arbitrary Markov times as well. The fundamental notion of the proof was to represent the excessive function f as the sum of a constant, for which (3) is obvious, and the potential

$$\begin{aligned} G\varphi(x) &= \varphi(x)+P\varphi(x)+P^2\varphi(x)+\ \ldots \\ &= \mathbf{M}_x[\varphi(x(0))+\varphi(x(1))+\ \ldots] \end{aligned} \tag{6}$$

of the nonnegative function $\varphi=f-Pf$. For the potential the inequality (3) was derived from the relation

$$\mathbf{M}_x G\varphi(x(\tau))=\mathbf{M}_x[\varphi(x(\tau))+\varphi(x(\tau+1))+\ \ldots], \tag{7}$$

* This fact (in a more general situation) was established by Hunt [5].

the right side of which is less than or equal to the right side of Eq. (6).

In the case of an arbitrary Markov chain the series (6) can diverge. We cope with this difficulty by introducing a "correction factor" $\alpha < 1$ and then letting α tend to unity.

Setting $\varphi(x) = f(x) - \alpha Pf(x)$, $0 < \alpha < 1$, we write the obvious identity

$$f = \varphi + \alpha P\varphi + \alpha^2 P^2 \varphi + \ldots + \alpha^n P^n \varphi + \alpha^{n+1} P^{n+1} f,$$

where, by virtue of (5), once again $\varphi \geq 0$. Utilizing the fact that $0 \leq P^n f = P^{n-1}(Pf) \leq P^{n-1} f$, so that $\alpha^n P^n f \to 0$ as $n \to \infty$, we obtain a representation of f as an infinite series:

$$\begin{aligned} f(x) &= \varphi(x) + \alpha P\varphi(x) + \alpha^2 P^2 \varphi(x) + \ldots \\ &= \mathbf{M}_x [\varphi(x(0)) + \alpha\varphi(x(1)) + \alpha^2 \varphi(x(2)) + \ldots] \end{aligned} \tag{8}$$

[in the general case the relation $P^n \varphi(x) = \mathbf{M}_x \varphi(x(n))$ is derived in exactly the same manner as for a random walk on a lattice]. Just as the relation (6) is implied by (7), it also follows from (8) that

$$\mathbf{M}_x \alpha^\tau f(x(\tau)) = \mathbf{M}_x [\alpha^\tau \varphi(x(\tau)) + \alpha^{\tau+1} \varphi(x(\tau+1)) + \ldots] \tag{9}$$

(we leave the verification of this relation to the reader). We infer from a comparison of (8) and (9) that

$$f(x) \geqslant \mathbf{M}_x \alpha^\tau f(x(\tau)).$$

In order to obtain the inequality from (3) from this, all we need is to let α tend to one.*

The following more general property of excessive functions is proved analogously: *If f is excessive and $\tau' \geq \tau$ are two Markov times, then*

$$\mathbf{M}_x f(x(\tau)) \geqslant \mathbf{M}_x f(x(\tau')) \qquad (x \in E). \tag{10}$$

*Passing to the limit too hastily in the argument of the expectation can result in invalid equations. However, $\xi_\alpha \to \xi$ implies $\mathbf{M}\xi_\alpha \to \mathbf{M}\xi$ in the following two important cases:

1) when $|\xi_\alpha| < \eta$ for all α and $\mathbf{M}\eta < \infty$;

2) when $\xi_\alpha \geq 0$ and $\xi_\alpha \to \xi$, increasing monotonically.

For the proof of this property it is required to write Eq. (9) for each of the times τ and τ'. Since $\tau \leq \tau'$, the series (9) for τ will contain all the same terms as the series (9) for τ' and possibly some additional positive terms. Consequently, for $0 < \alpha < 1$

$$M_x \alpha^{\tau} f(x(\tau)) \geqslant M_x \alpha^{\tau'} f(x(\tau')).$$

As $\alpha \to 1$, we obtain Eq. (10) from the latter.

It is readily deduced from the inequality (10) that *if a function v is excessive and τ is the time of first visit to some set Γ, then the function*

$$h(x) = M_x v(x(\tau))$$

is also excessive.

To prove this assertion, we denote by τ' the first time $t \geq 1$ at which the particle is situated in the set Γ. It is clear that $\tau' \geq \tau$, and hence that

$$M_x f(x(\tau')) \leqslant M_x f(x(\tau)) = h(x).$$

But if the first step has brought the particle from x to y, then under this condition $M_x f(x(\tau'))$ is equal to $M_y f(x(\tau)) = h(y)$. Consequently,

$$M_x f(x(\tau')) = \sum_{y \in E} p(x, y) h(y) = Ph(x).$$

Thus, $Ph \leq h$.

§4. The Value of a Game

If the payoff function f is excessive, then, as we readily perceive, the value of the game v coincides with f.

We note in the general case that *if an excessive function g majorizes the payoff function f, it also majorizes the value of the game v.*

In fact, if $g \geq f$ and g is excessive, then for any strategy τ

$$M_x f(x(\tau)) \leqslant M_x g(x(\tau)) \leqslant g(x)$$

and, hence,

$$v(x)=\sup_{\tau} \mathbf{M}_x f(x(\tau)) \leqslant g(x)$$

We show next that **the value of the game v itself is an excessive function.**

Clearly, the function v is nonnegative; zero payoff can always be obtained with the strategy $\tau = \infty$.

In order to test the condition $Pv \leq v$, we formulate a strategy τ yielding an average payoff $\mathbf{M}_x f(x(\tau))$ arbitrarily close to $Pv(x)$, then we make use of the inequality $\mathbf{M}_x f(x(\tau)) \leq v(x)$.

We pick an arbitrary number $\varepsilon > 0$ and denote by $\tau_{\varepsilon, y}$ the strategy for which

$$\mathbf{M}_y f(x(\tau_{\varepsilon, y})) \geqslant v(y) - \varepsilon \qquad (y \in E).$$

(The existence of a Markov time $\tau_{\varepsilon, y}$ for any y follows from the actual definition of the value of a game.) Let the strategy τ consist in first making one step, then, if this step brings the particle to the state y, using the strategy $\tau_{\varepsilon, y}$. More precisely, if $x(1) = y$, then $\tau = 1 + \tau_{\varepsilon, y}$, where $\tau_{\varepsilon, y}$ is found from the trajectory $x(1), x(2), \ldots$, beginning at the time 1 rather than the time 0. It is readily agreed that τ is a Markov time. For this τ we have

$$\mathbf{M}_x f(x(\tau)) = \sum_{y \in E} p(x, y) \mathbf{M}_y f(x(\tau_{\varepsilon, y}))$$

$$\geqslant \sum_{y \in E} p(x, y)[v(y) - \varepsilon] = Pv(x) - \varepsilon \sum_{y \in E} p(x, y) \geqslant Pv(x) - \varepsilon.$$

Consequently, $v(x) \geq Pv(x) - \varepsilon$ for any $\varepsilon > 0$, hence $Pv(x) \leq v(x)$. This proves the excessiveness of the function v.

Since one possible strategy is instantaneous stopping, $v(x) \geq f(x)$.

Summarizing, we have deduced the fact that **the value of a game v is the minimum excessive function greater than or equal to the payoff function f** (and is logically called the **excessive majorant of f**).

As a by-product, we have also proved the existence of an excessive majorant for any function f (this is not evident a priori).

This result permits the value of the game to be found by linear programming methods in the case of a finite number of states. In fact, the value of the game v(x) is the minimum function satisfying the following set of 3n linear inequalities:

$$\left.\begin{aligned} v(x) &\geqslant \sum_{y\in E} p(x, y)v(y), \\ v(x) &\geqslant f(x), \\ v(x) &\geqslant 0, \end{aligned}\right\} (x \in E),$$

where n is the number of states of the Markov chain.

§5. The Optimal Strategy

We denote by Γ the set of all states x in which the payoff function f(x) is equal to its excessive majorant v(x). We call this set the **support set** (in Fig. 24 the support set comprises the points 0, 9, 10, 11, and 12; at these points the graph of the function f "supports" the line representing the function v).

Let a particle begin to move at a point x of the support set. Immediate stopping at this point yields a payoff equal to v(x), and it is not possible to give a better strategy. On the other hand, stopping at an initial state x outside Γ results in a payoff f(x) that is strictly less than the value of the game v(x). Therefore, had we known beforehand, first, that an optimal strategy existed and, second, that this strategy prescribed stopping or continuing to scan solely as a function of the position of the particle at the current instant (as is specifically the situation in the optimal choice problem), we could have inferred that the optimal strategy is given by the time τ of first visit of the particle to Γ. So far, however, we can only adopt this as a reasonably plausible hypothesis.

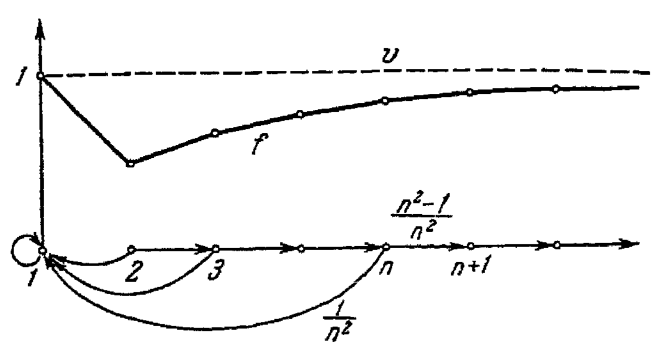

Fig. 25

Not always, however, does this hypothesis turn out to be true. Let us consider, for example, a Markov chain with an infinite number of states 1, 2, ..., n, ..., in which the particle has a probability $1/n^2$ of going from the point n to the point 1 and a probability $(n^2-1)/n^2$ of going from the same point to the point n+1 (Fig. 25). Let $f(n) = 1 - 1/n$ for $n > 1$, and let $f(1) = 1$. Obviously, in this situation it is always possible to expect a payoff arbitrarily close to, but never greater than, one, hence $v(n) = 1$. The support set Γ in this example consists of the single point 1. Inasmuch as $f(1) = 1$, for the time τ of first visit to Γ the average payoff $M_n f(x(\tau))$ is equal to the probability $\pi(n)$ of leaving n and arriving sometime at 1. The probability of the converse event, i.e., of the particle departing to infinity on the right, is equal to

$$\prod_{k=n}^{\infty} \frac{k^2-1}{k^2}. \tag{11}$$

Since

$$\prod_{k=n}^{m} \frac{k^2-1}{k^2} = \prod_{k=n}^{m} \frac{(k-1)(k+1)}{k \cdot k} = \frac{(n-1)(m+1)}{nm},$$

the infinite product (11) converges and is equal to $(n-1)/n$. Therefore, $\pi(n) = 1/n$, whereas $v(n) = 1$.

The violation of our hypothesis in this example is connected with the fact that the phase space is infinite. We will show that *the time τ_0 of first visit to the support set in the case of a finite phase space is an optimal strategy.*

Let us examine the average payoff

$$h(x) = M_x f(x(\tau_0)), \tag{12}$$

which corresponds to the strategy τ_0. It is required to prove that $h = v$. According to the actual definition of the value of a game, $h \le v$. Inasmuch as $x(\tau_0) \in \Gamma$ while f and v coincide on Γ, the function f may be replaced in Eq. (12) by the excessive function v; then it follows from this formula that h is also excessive (see §3). Since v is the minimum excessive function majorizing f, in order to obtain the inverse inequality $h \ge v$, it is sufficient to verify that $h \ge f$.

At points of the support set Γ we have $h(x) = f(x)$, because the strategy τ_0 prescribes immediate stopping at these points. Let us assume that the inequality $h(x) < f(x)$ is satisfied somewhere outside Γ. We denote by a the point at which the difference $f(x) - h(x)$ attains a maximum. Then the function $h_1(x) = h(x) + [f(a) - h(a)]$ majorizes f, coincides with f at the point a, and, as the sum of the excessive function $h(x)$ and the positive constant $f(a) - h(a)$, is also excessive. Consequently, h_1 majorizes v, and $f(a) = h_1(a) \geq v(a)$. This means that the point a chosen outside the support set Γ belongs to Γ. The ensuing contradiction reveals that the inequality $h(x) < f(x)$ is inadmissible. The optimality of the strategy τ_0 is thus proved.

We turn next to the case of a Markov chain with a denumerable phase space. Here, as we are aware, stopping at the time of first visit to the reference set Γ can prove to be a highly inauspicious strategy. It can be shown, however, that if we adopt in place of the set $\Gamma = \{x: f(x) = v(x)\}$ an "ε-support" set $\Gamma_\varepsilon = \{x: v(x) - f(x) \leq \varepsilon\}$ and investigate the time τ_ε of first visit to Γ_ε, we have for any $\varepsilon > 0$

$$M_x f(x(\tau_\varepsilon)) \geqslant v(x) - \varepsilon. \tag{13}$$

Consequently, the ε-support sets enable one to find strategies affording a payoff arbitrarily close to the value of the game.

The proof of the inequality (13) follows the same plan, with slight modifications, as in the case of a finite phase space, when $\varepsilon = 0$. Inasmuch as $f(x) \geq v(x) - \varepsilon$ on Γ_ε, we have

$$\mathbf{M}_x f(x(\tau_\varepsilon)) \geqslant \mathbf{M}_x v(x(\tau_\varepsilon)) - \varepsilon \mathbf{P}_x\{\tau_\varepsilon < \infty\} \geqslant \mathbf{M}_x v(x(\tau_\varepsilon)) - \varepsilon.$$

The function $h(x) = \mathbf{M}_x v(x(\tau_\varepsilon))$ is excessive, along with v. We will show that $h(x) \geq f(x)$. If $\sup[f(x) - h(x)] = c > 0$, the function $h(x) + c$ is excessive and majorizes $f(x)$. Consequently, $h(x) + c \geq v(x)$ for all x. Since $c > 0$, there exists a state a in which $f(a) - h(a) > 0$ and simultaneously $f(a) - h(a) > c - \varepsilon$. Then $f(a) = f(a) - h(a) + h(a) \geq c - \varepsilon + v(a) - c = v(a) - \varepsilon$, hence $a \in \Gamma_\varepsilon$. At points belonging to Γ_ε, however, the functions h and v are equal, so that $h(a) = v(a) \geq f(a)$. This contradicts the inequality $f(a) - h(a) > 0$. Therefore, c cannot be positive, hence $h(x)$ majorizes $f(x)$. But then the excessive function $h(x)$ also majorizes v(x), and consequently

$$\mathbf{M}_x f(x(\tau_\varepsilon)) \geqslant h(x) - \varepsilon \geqslant v(x) - \varepsilon.$$

§6. Application to a Random Walk with Absorption and to the Optimal Choice Problem

In a random walk along a line segment $[0, a]$ with absorption at the end points, a particle situated at any of the points 1, 2, ..., $a-1$ has a probability 1/2 of shifting one unit to the left or right in a single step, but if it arrives at the point 0 or a, it always stays there (see Fig. 24, where $a = 12$).

The solution to the optimal stopping problem for this kind of Markov chain was given without proof at the end of §2. In correspondence with the general formulations of §§3-5, all that is needed for the justification of this solution is to verify that the excessive functions are nonnegative concave functions.

By definition, a function f is excessive if $f \geq 0$ and $Pf \leq f$. The condition $Pf \leq f$ reduces in the present case to the inequalities

$$\frac{f(x-1)+f(x+1)}{2} \leqslant f(x) \qquad (x = 1, 2, \ldots, a-1) \tag{14}$$

and the trivial relations

$$f(0) \leqslant f(0), \qquad f(a) \leqslant f(a).$$

The inequalities (14) signify that if the adjacent points of the graph of the function $f(x)$ are joined by segments, the vertex of the resulting polygonal curve at any interior point x will be situated no lower than the chord connecting the vertices at the points $x-1$ and $x+1$ (Fig. 26). Consequently, the condition $Pf \leq f$ is tantamount to concavity of the function $f(x)$, which it was required to prove.

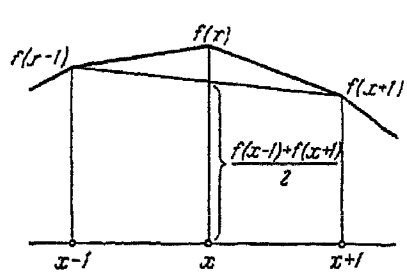

Fig. 26

Let us investigate how the concepts we have introduced operate in the optimal choice problem. As we know, this problem reduces to the optimal stopping of a chain characterized by states 1, 2, ..., n, transition probabilities

$$p(k, l) = \begin{cases} \dfrac{k}{l(l-1)} & \text{for} \quad l > k, \\ 0 & \text{for} \quad l \leqslant k, \end{cases}$$

and payoff functions $f(k) = k/n$ (see §2).

We find the excessive majorant v(k) of the payoff function f(k) and the reference set $\Gamma = \{k\colon f(k) = v(k)\}$. By definition, v is the minimum function satisfying the inequalities $v \geq f$, $Pv \geq v$, and $v \geq 0$. In the given case these inequalities assume the form

$$v(k) \geqslant \frac{k}{n},$$

$$v(k) \geqslant \sum_{l=k-1}^{n} \frac{k}{l(l-1)} v(l),$$

$$k = 1, 2, \ldots, n.$$

Hence, if $v(l)$, $l > k$, is already known, then

$$v(k) = \max\left\{\frac{k}{n},\ k \sum_{l=k+1}^{n} \frac{v(l)}{l(l-1)}\right\}.$$

We have obtained a recursion formula for the determination of v(k). On the basis of this formula we successively find

$$v(n) = \max\left\{\frac{n}{n}\right\} = 1 = f(n),$$

$$v(n-1) = \max\left\{\frac{n-1}{n};\ (n-1)\frac{1}{n(n-1)}\right\}$$

$$= \max\left\{\frac{n-1}{n},\ \frac{1}{n}\right\} = \frac{n-1}{n} = f(n-1),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$v(k) = \max\left\{\frac{k}{h},\ k\left[\frac{\frac{k+1}{n}}{(k+1)k} + \frac{\frac{k+2}{n}}{(k+2)(k+1)} + \cdots\right.\right.$$

$$\ldots + \frac{\frac{n}{n}}{n(n-1)}\Bigg]\Bigg\} = \max\left\{\frac{k}{n},\ \frac{k}{n}\left(\frac{1}{k} + \frac{1}{k+1} + \ldots + \frac{1}{n-1}\right)\right\} = \frac{k}{n} = f(k)$$

as long as the following inequality remains in force:

$$\frac{1}{k} + \frac{1}{k+1} + \ldots + \frac{1}{n-1} \leqslant 1. \qquad (15)$$

As soon as the sum $1/k + \ldots + 1/(n-1)$ becomes greater than one with diminishing k, v(k) turns out to be strictly greater than $k/n = f(k)$. With a further reduction of k the sum $1/k + \ldots + 1/(n-1)$ remains greater than one, so that at these points

$$v(k) \geqslant k \sum_{l=k+1}^{n} \frac{v(l)}{l(l-1)} \geqslant k \sum_{l=k+1}^{n} \frac{f(l)}{l(l-1)}$$

$$= \frac{k}{n}\left(\frac{1}{k} + \frac{1}{k+1} + \ldots + \frac{1}{n-1}\right) > \frac{k}{n} = f(k).$$

Hence, the support set Γ has the form $\{k_n, k_n + 1, \ldots, n\}$, where k_n is the smallest integer satisfying the inequality (15). We are already familiar with this result.

For $k \geq k_n$ the value of the game is equal to $v(k) = f(k) = k/n$, and for $k < k_n$ it is calculated in succession according to the relation*

$$v(k) = k \sum_{k=l+1}^{n} \frac{v(l)}{l(l-1)}.$$

§7. Optimal Stopping of a Wiener Process

The optimal stopping problem can be analyzed not only for Markov chains, but also for processes involving a nondenumerable phase space and continuous time. We propose to investigate one of the most elementary processes of this type, namely, a *Wiener process* x(t) *on the interval* $[0, a]$ *with absorption at the points*. By definition, given any initial position x, $0 \leq x \leq a$, the

*It is not difficult to show that v(k) is in fact independent of k for $k < k_n$ and is found from Eq. (2).

particle executes exactly the same motion as in an ordinary Wiener process on an infinite line until the first time it hits the end of the interval; on hitting the point 0 or the point a, the particle always becomes trapped at that point.*

Let the payoff function $f(x)$ be specified on the interval $[0, a]$. It is required to find the value of the game

$$v(x) = \sup_{\tau} \mathbf{M}_x f(x(\tau)) \quad (0 \leqslant x \leqslant a),$$

where τ represents all the possible Markov times, and to formulate the particular Markov time τ_0 at which

$$\mathbf{M}_x f(x(\tau_0)) = v(x)$$

(i.e., to find the optimal strategy).

The process of interest here is the continuous analog of a symmetric random walk on a line segment with absorption at the ends, i.e., the problem discussed in §§2 and 6. We see that the solution of the problem in the continuous case remains the same, except that instead of concave functions of an integer-valued argument it is necessary to use concave functions specified on the entire interval $[0, a]$.

We recall that a function $f(x)$ given on the interval $[0, a]$ is called *concave* if the entire chord connecting any two points of the graph of the function f is situated no higher than the graph f (Fig. 27). We note that *a function concave on an interval is continuous inside the interval* and at the ends of the interval has finite limits no smaller than the values of the function at the end points (see the Appendix, §2). For example, in Fig. 27

$$\lim_{x \to 0} f(x) = f(0), \quad \lim_{x \to a} f(x) > f(a).$$

---

* We are not concerned with a Wiener process on the entire infinite line, because in this case the particle has a probability one of hitting any point, and the optimal stopping problem has the same trivial solution as for a recurrent Markov chain.

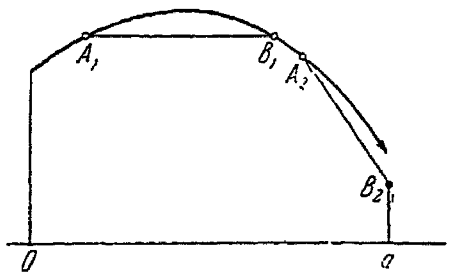

Fig. 27

The special role played by concave functions for our process is explained by the fact that **nonnegative concave functions (and only those functions) satisfy the inequality**

$$M_x f(x(\tau)) \leqslant f(x) \qquad (16)$$

at any Markov time τ. The proof of this statement is rather intricate, and we will save it for a special section (§8).

After having described the class of functions satisfying the condition (16), the value of the game and the optimal strategy are found in approximately the same manner as in §§4 and 5 for an arbitrary Markov chain.

We first calculate the probability $q(x) = q(x; x_1, x_2)$, on starting from x, of hitting the point x_1 before x_2, as well as the probability $p(x) = p(x; x_1, x_2)$ of hitting x_2 before x_1 $(0 \leq x_1 \leq x \leq x_2 \leq a)$. It follows from the results of Chapt. II that the function q(x) is a solution of the Dirichlet problem on the interval $[x_1, x_2]$ and assumes a value of one at the point x_1 and a value of zero at the point x_2. Inasmuch as the Laplace equation $\Delta q = 0$ assumes the form $q'' = 0$ in the one-dimensional case, all of its solutions are linear, i.e., they have the form $q(x) = cx + d$. Determining the values of the constants c and d from the boundary conditions $q(x_1) = 1$, $q(x_2) = 0$, we obtain

$$q(x;\ x_1,\ x_2) = \frac{x_2 - x}{x_2 - x_1},$$

$$p(x;\ x_1,\ x_2) = 1 - q(x;\ x_1,\ x_2) = \frac{x - x_1}{x_2 - x_1}. \qquad (17)$$

We now find the value of the game v(x), regarding the payoff function $f(x)$ for the time being only as bounded, but not necessarily continuous.* We note that if g(x) is a nonnegative concave function majorizing $f(x)$, then for any τ

$$M_x f(x(\tau)) \leqslant M_x g(x(\tau)) \leqslant g(x)$$

and, hence, g(x) majorizes v(x).

* The boundedness of the function c (along with measurability, which we agreed earlier not to discuss) guarantees the existence of the expectation $M_X f(x(\tau))$.

The function v(x) itself is nonnegative (as there exists a strategy $\tau = \infty$ yielding zero payoff) and is also concave. Thus, we let $[x_1, x_2]$ be some subinterval contained in $[0, a]$ and let τ_1 and τ_2 be strategies yielding average payoffs greater than $v(x_1) - \varepsilon$ and $v(x_2) - \varepsilon$ in the respective initial states x_1 and x_2 (the existence of these strategies for any $\varepsilon > 0$ ensues from the actual concept of upper bounds). Let us examine the strategy τ, whereby we first wait for the first hitting time at one of the points x_1 or x_2, then use the corresponding strategy τ_1 or τ_2. Here, according to Eq. (17),

$$\begin{aligned} M_x f(x(\tau)) &= \frac{x_2 - x}{x_2 - x_1} M_{x_1} f(x(\tau_1)) + \frac{x - x_1}{x_2 - x_1} M_{x_2} f(x(\tau_2)) \\ &\geqslant \frac{x_2 - x}{x_2 - x_1}\left[v(x_1) - \varepsilon\right] + \frac{x - x_1}{x_2 - x_1}\left[v(x_2) - \varepsilon\right] \\ &= \frac{x_2 - x}{x_2 - x_1} v(x_1) + \frac{x - x_1}{x_2 - x_1} v(x_2) - \varepsilon, \end{aligned}$$

hence

$$v(x) \geqslant \frac{(x_2 - x)\, v(x_1) + (x - x_1)\, v(x_2)}{x_2 - x_1} - \varepsilon \qquad (x_1 \leqslant x \leqslant x_2).$$

Inasmuch as ε can be an arbitrarily small positive number, the above inequality is also true for $\varepsilon = 0$. Since the function

$$\frac{(x_2 - x)\, v(x_1) + (x - x_1)\, v(x_2)}{x_2 - x_1}$$

is linear on $[x_1, x_2]$ and coincides with v at the points x_1 and x_2, this means that the graph of v on the interval $[x_1, x_2]$ does not pass any lower than the chord spanning it. Consequently, the function v(x) is concave.

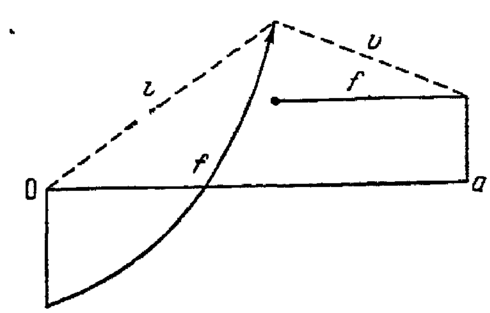

Fig. 28

Therefore, the value of the game is the minimum nonnegative concave function greater than or equal to the payoff function f, or, more concisely, the value of the game v is the nonnegative concave majorant of the function f (see Fig. 28, which illustrates a discontinuous function f).

We next show that if the function f is continuous, then, as in the discrete case, an optimum strategy is to stop the process at the time τ_0 of first visit to the support set Γ, on which $f(x) = v(x)$. We note that this statement no longer holds for a discontinuous payoff function f. Thus, in the example illustrated in Fig. 28 the set Γ comprises the single point a. This means that if we wait for arrival in Γ, we will never obtain a payoff greater than $f(a)$, whereas $v(x)$ is much larger than $f(a)$ at some points.

We first verify the fact that continuity of the payoff function f implies continuity of the value of the game v. Inasmuch as the function v is concave, it is continuous at all interior points of the interval $[0, a]$, and $\lim_{x\to 0} v(x) \geqslant v(0)$, $\lim_{x\to a} v(x) \geqslant v(a)$. We examine the point 0 for definiteness and show that

$$\lim_{x\to 0} v(x) \leqslant v(0). \tag{18}$$

We set $c(u) = \max_{0\leqslant x\leqslant u} f(x)$, $0 \leq u \leq a$. It is apparent that the function $c(u)$ is continuous, along with $f(x)$. For $x(\tau) < u$ the payoff, clearly, cannot exceed the value of $c(u)$, while for $x(\tau) \geq u$ it cannot exceed $c(a)$. Moreover, the inequality $x(\tau) \geq u$ for $x = x(0) < u$ can only occur in the event that the particle arrives from the point x at u before it arrives at the point 0. The probability of this event, according to Eq. (17), is equal to x/u. Conse-

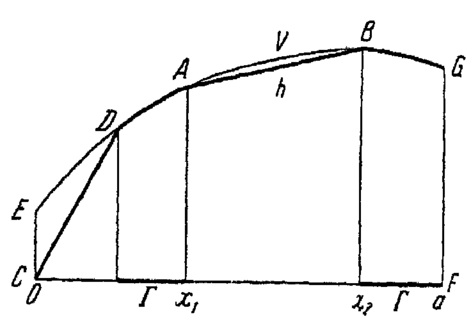

Fig. 29

quently, for $0 < x < u$ and any τ

$$\mathbf{M}_x f(x(\tau)) \leqslant c(u)\mathbf{P}_x\{x(\tau) < u\} + c(a)\cdot\frac{x}{u}.$$

If $c(u) \geq 0$, the first term here does not exceed $c(u)$, but if $c(u) < 0$, it does not exceed 0; hence, in any case

$$\mathbf{M}_x f(x(\tau)) \leqslant \max[c(u),\ 0] + c(a)\frac{x}{u},$$

and, consequently,

$$v(x) \leqslant \max[c(u),\ 0] + c(a)\frac{x}{u}.$$

Letting $x \to 0$ here, we obtain

$$\lim_{x\to 0} v(x) \leqslant \max[c(u),\ 0] \qquad (u > 0),$$

and then letting u tend to 0, we find

$$\lim_{x\to 0} v(x) \leqslant \max[c(0),\ 0] = \max[f(0),\ 0].$$

Inasmuch as $0 \leq v(0)$ and $f(0) \leq v(0)$, the inequality (18) is proved.

Since both of the functions f and v are continuous, the support set Γ, comprising those points x at which $f(x) = v(x)$, is closed (a priori, Γ can also be an empty set). Let τ be the time of first visit to Γ, and let

$$h(x) = \mathbf{M}_x f(x(\tau))$$

be the average payoff for the strategy τ. Since $f = v$ on Γ,

$$h(x) = \mathbf{M}_x v(x(\tau)). \tag{19}$$

We see now that the function h defined by Eq. (19), like v, is concave, continuous, and nonnegative. In fact, if $x = x(0) \in \Gamma$, then $\tau = 0$, and $h(x) = v(x)$. The points x not belonging to the closed set Γ form a system of intervals, the ends of which either belong to Γ or coincide with one of the points 0, a (Fig. 29). If the end points x_1 and x_2 of such an interval belong to Γ, then on the interval $[x_1, x_2]$, according to Eqs. (17), the function h is equal to

$$h(x)=\frac{x_2-x}{x_2-x_1}v(x_1)+\frac{x-x_1}{x_2-x_1}v(x_2). \tag{20}$$

It is apparent from this formula that the graph of h on the interval $[x_1, x_2]$ is the chord AB subtending the points of the graph of the function v. If, on the other hand, one end of the interval (x_1, x_2) coincides with an end of the segment $[0, a]$ and does not belong to Γ, the value of v at the corresponding point x_1 or x_2 in Eq. (20) is replaced by zero, and the graph of h is a line segment such as CD in Fig. 29. This segment may also be called a chord of the graph of v, provided the vertical sections CE and FG are included in this graph. Thus, the graph of h is obtained from the graph of v by "cutting off the convexities" with chords on some systems of intervals. It is geometrically evident that this operation again produces a graph of a continuous concave nonnegative function (see the Appendix).

Inasmuch as v is the smallest of the nonnegative concave functions majorizing f, for the proof of the inequality $h \geq v$ (and, hence, of the optimality of the strategy τ) it is sufficient to verify the fact that $h \geq f$. Let us assume that the difference $f - h$ acquires a positive value somewhere. Then the continuous function $f - h$ must reach its maximum value $c > 0$ at some point x_0. The nonnegative concave function $h(x) + c$ majorizes f, which means that it also majorizes v. Consequently, $h(x_0) + c \geq v(x_0)$, which combines with the equation $c = f(x_0) - h(x_0)$ to yield the relation $f(x_0) \geq v(x_0)$. Therefore, $x_0 \in \Gamma$, whence $h(x_0) = v(x_0) = f(x_0)$ and $c = f(x_0) - h(x_0) = 0$. This contradicts the premise that $c > 0$. Thus the optimality of the strategy τ is proved.

We conclude with a few remarks about the multidimensional case. Consider the optimal stopping problem of an l-dimensional Wiener process in a closed domain G with absorption at the boundary. The value of the game is found as in the one-dimensional case, except that the nonnegative concave functions must be replaced by nonnegative functions f satisfying the two following conditions:

1) For any l-dimensional sphere $S \subset G$ with center x the mean value of f on S does not exceed $f(x)$.

2) For any $x \in G$ and any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$f(y) \geqslant f(x) - \varepsilon,$$

provided only that

$$|y - x| < \delta,\ y \in G.$$

[We point out that the condition 1 is a special case of the inequality $M_x f(x(\tau)) \leq f(x)$ when τ is the time of first exit from S.] The conditions 1 and 2 form the definition adopted in modern potential theory for a *superharmonic* function in the domain G.* Consequently, it may be stated that the value of the game is the nonnegative superharmonic majorant of the payoff function. As far as the optimal strategy is concerned, it far from always exists. In any case, however, it is possible to formulate ε-optimal strategies by means of ε-support sets, as was done for denumerable Markov chains at the end of §6.

We note further that inasmuch as for $l \geq 3$ a Wiener path no longer has a probability one of entering any arbitrary domain, for $l \geq 3$ the optimal stopping problem is important in the special case when G is the entire space (see the footnote on page 113).

§8. Proof of the Fundamental Property of Concave Functions

It remains for us to prove that in the case of a Wiener process on an interval $[0, a]$ with absorption at the end points the class of functions $f(x)$, $x \in [0, a]$, satisfying the condition

$$f(x) \geqslant M_x f(x(\tau)) \tag{21}$$

for any Markov time τ coincides with the class of nonnegative concave functions.

This is a very simple matter in one direction. Letting $\tau = \infty$ in (21), we find that $f \geq 0$. Moreover, let the subinterval $[x_1, x_2]$ be

* If the function f is continuous and has continuous second partial derivatives, 2 is fulfilled automatically, and 1 reduces to the inequality $\Delta f \leq 0$, where Δ is the Laplace operator (cf. the derivation of the equation $\Delta f = 0$ in Chapt. II, §4).

contained in $[0, a]$, and let τ be the time of first exit of $x(t)$ from $[x_1, x_2]$. According to Eq. (17), for this τ

$$\mathbf{M}_x f(x(\tau)) = f(x_1)\frac{x_2 - x}{x_2 - x_1} + f(x_2)\frac{x - x_1}{x_2 - x_1}$$

for $x_1 \leq x \leq x_2$. Consequently, the graph of the function $\mathbf{M}_x f(x(\tau))$ for $x \in [x_1, x_2]$ is a line segment connecting the points with abscissas x_1 and x_2 on the graph of the function $f(x)$. It follows from the inequality (21), therefore, that any chord of the graph of f does not lie higher than this graph, i.e., the function f is concave.

It is a much more complex task to show that every concave nonnegative function satisfies the condition (21), although basically the argumentation remains the same as in the derivation of the condition (21) in the discrete case for excessive functions. We analyze the proof in six parts.

1°. We define an operator P_t $(t > 0)$ on bounded functions $f(x)$, $0 \leq x \leq a$, by the formula

$$P_t f(x) = \mathbf{M}_x f(x(t)) = \int_0^a f(y)\,\mu_t(dy), \tag{22}$$

where $\mu_t(\Gamma) = \mathbf{P}_x\{x(t) \in \Gamma\}$, and we let

$$P_\infty f(x) = \lim_{t \to \infty} P_t f(x).$$

By virtue of the Markov property, the process $y(s) = x(s+t)$ for any fixed $t > 0$ is a Wiener process with absorption at the end points and an initial distribution $\mu_t(\Gamma)$. Therefore, applying Eq. (22) twice, we write

$$\mathbf{M}_x f(y(s)) = \int_0^a \mathbf{M}_y f(x(s))\,\mu_t(dy) = \int_0^a P_s f(y)\,\mu_t(dy) = P_t P_s f(x).$$

On the other hand,

$$\mathbf{M}_x f(y(s)) = \mathbf{M}_x f(x(t+s)) = P_{t+s} f(s).$$

Consequently, the *operators* P_t *are multiplied according to the rule*

$$P_t P_s = P_{t+s}, \tag{23}$$

We recall for comparison that in the case of a discrete Markov chain

$$\mathbf{M}_x f(x(n)) = P^n f(x),$$

where P is the one-step shift operator. Consequently, in the discrete-time case Eq. (23) reduces to the ordinary rule for the multiplication of powers. We note that families of operators P_t $(t > 0)$ which are multiplied according to Eq. (23) are called *one-parameter subgroups*.

It is immediately evident from the definition of the operator P_t that if $f \geq 0$, then $P_t f \geq 0$ also (the operator P_t is *positive*). Applying this property to the difference function $f - g$, we deduce that if $f \geq g$, then $P_t f \geq P_t g$ also (the operator P_t preserves the inequality between functions).

We next calculate $P_\infty f(x)$. We know that a particle leaving any point of the interval has a probability one of sooner or later arriving at an end point of the interval, where it will always remain. Consequently, as $t \to \infty$ the measure μ_t of the interval $(0, a)$ tends to zero, while the measure μ_t of the points 0 and a tends to $\mathbf{P}_x\{x(\tau) = 0\}$ and $\mathbf{P}_x\{x(\tau) = a\}$, where τ is the time of first exit of the path from the interval $(0, a)$. Therefore,

$$P_\infty f(x) = f(0) \cdot \mathbf{P}_x\{x(\tau) = 0\} + f(a)\mathbf{P}_x\{x(\tau) = a\}$$

or, according to Eq. (17),

$$P_\infty f(x) = f(0)\frac{a-x}{a} + f(a)\frac{x}{a}.$$

It is apparent from the above expression that $P_\infty f$ *is a linear function whose values at the points* 0 *and* a *coincide with the values of* f *at those points* (Fig. 30).

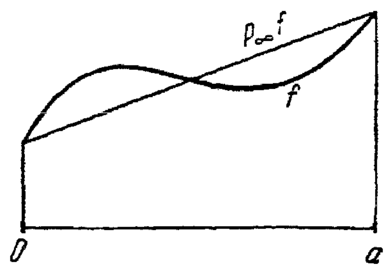

Fig. 30

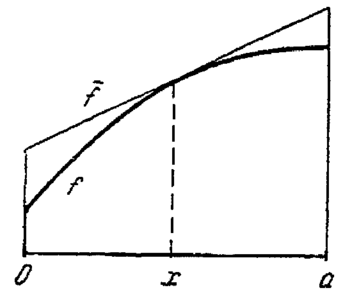

Fig. 31

2°. For linear functions f

$$P_t f = f. \tag{24}$$

According to part 1°, if f is linear, $f = P_\infty f$. Passing to the limit in the identity*

$$P_t(P_s f) = P_{t+s} f$$

as $s \to \infty$, we arrive at Eq. (24). It is easily shown that the converse is also true (we leave the proof of this to the reader).

3°. If the function f is concave,

$$P_t f \leqslant f.$$

For $x = 0$ and $x = a$ the probability is one that $x(t) = x(0)$, so that

$$P_t f(x) = \mathbf{M}_x f(x(t)) = \mathbf{M}_x f(x(0)) = f(x).$$

Let x be an interior point of the interval $[0, a]$. Inasmuch as the function f is concave, it is possible to formulate a linear function $\bar{f}$ such that $\bar{f}(x) = f(x)$ at the given point x, and at all other points $f \geq \bar{f}$ (Fig. 31) (the proof of this property of concave functions is given in § 2 of the Appendix). According to part 2°,

$$P_t \bar{f} = \bar{f}.$$

Since $\bar{f} \geq f$ on the entire interval, while at the point x the values of

*See the footnote on page 104.

$\bar{f}$ and f are equal, we have

$$P_t f(x) \leqslant P_t \bar{f}(x) = \bar{f}(x) = f(x).$$

4°. Let α be some number from the interval (0, 1). We agree to call functions h represented in the form

$$h(x) = \int_0^\infty \alpha^t P_t g(x)\, dt = \mathbf{M}_x \int_0^\infty \alpha^t g(x(t))\, dt,$$

where $g \geq 0$, *α-potentials* [the α-potentials play the same role in the continuous case as series of the form (8) in the discrete case].

We now show that if $f \geq 0$, $P_t f \leq f$ *for all t, and the function f is continuous at interior points of the interval $[0, a]$, then no matter what α is, $0 < \alpha < 1$,* and *the function f may be represented as the limit of nondecreasing α-potentials.*

Making use of the identity $P_t P_s = P_{t+s}$, we write

$$\int_0^s \alpha^t P_t f\, dt = \int_0^\infty \alpha^t P_t f\, dt - \int_s^\infty \alpha^t P_t f\, dt$$

$$= \int_0^\infty \alpha^t P_t f\, dt - \int_0^\infty \alpha^{s+t} P_{s+t} f\, dt = \int_0^\infty \alpha^t P_t (f - \alpha^s P_s f)\, dt,$$

or

$$\frac{1}{s} \int_0^s \alpha^t P_t f\, dt = \int_0^\infty \alpha^t P_t g\, dt, \tag{25}$$

where

$$g = \frac{f - \alpha^s P_s f}{s}; \tag{26}$$

these integrals converge, since $|\alpha| < 1$ and $|P_t f(x)| = |\mathbf{M}_x f(x(t))|$ is bounded by the number $\sup_x |f(x)|$. Inasmuch as $0 \leq P_t f \leq f$ and

$0 < \alpha < 1$, it follows from (26) that $g \geq 0$. Consequently, an α-potential stands on the right side of Eq. (25). We can establish the fact that this α-potential, without decreasing, converges to f as $s \to 0$ if we verify that

$$\lim_{t \to 0} P_t f = f \tag{27}$$

and that $\alpha^t P_t f$ is a nonincreasing function of the argument t. The required monotonicity of $\alpha^t P_t f$ follows from the sequence of relations

$$\alpha^{t+u} P_{t+u} f \leqslant \alpha^t P_{t+u} f = \alpha^t P_t (P_u f) \leqslant \alpha^t P_t f$$
$$(u > 0).$$

In order to demonstrate (27), we recall that

$$P_t f(x) = \mathbf{M}_x f(x(t)).$$

As $t \to 0$, the probability is one that $x(t) \to x$, because the paths of the process $x(t)$ are continuous. This means that $f(x(t)) \to f(x)$ also with probability one at those points x where f is continuous, i.e., at all interior points of $(0, a)$. But if the random variable $f(x(t))$ converges with probability one to a constant $f(x)$, its expectation converges to the expectation of the constant $f(x)$, i.e., to the number $f(x)$ itself [passage to the limit in the argument of the expectation is legitimate, insofar as the random variable $f(x(t))$ is bounded for any t by the same number $k = \sup_x |f(x)|$]. Consequently,

$$\lim_{t \to 0} \alpha^t P_t f(x) = \lim_{t \to 0} \alpha^t \cdot \lim_{t \to 0} \mathbf{M}_x f(x(t)) = f(x)$$
$$(0 < x < a).$$

As for the points $x = 0$ and $x = a$, where f can suffer a discontinuity, there $P_t f(x) = f(x)$ for all t; Eq. (27) follows at once from this.

5°. If $h(x)$ is an α-potential and τ is any Markov time,

$$\mathbf{M}_x \alpha^\tau h(x(\tau)) \leqslant h(x).$$

We have the condition

$$h(x) = \mathrm{M}_x \int_0^\infty \alpha^t g(x(t))\, dt,$$

where $g \geq 0$. Therefore,

$$h(x) \geqslant \mathrm{M}_x \int_\tau^\infty \alpha^t g(x(t))\, dt = \mathrm{M}_x \alpha^\tau \int_0^\infty \alpha^s g(x(\tau + s))\, ds = \mathrm{M}_x \alpha^\tau \int_0^\infty \alpha^s g(y(s))\, ds, \tag{28}$$

where $y(s) = x(\tau + s)$. According to the strong Markov property, the process $y(s)$ under the condition $\tau = t$, $x(\tau) = y$ is exactly the same process as $x(s)$ beginning at the point y.* Hence

$$\mathrm{M}_x (\alpha^\tau \int_0^\infty \alpha^s g(y(s))\, ds \mid \tau = t,\ x(\tau) = y) = \alpha^t \mathrm{M}_y \int_0^\infty \alpha^s g(x(s))\, ds = \alpha^t h(y).$$

Denoting by $F(t, y)$ the joint distribution function of the pair of random variables τ and $x(\tau)$, we then write

$$\mathrm{M}_x \alpha^\tau \int_0^\infty \alpha^s g(y(s))\, ds = \int_0^\infty \int_0^a \alpha^t h(y)\, dF(t,\ y) = \mathrm{M}_x \alpha^\tau h(x(\tau)).$$

Subsituting this value into Eq. (28), we obtain the desired result.

6°. Finally, we prove that a nonnegative concave function f satisfies the condition (21). It follows from the continuity of a concave function inside the interval $[0, a]$ and parts 3° and 4° that for any $\alpha \in (0, 1)$ f is the limit of a nondecreasing sequence of α-potentials $h_1, h_2, \ldots, h_n, \ldots$. According to part 5°, for any Markov time τ

$$\mathrm{M}_x \alpha^\tau h_n(x(\tau)) \leqslant h_n(x) \leqslant f(x).$$

* The intuitively justified but rather loose arguments presented here with regard to what happens under the condition $\tau = t$, $x(\tau) = y$, which has a probability zero, can be translated into completely rigorous form.

Since h_n converges monotonically to f, it is permissible to pass to the limit in this inequality as $n \to \infty$ in the argument of M_x. Thus

$$M_x \alpha^\tau f(x(\tau)) \leqslant f(x)$$

for any positive $\alpha < 1$. Passing to the limit again as $\alpha \to 1$, we obtain $\mathbf{M}_x f(x(\tau)) \leq f(x)$.

To what extent does the given proof extend to a multidimensional Wiener process? As already stated at the end of §7, in general the role of the concave functions is taken by superharmonic functions. Defining the operator P_t as before by the formula

$$P_t f(x) = \mathbf{M}_x f(x(t)),$$

we call nonnegative functions f satisfying the conditions

$$P_t f \leqslant f, \quad \lim_{t \to 0} P_t f = f \tag{29}$$

excessive functions (cf. the definition of excessive functions for Markov chains in §3). In essence, we began in the present section by demonstrating that nonnegative concave functions are excessive, then we established the fact that excessive functions satisfy the inequality

$$\mathbf{M}_x f(x(\tau)) \leqslant f(x)$$

for any Markov time τ. The reader can easily verify that this second half of the proof has a completely general character and is equally applicable to the multidimensional case. On the other hand, the proof that superharmonic nonnegative functions are excessive in the multidimensional case is more complicated than in the one-dimensional case (on more than one occasion we made use of the special properties of concave functions, for example, their continuity at interior points of the interval). Moreover, the one-dimensional nature of the problem made it possible for us in §7 to circumvent the problems associated with the measurability of the value of a game.

PROBLEMS

Choosing One of the Two Best Objects

Suppose that it is required to choose one of the two best objects (it being immaterial exactly which one) among n objects. As in the case analyzed in §1, the problem reduces to the optimal stopping of a certain Markov chain x(0), x(1), x(2), ... In §1 the elements of this chain stand for the indices of the maximal objects (points), i.e., objects better than all those already scanned. It is clear that maximality in the new problem must be given a weaker interpretation, regarding an object a_k as "maximal" if it is the best or second best of all the objects $a_1, a_2, ..., a_k$ already scanned. This is a small matter, however The value of x(i) must indicate not only the order number (index) of the corresponding "maximal" object, but also whether this object is the best (i.e., maximal in the previous sense) or second best. The phase space of the chain x(i) is therefore conveniently represented as two parallels rows of n points, regarding the upper row as representative of objects better than all the preceding ones and the lower row as representative of objects inferior to just one of the preceding ones (Fig. 32).

1. Find the transition probabilities of the chain x(i).

Answer. Regardless of whether the points k and l are situated in the upper or lower row,

$$p(k, l)=\frac{k(k-1)}{l(l-1)(l-2)} \qquad (l>k)$$

[for $l=2$, $k=1$ the fraction is to be assumed equal to $k/[l(l-1)]$].

2. Find the probability of success (payoff function) f for stopping of the chain at a given point.

Answer. Letting the subscript 1 refer to points in the upper, the subscript 2 to points in the lower row, we have

$$f_1(k)=\frac{k(2n-k-1)}{n(n-1)},$$

$$f_2(k)=\frac{k(k-1)}{n(n-1)}.$$

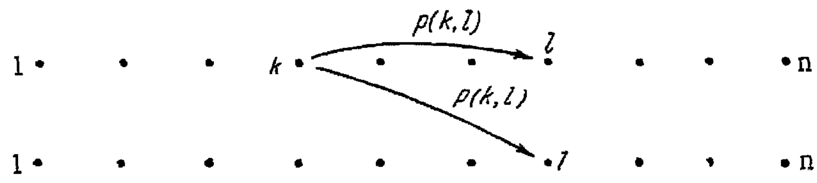

Fig. 32

Arguments similar to those used in §6 indicate that the value of the game v is found successively according to the relation

$$v_j(k) = \max\left\{f_j(k),\ \sum_{l=k+1}^{n} p(k, l)[v_1(l) + v_2(l)]\right\} \tag{30}$$

(v_1 corresponds to upper points, v_2 to lower points). We denote by Γ_j the set of points of the jth row in which the functions f and v coincide ($j = 1, 2$).

3. The set Γ_2 has the form $\{m_2, m_2+1, \ldots, n\}$, where m_2 is the smallest integer greater than or equal to $(2n+1)/3$.

The set Γ_1 also contains all the numbers $m_2, m_2+1, \ldots, n$.

Hint. Verify the fact that

$$\sum_{l=k+1}^{n} p(k, l)[f_1(l) + f_2(l)] = \frac{2k(n-k)}{n(n-1)}$$

and apply Eq. (30).

We denote by B_k the set comprising Γ_2 plus the points $k+1$, $k+2, \ldots, n$ of the upper column ($k < m_2$), and we let τ_k denote the time of first visit to B_k.

4. If $f_1(k) < \mathbf{M}_x f(x(\tau_k))$, k does not belong to Γ_1. If $k+1$, $k+2, \ldots, n$ belong to Γ_1 and $f_1(k) \geq \mathbf{M}_k f(x(\tau))$, k belongs to Γ_1.

Hint. Given any initial state, stopping at the time of first visit to $\Gamma_1 \cup \Gamma_2$ is an optimal strategy (see §5).

5. Find the distribution of $x(\tau_k)$ for an initial state k.

Hint. Describe the event $x(\tau_k) = l$ in terms of the objects $a_{k+1}, a_{k+2}, \ldots, a_l$. For $k < l < m_2$ we have $\mathbf{P}_k\{x(\tau_k) = l\} = \frac{k}{l(l-1)}$

for points l of the upper row, and for $m_2 \leq l \leq n$ we have $P_k\{x(\tau_k)=l\} = \frac{k}{m_2-1} p(m_2-1, l)$ for points l of both rows.

6. The set Γ_1 has the form $\{m_1, m_1+1, \ldots, n\}$, where m_1 is the smallest positive integer for which

$$\left(\frac{1}{m_1} + \frac{1}{m_1+1} + \cdots + \frac{1}{m_2-2}\right) \leqslant \frac{3m_2-2m_1-4}{2(n-1)}.$$

7. If the number of objects n grows indefinitely,

$$\lim \frac{m_1}{n} = \alpha, \quad \lim \frac{m_2}{n} = \frac{2}{3},$$

where α is the root of the equation $\alpha - \ln \alpha = 1 + \ln(3/2)$ and is smaller than one ($\alpha \approx 0.347$).

8. The probability of success with an optimal strategy tends to $\alpha(2-\alpha) \approx 0.574$ as $n \to \infty$.

Hint. The distribution at the time of first visit to the set $\Gamma_1 \cup \Gamma_2$ for any initial state $s < m_1$ will be the same as the distribution of $x(\tau_k)$ for $k = m_1 - 1$ in Problem 5.

Further Generalization of the Choice Problem

Now let it be required to choose with maximum probability one of the first s objects in order of quality [for a total number of objects n $(s<n)$]. The phase space of the chain x(i) consists in this case of s rows involving n points, and the arrival of the particle at a point k of the jth row means that the object a_k is ranked in jth place according to quality in the group $a_1, a_2, \ldots, a_k$. We denote by $f_j(k)$ the payoff function (probability of success) for stopping the chain at the point k of the jth row, by $v_j(k)$ the value of the game at that point, and by Γ_j the part of the support set Γ located in the jth row.

9. The transition probabilities $p(k, l)$ of the chain x(i) do not depend on which rows the points indexed by k and l are located in.

It is easily shown that

$$v_j(k) = \max\left\{f_j(k),\ \sum_{l=k+1}^{n} p(k,\ l) \sum_{i=1}^{s} v_i(l)\right\} \tag{31}$$

[cf. Eq. (30)].

10. The function $f_j(k)$ increases monotonically with respect to the argument k and decreases monotonically with respect to the argument j.

11. The double sum in Eq. (31) decreases monotonically with increasing k.

Hint. This sum is equal to the expected payoff for the optimal strategy if stopping is forbidden at the first k objects.

12. The set Γ_j has the form $m_j, m_{j+1}, \ldots, n$, where $1 \le m_1 \le m_2 \le m_s \le n$.

13. Calculate $\sum_{j=1}^{s} f_j(k)$.

Hint. We introduce the following symbolic events:

$A = \{a_k$ is one of the s best objects$\}$,

$B_j = \{a_k$ is the jth in quality of the objects $a_1, a_2, \ldots, a_k\}$.

Then

$$\sum_{j=1}^{s} f_j(k) = \sum_{j=1}^{s} \mathbf{P}\{A/B_j\} = k \sum_{j=1}^{s} \mathbf{P}\{A/B_i\}\, \mathbf{P}\{B_i\} = k\mathbf{P}(A) = \frac{ks}{n}.$$

14. In the notation of Problem 12, for $s \ge 2$

$$\lim_{n\to\infty} \frac{m_s}{n} = \sqrt[s-1]{\frac{s}{2s-1}}.$$

Hint. After computing

$$f_s(k) = \frac{k(k-1)\ldots(k-s+1)}{n(n-1)\ldots(n-s+1)}$$

and

$$p(k, l) = \frac{k(k-1)\dots(k-s+1)}{l(l-1)\dots(l-s)},$$

use Eq. (31) and Problem 12 (cf. the cases s = 1 and s = 2). For the calculation of the sum in Eq. (31) use the identity

$$\sum_{l=k+1}^{\infty} \frac{1}{(l-1)(l-2)\dots(l-s)} = \frac{1}{s-1} \cdot \frac{1}{(k-1)(k-2)\dots(k-s+1)},$$

which is valid for $s \geq 2$.

Optimal Rule for the Stopping of a Sequence of Independent Random Variables

Let $\xi_1, \xi_2, \dots, \xi_n$ be independent random variables, which take values from a certain number set X, and let $f(k, x)$ ($k = 1, 2, \dots, n$; $x \in X$) be a nonnegative function. We identify ξ_1 first, then ξ_2, ξ_3, etc. The observations may be terminated at any time k. The gain in this case is $f(k, \xi_k)$. It is required to find the optimal stopping rule such that the average payoff is maximized.

As in the optimal choice problem, it is possible by retrograde induction to formulate the value of the game v(k, x) and to verify that the optimal strategy is to stop at the time of first visit of the point (k, ξ_k) to the support set Γ consisting of those pairs (k, x) for which $f(k, x) = v(k, x)$.

The statement of the problem is preserved intact for dependent random variables, but the solution is greatly complicated by the fact that the optimal stopping rule, in general, requires inclusion of all the values observed, rather than the last one only. It is interesting that the optimal stopping problem was in fact first formulated for the dependent case. In particular, A. Cayley posed the following problem in 1874 (see [20] and the solution in [21]):

"A lottery is arranged as follows: There are k tickets representing a, b, c, ... pounds, respectively. A person draws once; looks at his ticket; and, if he pleases, draws again (out of the remaining $k-1$ tickets); looks at his ticket; and, if he pleases, draws again (out of the remaining $k-2$ tickets); and so on, drawing in all not more than n times; and he receives the value of the last drawn ticket. Suppose he regulates his drawings in the manner most advantageous to him according to the theory of probabilities, what is the value of his expectations?"

For his solution of the problem Cayley formulated an algorithm incorporating retrograde induction and calculated the answer for the case $k=4$, $a=1$, $b=2$, $c=3$, $d=4$, and $n=1, 2, 3, 4$.

The choice of one of the first s objects according to quality (see Problems 9-12) is reduced as follows to a choice from among a sequence of independent random variables.*

15. If an object a_k occupies the jth place in quality in the group $a_1, a_2, \ldots, a_k$, we put

$$\xi_k = \begin{cases} j, & 1 \leqslant j \leqslant s, \\ s+1, & s+1 \leqslant j. \end{cases}$$

The random variables $\xi_1, \xi_2, \ldots, \xi_n$ are independent, and the probability $f(k, j)$ of success in choosing the object a_k under the condition $\xi_k = j$ is equal to

$$f(k, j) = \begin{cases} f_j(k), & 1 \leqslant j \leqslant s, \\ 0, & s+1 \leqslant j, \end{cases}$$

where $f_j(k)$ is the function from Problem 10.

16. If $f(k, x)$ is a nondecreasing function of the argument k, and $f > 0$, there exists an integer-valued function $m(x)$, $x \in X$, such that the set Γ is specified by the inequalities $m(x) \leq k \leq n$. If, in addition, $f(k, x)$ is a nonincreasing (nondecreasing) function of x, $m(x)$ is a nondecreasing (nonincreasing) function of x.

17. (See [23]). Let ξ_k be distributed uniformly on the interval [0, 1] and let $f(k, x) = x$.

*See [22]; also obtained in this paper are some results that were presented in another manner in the preceding sets of problems.

Then

$$m(x) = n - k \qquad \text{for} \quad x_k \leqslant x < x_{k+1},$$

where the numbers x_k are found from the relations

$$x_{k+1} = \frac{1 + x_k^2}{2}, \qquad x_0 = 0.$$

<u>Hint.</u> It can be shown by induction on k that

$$v(k, x) = \begin{cases} x_k, & 0 \leqslant x \leqslant x_k, \\ x, & x_k \leqslant x \leqslant 1. \end{cases}$$

18. In the preceding problem, as $k \to \infty$,

$$1 - x_k \sim \frac{2}{k}.$$

<u>Hint.</u> Letting

$$x_k = 1 - \frac{2}{\alpha_k},$$

we find that

$$\alpha_{k+1} = \alpha_k + 1 + \frac{1}{\alpha_k - 1}, \qquad \alpha_0 = 2.$$

We find in succession, therefore, that $\alpha_k \to \infty$, $\alpha_{k+1} - \alpha_k \to 1$, $\alpha_k / k \to 1$. A better estimate

$$k + \left(1 + \tfrac{1}{2} + \ldots + \tfrac{1}{k}\right) + 1 < \alpha_k \leqslant k + \left(1 + \tfrac{1}{2} + \ldots + \tfrac{1}{k}\right) + 2$$

and further refinements may be found in the paper by Moser [23].

<u>Optimal Stopping of a General Markov Chain</u>

19. If in a chain with a denumerable infinity of states the support set Γ is accessible with probability one from any

state x, stopping at the time of first visit to Γ is an optimal strategy.

Hint. Investigate the time τ_ε of first visit to the ε-support set Γ_ε and let ε tend to zero.

20. A state a belongs to the support set Γ if and only if there exists an excessive function h everywhere greater than or equal to the payoff function f and coinciding with f at the point a.

21. (Method of successive approximations.*) Let f^+ be a function equal to the payoff function f wherever $f \geq 0$ and equal to zero wherever $f < 0$, and let the operator Q be given by the equation

$$Qf(x) = \max\{f(x),\ Pf(x)\}.$$

Then $Q^n f^+$ converges monotonically to the value of the game v as $n \to \infty$.

Hint. The function $Q^\infty f = \lim_{n\to\infty} Q^n f$ is the excessive majorant of f.

Fee for a Game

Suppose that after every transition from x to y a fee $\Phi(x, y)$ is collected. If for any initial state x the expectation of the fee up to the instant of termination of the chain ζ

$$F(x) = \mathbf{M}_x \sum_{t=1}^{\zeta-1} \Phi(x(t-1),\ x(t))$$

is finite, the optimal stopping problem reduces to the case in which there is no fee for the game.

22. For any Markov time τ

$$F(x) = \mathbf{M}_x \sum_{t=1}^{\tau} \Phi(x(t-1),\ x(t)) + \mathbf{M}_x F(x(\tau)).$$

Hint. Compare the proof of Eq. (24) from Chapt. I, §5.

*Proposed by A. D. Venttsel'.

23. The quantity

$$M_x\left[f(x(\tau))-\sum_{t=1}^{\tau}\Phi(x(t-1),\ x(t))\right]$$

attains its maximum value at the Markov time τ when and only when τ is the optimal strategy in the stopping problem for a chain $x(t)$ with the payoff function $f(x)+F(x)$.

Unbounded Payoff Functions

It was postulated in Chapt. III that the payoff function f was bounded. Now we lift this assumption, assuming that f is nonnegative [so that there always exists a finite or infinite expectation $M_x f(x(\tau))$]. We define the value of the game and the class of excessive functions in the same manner as in §§2 and 3, except that now we admit the value of $+\infty$ for these functions.

24. Any excessive function f is the limit of a nondecreasing sequence of bounded excessive functions.

Hint. Investigate $f_n(x)=\min\{n, f(x)\}$.

25. Extend the inequality $M_x f(x(\tau))\leq f(x)$ (τ is any Markov time) to excessive functions admitting the value $+\infty$.

26. The value of the game v is the excessive majorant of the payoff function f.

Hint. The function v is the limit of a nondecreasing sequence $\{v_n\}$, where v_n is the value of the game corresponding to the payoff function f_n of Problem 24.

27. The value of the game v can be infinite for a finite payoff function f.

Hint. Investigate a random walk on the integer points of the line $x\geq 0$ with absorption at zero and assume a payoff function $f(0)=1, f(k)=k$ $(k\geq 1)$.

28. The average payoff for stopping at the time of first visit to the ε-support set Γ_ε optionally tends to the value of the game as $\varepsilon\downarrow 0$, when the value of the game is finite.

The Martin Boundary

The method of Martin (developed later by Doob [24]) provides a means for exhibiting the structure of the set of all excessive functions associated with a denumerable Markov chain.

Let x(t) be a Markov chain on a denumerable infinite phase space E, such that for any initial state x the probability of returning to x is smaller than one. We denote by g(x, y) the expectation of the number of hits on the point y for an initial state x (Green's function, cf. Chapt. I, §5).

29. Prove that

$$g(x, y) = \pi_y(x) g(y, y),$$

where $\pi_y(x)$ is the probability, on leaving x, of arriving sometime at y.

It follows from Problems 29 and 2 of Chapt. I that

$$g(x, y) < \infty$$

for any x, y.

Let us extend the definitions given in Chapt. I for the potential and harmonic function in the case of a symmetric random walk to the case of the chain x(t); the potential of a nonnegative function φ refers to the function

$$G\varphi = \varphi + P\varphi + P^2\varphi + \ldots + P^n\varphi + \ldots,$$

and a harmonic function is a function h for which Ph = h.

As in Chapt. I, §5, we establish the fact that

$$G\varphi(x) = \sum_{y \in E} g(x, y)\varphi(y),$$

that the potential is excessive, and that any excessive function f is described in the form $G\varphi + h$, where $\varphi = f - Pf$, $h = \lim_{n\to\infty} P^n f$ is a nonnegative harmonic function.

30. An excessive function f is a potential when and only when $P^n f \to 0$ as $n \to \infty$.

31. The minimum of an excessive function and a potential is a potential.

32. Any excessive function is the limit of a nondecreasing sequence of potentials.

Hint. We number the points of the space E and denote by B_n the set of the first n points. Then the functions

$$f_n = \min\{nG\chi_{B_n}, f\}$$

form the required sequence of potentials (χ_B is the characteristic function of the set B).

We assume in addition that for some state $0 \in E$ the probability $\pi_y(0)$ is positive for all $y \in E$.* Then $g(0, y) > 0$ also. According to Problem 32, for an excessive function f there exists a sequence of functions $\varphi_n \geq 0$ such that

$$f(x) = \lim_{n\to\infty} \sum_{y\in E} g(x, y)\varphi_n(y). \tag{32}$$

Introducing the Martin kernel

$$k(x, y) = \frac{g(x, y)}{g(0, y)} = \frac{\pi_y(x)}{\pi_y(0)}$$

(see Problem 29), we rewrite (32) in the form

$$f(x) = \lim_{n\to\infty} \sum_{y\in E} k(x, y)\mu_n(y), \tag{33}$$

where μ_n is a sequence of measures on E described by the equation

$$\mu_n(y) = g(0, y)\varphi_n(y). \tag{34}$$

---

*In general the same formulations are applicable to a Markov chain as are obtained when the set S of states accessible from the state 0 is bounded (clearly, it is impossible to go from S into $E \setminus S$).

In cases requiring emphasis of the fact that k(x, y) is regarded as a function of $x \in E$ for a fixed value of y we write $k_y(x)$ instead of k(x, y).

33. Different states $y \in E$ correspond to different functions $k_y(x)$.

Hint. The function $k_y(x) - Pk_y(x)$ has a nonzero value at the single point x = y.

34. The values of all the functions $k_y(x)$ at a given point $x \in E$ are bounded by the number $1/\pi_x(0)$.

Hint. Make use of Problem 29 and the inequality $\pi_y(0) \geq \pi_x(0)\ \pi_y(x)$.

Problem 33 shows that the functions $k_y(x)$ $(y \in E)$ stand in one-to-one correspondence with points y of the space E. We affix to the family of functions $\{k_y\}$ all the possible limits of these functions (in other words, we close the set of functions k_y, using coordinatewise convergence). According to Problem 34 here and Problem 4 of Chapt. I, the resulting set of functions K is compact. Identifying the points $y \in E$ with their corresponding functions k_y, we say that the space E is embedded in the compactum K. The set $B = K \setminus E$ is called the Martin boundary for the Markov chain x(t). The elements of the set B, like those of E, are represented either by the letter y or, if it is to be stressed that they are functions on the space E, by the symbol $k_y(x)$.

35. The function $k_y(x)$ is excessive for any $y \in K$.

If for every x the function p(x, y) has a nonzero value only for a finite number of values of y, then $k_y(x)$ is a harmonic function for $y \in B$.

Hint. Examine the case $y \in B$. If $y = \lim_{n \to \infty} y_n$ $(y_n \in E)$, then, according to the hint to Problem 33, for any $x \in E$ we have

$$k_y(x) = \lim_{n \to \infty} k_{y_n}(x) = \lim_{n \to \infty} Pk_{y_n}(x) = \lim_{n \to \infty} \sum_{z \in E} p(x, z)\, k_{y_n}(z)$$

$$\geqslant \sum_{z \in E} \lim_{n \to \infty} p(x, z)\, k_{y_n}(z) = Pk_y(x).$$

[It is easily verified that if the variables $u_n(z)$ are nonnegative and

$u_n(z) \to u(z)$, then

$$\lim_{n\to\infty} \sum_z u_n(z) \geqslant \sum_z u(z).]$$

If the sums are finite, the equality sign holds.

36. For the measures μ_n given by Eq. (34) the sequence $\mu_n(E)$ is bounded.

Hint. Set $x=0$ in Eq. (33).

Let us continue the measures μ_n over the entire compactum K, setting $\mu_n(B)=0$. Then Eq. (33) may be rewritten in the form

$$f(x)=\lim_{n\to\infty}\left[\sum_{y\in E} k(x, y)\mu_n(y)+\int_B k(x, y)\mu_n(dy)\right]$$

$$=\lim_{n\to\infty}\int_K k(x, y)\mu_n(dy), \tag{35}$$

where $k(x, y)=k_y(x)$ $(x\in E, y\in K)$.

In the actual structure of the compactum K the function k(x, y) is continuous with respect to y for any x. According to a theorem of Helly,* if $\{\mu_n\}$ is a sequence of measures on the compactum K, such that the values of $\mu_n(K)$ are bounded, it is possible to construct a measure μ on K and to pick out from $\{\mu_n\}$ a subsequence $\{\mu_{n_k}\}$ such that for any continuous function $F(y)$ $(y\in K)$

$$\lim_{k\to\infty}\int_K F(y)\mu_{n_k}(dy)=\int_K F(y)\mu(dy).$$

* Helly proved this theorem for the case when K is a line segment. The proof is available in any standard text on probability theory (see, e.g., [10], Chapt. IV, §11.2). A general proof is easily obtained by comparing the following two facts: 1) In the Banach space C of all continuous functions on the compactum K any nonegative linear functional l is expressed as an integral over some finite measure ν; here $\|l\|=\nu(K)$ (see, e.g., [25], §56); 2) it is possible from every sequence of linear functionals with bounded norms to pick out a weakly convergent subsequence (see, e.g., [26], Chapt. III, §24).

In application to Eq. (35) this leads to the equation

$$f(x) = \int_K k(x, y)\mu(dy), \tag{36}$$

where μ is a finite measure on K depending on the excessive function f.

37. Any function f representable in the integral form (36) with $\mu(K) < \infty$ is excessive.

Hint. In the case of nonnegative functions it is permissible to change the order of summation and integration.

We denote by V the set of all excessive functions satisfying the condition $f(0) = 1$. It is readily seen that V is a convex set (see the problems to Chapt. I).

38. Any excessive function, unless identically equal to zero, is specified in the form $cf(x)$, where $f \in V$, $c > 0$.

Hint. It is required to verify the fact that if f is excessive and $f(0) = 0$, then $f = 0$ everywhere. This is easily deduced from the accessibility of all states from zero and the inequality $M_x f(x(\tau)) \leq f(x)$ (τ is any Markov time).

39. All extremal points of the set V are included among the functions $k_y(x)$ $(y \in K)$.

Hint. Let f be an extremal point of the set V. Setting $x = 0$ in (36), we find that $\mu(K) = 1$. Inasmuch as K is a compact, there exists a point $z \in K$ such that $\mu(U) > 0$ for any neighborhood U of the point z. If $\mu(U) < 1$, it follows from the representation

$$f(x) = \mu(U)\frac{\int_U k(x, y)\mu(dy)}{\mu(U)} + \mu(K \setminus U)\frac{\int_{K \setminus U} k(x, y)\mu(dy)}{\mu(K - U)}$$

that

$$f(x) = \frac{1}{\mu(U)}\int_U k(x, y)\mu(dy)$$

(see Problem 37). Clearly, this is equally true for $\mu(U) = 1$.

Shrinking U to the point z, we deduce that $f(x) = k_z(x)$.

40. For $y \in E$ the function $k_y(x)$ is an extremal point of the set V.

Hint. For $y \in E$

$$k_y(x) = G\varphi(x),$$

where $\varphi(x)$ has a nonzero point only at the one point y. If

$$k_y(x) = \alpha f_1(x) + \beta f_2(x),$$

where $f_1, f_2 \in V$, $\alpha > 0$, $\beta > 0$, $\alpha + \beta = 1$, then f_1 and f_2 are also potentials of some functions $\varphi_1 \geq 0$ and $\varphi_2 \geq 0$ (see Problem 31). It is easily verified that $\alpha\varphi_1 + \beta\varphi_2 = \varphi$, whence it follows that φ_1 and φ_2 are proportional to φ and, therefore, $f_1 = f_2 = k_y$.

It is readily seen that the subset H of harmonic functions from V is also a convex set.

41. All extremal points of the set H are included among the functions $k_y(x)$ $(y \in B)$.

Hint. From the representation of the excessive function in the form $G\varphi + h$ deduce the fact that an extremal point of the set H is also an extremal point of the set V.

We denote by B_e the set of points of the boundary B that correspond to extremal functions of H. According to a theorem of Choquet,* if H is a compact convex set in the space of sequences and B_e is the set of extremal points of H, then any element $h \in H$ is represented in the form of an integral of the extremal functions according to some finite measure ν on B_e.

Consequently, any positive harmonic function h is specified in the form

$$h(x) = \int_{B_e} k(x, y)\,\nu(dy). \tag{37}$$

*See, for example, [27]; in this paper the theorem is proved for any locally convex linear topological space.

Specifying the potential $G\varphi$ ($\varphi \geq 0$) in the form

$$G\varphi(x) = \sum_{y \in E} g(x, y)\varphi(y) = \sum_{y \in E} k(x, y)\nu(y),$$

where $\nu(y) = g(0, y)\varphi(y)$, we obtain the following representation for an arbitrary excessive function $f = G\varphi + h$:

$$f(x) = \sum_{y \in E} k(x, y)\nu(y) + \int_{B_e} k(x, y)\nu(dy) = \int_{E \cup B_e} k(x, y)\nu(dy).$$

It is inferred from another theorem of Choquet that the representation obtained for $f(x)$ is unique.

In essence we were dealing with Martin boundaries in the problems to Chapt. I, where we computed the set B_e for a symmetric random walk on a plane (see Problems 42-47). Another instructive example of the calculation of a Martin boundary is offered in the next set of problems.

Random Walk on a Free Group with a Finite Number of Generators [28]

A free group G with generators $a_1, a_2, \ldots, a_m$ is constructed as follows. We consider a word $a_{i_1}a_{i_2}\ldots a_{i_n}$ of arbitrary length n, where the indices assume values of ±1, ±2, ..., ±m. Adjoining one word to another, we obtain the product of these words. The inverse element is defined by the relation $(a_{i_1}a_{i_2}\ldots a_{i_n})^{-1} = a_{-i_n}\ldots a_{-i_2}a_{-i_1}$. The identity element is the "word" e containing no letters. Two words specify one and the same element of a group when and only when one of them can be derived from the other by the insertion or deletion of a product of the form $a_j a_{-j}$ an arbitrary number of times. For every element there exists a uniquely defined notation comprising a minimum number of letters.

Let $p_1, \ldots, p_m, p_{-1}, \ldots, p_{-m}$ represent positive numbers which sum to unity. We assume that during unit time the word g is transformed with probability p_i to the word ga_i (if $g = a_{i_1} \ldots a_{i_n}$, then $ga_i = a_{i_1} \ldots a_{i_{n-1}}$ for $i = -i_n$). The Markov chain thus defined is called a random walk on the group G.

42. The probability $r(x)$, on starting from x, of returning at some time to this state is the same for all $x \in G$.

43. If $p_i \neq p_{-i}$ for at least one i, then $r(x) < 1$.

Hint. If $p_i > p_{-i}$, then there is a probability one, beginning with a certain time, that the number of occurrences of the letter a_i will exceed the number of occurrences of the letter a_{-i} (this follows from the irreversibility of an asymmetric random walk on a line; see Chapt. IV, §4).

44. If all the p_i are equal and the number of generators $m \geq 2$, then $r(x) < 1$.

Hint. The probability of a minimal notation of a word $x \neq e$ being lengthened by one letter is $(2m - 1)/2m$, the probability of its being shortened by one letter is $1/2m$, and the affair reduces to an asymmetric random walk on a half-line (see Chapt. IV, §4).

Subtler considerations reveal that $r(x) = 1$ in the unique case $m = 1$, $p_1 = p_{-1} = 1/2$. Henceforth we postulate that $r(x) < 1$ and use only the minimal notation of the elements of the group.

45. Express the Martin kernel $k(x, y) = k_y(x)$ in terms of the probabilities u_i of arriving sometime at a_i from e ($i = \pm 1, \pm 2, \ldots, \pm m$).

Answer. If $x = a_{i_1} \ldots a_{i_n}$, $y = a_{j_1} \ldots a_{j_s}$ and the letters from the first to the kth coincide in these two words, while $i_{k+1} \neq j_{k+1}$, then

$$k(x, y) = \frac{u_{-i_{k+1}} \cdots u_{-i_n}}{u_{j_1} \cdots u_{j_k}}. \tag{38}$$

46. The sequence $k_{y_1}(x), k_{y_2}(x), \ldots, k_{y_n}(x), \ldots$ converges for every $x \in G$ if and only if the number of letters coinciding from the beginning in the words $y_n, y_{n+1}, y_{n+2}, \ldots$ tends to infinity as $n \to \infty$.

Hint. Verify first that $u_i u_{-i} < 1$ ($i = 1, \ldots, m$).

By virtue of Problem 46, the points of a Martin boundary are uniquely identified with infinite words $y = a_{i_1} a_{i_2} \ldots a_{i_n} \ldots$, where $i_k + i_{k+1} \neq 0$ ($k = 1, 2, \ldots$). The Martin boundary B consists of all such words. By virtue of Problem 35, the function $k_y(x)$ is harmonic for $y \in B$.

47. The functions $k_y(x)$ for $y \in B$ are extremal points of the set H (see Problem 41).

Hint. Let $y = a_{j_1} a_{j_2} \ldots a_{j_s} \ldots$, and let

$$k_y(x) = \alpha f_1(x) + \beta f_2(x) \qquad (x \in G),$$

where $f_1, f_2 \in H$, $\alpha > 0$, $\beta > 0$, $\alpha + \beta = 1$. We set $y_s = a_{j_1} a_{j_2} \ldots a_{j_s}$ ($s = 1, 2, \ldots$). It follows from the inequality $f_i(x) \geq \mathbf{M}_x f_i(x(\tau))$ (see §3) that

$$f_i(x) \geqslant f_i(y_s)\, \pi_{y_s}(x) \qquad (i = 1, 2), \tag{39}$$

where $\pi_z(x)$ is the probability of arriving sometime from x at z $(x, z \in G)$. If the word x contains n letters and $n \leq s$, then, by virtue of (38),

$$k_y(x) = \pi_{y_s}(x)\, k_y(y_s). \tag{40}$$

Therefore, for $n \leq s$

$$k_y(x) = \alpha f_1(x) + \beta f_2(x) \geqslant \pi_{y_s}(x)\,[\alpha f_1(y_s) + \beta f_2(y_s)] = \pi_{y_s}(x)\, k_y(y_s) = k_y(x),$$

hence the equality sign is indeed valid in (39) for $n \leq s$. Combined with (40), this yields the proportionality

$$\frac{f_i(x)}{k_y(x)} = \frac{f_i(y_s)}{k_y(y_s)} \qquad (n \leqslant s),$$

whence it is readily inferred that $f_1(x) = f_2(x) = k_y(x)$. In the case considered, therefore, $B_e = B$.

48. All positive harmonic functions are obtained according to the relations

$$f(e) = \nu,$$

$$f(a_{i_1} \ldots a_{i_n}) = \frac{\nu(i_1, \ldots, i_n)}{u_{i_1} \ldots u_{i_n}}$$

$$+ \sum_{k=0}^{n-1} \frac{u_{-i_{k+1}} \ldots u_{-i_n}}{u_{i_1} \ldots u_{i_k}} [\nu(i_1, \ldots, i_k) - \nu(i_1, \ldots, i_k, i_{k+1})],$$

where ν and $\nu(i_1, \ldots, i_n)$ are arbitrary nonnegative numbers satisfying the relations

$$\nu(i_1, \ldots, i_n) = \sum_{i_{n+1}=1}^{m} \nu(i_1, \ldots, i_n, i_{n+1}) + \sum_{i_{n+1}=-1}^{-m} \nu(i_1, \ldots, i_n, i_{n+1})$$

$$(n = 0, 1, 2, \ldots)$$

[for n = 0 we interpret $\nu(i_1, \ldots, i_n)$ as the number ν].

Hint. Use Eqs. (37) and (38).

Chapter IV

Boundary Conditions

§1. Introduction

The probabilistic approach to analysis problems has proven extremely fruitful in the study of one of the basic problem areas in the theory of differential equations, namely, the boundary-value problems for these equations. From the probabilistic point of view we are concerned with the behavior of the paths of diffusion processes on the boundary of a domain. In order to get to the heart of the matter, we analyze a Wiener process in a plane domain G bounded by a certain curve (Fig. 33). As long as the particle remains inside the domain G, its motion is controlled by the characteristic operator $\frac{1}{2}\Delta$, where Δ is the Laplace operator (see Chapt. II, §9). What happens when the particle escapes outside the boundary of the domain?

Conceivably, on reaching the boundary, the particle might be thought of as rebounding back into the domain at a fixed point y. The further progress of the process is completely determined, since the law of motion inside G is known. An obvious generalization of this case is the rebound of the particle inside G with a certain probability π (depending in general on the boundary point r). Other possibilities, of which we are already cognizant from Chapt. III, are absorption (the particle always remaining at the point of the boundary at which it first arrives) and extinction at the time of first visit to the boundary. Finally, one of the fundamental boundary effects is reflection. The simplest reflection is defined for a process occurring in a half-plane. For example, if G is the half-plane $x_1 > 0$, it is necessary first of all to analyze

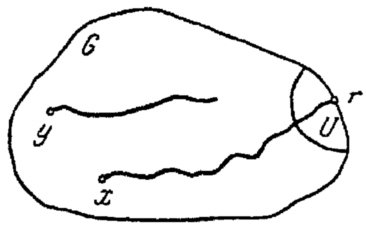

Fig. 33

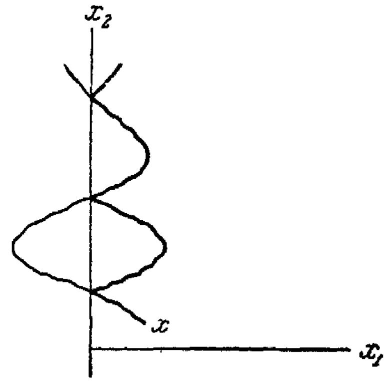

Fig. 34

the path of x(t) on the entire plane, then to reflect the part of the path lying in the left half-plane $x_1 > 0$ symmetrically about the x_2 axis (Fig. 34).

Whereas inside the domain G the process is specified by the operator $\frac{1}{2}\Delta$, the various continuations of the process after arrival at the boundary are described analytically by means of the boundary conditions. These conditions occur in the calculation of the characteristic operator $\mathfrak{A}$ at the boundary points of the domain.

We recall that by definition

$$\mathfrak{A}f(x) = \lim_{U \downarrow x} \frac{M_x f(x(\tau)) - f(x)}{M_x(\tau)}, \tag{1}$$

where τ is the time of first exit from the neighborhood U of the point x.

If absorption takes place at the boundary r, at x = r the denominator in Eq. (1) goes to ∞, and we obtain the boundary condition

$$\mathfrak{A}f(r) = 0. \tag{2}$$

In order to obtain the boundary condition for rebound at the point y, we first assume that before the jump at y the particle spends a random time ξ at the boundary point r with a distribution $\mathbf{P}_r\{\xi > t\} = e^{-at}$.* Then for a neighborhood U of the point r not con-

* Only with this type of distribution can the process have the Markov property; see §2 (pp. 152-153).

taining y the time τ coincides with ξ, and

$$M_r\tau = M_r\xi = \frac{1}{a}.$$

Here $f(x(\tau)) = f(y)$, and it follows from Eq. (1) that

$$\frac{1}{a}\mathfrak{A}f(r) = f(y) - f(r).$$

If a is made to tend to infinity, ξ goes to zero in this limit, and we obtain an instantaneous rebound from r to y. The preceding equation then goes over to the boundary condition

$$f(r) - f(y) = 0.$$

If the rebound from r occurs to a random point y with a distribution $\pi(\Gamma)$, then analogous considerations lead to the more general boundary condition

$$f(r) - \int_G f(y)\,\pi(dy) = 0, \tag{3}$$

or, equivalently,

$$\int_G [f(r) - f(y)]\,\pi(dy) = 0. \tag{4}$$

A process with extinction at the boundary may be regarded as a degenerate case of a rebound process when the measure π is equal to zero. From (3) we obtain the following boundary condition for this case:

$$f(r) = 0. \tag{5}$$

Finally, reflection at the boundary $x_1 = 0$ of the domain G (Fig. 34) corresponds to the boundary condition

$$\frac{\partial f}{\partial x_1}(r) = 0. \tag{6}$$

In order to understand this condition better, we observe that reflection may be obtained by passing to the limit from a rebound, considering that the particle rebounds from the boundary point r to a distance h in the direction of the x_1 axis, then letting h tend to zero. For a right rebound to a distance h the boundary condition (3) assumes the form

$$f(h, x_2) - f(0, x_2) = 0,$$

where x_2 is the ordinate of the boundary point r. Dividing this equation by h and passing to the limit as $h \downarrow 0$, we obtain the boundary condition (6).

In addition to the foregoing simplified boundary conditions, we can also introduce more complex conditions. The problem of describing the most general boundary conditions has made considerable headway so far. Its exhaustive solution, however, has only been obtained in the one-dimensional case. Let x(t) be a Wiener process on the half-line $(-\infty, 0]$. The most general condition for x(t) at the point 0 has the form

$$\beta \mathfrak{A} f(0) + \alpha f'(0) + \gamma f(0) + \int_{-\infty}^{0} [f(0) - f(y)]\, \pi(dy) = 0, \tag{7}$$

where α, β, and γ are nonnegative constants, and π is a measure on the half-line $(-\infty, 0)$ such that

$$\pi((-\infty, -1)) - \int_{-1}^{0} y\pi(dy) < \infty.$$

The numbers α, β, γ, and $\delta = \pi((-\infty, 0)$ must not be equal to zero at the same time. If $\alpha = \gamma = \delta = 0$, $\beta \neq 0$, then (7) reverts to the condition (2), and we have absorption. For $\alpha \neq 0$, $\beta = \gamma = \delta = 0$ we obtain a condition analogous to (6), i.e., reflection. The case $\gamma \neq 0$, $\alpha = \beta = \delta = 0$ reduces to a condition of the form (5) and corresponds to extinction. If $\alpha = \beta = \gamma = 0$ and $0 < \delta < \infty$, then, on dividing Eq. (7) by δ, we arrive at a condition of the form (4), i.e., we obtain a rebound with the distribution π/δ. In the general case we have some combination of all these effects. (A graphic interpretation of a boundary condition with infinite measure will be given in §8).

The transition from a Wiener process to a diffusion process corresponding to the operator

$$L = a(x)\frac{\partial^2}{\partial x^2} + b(x)\frac{\partial}{\partial x}$$

(cf. Chapt. II, §9) produces some new and interesting effects; certain now of the boundary conditions (7) become inadmissible, depending on the behavior of the coefficients $a(x)$ and $b(x)$. From the probabilistic point of view the situation is such that the point 0 can turn out to be inaccessible for the path, in which case the process is uniquely determined by the operator $\mathfrak{A}$ without any boundary conditions. But even if the boundary is accessible, reflection is not always possible.

In this chapter we analyze the boundary conditions, not for diffusion processes, but for their discrete analog, namely birth and death processes. Without altering the conceptual picture, this enables us to investigate the paths of a process by more elementary devices (instead of differential operators, we deal with difference operators).*

§2. The Birth and Death Process

A typical discrete model of a Wiener process and one with which we are well acquainted from the first chapter is a symmetric random walk on an integer-valued lattice. In order to obtain the discrete analog $x(t)$ of an arbitrary one-dimensional diffusion process, we need to discard symmetry on the part of the walk and assume that the probabilities of jumps from a point n to the neighboring states $n-1$ and $n+1$ are not equal to $1/2$, but to arbitrary numbers q_n and p_n, which satisfy the conditions

$$p_n \geqslant 0, \quad q_n \geqslant 0, \quad p_n + q_n = 1 \tag{8}$$

(Fig. 35). However, even with this assumption a model in which the time between two successive displacements of the particle is fixed and equal to one is still too crude to impart a number of the

*This same problem has been treated from an alternate point of view in [29]. In an unpublished dissertation Wang Tzŭ-k'uang has effected the formulation of these processes by passing to the limit from the case of rebound with a given distribution.

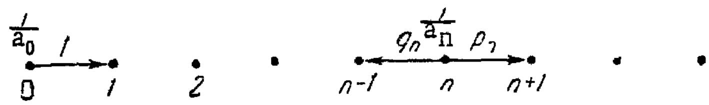

Fig. 35

important effects encountered in diffusion theory. We postulate, therefore, that the time parameter t varies continuously, as in the case of a Wiener process, and that it does not assume only integer values.

It can be shown that if the time ξ_n from arrival at a point to exit from this point is distributed according to the exponential law

$$P\{\xi_n \geqslant t\} = e^{-a_n t} \qquad (t \geqslant 0), \tag{9}$$

where

$$0 < a_n < \infty, \tag{10}$$

(and it is independent of the previous history of the particle), then the random path x(t) has the Markov property. This means that under the condition $x(s) = n$ the process $y(t) \equiv x(s+t)$ does not depend on the behavior of x(t) prior to the time s and has the same probability distribution as the process x(t), starting with the time zero in the state n.

The reader may inquire whether it is not possible to consider another distribution law for the time ξ_n to exit from the state n under the condition $x(0) = n$. We will show that if the process x(t) is Markovian, ξ_n is necessarily distributed according to an exponential law.

In fact, let p(t) $(t > 0)$ be the probability of the event $\{x(u) \equiv n, 0 \leq u \leq t\}$ under the condition $x(0) = n$. For any $s, t > 0$ the event $\{x(u) \equiv n, 0 \leq u \leq s+t\}$ is equal to the intersection of the events $\{x(u) \equiv n, 0 \leq u \leq s\}$ and $\{y(u) \equiv n, 0 \leq u \leq t\}$, where $y(u) = x(s+u)$. Therefore, if the Markov property holds,

$$p(s+t) = p(s)\,p(t) \qquad (s, t > 0). \tag{11}$$

All bounded solutions of Eq. (11) have the form

$$p(t) = e^{-a_n t},$$

where a_n is a constant, $0 \leq a_n \leq +\infty$ (see the Appendix, §3). It follows from the embedding of the events

$$\{x(u) \equiv n,\ 0 \leqslant u \leqslant t\} \subset \{\xi_n \geqslant t\} \subset \{x(u) \equiv n,\ 0 \leqslant u \leqslant t-h\}$$
$$(0 < h < t)$$

that

$$p(t) \leqslant \mathbf{P}_n\{\xi_n \geqslant t\} \leqslant p(t-h) \qquad (0 < h < t),$$

whence we obtain $\mathbf{P}_n\{\xi_n \geq t\} = p(t)$ for $h \downarrow 0$. Thus, ξ_n has the exponential distribution (9) with $0 \leq a_n \leq +\infty$.

If $a_n = +\infty$, the probability is one that the time $\xi_n = 0$, and the particle jumps instantaneously from the state n. Conversely, if $a_n = 0$, the probability is one that the time $\xi_n = +\infty$, and the particle sometimes does not escape from n. Both of these extreme cases are possible, but they are not of interest and will therefore be excluded from further discussion.

It follows from (9) that the expectation of the time ξ_n is equal to $1/a_n$.

In order not to be involved with combinations of two boundary conditions (at the points $-\infty$ and $+\infty$), we assume that the walk proceeds only via the states 0, 1, 2, ..., n, ... such that

$$q_0 = 0,\ p_0 = 1\,. \tag{12}$$

Moreover, we assume in place of (8) that

$$q_n > 0,\ p_n > 0,\ p_n + q_n = 1 \qquad (n > 0), \tag{13}$$

so that it is possible to go from any state other than 0 either to the left or to the right.

Finally, we say that at the actual instant of transition from n to n ± 1 the particle is situated at the point n ± 1, not at the point n. This means that the path x(t) of the particle is regarded as a right continuous function of t. It is shown that in this case our process not only has the Markov property, it in fact has the strong Markov property; the latter is obtained by replacing the fixed time s in the Markov property with any Markov time τ (cf. the corresponding definitions in Chapt. II, §3 and Chapt. III, §2).

Thus, the process x(t) in which we are interested is specified by a set of constants p_n, q_n, and a_n (n = 0, 1, 2, ...) satisfying the conditions (10), (12), and (13), and takes the following form. At the starting time the particle is situated in some state x_0. At a time τ_1, which has an exponential distribution with parameter a_{x_0}, the particle transfers to the new state x_1. Here $x_1 = x_0 - 1$ with a probability q_{x_0}, and $x_1 = x_0 + 1$ with a probability p_{x_0}. The particle remains in the state x_1 until τ_2, at which time it transfers to the next state x_2. Now $\tau_2 - \tau_1$ has an exponential distribution with the parameter a_{x_1}, $x_2 = x_1 - 1$ with a probability q_{x_1}, and $x_2 = x_1 + 1$ with a probability p_{x_1}. In general, at the time τ_n the particle transfers from x_{n-1} to x_n. Then $\tau_n - \tau_{n-1}$ has an exponential distribution with the parameter $a_{x_{n-1}}$, $x_n = x_{n-1} - 1$ with a probability $q_{x_{n-1}}$, and $x_n = x_{n-1} + 1$ with a probability $p_{x_{n-1}}$, and so on to infinity. The path x(t) thus constructed is defined on the random half-interval [0, T), where

$$T = \lim_{n \to \infty} \tau_n \tag{14}$$

is the cumulative jump time.

The process x(t) just described is called a *birth and death process*. This name ties in with the interpretation of x(t) as the number of singularities in the time t. A jump of the particle one unit to the right corresponds to the birth of a new singularity, and a jump one unit to the left corresponds to the death of one singularity. Since the only jumps possible from a state n are those to the neighboring states n + 1 and n − 1, in such a model the simultaneous birth (or death) of two or more singularities has a probability zero. From the point of view of applications to biology, this limitation may prove rather confining. However, it is the condition which provides the analogy between a process with a state space {0, 1, 2, ..., n, ...} and a diffusion process on the half-line [0, ∞) with continuous paths; a transition from point x to point y is only possible via all the intermediate states.

Disregarding the biological implications of the birth and death process, we can shift the state of this process from the integer points 0, 1, 2, ..., n, ... to the points of any monotonic sequence $u_0 \leq u_1 \leq \ldots \leq u_n \leq \ldots$. Then the birth and death process turns out to be the discrete analog of a diffusion process on the interval $[u_0, r)$, where $r = \lim_{n \to \infty} u_n$. Now the point r plays the role of a single boundary point of the phase space $\{u_0, u_1, \ldots, u_n, \ldots\}$.

We seek the continuations of the process x(t) after the time T when the jumps cluster (in the case when $T < \infty$). The problem amounts to an investigation of the boundary conditions at the point r. But first we have to find out in which cases $T < \infty$ with positive probability. In the earlier sections we showed how to calculate the escape probabilities and average exit times of a particle from intervals; this is also useful for the investigation of boundary conditions.

§3. The Canonical Scale and Escape Probabilities

From the description given of the birth and death process in §2 it is evident that the sequence of states $x_0, x_1, \ldots, x_n, \ldots$ occupied in succession by the particle comprises a Markov chain with phase space $\{0, 1, 2, \ldots, n, \ldots\}$ and transition probabilities

$$\begin{aligned} p(n, n-1) &= q_n, \\ p(n, n+1) &= p_n, \\ p(n, m) &= 0 \qquad (|m-n| \neq 1). \end{aligned}$$

If we study the structure of this chain, we then find ourselves in a position to answer a great many of the problems bearing on the birth and death process x(t). Thus, we ascertain the probability of arriving at one state before another, the probability of ever reaching a given state, etc. Of course, all the problems touching on the rate of displacement of the particle among the states re-

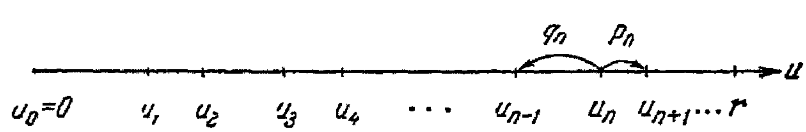

Fig. 36

main open; their answer requires consideration of the values of the constants a_n as well.

Let us suppose that $p_n = q_n = 1/2$ for all $n \geq 1$. Then the chain x_n represents the discrete counterpart of a Wiener process on the half-line $[0, \infty)$ with reflection at zero. [In order to generate a Wiener process with reflection at zero, it is sufficient to consider the process $|x(t)|$, where $x(t)$ is the Wiener process on the entire line.] Moreover, the chain x_n can also be obtained directly from a Wiener process with reflection; precisely the same chain is formed by the successive distinct integer-valued points visited by the Wiener path. In fact, after any integer-valued point $n \geq 1$ the Wiener path has a probability 1/2 of arriving at the integer-valued point $n-1$ and a probability 1/2 of arriving at the point $n+1$ (and it goes from zero to one with absolute certainty). It is not necessary to analyze the Wiener path specifically at integer-valued points; a Markov chain with the same transition probabilities is obtained if we analyze the arrivals of the Wiener particle at any equidistant points $u_0 = 0 < u_1 < u_2 < \ldots < u_n < \ldots$.

What happens if these points are chosen at unequal distances (Fig. 36)? The end result is a Markov chain with a phase space $u_0 = 0, u_1, \ldots, u_n, \ldots$, and the probabilities q_n and p_n of transitions to the left and right from u_n are equal to, respectively,

$$q_n = \frac{u_{n+1} - u_n}{u_{n+1} - u_{n-1}},$$
$$p_n = \frac{u_n - u_{n-1}}{u_{n+1} - u_{n-1}} \qquad (15)$$

[see Eq. (17) from Chapt. III, §7]. If we arrange the points u_n such that the probabilities q_n and p_n calculated according to Eqs. (15) coincide with the probabilities q_n and p_n in the given birth and death process, the Wiener process with reflection will induce at the states $\{u_0, u_1, \ldots, u_n, \ldots\}$ exactly the same Markov chain as generated by the birth and death process at the states $\{0, 1, \ldots, n, \ldots\}$.

Letting $u_1 = 1$ for definiteness, we successively find the u_n from Eqs. (15). Let

$$\delta_n = u_{n+1} - u_n \qquad (n \geqslant 0) \qquad (16)$$

be the distance between two successive points. It follows from (15) that

$$\delta_n = \frac{q_n}{p_n}\delta_{n-1} \qquad (n \geqslant 1)$$

and, hence,

$$\delta_n = \frac{q_1 \ldots q_n}{p_1 \ldots p_n} \qquad (n \geqslant 1). \tag{17}$$

Consequently,

$$u_0 = 0,\ u_1 = 1,\ u_n = \delta_0 + \ldots + \delta_{n-1} = 1 + \frac{q_1}{p_1} + \ldots + \frac{q_1 \ldots q_{n-1}}{p_1 \ldots p_{n-1}} \qquad (n \geqslant 1). \tag{18}$$

We call the number u_n the c a n o n i c a l c o o r d i n a t e of the state n and call the u axis with the points $u_0, u_1, \ldots, u_n, \ldots$ set off on it the c a n o n i c a l s c a l e of the given birth and death process. At first we examine this process only in the canonical scale and consider that the particle x(t) or x_n does not move via integer-valued points, but via the points u_n. We agree to let E represent the phase space $\{u_0, u_1, \ldots, u_n, \ldots\}$. We note that specification of the canonical scale is equivalent to specification of the constants q_n and p_n.

In the previous integer-valued scale the states had a limit point $+\infty$. In the canonical scale this limit corresponds to the number

$$r = \lim_{n\to\infty} u_n = 1 + \sum_{n=1}^{\infty} \frac{q_1 \ldots q_n}{p_1 \ldots p_n}, \tag{19}$$

which we call the b o u n d a r y of the phase space E. The boundary r can be either infinite or finite.

For example, if $p_n = p$, $q_n = q$ for all $n \geq 1$, then the series (19) comprises a geometric progression and, hence, $r = p/(p-q)$ for $p > q$, and $r = \infty$ for $p \leq q$. In the original scale, for $p < q$ the particle

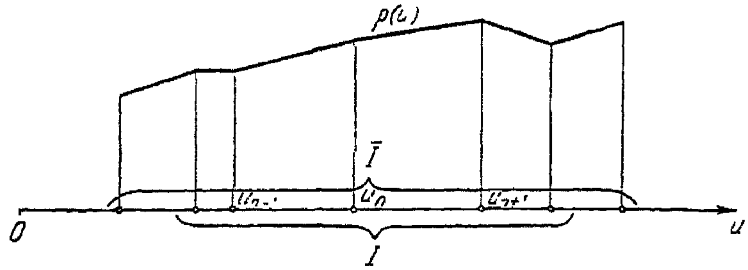

Fig. 37

had a tendency to drift to the left, while for $p > q$ it tended to drift to the right. In the canonical scale the particle executes equal oscillations to the right and left on the average in both cases, but clearly for $p > q$ the states are more densely populated on the right, for $p < q$ on the left.

We turn our attention now to finding the escape probabilities for a Markov chain $x_0, x_1, \ldots, x_n, \ldots$. Of course, they are easily obtained once the escape probabilities from the interval for the Wiener process are known. However, we wish to calculate these probabilities independently of the Wiener process, because the general solution thus obtained for Eq. (20) will meet our needs later on.

We agree to consider any set of states satisfying an inequality of the type $\alpha < u_k < \beta$ (α and β are given numbers) as an interval I in the phase space E. We interpret an expanded interval $\bar{I}$ as the interval generated by the attachment to I of one contiguous state each to the right and left of I (provided such states exist). For example, if $I = \{u_3, u_4\}$, then $\bar{I} = \{u_2, u_3, u_4, u_5\}$; if $I = \{u_0, u_1\}$, then $\bar{I} = \{u_0, u_1, u_2\}$; and if I is congruent with the total phase space E, then $\bar{I} = I$.

Let $p(u)\,(u \in \bar{I})$ be the probability that the particle, on initiating its motion at the state u, will occupy a certain fixed state at the time of first exit from the interval I. According to the total probability formula,

$$p(u_n) = q_n p(u_{n-1}) + p_n p(u_{n+1}) \qquad (u_n \in I). \tag{20}$$

We now investigate the solutions of Eq. (20). If $n \neq 0$, then, according to Eqs. (15) and (16),

$$q_n = \frac{\delta_n}{\delta_{n-1} + \delta_n}, \qquad p_n = \frac{\delta_{n-1}}{\delta_{n-1} + \delta_n}, \tag{21}$$

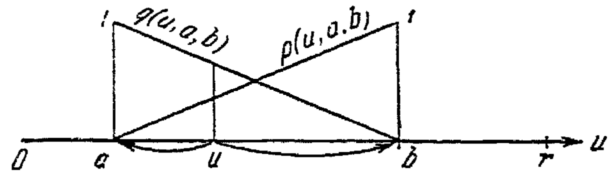

Fig. 38

and Eq. (20) assumes the form

$$\frac{p(u_{n+1}) - p(u_n)}{\delta_n} = \frac{p(u_n) - p(u_{n-1})}{\delta_{n-1}}; \qquad (22)$$

for n = 0 we have $q_n = 0$, $p_n = 1$, and, consequently, $p(u_1) = p(u_0)$ or, for uniformity,

$$\frac{p(u_1) - p(u_0)}{\delta_0} = 0. \qquad (23)$$

The function p(u) is defined only at the discrete sequence of points $u_n \in \bar{I}$. We augment the definition of this function linearly on each of the intervals (u_n, u_{n+1}), where $u_n, u_{n+1} \in \bar{I}$ (Fig. 37). Then the fraction $[p(u_{n+1}) - p(u_n)]/\delta_n$ acquires a simple geometric interpretation; it is equal to the slope of the graph of p(u) on the interval (u_n, u_{n+1}), i.e., it is equal to p'(u) for $u \in (u_n, u_{n+1})$. Stressing the dependence of this derivative on the canonical scale u, we conditionally designate it $D_u p$. The derivative $D_u p$ is not defined at points $u = u_n$.

Equation (22) implies that the adjacent segments meeting at the point on the graph corresponding to u_n have equal slope. Equation (23) tells us that the first segment of this polygonal graph is horizontal. Consequently, if $u_0 = 0 \bar{\in} I$, then all functions linear on $\bar{I}$ are solutions of Eq. (20), and if $0 \in I$ then the solutions are functions constant on $\bar{I}$.

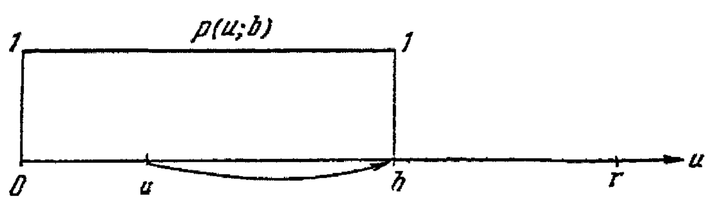

Fig. 39

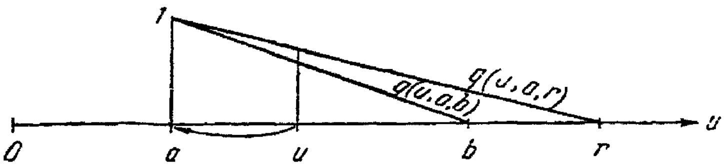

Fig. 40

Let the interval I not contain zeros and consist of a finite number of states. Then at the time τ of exit from I the particle can be situated in either of the two state $a < b$ adjoining I and forming together with I the expanded interval $\bar{I}$ (we assume that the initial state u belongs to $\bar{I}$). The probabilities of the events $x(\tau) = a$ and $x(\tau) = b$ are designated, respectively, $q(u; a, b)$ and $p(u; a, b)$. According to what has been demonstrated thus far, both of these functions are linear on the interval $a \leq u \leq b$. Moreover, their values at the end points of that interval are known: $q(a; a, b) = p(b; a, b) = 1$, $q(b; a, b) = p(a; a, b) = 0$. Therefore,

$$q(u;\ a,\ b) = \frac{b-u}{b-a}, \qquad p(u;\ a,\ b) = \frac{u-a}{b-a} \qquad (a \leqslant u \leqslant b) \tag{24}$$

(Fig. 38). As was to be expected, we have obtained the same relations as for a Wiener process on a line.

If the interval I contains the state 0 and is bounded on the right by the state b, the probability $p(u; b)$ of escape at the point b is constant on the interval $0 \leq u \leq b$. Since it becomes one at $u = b$, we have

$$p(u;\ b) = 1 \tag{25}$$

(Fig. 39). Consequently, the particle will reach with probability one arbitrarily distant states b.

Finally, we examine an interval I including all the states situated to the right of a given state a, i.e., the interval bounded by the state a and the limit point of the phase space, r. For the probability $q(u; a, r)$ of escape from I at a we have here only one boundary condition at the point a.* Consequently, we find $q(u; a, r)$

*If $r = \infty$, the role of the boundary condition at the point r is taken by the requirement of boundedness on the part of the function q; for $r < \infty$, however, boundedness is not adequate for the unique definition of q.

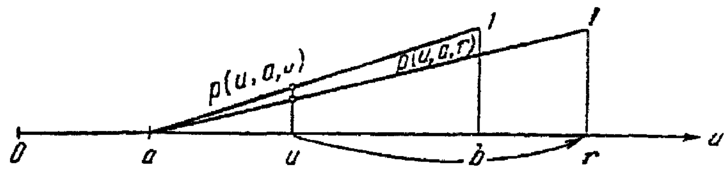

Fig. 41

by passage to the limit. We observe that the events A_b = { In the sequence $x_0 = u, x_1, x_2, \ldots, x_n, \ldots$ the state a is encountered before b} expand with increasing b, and their union yields the event A = { In the sequence $x_0 = u, x_1, x_2, \ldots, x_n, \ldots$ the state a is encountered}. Therefore, $q(u;\ a, r) = \mathrm{P}_u\{A\} = \lim_{b \uparrow r} P_u\{A_b\} = \lim_{b \uparrow r} q(u;\ a, b)$, and from (24) we obtain*

$$q(u;\ a,\ r) = \begin{cases} \dfrac{r-u}{r-a} & \text{for } r < \infty, \\ 1 & \text{for } r = \infty \end{cases} \tag{26}$$

(Fig. 40). Completely analogously, if we interpret p(u; a, r) as the probability, after leaving u, of being situated in arbitrarily distant states without once visiting a, then $p(u;\ a,\ r) = \lim_{b \uparrow r} p(u;\ a,\ b)$; hence

$$p(u;\ a,\ r) = \begin{cases} \dfrac{u-a}{r-a} & \text{for } r < \infty, \\ 0 & \text{for } r = \infty \end{cases} \tag{27}$$

(Fig. 41).

§ 4. Repelling and Attracting Boundaries

The behavior of a Markov chain

$$x_0,\ x_1,\ x_2,\ \ldots,\ x_n,\ \ldots \tag{28}$$

as $n \to \infty$ [and, consequently, the path-of the birth and death process x(t) as $t \uparrow T$] is largely dependent on whether the number r is finite or infinite.

First let $r = \infty$. Then it follows from Eqs. (25) and (26) that from any state u the particle has a probability one of arriving

* As usual, P_u and M_u denote the probability and mathematical expectation for the initial state u.

sometime at any other state. On going from u to v, the particle then has a probability one of going from v to u. Consequently, the probability of returning from u to u is equal to one (it is customarily said that the chain is *recurrent*). Clearly, if the chain is recurrent, any fixed state, for example, $u_0 = 0$, will be visited with probability one in the sequence (28) an infinite number of times (every exit from 0 is followed by a return to 0). Inasmuch as an infinite number of returns to 0 prevents the sequence (28) from tending to the limit r, the probability of x_n tending to r is equal to zero. We say that the boundary in this case *repels* the particle.

If $r < \infty$, it then follows from Eq. (26) that the probability of visiting a given state a from any state to the right of a is less than one. Since it is possible with positive probability to go from a to the right, this means that the probability β of returning to a from a is less than one (the chain $\{x_n\}$ is nonrecurrent). The probability of returning to a from a at least m times is equal to β^m (the probabilities are multiplicative because the corresponding events are independent). Consequently, the probability of an infinite number of returns to a is equal to $\lim_{m\to\infty} \beta^m = 0$. Therefore, the state a is visited with probability one in the sequence (28) a finite number of times. Since the number of states is denumerable, this is true with probability one for all states. But if every state is visited in the sequence (28) a finite number of times, this sequence inevitably tends to the limit r. Hence,

$$\mathbf{P}\{\lim_{n\to\infty} x_n = r\} = 1, \tag{29}$$

or, so to speak, the boundary r *attracts* the particle.

Thus, for $r = \infty$ the process is recurrent and has a repelling boundary, while for $r < \infty$ it is nonrecurrent and has an attracting boundary.

The further classification of boundaries now depends on the rate of displacement of the particle among the states. We must therefore return from the chain x_n ($n = 0, 1, 2, \ldots$) to the original birth and death process $x(t)$ $(0 \leq t < T)$.

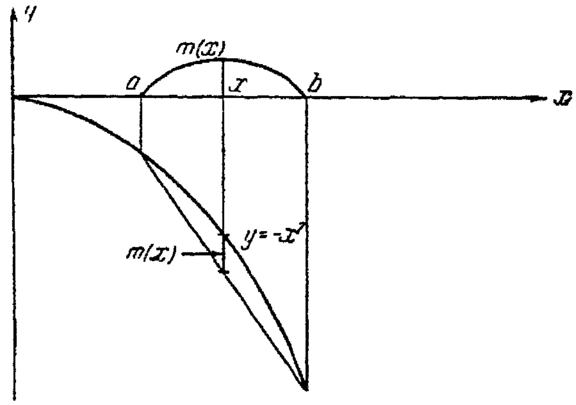

Fig. 42

§5. The Characteristic, Average Exit Time, and Velocity Measure

We now recall how the average exit time from an interval is determined in the case of a Wiener process on a line. According to the results of Chapt. II, §8, the average time $m(x; a, b)$ to exit from x at either of the end points a or b ($a \leq x \leq b$) is obtained by adding to the function $-x^2$, which is chosen once and for all, a linear function such that $m(x; a, b)$ goes to zero at the points $x=a$ and $x=b$. This time is expressed geometrically by a vertical line segment between the parabola $y=-x^2$ and the chord linking those points of the parabola having abscissas $x=a$ and $x=b$ (Fig. 42). We can also say that $m(x; a, b)$ is a solution of the Poisson equation

$$\frac{d^2}{dx^2} m = -2$$

with zero boundary conditions at the points a and b (in the one-dimensional case $\Delta = d^2/dx^2$). We will see that the situation is analogous in the case of a birth and death process, except that the role of the parabola $y=-x^2$ is taken by certain broken lines, which differ for different processes.

Let I be some interval of states of a birth and death process $x(t)$, and let $m(u)$ be the expectation of the time τ to exit from I under the condition that the particle is situated in the state u at the starting time. We assume in this case that $\tau=0$ if $x(0) \notin I$ and that $\tau=T$ if $x(t) \in I$ for all $t<T$. In particular, T is the time of first exit from the total phase space $E=\{u_0, u_1, \ldots, u_n, \ldots\}$.

We begin by showing that **the average time m(u) is finite for an interval I containing a finite number of states.** Thus, during any fixed time $t > 0$ it is possible with positive probability to go from any given state u to any other state a (this follows from the definition of the birth and death process in §2). Choosing a outside I, we deduce that for every $u \in I$

$$\mathbf{P}_u\{\tau < t\} > 0$$

Since the number of states of the interval I is finite, we have

$$\alpha = \min_{u \in I} \mathbf{P}_u\{\tau < t\} > 0.$$

The inequality $m(u) < \infty$ now ensues from the following general remark.

Let I be any set of states, and let τ be the time of first exit from I. If for some $t < \infty$

$$\mathbf{P}_u\{\tau < t\} \geqslant \alpha > 0$$

for all states $u \in I$, then $\mathbf{P}_u\{\tau < \infty\} = 1$, and $\mathbf{M}_u\tau < \infty$ for $u \in I$

For the proof of this assertion we let $p(u, v)$ $(u, v \in I)$ represent the probability of a transition from u to v in the time t such that the particle remains inside the set I throughout this entire time. By hypothesis,

$$\sum_{v \in I} p(u, v) = \mathbf{P}_u\{\tau > t\} \leqslant 1 - \alpha \qquad (u \in I)$$

Using this estimate, we obtain

$$\mathbf{P}_u\{\tau > 2t\} = \sum_{v \in I} p(u, v)\mathbf{P}_v\{\tau > t\} \leqslant (1-\alpha)\sum_{v \in I} p(u, v) \leqslant (1-\alpha)^2$$

and in general, by induction,

$$\mathbf{P}_u\{\tau > nt\} \leqslant (1-\alpha)^n \qquad (u \in I). \tag{30}$$

Since $\alpha > 0$, it follows from (30) for $n \to \infty$ that $\mathbf{P}_u\{\tau = \infty\} = 0$. Furthermore,

$$\mathbf{M}_u\tau \leqslant \sum_{n=0}^{\infty} (n+1)\, t\mathbf{P}_n \{nt < \tau \leqslant (n+1)t\}$$

$$\leqslant t \sum_{n=0}^{\infty} (n+1)\, \mathbf{P}_u \{nt < \tau\} \leqslant t \sum_{n=0}^{\infty} (n+1)(1-\alpha)^n < \infty$$

$$(u \in I)$$

We now derive an equation for the function m(u), assuming that $m(u) < \infty$. Inasmuch as the average sojourn time in u_n is equal to $1/a_n$, after which time the particle transfers with probabilities q_n and p_n to u_{n-1} and u_{n+1}, respectively, and then the process behaves as if it had begun at these states, we have

$$m(u_n) = \frac{1}{a_n} + q_n m(u_{n-1}) + p_n m(u_{n+1}) \tag{31}$$

for every $u_n \in I$.

Obviously, the difference between the two solutions of Eq. (31) on the interval I satisfies the appropriate homogeneous equation on I, namely Eq. (20), which was studied in §3. We recall that the solutions of the latter equation are functions linear on the expanded interval and, if the interval I contains the point 0, then only functions constant on $\bar{I}$. It is sufficient therefore to find some solution of Eq. (31) on the entire phase space E in order to know all of its solutions on any interval I.

Let us examine in closer detail the solution $S_n = S(u_n)$ of Eq. (31) such that the following initial condition is satisfied:

$$S_0 = 0. \tag{32}$$

We rewrite Eq. (31) for S_n in the form

$$(S_{n+1} - S_n)\, p_n = (S_n - S_{n-1})\, q_n - \frac{1}{a_n}. \tag{33}$$

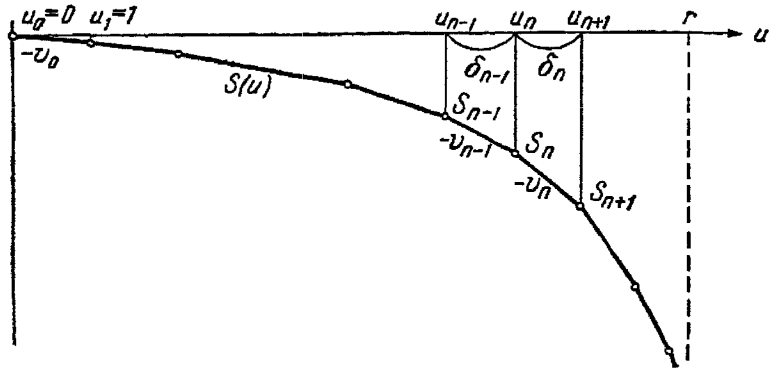

Fig. 43

For $n \neq 0$ the probabilities q_n and p_n are expressed in terms of the distances $\delta_n = u_{n+1} - u_n$ as follows:

$$p_n = \frac{\delta_{n-1}}{\delta_{n-1} + \delta_n}, \quad q_n = \frac{\delta_n}{\delta_{n-1} + \delta_n} \tag{34}$$

[see Eqs. (21) from §3]. Substituting these expressions into (33), we obtain

$$\frac{S_{n+1} - S_n}{\delta_n} = \frac{S_n - S_{n-1}}{\delta_{n-1}} - \frac{1}{a_n} \frac{\delta_{n-1} + \delta_n}{\delta_{n-1}\delta_n}. \tag{35}$$

The first two relations entering into Eq. (35) represent the values of the derivative $D_u S$ on the adjacent intervals (u_n, u_{n+1}) and (u_{n-1}, u_n) [as in §3, we assume that the definitions of the functions $m(u)$ and $S(u)$ are extended by linear interpolation between the states] (Fig. 43). To shorten the notation, we set

$$D_u S(u) = -v(u) \tag{36}$$

and denote the value of $v(u)$ on the interval (u_n, u_{n+1}) by v_n. Equation (35) implies that

$$v_n = v_{n-1} + 2\mu_n \qquad (n \geqslant 1), \tag{37}$$

where

$$2\mu_n = \frac{1}{a_n} \frac{\delta_{n-1} + \delta_n}{\delta_{n-1}\delta_n} \tag{38}$$

is a known quantity (the coefficient 2 is inserted in order to impart a simpler and more straightforward interpretation to μ_n later on). In exactly the same manner we obtain the following from (33) for $n=0$:

$$v_0 = 2\mu_0, \tag{39}$$

where

$$2\mu_0 = \frac{1}{a_0}. \tag{40}$$

It follows from Eqs. (37) and (36) and the initial conditions (39) and (32) that

$$v_n = 2\sum_{k=0}^{n} \mu_k \qquad (n \geqslant 0),$$

$$S_n = -\sum_{m=0}^{n-1} v_m \delta_m = -2 \sum_{0 \leqslant k \leqslant m \leqslant n-1} \mu_k \delta_m \qquad (n \geqslant 1).$$

Since

$$\delta_0 = 1, \quad \delta_m = \frac{q_1 \cdots q_m}{p_1 \cdots p_m} \qquad (m \geqslant 1)$$

[see Eq. (17) of §3], we have by virtue of (38)

$$2\mu_k = \frac{1}{a_k} \frac{p_1 \cdots p_{k-1}}{q_1 \cdots q_k} \qquad (k \geqslant 1)$$

and, finally,

$$v_n = \frac{1}{a_0} + \sum_{k=1}^{n} \frac{1}{a_k} \frac{p_1 \cdots p_{k-1}}{q_1 \cdots q_k} \qquad (n \geqslant 0), \tag{41}$$

$$S_n = -\sum_{0 \leqslant k \leqslant m \leqslant n-1} \frac{1}{a_k} \frac{q_{k+1} \cdots q_m}{p_k \cdots p_m} \qquad (n \geqslant 1). \tag{42}$$

We call the function S(u) the *characteristic* of the birth and death process x(t). It is apparent from Eq. (42) that S(u) is

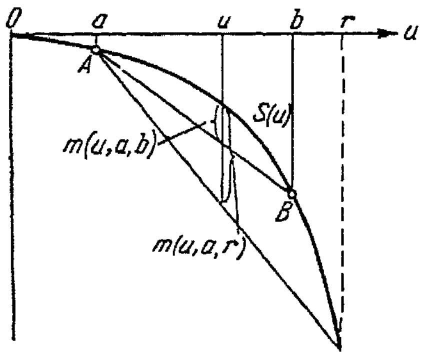

Fig. 44

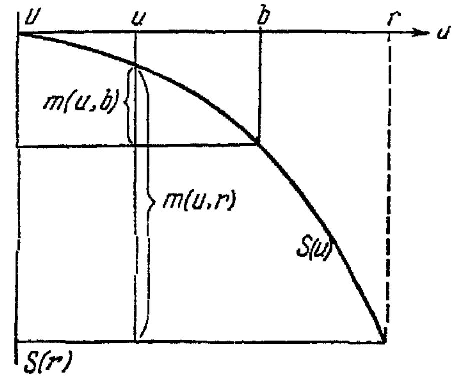

Fig. 45

negative for u > 0 and decreases monotonically. Equation (37) shows that its derivative −v(u) also decreases on passing through each state u_n, so that S(u) is concave.

Inasmuch as the functions S(u) and v(u) are monotonic, they have finite or infinite limits as $u \uparrow r$. These limits are designated S(r) and v(r). It is geometrically apparent that if r = ∞, S(r) = − ∞ [this is also readily obvious from a comparison of Eqs. (41) and (42)].

We denote by m(u; a, b) the average time to exit from the state u to one of the states a or b ($a \leq u \leq b$) and by m(u; b) the average time to arrival of the particle from the state u to the state b ($u \leq b$). The function m(u; a, b) satisfies Eq. (31) for $a < u = u_n < b$ and goes to zero at the points a and b. Due to the properties of the same equation, the difference m(u; a, b) − S(u) is linear on the expanded interval $a \leq u \leq b$. At the points a and b this difference is equal to −S(a) and −S(b), respectively. Constructing such a linear function, we obtain

$$m(u;\ a,\ b) = S(u) - \frac{(b-u)\,S(a) + (u-a)\,S(b)}{b-a} \qquad (43)$$

$$(a \leqslant u \leqslant b).$$

Geometrically, m(u; a, b) is expressed as the distance along the vertical between the graph of the characteristic S(u) and its chord AB (Fig. 44*). The function m(u; b) satisfies Eq. (31) for

*In Figs. 44 and 45 the characteristic is not depicted as a broken line, but as a smooth curve, because the overall picture is more important for our purposes than the details of the broken-line segments.

all $u < b$, including the point $u = 0$, and therefore differs from $S(u)$ on the expanded interval $0 \leq u \leq b$ by a constant. Since $m(b; b) = 0$, we have

$$m(u;\ b) = S(u) - S(b) \qquad (u \leqslant b) \tag{44}$$

(Fig. 45).

If the interval I contains infinitely many states (i.e., if its right end is the limit point r), then the immediately preceding arguments are inapplicable, because we do not know whether the average exit time $m(u)$ is finite. In this case it is necessary to pass to the limit as $b \uparrow r$ in Eqs. (43) and (44). We will show that, given this transition to the limit, the average time $m(u; r) = \mathbf{M}_u T$ to exit from the entire phase space is indeed obtained from Eq. (44) [the passage to the limit in Eq. (43) is justified analogously]. Let τ_b be the time of first arrival of the particle at the point b. Then $m(u; b) = \mathbf{M}_u \tau_b$ for $u \leq b$. The quantity τ_b increases monotonically with b. It is legitimate therefore to pass to the limit in the argument of the expectation,* and we obtain

$$\lim_{b \uparrow r} \mathbf{M}_u \tau_b = \mathbf{M}_u \left\{ \lim_{b \uparrow r} \tau_b \right\}.$$

According to Eq. (25), $\tau_b < T$ with probability one for all $b > u$. Yet we have $\lim\limits_{b \uparrow r} \tau_b \geqslant T$, because up to the time $\lim\limits_{b \uparrow r} \tau_b$ the particle has to make infinitely many jumps. Consequently, $\lim\limits_{b \uparrow r} \tau_b = T$ with probability one. Therefore,

$$\lim_{b \uparrow r} m(u;\ b) = m(u;\ r),$$

and we obtain

$$m(u;\ r) = S(u) - S(r). \tag{45}$$

Passing to the limit in Eq. (43), we find that for $r < \infty$, when the point r attracts the particle, the average time to arrival at *a* from

*See the footnote on p. 104.

u or to the cumulative jump time T is equal to

$$m(u;\ a,\ r) = S(u) - \frac{(r-u)\,S(a) + (u-a)\,S(r)}{r-a}. \tag{46}$$

If $r=\infty$ (the boundary repelling the particle), it is necessary in the calculation of $m(u;\ a,\ r)$ to apply the l'Hôpital rule.

The number μ_n investigated in the formulation of the characteristic S is conveniently regarded as a measure (mass) concentrated at the point u_n. This measure characterizes the rate of displacement of the particle and is therefore called the *velocity measure*. Thus, the particle resides on the average for a period of time $1/a_n$ in the state u_n, after which it is situated with a probability q_n at a distance δ_{n-1} and with a probability p_n at a distance δ_n from the point u_n. Hence, on the average, the particle covers a path $q_n\delta_{n-1}+p_n\delta_n$ during the period $1/a_n$. Dividing the average path by the average time, we obtain, taking account of Eqs (34),

$$a_n(q_n\delta_{n-1} + p_n\delta_n) = \frac{2a_n\delta_{n-1}\delta_n}{\delta_{n-1}+\delta_n} = \frac{1}{\mu_n}.$$

This means that the larger $\boldsymbol{\mu} = \{\mu_0,\ \mu_1,\ \ldots,\ \mu_n,\ \ldots\}$ is, the slower the particle moves along the canonical scale.

Equation (35) can be given a simpler form if we introduce the notion of the *derivative* D_μ *with respect to the measure* μ. Let v(u) be a function constant on each of the intervals $(u_n,\ u_{n+1})$ and unbounded at the points u_n. The derivative of this function with respect to the measure $\boldsymbol{\mu}$ at a point u_n is defined as the number

$$D_\mu v(u_n) = \frac{v(u'') - v(u')}{\mu_n},$$

where

$$u' \in (u_{n-1},\ u_n), \qquad u'' \in (u_n,\ u_{n+1}).$$

In the new notation Eq. (37) assumes the form $D_\mu v = 2$, and since $v = -D_u S$, we obtain the following equation for S·

$$D_\mu D_u S(u_n) = -2.$$

The reader may easily verify that if he sets $D_u S(u) = 0$ for $u < 0$ [i.e., regards the graph of S(u) as horizontal to the left of 0], then this equation will be valid not only for $n \geq 1$, but for $n = 0$ as well.

The same Eq. (31) for m(u) leads to the representation

$$D_\mu D_u m(u_n) = -2. \qquad (u_n \in I).$$

Comparing this equation with the Poisson equation $\Delta m = -2$ for the average time m(x) in the case of a Wiener process, we see that the operator $D_\mu D_u$ is the analog of the Laplace operator Δ.

As noted in §3, specification of the canonical scale $\{u_0 = 0 < u_1 < u_2 < \ldots < u_n < \ldots\}$ is equivalent to specification of the jump probabilities q_n and p_n. It is apparent from the relation defining μ_n that the parameter a_n is uniquely specified by μ_n for known numbers q_k and p_k. Consequently, the canonical scale and velocity measure may be specified instead of the constants q_n, p_n, and a_n of the definition of the birth and death process. An arbitrary increasing sequence $0 = u_0 < u_1 < \ldots < u_n < \ldots$ may be used as the canonical scale, and an arbitrary sequence of positive numbers μ_0, μ_1, ..., μ_n, ... may be used as the velocity measure.

Inasmuch as the numbers of the canonical scale are the points at which sharp bends occur in the graph of S(u), the canonical scale of the process is found from the known characteristic; then the constants a_n are determined. Thus, the birth and death process is also completely specified by its characteristic S(u). Any diminishing concave piecewise-linear continuous function on the half-interval [0, r), equal to zero at $u = 0$ and having a denumerable number of segments and a unique limit point r for the sharp bends, may serve as the characteristic.

§6. Accessible and Inaccessible Boundaries

Let us see how the path of a birth and death process x(t) depends on the finiteness or nonfiniteness of the variable S(r).

If $S(r) > -\infty$, $\mathbf{M}_u T$ is also finite for any initial state u, according to Eq. (45). Consequently,

$$\mathbf{P}_u\{T < \infty\} = 1 \text{ for all } u.$$

Now let $S(r) = -\infty$. We will show that in this case

$$\mathbf{P}_u\{T = \infty\} = 1 \quad \text{for all } u. \tag{47}$$

Suppose that for some u

$$\mathbf{P}_u\{T < \infty\} > 0. \tag{48}$$

Since it is possible to arrive at u from the state $u_0 = 0$ in a finite time, we then have

$$\mathbf{P}_0\{T < \infty\} > 0.$$

Consequently, there exists a $t < \infty$ such that

$$\alpha = \mathbf{P}_0\{T < t\} > 0.$$

Inasmuch as the probability of visiting any state u from 0 is equal to one, the $\mathbf{P}_0$-probability is one that

$$T = \tau + T',$$

where τ is the time to arrival at u from 0, and T' is the time to the accumulation of an infinite number of jumps, beginning with the time of arrival at u. According to the strong Markov property, T' does not depend on τ and has the same distribution as T, given the condition $x(0) = u$. Therefore, denoting by F(s) the distribution function of the random variable τ, we write

$$\alpha = \mathbf{P}_0\{T < t\} = \int_0^t \mathbf{P}_u\{T < t - s\}\, dF(s)$$

$$\leqslant \int_0^t \mathbf{P}_u\{T < t\}\, dF(s) \leqslant \mathbf{P}_u\{T < t\}. \tag{49}$$

According to the remark made at the beginning of §5, the estimate (49) tells us that $M_u T < \infty$ (for all u). It follows from the assumption $S(r) = -\infty$ and Eq. (45), on the other hand, that $\mathbf{M}_u T = \infty$. Hence, the assumption (48) contradicts the condition $S(r) = -\infty$, thus proving Eq. (47).

We have found, therefore, that the processes in question fall into two classes, depending on whether S(r) is finite or infinite: in one class the cumulative jump time T is finite with probability one for any initial state u; in the other the time T is infinite with probability one for any u. We say then that the boundary r is accesible in the former instance and inaccessible in the latter.

Let us set this classification in juxtaposition with the separation into attracting and repelling boundaries (see §4). The condition $r = \infty$ implies that $S(r) = -\infty$, in consequence of which a repelling boundary is inaccessible. Therefore, we have in fact three distinct types of boundaries:

I. An accessible boundary $[r < \infty, S(r) > -\infty]$. With probability one the time T is finite, and $\lim_{t \uparrow T} x(t) = r$.

II. An attracting inaccessible boundary $[r < \infty, S(r) = -\infty]$. With probability one $T = \infty$, and $\lim_{t \uparrow T} x(t) = r$.

III. A repelling boundary $[r = \infty, S(r) = -\infty]$. With probability one $T = \infty$, and as $t \uparrow T$ the particle traverses all states infinitely many times.

Analytically, the condition of accessibility of a boundary is contained in the convergence of the double series

$$-S(r) = \sum_{0 \leqslant k \leqslant m < \infty} \frac{1}{a_k} \frac{q_{k+1} \cdots q_m}{p_k \cdots p_m},$$

which is derived from the expression (42) for S_n as $n \to \infty$. For example, let $p_n = p$, $q_n = q$ be independent of n $(n \geq 1)$. As already mentioned in §3, for $p \leq 1/2$ we have $r = \infty$, and for $p > 1/2$ we have $r = p/(p-q)$. Therefore, if $p \leq 1/2$, we have a repelling boundary, but if $p > 1/2$, we have an attracting boundary. Let $p > 1/2$, and let us ascertain for which a_n the boundary r is accessible. Since in this case, fixed k,

$$\sum_{m=k}^{\infty} \frac{1}{a_k} \frac{q_{k+1}\cdots q_m}{p_k\cdots p_m} = \frac{1}{a_k p} \sum_{l=0}^{\infty} \left(\frac{q}{p}\right)^l = \frac{1}{a_k(p-q)},$$

we have

$$S(r) = \frac{-1}{p-q} \sum_{k=0}^{\infty} \frac{1}{a_k}.$$

Consequently, in order for the boundary to be accessible, it is necessary and sufficient that the series $\sum \frac{1}{a_k}$ converge. It is easily verified that this condition also prevails in the degenerate case when q = 0, p = 1.*

§7. Continuations of the Birth and Death Process; Statement of the Problem

We are now prepared to study the continuations of the birth and death process after the time T of first accumulation of jumps. Clearly, the problem is devoid of meaning when the boundary r is inaccessible (in this case $T = \infty$ with probability one). We take it for granted, therefore, that r and S(r) are finite. Then

$$\mathbf{P}_u\left\{T < \infty, \lim_{t \uparrow T} x(t) = r\right\} = 1, \tag{50}$$

i.e., almost all paths of the process arrive at the point r at a time $T < \infty$. We can neglect the exceptional event of probability zero, assuming that this property is true for all paths of the process.

We adjoin the point r to the phase space. Thus, our phase space E now consists of the sequence of isolated points $0 = u_0 < u_1 < \ldots < u_n < \ldots$ and the point r, comprising their limit. The function $f(u)$ $(u \in E)$ is continuous when and only when $f(u_n) \to f(r)$.

Let us suppose that there exists some continuation of the birth and death process. This implies the following:

* This example is analyzed in detail in Feller's book [1].

1) The path x(t), originally defined for $t \in [0, T)$, is continued onto some interval $[0, \zeta)$ $(T \leq \zeta \leq \infty)$.* Here the values of x(t) can be the points u_n, as well as the point r.

2) The probability distributions $\mathbf{P}_u$ are extended over a broader class of events, which are specified by the entire behavior of the continued process. They remain unchanged for events that are defined in terms of the path up to the instant T. Moreover, an auxiliary distribution P_r is introduced, corresponding to the initial point r.

3) The continued process is strongly Markovian, i.e., for any Markov time $\tau < \zeta$ the process $y(t) \equiv x(\tau + t)$ under the conditions $x(\tau) = u$ $(u \in E)$ does not depend on the behavior of x(t) before the time τ and has the same probability distribution as the process x(t), beginning with the time 0 at the point u.

We narrow the problem somewhat with the following additional assumption:

4) The path of the continued process remains right continuous, i.e., $x(t+h) \to x(t)$ as $h \downarrow 0$. (This means that a particle situated at time t in a state $u \neq r$ stays at the point u for some positive period of time; a particle situated at the point r cannot travel far from r in a short period of time.)

5) $x(T) = r$ [this is entirely reasonable on the basis of (50)].

We note that the assumption 5 and the condition $T < \infty$ convert the probabilities

$$q(u;\ a,\ r) = \frac{r-u}{r-a}, \qquad p(u;\ a,\ r) = \frac{u-a}{r-a}, \tag{51}$$

which were found at the end of §3, into the probabilities of a path initiated from u arriving at the point a before r, and at the point r before a, respectively.

We sometimes refer to processes continued after the time T with observance of the conditions 1 - 5 as class A processes. It turns out that a class A process is uniquely speci-

* The time ζ represents the instant of termination of the path (extinction of a particle executing a random walk). If the process does not terminate, $\zeta = \infty$.

fied by its characteristic operator $\mathfrak{A}$ (The proof of this assertion is given in §11 with reference to the general theory of Markov processes). We see that in all states except r the operator $\mathfrak{A}$ is completely defined in terms of the behavior of the process up to the time T. Consequently, different continuations of the process x(t) are uniquely described by the form of the characteristic operator at the boundary point (boundary condition).

As already mentioned, the characteristic operator $\mathfrak{A}$ is defined by the equation

$$\mathfrak{A} f(u) = \lim_{U \downarrow u} \frac{M_u f(x(\tau)) - f(u)}{M_u \tau} \qquad (u \in E), \tag{52}$$

where U is a shrinking neighborhood about the state u, and τ is the time of first exit of the path from U. Here, if $x(t) \in U$ for all $t < \zeta$, then τ is set equal to ζ, and this value is included in the calculation of $M_u \tau$; in the calculation of $M_u f(x(\tau))$, on the other hand, paths for which $\tau = \zeta$ are not taken into account, because $x(\zeta)$ is not meaningful. In the event $M_u \tau = \infty$, the entire fraction in Eq. (52) is assumed equal to zero. We assume that the operator $\mathfrak{A}$ is defined for all bounded functions f(u) on E for which the limit on the right side of Eq. (52) exists and is finite for any $u \in E$.*

For any u_n the neighborhood U can be made small enough to exclude states other than u_n. Then τ is the time of first exit from u_n, the mathematical expectation of which is equal to $1/a_n$, and $x(\tau)$ is the point u_{n-1} with probability q_n and the point u_{n+1} with probability p_n. The fraction in Eq. (52) then becomes

$$\frac{q_n f(u_{n-1}) + p_n f(u_{n+1}) - f(u_n)}{\frac{1}{a_n}} \tag{53}$$

and does not change with a further shrinkage of the neighborhood U. Consequently, the expression (53) is then $\mathfrak{A} f(u_n)$. The equation for $\mathfrak{A} f(u_n)$ can be written in a very nice form by invoking the derivatives D_u and D_μ introduced in §§3 and 5. Inasmuch as

$$q_n = \frac{\delta_n}{\delta_{n-1} + \delta_n}, \quad p_n = \frac{\delta_{n-1}}{\delta_{n-1} + \delta_n} \quad (n \geqslant 1),$$

*See the footnote on page 72.

we find from (53), after a few straightforward manipulations,

$$\mathfrak{A}f(u_n) = \frac{\dfrac{f(u_{n+1}) - f(u_n)}{\delta_n} - \dfrac{f(u_n) - f(u_{n-1})}{\delta_{n-1}}}{\dfrac{1}{a_n}\dfrac{\delta_{n-1}+\delta_n}{\delta_{n-1}\delta_n}}$$

$$= \frac{D_u f(u'') - D_u f(u')}{\dfrac{1}{a_n}\dfrac{\delta_{n-1}+\delta_n}{\delta_{n-1}\delta_n}},$$

where

$$u' \in (u_{n-1}, u_n), \quad u'' \in (u_n, u_{n+1}).$$

Having obtained $2\mu_n$ in the denominator [see Eq. (38)], we have finally

$$\mathfrak{A}f(u_n) = \frac{1}{2} D_\mu D_u f(u_n). \tag{54}$$

This equation is equally valid for $n = 0$ if we set $D_u f(u) = 0$ for $u < 0$, i.e., regard the function f as constant to the left of zero.

Now we are in a position to graph the analogy between the Laplace operator Δ and the operator $D_\mu D_u$, as remarked in §5 while determining the average exit time $m(u)$. In fact, the equations $\Delta m = -2$ and $D_\mu D_u m = -2$, which we had in the case of the Wiener process and in the case of the birth and death process, may be unified into the form

$$\mathfrak{A}m = -1. \tag{55}$$

Equation (55) is valid under very general assumptions regarding the process $x(t)$. Thus, let $m(x) = \mathbf{M}_x \tau$, where τ is the time of first exit of the path from a certain set I. Let us suppose that $m(y) < \infty$ for all y and that x is an interior point of the set I. Exluding x in a neighborhood U contained in I, we write

$$\tau = \tau_U + \tau', \tag{56}$$

where τ_U is the time of first exit from U, and τ' is the time from the first exit from U to the first exit from I. If the process has the strong Markov property, then under the condition $x(\tau_U) = y$ the conditional expectation of τ' is equal to $m(y)=m(x(\tau_U))$.* Therefore $M_x\tau'=M_x m(x(\tau_U))$, and from (56) we obtain

$$m(x)=M_x\tau_U+M_x m(x(\tau_U)).$$

Hence, by definition of the characteristic operator,

$$\mathfrak{A}m(x)=\lim_{U\downarrow x}\frac{M_x m(x(\tau_U))-m(x)}{M_x\tau_U}=\lim_{U\downarrow x}(-1)=-1.$$

We observe that the operator

$$L=a(x)\frac{\partial^2}{\partial x^2}+b(x)\frac{\partial}{\partial x}, \tag{57}$$

which corresponds to an arbitrary diffusion process on a line, also reduces to the form $\frac{1}{2}D_\mu D_u$ (we recall that we introduced the birth and death process as a discrete analog of the diffusion process). We denote by $D_u f$ the derivative of the function $f(x)$ with respect to the increasing function $u(x)$:

$$D_u f(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{u(x+h)-u(x)}.$$

Also, we let $D_\mu f$ represent the derivative of the function $f(x)$ with respect to the function $v(x)$ specified by the relation

$$v(x)=\int_0^x \mu(y)\,dy,\quad \mu(y)>0.$$

Let us assume that the functions u and f are twice differentiable and that the density μ is continuous. Then

$$D_u f(x)=\frac{f'(x)}{u'(x)},\qquad D_\mu f(x)=\frac{f'(x)}{\mu(x)}.$$

---

*See the footnote on page 125.

Therefore

$$\frac{1}{2} D_\mu D_u f = \frac{u' f'' - u'' f'}{2\mu u'^2},$$

and in order to satisfy the equation

$$af'' + bf' = \frac{1}{2} D_\mu D_u f,$$

it is sufficient to set

$$u(x) = \int_0^x e^{-\int_0^y \frac{b(z)}{a(z)} dz} dy,$$

$$\mu(x) = \frac{1}{2a(x)} e^{\int_0^x \frac{b(y)}{a(y)} dy}$$

If the functions u(x) and v(x) are not differentiable, the operator $\frac{1}{2} D_V D_U$ is not written in the form (57). However, even in this case the same operator corresponds to a Markov process with continuous paths (it is sufficient to require that u and v be definitely increasing and that u be continuous).

The escape probabilities from the interval and the average exit time are expressed in terms of the functions u(x) and

$$S(x) = -\int_0^x v(y)\, du(y)$$

in the continuous case by the same relations as in the case of the birth and death process.

We now concern ourselves with an investigation of the form of the characteristic operator at the boundary point r. We give a complete analysis of all the possibilities in §§8–10. Here we merely bring out some preliminary considerations.

The neighborhood U of the point r is conveniently specified by indicating the rightmost state u_n not contained in U. We denote this state by y, the corresponding neighborhood of the point r by U_y, and the time of first exit from U_y by τ_y; we also let

$$\pi_y(u) = \mathbf{P}_r\{x(\tau_y) = u\}, \qquad m(y) = \mathbf{M}_r\tau_y. \tag{58}$$

Then the equation for the characteristic operator at the point r assumes the form

$$\mathfrak{A}f(r) = \lim_{y \uparrow r} \frac{\sum\limits_{0 \leqslant u \leqslant y} f(u)\,\pi_y(u) - f(r)}{m(y)}. \tag{59}$$

Consequently, the problem of finding all possible forms of the operator $\mathfrak{A}$ at the point r and, hence, all the processes that interest us is solved if we find all the distributions

$$\pi_y = \{\pi_y(u_0),\ \pi_y(u_1),\ \ldots,\ \pi_y(y)\}$$

and all the average times m(y) compatible with the requirements 1 - 5 imposed on the continued process.

A rather singular situation occurs in the case when the boundary r is absorbing, i.e., when a particle arriving at r never exists beyond r. In this case $m(y) = +\infty$, $\pi_y(u) = 0$, and, according to Eq. (59),

$$\mathfrak{A}f(r) = 0. \tag{60}$$

We will show that if the boundary r is not absorbing, then for all y

$$\mathbf{P}_r\{\tau_y < \infty\} = 1$$

and

$$\mathbf{M}_r\tau_y < \infty.$$

Clearly, it is sufficient to regard only the time τ_0, because $\tau_y \leq \tau_0$ for all y. We once again invoke the remark made at the beginning of §5. According to Eq. (45), the average time to arrival at r from u is equal to

$$m(u;\ r)=S(u)-S(r)=|S(r)|-|S(u)|\leqslant|S(r)|$$

[we recall that $S(u)\leq 0$]. According to the Chebyshev inequality, it follows that

$$\mathbf{P}_u\{T>t\}\leqslant\frac{|S(r)|}{t},$$

so that for some $t_0<\infty$

$$\mathbf{P}_u\{T<t_0\}>\frac{1}{2} \qquad (61)$$

simultaneously for all u. If the boundary r is not absorbing, the particle has a positive probability of exiting from r and, therefore, of escaping from some neighborhood U_y of the point r. At the time of exit from U_y the particle is either situated in one of the states $u\leq y$ or becomes extinct. In the former case the particle can go from y to 0 with positive probability, thus emerging from U_0, while in the latter case it simultaneously escapes from both U_y and from U_0. Consequently, if the boundary r does not absorb the particle, the probability of escape from U_0 is positive, so that

$$\mathbf{P}_r\{\tau_0<t_1\}=\alpha>0 \qquad (62)$$

for some $t_1<\infty$. It follows from (61) and (62) that wherever the particle happens to be situated inside U_0, the probability is greater than $\alpha/2$ that it will be able to arrive at r in the period t_0 and that after arrival at r it will be able in the period t_1 to escape from U_0. But in this case, having initiated its motion at the point u, it succeeds in the time t_0+t_1 to escape from U_0, and we find that for all $u\in U_0$

$$\mathbf{P}_u\{\tau_0<t_0+t_1\}>\frac{\alpha}{2}>0,$$

and the truth of our assertion is implied by this estimate.

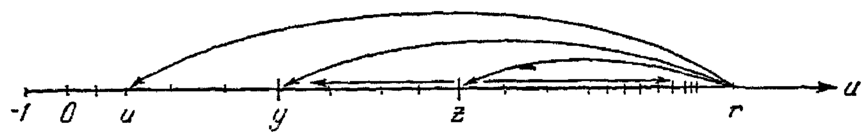

Fig. 46

§8. The Jump Measure and Reflection Coefficient

Our first concern is with the distribution π_y. This is the distribution of the point $x(\tau_y)$, where τ_y is the time of first exit of the path from U_y.

By definition, at the time τ_y the particle is either situated in one of the states $u_0 = 0, u_1, \ldots, y$ or it vanishes from the phase space E (the latter case occurs if the particle remains the whole time to the right of the point y until the termination time ζ). In order to avoid tiresome repetitions and to simplify the notation of the equations, we say that the particle does not become extinct at the termination time ζ of the path, but that it enters a fictitious state -1, from which it can never escape. In accordance with this, we let

$$\pi_y(-1) = \mathbf{P}_r\{\tau_y = \zeta\}$$

and include the probability $\pi_y(-1)$ in the set of probabilities π_y.

We assume that the particle, having sojourned for a certain time at the point r,* jumps at the time ξ from this point with a distribution

$$\pi = \{\pi(-1),\ \pi(0),\ \ldots,\ \pi(u),\ \ldots\},$$

where

$$\sum_u \pi(u) = 1. \tag{63}$$

We will show that all distributions π_y are uniquely expressed in terms of π.

---

*As is shown in §2 (pp. 152-153), this time can be distributed according to a power law.

If $x(\xi) = u \leq y$, then $\tau_y = \xi$, and $x(\tau_y) = u$ (Fig. 46). But if $x(\xi) = z > y$, then the particle can go from the point z either to y first or to r first. In the former case $x(\tau_y) = y$, while in the latter, according to the strong Markov property, the story begins all over again. The probabilities of going from z to y or to r are known; they are equal to $(r - z)/(r - y)$ and $(z - y)/(r - y)$, respectively [see Eq. (51)]. With all this in mind, we obtain

$$\pi_y(u) = \pi(u) + \sum_{y<z<r} \pi(z) \frac{z-y}{r-y} \pi_y(u) \qquad (u < y), \tag{64}$$

$$\pi_y(y) = \pi(y) + \sum_{y<z<r} \pi(z) \left[\frac{z-y}{r-y} \pi_y(y) + \frac{r-z}{r-y}\right]. \tag{65}$$

From (64) and (65) we find

$$\pi_y(u) = \frac{\pi(u)}{1 - \frac{1}{r-y} \sum\limits_{y<z<r} (z-y)\pi(z)} \qquad (u < y), \tag{66}$$

$$\pi_y(y) = \frac{\pi(y) + \alpha(y)}{1 - \frac{1}{r-y} \sum\limits_{y<z<r} (z-y)\pi(z)}, \tag{67}$$

where

$$\alpha(y) = \frac{1}{r-y} \sum_{y<z<r} (r-z)\pi(z). \tag{68}$$

If $\alpha(y)$ is subtracted from the denominator of Eqs. (66) and (67), then on the strength of (63) we obtain

$$1 - \frac{1}{r-y} \sum_{y<z<r} (r-y)\pi(z) = 1 - \sum_{y<z<r} \pi(z) = \sum_{u \leqslant y} \pi(u).$$

Consequently, the equations for π_y may be rewritten in the more compact form

$$\pi_y(u) = \frac{\pi(u)}{\sum\limits_{u \leqslant y} \pi(u) + \alpha(y)} \qquad (u < y), \tag{69}$$

$$\pi_y(y) = \frac{\pi(y) + \alpha(y)}{\sum\limits_{u \leqslant y} \pi(u) + \alpha(y)}. \tag{70}$$

In order to clarify the meaning of the relations derived above, we consider the following event C: $\{$A particle, on leaving the point r, does not return to it without first going outside the set $U_y\}$. The denominator of Eqs. (69)-(70) contains the probability $\mathbf{P}_r^y\{C\}$, and these equations imply that the unconditional distribution of the point $x(\tau_y)$ coincides with the conditional distribution associated with the condition C.

Choosing different distributions $\boldsymbol{\pi}$, we obtain different continued processes x(t). However, this does not exhaust all the possibilities. As a matter of fact, the first jump taking the particle from the state r might not exist; another case is possible in which the particle executes infinitely many jumps in an arbitrarily small time t, returning each time to the original state (of course, for every $\varepsilon > 0$ there can only be a finite number of jumps exceeding ε). The foregoing arguments are certainly not applicable to this case. Nevertheless, Eqs. (69)-(70) remain in force in the general case as well, but $\boldsymbol{\pi}$ is no longer a probability measure [the series (63) diverges], and the transparent interpretation of $\boldsymbol{\pi}$ becomes more complex. Moreover, Eq. (68) contains the additional term α, which characterizes reflection.

We note that if we multiply all the numbers $\pi(u)$ by the same positive constant, Eqs. (68)-(70) still yield the same values for the distribution $\boldsymbol{\pi}_y$. It may be said, therefore, that $\boldsymbol{\pi}$ is unique up to a positive factor.

We now ascertain how the situation stands in the general case. *For any process of the class A there exist, correct to a positive multiplicative factor, a nonnegative constant α (reflection coefficient) and a nonnegative sequence $\boldsymbol{\pi} = \{\pi(-1), \pi(0), \ldots, \pi(u), \ldots\}$ (jump measure) such that*

1) $$\sum_u (r-u)\pi(u) < \infty; \tag{71}$$

2) *if at least one of the numbers α, $\pi(u)$ $(u = -1, 0, u_1, \ldots)$ is distinct from zero, then $\boldsymbol{\pi}_y$ is expressed*

for every $y \geq 0$ in terms of α and π according to Eqs. (69)-(70) with

$$\alpha(y) = \frac{1}{r-y}\left[\alpha + \sum_{y<z<r}(r-z)\pi(z)\right]; \tag{72}$$

3) if all the numbers α, $\pi(u)$ are equal to zero, then all the numbers $\pi_y(u)$ $(u \leq y)$ are also equal to zero.

In the case when the boundary r is absorbing it is sufficient to set $\alpha = 0$, $\pi = 0$. Then it is apparent from Eqs. (69), (70), and (72) that no other choice of numbers α and π is applicable. In the remaining cases, as shown at the end of the preceding section,

$$\pi_y(-1) + \pi_y(0) + \ldots + \pi_y(y) = 1 \tag{73}$$

for any state y from the interval $[0, r)$.

Assuming Eq. (73) is satisfied, we examine two neighborhoods of the point r, viz., U_x and U_y $(x > y)$. Then $\tau_y \geq \tau_x$, and if $x(\tau_x) \leq y$, then $x(\tau_y) = x(\tau_x)$. If, on the other hand, $x(\tau_x) = z > y$, then it is possible to go from z either to y first (in which case the time τ_y occurs) or to r first. After returning to r, the random variable $x(\tau_y)$ again has the distribution π_y. For $\pi_y(u)$ $(u \leq y)$, therefore, we obtain equations entirely analogous to Eqs. (64)-(65), except that now $\pi_x(u)$ and $\pi_x(z)$ replace $\pi(u)$ and $\pi(z)$ in these equations, and the summation extends over the states z from the interval $y < z \leq x$. Reiterating the same computations as those which reduced Eqs. (64)-(65) to Eqs. (69)-(70), we obtain

$$\pi_y(u) = \frac{\pi_x(u)}{\sum_{u \leqslant y}\pi_x(u) + \alpha_x(y)} \qquad (u < y), \tag{74}$$

$$\pi_y(y) = \frac{\pi_x(y) + \alpha_x(y)}{\sum_{u \leqslant y}\pi_x(u) + \alpha_x(y)}, \tag{75}$$

where

$$\alpha_x(y) = \frac{1}{r-y}\sum_{y<z\leqslant x}(r-z)\pi_x(z) \tag{76}$$

[Eq. (73) is used in place of (63) for simplification of the denominator].

It follows from Eq. (74) that for any fixed states $y < x$ two finite sets of nonnegative numbers

$$\pi_y(u) \qquad (u < y) \tag{77a}$$

and

$$\pi_x(u) \qquad (u < y) \tag{77b}$$

differ only by a positive factor. Consequently, there exists an infinite sequence

$$\pi = \{\pi(-1),\ \pi(0),\ \ldots,\ \pi(u),\ \ldots\} \tag{78}$$

of nonnegative numbers such that each of the sets (77a,b) is obtained from the corresponding segment of the sequence π by multiplying by a positive constant.

If all the sets (77a,b) consist of zeros, then it is required to take $\pi = 0$. If, however, $\pi_w(v) > 0$ for some pair of states $v < w$, then it is sufficient to consider $\pi_x(u)/\pi_x(v)$, where $x > u$, $x > v$, and to notice that this ratio does not depend on x (the denominator π_x, insofar as it has a nonzero value for $x = w$, also has a nonzero value for any $x > v$).

In the special case when $\pi(u) = 0$ for all u, we have $\pi_y(u) = 0$ for any $u < y$, hence $\pi_y(y) = 1$ for all $y \geq 0$. Therefore, Eqs. (69)-(72) are valid if we assume $\alpha > 0$.

Let us now suppose that the measure $\boldsymbol{\pi}$ has a nonzero value and that v is the first state to the left for which $\pi(v) > 0$. Let $x > y > v$. Multiplying the numerators and denominators of the fractions in Eqs. (74)-(75) by the factor $\lambda(x)$ by which the set $\pi_x(u)$ differs from the corresponding segment of the sequence (78), we obtain

$$\pi_y(u) = \frac{\pi(u)}{\sum\limits_{u \leqslant y} \pi(u) + \alpha_x(y)\lambda(x)} \qquad (u < y), \tag{79}$$

$$\pi_y(y) = \frac{\pi(y) + \alpha_x(y)\lambda(x)}{\sum\limits_{u \leqslant y} \pi(u) + \alpha_x(y)\lambda(x)}. \tag{80}$$

By assumption, $\pi(v) \neq 0$, therefore the product $\alpha_x(y)\lambda(x)$ is uniquely determined from Eq. (79) for $u = v$. Consequently, $\alpha_x(y)\lambda(x)$ does not depend on x, and we are justified in setting

$$\alpha_x(y)\lambda(x) = \alpha(y) \qquad (x > y).$$

Then Eqs. (79)-(80) assume the required form (69)-(70). It is evident from these equations that the set $\pi(u)$ $(u < y)$ is obtained from the set $\pi_y(u)$ $(u < y)$ by multiplying by $\sum\limits_{u \leqslant y} \pi(u) + \alpha(y)$. Hence,

$$\lambda(y) = \sum_{u \leqslant y} \pi(u) + \alpha(y), \tag{81}$$

and Eqs. (69)-(70) may be rewritten in the form

$$\pi_y(u) = \frac{\pi(u)}{\lambda(y)} \qquad (u < y),$$
$$\pi_y(y) = \frac{\pi(y) + \alpha(y)}{\lambda(y)}.$$

Substituting these expressions with $y = x$ into Eq. (76) and multiplying both sides by $\lambda(x)$, we arrive at the equation

$$\alpha(y) = \frac{1}{r - y}\left[\alpha(x)(r - x) + \sum_{y < z \leqslant x} \pi(z)(r - z)\right]. \tag{82}$$

If x is increased, the sum of the terms $\pi(z)(r-z)$ in this equation does not decrease, because new nonnegative terms are added to it; on the other hand, it has the upper bound $\alpha(y)(r-y)$. Consequently, the series

$$\sum_z \pi(z)(r - z),$$

i.e., the series (71), converges, and it follows from (82) that the product $\alpha(x)(r-x)$ has a finite limit α as $x \uparrow r$.

This limit, clearly, is nonnegative; it is also the reflection coefficient. Passing to the limit in (82) as $x \uparrow r$, we arrive at Eq. (72)

Thus, for $y > v$ (v is the leftmost state in which the measure $\boldsymbol{\pi}$ is positive) the distribution $\boldsymbol{\pi}_y$ is written in terms of the parameters α and $\boldsymbol{\pi}$ in the form we require. For states y from the interval $0 \leq y \leq v$ (provided such states exist) Eqs. (69), (70), and (72) are also valid. In fact, for these y the distribution $\boldsymbol{\pi}_y$ is concentrated at the single point y, and it is readily seen that Eqs. (69), (70), and (72) reduce to the very same distribution. [The denominator of Eqs. (69) and (70) for $y \leq v$ include the term $\pi(v)$ with a positive coefficient and is therefore not equal to zero.] Thus, the representability of the distribution $\boldsymbol{\pi}_y$ in terms of α and $\boldsymbol{\pi}$ according to Eqs. (69)-(72) has been established in every case.

Multiplication of α and $\pi(u)$ by a common positive factor does not violate the relationship of these characteristics to the distribution $\boldsymbol{\pi}_y$, and, conversely, Eqs. (69), (70), and (72) imply that if the distributions $\boldsymbol{\pi}_y$ are obtained from two different pairs $\alpha^{(1)}$, $\boldsymbol{\pi}^{(1)}$ and $\alpha^{(2)}$, $\boldsymbol{\pi}^{(2)}$ according to these equations, then $\alpha^{(1)}$, $\boldsymbol{\pi}^{(1)}$ and $\alpha^{(2)}$, $\boldsymbol{\pi}^{(2)}$ are proportional. Hence, the jump measure $\boldsymbol{\pi}$ and the reflection coefficient are uniquely defined up to a positive factor.

We conclude with a discussion of the nature of the paths x(t) in various circumstances. If $\alpha = 0$ and the series $\Sigma\pi(u)$ converges, then, norming the measure $\boldsymbol{\pi}$, it may be assumed that $\Sigma\pi(u) = 1$. We then arrive at the situation analyzed at the beginning of the present section, when $\boldsymbol{\pi}$ is the distribution of the particle at the time of its first jump from r.

If $\alpha > 0$, $\boldsymbol{\pi} = 0$, jumps from r to other states are impossible, but the particle has a probability one of abandoning r. In this case the path of the particle does not have a discontinuity at the time of exit from r. For a continuous (for example, Wiener) process an analogous effect occurs in reflection. It is reasonable in our discrete case also to say that reflection takes place. A particle situated near r has a probability close to one of returning to r before going any substantial distance from r. It is therefore to be expected that upon reflection the particle will not depart monotonically from the point r, but will "pulsate" around the point r, returning to r an infinite number of times before arriving in a fixed state $u < r$. This is in fact what happens.

If the series $\Sigma\pi(u)$ diverges and $\alpha = 0$, infinitely many jumps to states lying between a fixed state u and r occur before jumping to u. The picture is similar to reflection in part, but the exits from r occur jumpwise, rather than continuously. The series $\Sigma\pi(u)$ cannot diverge too rapidly. In other words, before a particle from r can arrive in any fixed state $u<r$, it would have to execute jumps to nearer states, which would take an infinitely long time along with the return to r. The condition (71) is required in order to preclude that possibility. For certain birth and death processes it is also sufficient. In the general case it must be replaced by a stronger condition, to be enunciated in §9.

If α and $\boldsymbol{\pi}$ differ from zero, we have a combination of jumps from r and reflection. The comparative magnitude of α in relation to $\boldsymbol{\pi}$ characterizes the "specific weight" of the reflection among the jumps. To wit, as exemplified by Eqs. (69), (70), and (72), the escape probabilities from U_y as a result of a jump to the state u $(-1 \le u \le y)$, a jump to the state z $(y < z < r)$, and a reflection are proportional to the numbers $\pi(u)$, $\pi(z)[(r-z)/(r-y)]$, and $\alpha/(r-y)$ (see the problems).

§9. Absorption Coefficient; Inward Passability

In straightforward fashion the jump measure $\boldsymbol{\pi}$ and reflection coefficient α determine where the particle goes from the point r. For the description of a continued process x(t) it is also necessary to characterize the time of sojourn of the particle in r. For example, if $\alpha = 0$ and $\boldsymbol{\pi}$ is a probability measure, the particle jumps from r with a distribution $\boldsymbol{\pi}$. The time from the first arrival at r to the first exit from r is logically regarded as a power-law distribution, like the sojourn time in the remaining states. The parameter of this power law is arbitrary and represents a new characteristic of the continued process, and is necessarily added to the roster with α and $\boldsymbol{\pi}$. It is difficult to imagine any other possibilities for $\alpha = 0$ and $\Sigma\pi(u) < \infty$, and indeed there are none. The situation is roughly the same in the more general case when $\alpha \neq 0$ or $\Sigma\pi(u) = \infty$. Despite the fact that now the particle exits from r an infinite number of times during a finite period of time, its sojourn time at r is still described by one constant β. This constant makes its appearance in the investigation of the denominator $m(y) = \mathbf{M}_r \tau_y$ in the relation (59) for the characteristic operator $\mathfrak{A}$. Moreover, in the calculation of m(y) a new restriction is imposed on the measure $\boldsymbol{\pi}$, and it is found that reflection is not always possible from every boundary.

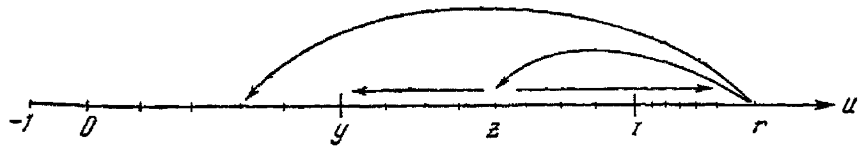

Fig. 47

Let x(t) be a process of the class A. We will show that for any $y \geq 0$

$$m(y) = \frac{\beta + \alpha v(r) + \sum_{y<z<r} \pi(z)[S(z) - S(r)] - \alpha(y)[S(y) - S(r)]}{\lambda(y)}, \quad (83)$$

where β is a nonnegative constant defined at a location with jump measure $\boldsymbol{\pi}$ and reflection coefficient α, correct to a common positive factor, and not equal to zero for $\boldsymbol{\pi} = 0, \alpha = 0$; S(u) and v(u) are the characteristic of the birth and death process and its derivative with respect to u;

$$\alpha v(r) = 0 \quad \text{for} \quad \alpha = 0, \ v(r) = \infty;$$

$\alpha(y)$ and $\lambda(y)$ are variables defined by Eqs. (72) and (81). Here .

$$\alpha v(r) < \infty \quad (84)$$

and

$$\sum_{u} \pi(u)[S(u) - S(r)] < \infty. \quad (85)$$

We agree to refer to the constant β as the absorption coefficient.

If r is an absorbing boundary, then $\alpha = 0$, $\boldsymbol{\pi} = 0$, and Eq. (83) yields the correct value of $+\infty$ for m(y) for any $\beta > 0$.

In the remaining cases, as shown at the end of §7, $m(y) < \infty$. Assuming this condition to be fulfilled, we once again examine two neighborhoods of the point r: U_x and U_y $(x > y)$. At the time τ_x of first exit from U_x the particle has the distribution $\boldsymbol{\pi}_x$. If $x(\tau_x) \leq y$, then $\tau_y = \tau_x$. But if $x(\tau_x) = z > y$, then $\tau_y > \tau_x$, in which case the particle can go from z and either hit y first or hit r first (Fig. 47). The average time required for this is known:

$$m(z;\ y,\ r) = S(z) - \frac{(r-z)\,S(y) + (z-y)\,S(r)}{r-y}$$
$$= [S(z) - S(r)] - \frac{r-z}{r-y}[S(y) - S(r)] \tag{86}$$

[see Eq. (46) in §5]. The hitting probabilities at y and r are equal to $(r-z)/(r-y)$ and $(z-y)/(r-y)$, respectively. The time τ_y occurs for arrival at y, but for arrival at r an average time m(y) is again required for exit from U_y. Thus, for the function m(y) we have the equation

$$m(y) = m(x) + \sum_{y<z\leqslant x} \pi_x(z)\left[m(z;\ y,\ r) + \frac{z-y}{r-y}\,m(y)\right]. \tag{87}$$

Expressing m(y) from this equation and replacing the numbers $\pi_x(z)$ with their values taken from Eqs. (69)-(70), we obtain

$$m(y) = \frac{\lambda(x)\,m(x) + \sum\limits_{y<z\leqslant x} \pi(z)\,m(z;\,y,\,r) + \alpha(x)\,m(x;\,y,\,r)}{\lambda(x) - \frac{1}{r-y}\left[\sum\limits_{y<z\leqslant x} \pi(z)(z-y) + \alpha(x)(x-y)\right]}, \tag{88}$$

where

$$\lambda(x) = \sum_{u\leqslant x} \pi(u) + \alpha(x).$$

The denominator of Eq. (88) is equal to

$$\sum_{u\leqslant x} \pi(u) + \alpha(x) - \frac{1}{r-y}\left[\sum_{y<z\leqslant x} \pi(z)(z-y) + \alpha(x)(x-y)\right]$$
$$= \sum_{u\leqslant y} \pi(u) + \frac{1}{r-y}\left[\sum_{y<z\leqslant x} \pi(z)(r-z) + \alpha(x)(r-x)\right]$$
$$= \sum_{u\leqslant y} \pi(u) + \frac{1}{r-y}\left[\sum_{y<z\leqslant x} \pi(z)(r-z) + \overline{\alpha} + \sum_{x<z<r} \pi(z)(r-z)\right]$$
$$= \sum_{u\leqslant y} \pi(u) + \alpha(y) = \lambda(y)$$

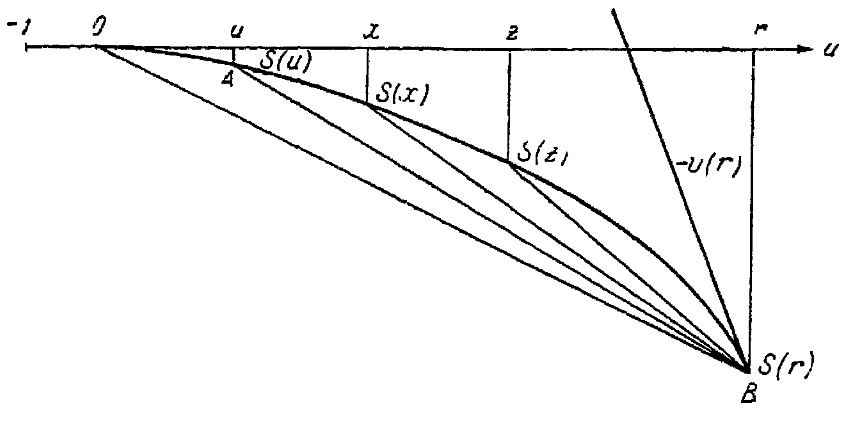

Fig. 48

[twice we have made use of Eq. (72), first for the state x, then for the state y]. Substituting into the numerator of Eq. (88) the values of $m(z; y, r)$ and $\alpha(x)$ taken from Eqs. (86) and (72), we find

$$\lambda(y)m(y)=\lambda(x)m(x)+\sum_{y<z\leqslant x}\pi(z)[S(z)-S(r)]$$
$$-\frac{S(y)-S(r)}{r-y}\sum_{y<z\leqslant x}\pi(z)(r-z)+\alpha(x)[S(x)-S(r)]$$
$$-\frac{S(y)-S(r)}{r-y}\left[\alpha+\sum_{x<z<r}\pi(z)(r-z)\right].$$

According to Eq. (72), the negative terms give the following in the sum here: $-\alpha(y)[S(y)-S(r)]$. Therefore,

$$\lambda(y)m(y)+\alpha(y)[S(y)-S(r)]=\lambda(x)m(x)$$
$$+\alpha(x)[S(x)-S(r)]+\sum_{y<z\leqslant x}\pi(z)[S(z)-S(r)]. \tag{89}$$

Let us pass to the limit in Eq. (89) as $x\uparrow r$. The left-hand side of this equation does not depend on x; all the terms on the right-hand side have the same sign. Consequently, the sum $\Sigma\pi(z)[S(z)-S(r)]$ remains bounded as $x\uparrow r$, so that the series (85) converges.

We next examine the term $\alpha(x)[S(x)-S(r)]$. According to (72), we have

$$\alpha(x)[S(x)-S(r)]=\alpha\frac{S(x)-S(r)}{r-x}+\frac{S(x)-S(r)}{r-x}\sum_{x<z<r}\pi(z)(r-z).$$

It follows from the convexity of the function S that the slope of the chord connecting the points of the graph of S with abscissas z and r increases in absolute value as z increases (Fig. 48). Hence,

$$\frac{S(x)-S(r)}{r-x} < \frac{S(z)-S(r)}{r-z}$$

for $x<z<r$, so that

$$0 \leqslant \frac{S(x)-S(r)}{r-x} \sum_{x<z<r} \pi(z)(r-z) \leqslant \sum_{x<z<r} \pi(z)\,[S(z)-S(r)].$$

On the right-hand side we have obtained the remainder of the convergent series (85), therefore

$$\lim_{x\uparrow r} \frac{S(x)-S(r)}{r-x} \sum_{x<z<r} \pi(z)(r-z) = 0.$$

Consequently,

$$\lim_{x\uparrow r} \alpha(x)\,[S(x)-S(r)] = \alpha \lim_{x\uparrow r} \frac{S(x)-S(r)}{r-x} = \alpha v(r), \tag{90}$$

where for $\alpha=0$ and $v(r)=\infty$ this limit is equal to zero.

Inasmuch as S(u) is a broken line, rather than a differentiable curve, the equation

$$\lim_{x\uparrow r} \frac{S(x)-S(r)}{r-x} = \lim_{n\to\infty} \frac{S_n-S_r}{r-u_n} = v(r) \tag{91}$$

in general is in need of proof. It follows from the relations of §5 that

$$S_n - S_m = \sum_{k=n}^{m-1} v_k \delta_k,$$

$$u_m - u_n = \sum_{k=n}^{m-1} \delta_k$$

Since v_k increases with k, we have

$$v_n(u_m-u_n) \leqslant S_n - S_m \leqslant v_{m-1}(u_m-u_n) \leqslant v_{m-1}(r-u_n).$$

From this we find that as $m \to \infty$

$$v_n (r - u_n) \leqslant S_n - S(r) \leqslant v(r)(r - u_n)$$

or

$$v_n \leqslant \frac{S_n - S(r)}{r - u_n} \leqslant v(r).$$

This proves Eq. (91), because $v(r) = \lim_{n \to \infty} v_n$.

Finally, since the remaining terms of Eq. (89) have finite limits as $x \uparrow r$, there also exists a finite nonnegative limit

$$\beta = \lim_{x \uparrow r} \lambda(x) m(x). \tag{92}$$

Taking (90) and (92) into account, from Eq. (89) in the limit we obtain the required Eq. (83). The condition (84) follows from (83) and the finiteness of m(y) for $\alpha > 0$. It is apparent from Eq. (83) that for given α and $\boldsymbol{\pi}$, not simultaneously equal to zero, the coefficient β is uniquely defined and that multiplication of α and $\boldsymbol{\pi}$ by a positive factor causes β to be multiplied by the same factor (or to remain equal to zero if it was equal to zero). If $\alpha = 0$ and $\boldsymbol{\pi} = 0$, as already mentioned, β is an arbitrary positive number. Consequently, in any case the coefficients α and β and the measure $\boldsymbol{\pi}$ are uniquely defined correct to a common positive multiplier.

We now need to find how the absorption coefficient β is related to the sojourn time of the particle at the boundary point r. Let us suppose that the boundary r is nonabsorbing. We note that the time τ_y to exit from U_y is made up of the time ξ_y spent at the point r plus the time σ_y spent in states z of the interval $y < z < r$. Therefore,

$$m(y) = \mathrm{M}_r \xi_y + M_r \sigma_y. \tag{93}$$

In order to find $\mathrm{M}_r \sigma_y$, we investigate a neighborhood U_x of the point r ($y < x < r$). Up to the time τ_y the particle can exit from U_x several times. Every time it escapes U_x but remains inside U_y the particle spends a certain time returning to r or hitting the point y. Let η_y^x denote the sum of all these time intervals. Clearly, then

$\sigma_y^x \leq \eta_y^x \leq \sigma_y$, where σ_y^x is the time spent by the particle up to the time τ_y in the states z of the arrival $y < z \leq x$. Since σ_y^x, on increasing, converges to σ_y as $x \uparrow r$, we have $M_r\sigma_y^x \uparrow M_r\sigma_y$;hence*

$$M_r\sigma_y = \lim_{x \uparrow r} M_r\eta_y^x. \tag{94}$$

The quantity $M_r\eta_y^x$ is easily calculated. In fact, arguing as in the derivation of Eq. (87), we obtain

$$M_r\eta_y^x = \sum_{y<z\leqslant x} \pi_x(z)\left[m(z;\ y,\ r) + \frac{z-y}{r-y} M_r\eta_y^x\right]. \tag{95}$$

The only distinction between the latter equation and the equation (87) for m(y) is the fact that the term m(x) on the right-hand side has been replaced by zero. Consequently, reiterating the same transformations with Eq. (95) as with (87), we obtain the same relation for the quantity $\lim_{x \uparrow r} M_r\eta_y^x = M_r\sigma_y$ as the equation (83) for m(y), but without the term β. Comparing (83) and (93), we find

$$M_r\xi_y = \frac{\beta}{\lambda(y)}. \tag{96}$$

Consequently, the expectation of the sojourn time at the point r prior to exit from the given neighborhood is proportional to β.

As $y \uparrow r$ the random variables ξ_y, on diminishing, converge to the time ξ of continuous sojourn of the particle at the point r. It is evident from Eqs. (72) and (81) that

$$\lim_{y \uparrow r} \lambda(y) = \begin{cases} \sum_u \pi(u) & \text{for } \alpha = 0, \\ \infty & \text{for } \alpha > 0. \end{cases} \tag{97}$$

Thus, if $\alpha = 0$ and the series $\Sigma\pi(u)$ converges, then

$$M_r\xi = \frac{\beta}{\sum_u \pi(u)} > 0.$$

---

* See the footnote on page 104.

The state r is similar in this case to all the other states; the sojourn time in it is distributed according to a power law with the parameter

$$a_\infty = \frac{\sum_u \pi(u)}{\beta} < \infty.$$

If $\alpha > 0$ or the series $\Sigma\pi(u)$ diverges, it follows from (97) that $\mathbf{M}_r \xi = 0$. In this case the set R of times t for which $x(t) = r$ does not contain a single interval of positive length. Nevertheless, for $\beta > 0$ the total sojourn time of the particle in r is positive. The set R is structured in this case like a Cantor perfect set of positive measure (see, e.g., [25], §15).

For an absorbing boundary $\lambda(y) = 0$ for all $y < r$, and Eq. (96) is also valid.

Let us now explore the conditions (84) and (85) in further detail. It follows from (84) that the reflection coefficient α can only have nonzero values as long as $v(r) < \infty$. Obviously, in the case of an attracting boundary ($r < \infty$) the finiteness or infiniteness of $v(r)$ dictates the possibility of traveling from r to any other state y in a finite period of time (see the problems). We say, therefore, that for $v(r) < \infty$ an absorbing boundary r is *passable inward*, while for $v(r) = \infty$ it is *impassable inward*. *Thus, for reflection it is required that the boundary be passable inward.*

The relation between the conditions on the jump measure $\boldsymbol{\pi}$

$$\sum_u \pi(u)(r-u) < \infty \tag{98}$$

and

$$\sum_u \pi(u)[S(u) - S(r)] < \infty, \tag{99}$$

which was derived in the two preceding sections, depends on the inward passability of the boundary. It follows from the convexity of the function S that for $u < r$

$$\frac{S(u) - S(r)}{r-u} \geqslant \frac{|S(r)|}{r}$$

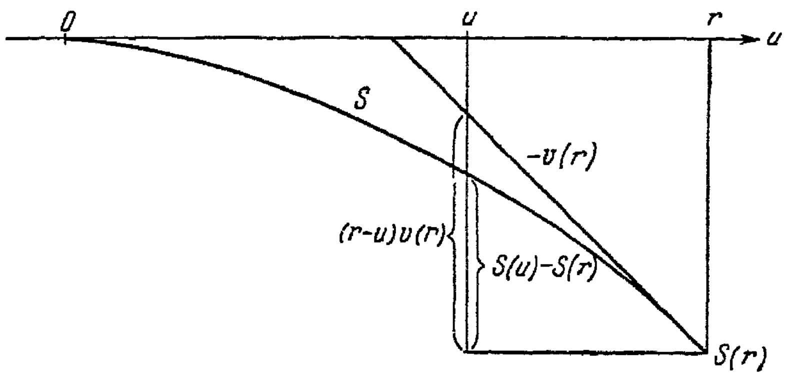

Fig. 49

(see Fig. 48, wherein the chord AB has a steeper slope than OB). Therefore,

$$r - u \leqslant \frac{r}{|S(r)|}[S(u) - S(r)]$$

and the convergence of the series (99) indicates convergence of the series (98). Consequently. in any case only the second of the two conditions (98) and (99) need be retained. However, if the boundary r is passable inward, then, on the other hand, (99) is a consequence of (98), and these conditions become equivalent. In fact, for $v(r) < \infty$ the geometrically obvious estimate $S(u) - S(r) \leq (r - u)v(r)$ (Fig. 49) may be used.

In the case of an accessible boundary, for which r and S(r) are finite, inward passability is a rather strong additional constraint. For example, if $p_n = p$, $q_n = q$ for all $n \geq 1$ and $p > q$, then, according to Eq. (41),

$$v(r) = \frac{1}{a_0} + p\sum_{n=1}^{\infty}\frac{1}{a_n}\left(\frac{p}{q}\right)^n. \tag{100}$$

The convergence of the resulting series imposes a much stiffer requirement on the parameter a_n than the convergence of the series

$$-S(r) = \frac{1}{p-q}\sum_{n=0}^{\infty}\frac{1}{a_n}$$

(see §6).

It is geometrically obvious (see Fig. 49) that the conditions $r<\infty$ and $v(r)<\infty$ imply finiteness of S(r). On the other hand, the condition $v(r)<\infty$ alone is insufficient for accessibility of the boundary; v(r) can be finite for infinite r and S(r) [for example, the series (100) may well converge for $p<q$]. Consequently, an accessible inward passable boundary is characterized by the condition

$$r<\infty, \quad v(r)<\infty.$$

(See the problems for more details regarding the relations between the various types of boundaries.)

§10. Boundary Conditions

We now draw certain conclusions. We showed in §§8 and 9 that every continuation, referred to the class A, of a birth and death process corresponds to a definite jump measure π, reflection coefficient α, and absorption coefficient β, correct to a common positive multiplier. Knowing α, β, and π, it is possible from Eqs. (69), (70), (72), and (83) to find the distribution π_y and average time m(y) for any $y\geq 0$; hence the characteristic operator at the point r is equal to

$$\mathfrak{A}f(r)=\lim_{y\uparrow r}\frac{\sum\limits_{0\leqslant u\leqslant y}\pi_y(u)\,f(u)-f(r)}{m(y)} \tag{101}$$

[see Eq. (59)]. At the remaining points the operator $\mathfrak{A}$ is known; according to Eq. (54),

$$\mathfrak{A}f(u_n)=\frac{1}{2}D_\mu D_u f(u_n) \quad (n=0,\ 1,\ 2,\ \ldots). \tag{102}$$

Inasmuch as the characteristic operator completely defines the class A process, the implication is that the given continuations of the birth and death process are uniquely determined by the jump measure π and the coefficients α and β.

We recall that in constructing the jump measure π we replaced the extinction of the particle by visitation to a fictitious state -1. We return now to the original terminology, introducing the following special notation for the number $\pi(-1)$:

$$\pi(-1)=\gamma \tag{103}$$

and calling γ the extinction coefficient. As before, we refer to the rest of the numbers $\pi(u)$ ($u = u_0, u_1, \ldots$) as the jump measure. Consequently, we now have the three coefficients α, β, and γ and the jump measure $\boldsymbol{\pi}$.

Instead of numbering each of the quantities α, β, γ, and $\boldsymbol{\pi}$ separately, we write one all-inclusive equation. This equation is obtained on passing to the limit in Eq. (101).

Let the boundary r be nonabsorbing. Substituting the values of the numbers $\pi_y(u)$ taken from Eqs. (69)-(70) into Eq. (101) and making use of the notation

$$\lambda(y)=\sum_{-1\leqslant u\leqslant y}\pi(u)+\alpha(y)=\gamma+\sum_{0\leqslant u\leqslant y}\pi(u)+\alpha(y)$$

[see (81)], we obtain

$$\mathfrak{A}f(r)=\lim_{y\uparrow r}\frac{\sum\limits_{0\leqslant u\leqslant y}\pi(u)f(u)+\alpha(y)f(y)-\lambda(y)f(r)}{\lambda(y)m(y)}$$

$$=\lim_{y\uparrow r}\frac{\sum\limits_{0\leqslant u\leqslant v}\pi(u)[f(u)-f(r)]+\alpha(y)[f(y)-f(r)]-\gamma f(r)}{\lambda(y)m(y)}. \tag{104}$$

According to Eq. (92), the denominator in the latter expression tends to the coefficient β as $y\uparrow r$. If the function f has a finite derivative $f'(r)$ at the point r, then, by virtue of Eq. (72) and the condition (71),

$$\lim_{y\uparrow r}\alpha(y)[f(y)-f(r)]=\lim_{y\uparrow r}\alpha(y)(r-y)\frac{f(y)-f(r)}{r-y}=-\alpha f'(r).$$

Furthermore, it follows from the existence of $f'(r)$ that $f(u)-f(r)\sim -f'(r)(r-u)$ for $u\uparrow r$; hence the convergence of the series $\Sigma\pi(u)(r-u)$ indicates convergence of the series $\Sigma\pi(u)[f(u)-f(r)]$. Consequently, the numerator in Eq. (104) tends to

$$\sum_{u}\pi(u)[f(u)-f(r)]-\alpha f'(r)-\gamma f(r). \tag{105}$$

Thus, if $\beta > 0$, we obtain

$$\mathfrak{A}f(r) = \frac{1}{\beta}\left\{\sum_u \pi(u)[f(u) - f(r)] - \alpha f'(r) - \gamma f(r)\right\},$$

or

$$\beta\mathfrak{A}f(r) + \alpha f'(r) + \gamma f(r) + \sum_u \pi(u)[f(r) - f(u)] = 0. \qquad (106)$$

If $\beta = 0$, the finiteness of $f'(r)$ still does not assure the existence of the limit (104), but for the finiteness of $\mathfrak{A}f(r)$ in any case it is required that the limit of the numerator be equal to zero. Setting the expression (105) equal to zero, we once again obtain Eq. (106) with $\beta = 0$. Finally, Eq. (106) is also true for an absorbing boundary r, because in this case

$$\alpha = \gamma = \pi(u) = \mathfrak{A}f(r) = 0.$$

Thus, in any case Eq. (106) is a necessary condition for a function f having a finite derivative at the point r to enter the domain of definition of the characteristic operator $\mathfrak{A}$. This equation is conventionally called the *boundary condition* for the process x(t) or for the operator $\mathfrak{A}$. Clearly, specification of an equation of the form (106) is equivalent to specification of the set of nonnegative numbers α, β, γ, and $\pi(u)$ correct to a positive multiplier. Consequently, *every continuation of a given birth and death process of the class A corresponds to a single definite boundary condition, and the continued process is uniquely reproduced according to this boundary condition.*

In other words, a process of the class A is described by the operator $^1\!/_2 D_\mu D_u$, defining the process at interior points, and by the boundary condition (106).

It was established in §§8 and 9 that the coefficients α, β, $\gamma = \pi(-1)$ and the measure π must satisfy the conditions

$$\alpha \geqslant 0,\ \ \beta \geqslant 0,\ \ \gamma \geqslant 0,\ \ \pi(u) \geqslant 0,\ \ \alpha^2 + \beta^2 + \gamma^2 + \sum_u \pi(u)^2 > 0,$$
$$\alpha\, v(r) < \infty,\ \ \sum_u \pi(u)[S(u) - S(r)] < \infty. \qquad (107)$$

We so far do not know whether these conditions are *sufficient* for the existence of the process x(t) with predetermined parameters α, β, γ, and π. In order to arrive at the answer to this question, we need to use another mathematical device, and we confine ourselves here solely to the formulation of the results and some of the more obvious interpretations. It turns out that the boundary condition (106) with coefficients obeying the inequalities (107) always correspond to a Markov process x(t), but sometimes this process does not belong to the class A. As a matter of fact, for certain values of the parameters α, β, γ, and π a particle hitting the boundary point r at the time T instantaneously jumps therefrom to a considerable distance. As a result, the requirement

$$\lim_{h \downarrow 0} x(T+h) = r, \tag{108}$$

which follows from the assumptions 4 and 5 made in §7 in the definition of the class A, is violated. If $\beta > 0$, a particle situated at r will spend (on the average) a positive time in the state r before escaping from any neighborhood $U \ni r$. The phrase "on the average" may be replaced by the phrase "with probability one," whereupon the condition (108) is fulfilled for $\beta > 0$. It is again fulfilled if the series $\Sigma\pi(u)$ diverges, because in this case the first jump from r beyond the limits of the fixed neighborhood U is preceded by infinitely many jumps to points nearer to r. If $\beta = 0$ and the series $\Sigma\pi(u)$ converges, then it is sufficient for the condition (108) that α be positive. Thus, it follows from Eq. (72) in this case that $\alpha(y) \to \infty$ as $y \uparrow r$, whereas $\sum_{u \leqslant y} \pi(u)$ remains bounded. Consequently, for y near to r, a direct jump from r to the states -1, 0, u_1, ..., y is far less probable than arrival at y from r via the state nearest to y on the right. It turns out in the limit that the first jump from r is preceded by multiple reflection at the point r, and the condition (108) is satisfied. The only case left is

$$\alpha = \beta = 0, \quad \sum_u \pi(u) < \infty. \tag{109}$$

In this case the particle executes a jump at the time T with the distribution π, and the condition (108) is violated. A boundary condition with these parameters does not fit any process of the class A. In the case (109) it is reasonable in general not to associate the point r with the phase space.

§11. The Uniqueness Theorem

We now prove the uniqueness theorem formulated in §7 with regard to a process of the class A with a given characteristic operator.

First we make a more precise statement of the problem. Consider the *transition function*

$$p(t, u, v) = \mathbf{P}_u\{x(t) = v\} \qquad (t \geqslant 0)$$

of the process x(t), where u and v are any points of the space $E = \{u_0, u_1, \ldots, u_n, \ldots, r\}$. The transition function p(t, u, v) plays the same role in the continuous-time case as the transition probabilities p(x, y) in the case of a Markov chain with discrete time. By virtue of the Markov property, the probability of the event

$$A = \{x(t_1) = v_1, \ x(t_2) = v_2, \ \ldots, \ x(t_n) = v_n\} \tag{110}$$

$$(0 \leqslant t_1 < t_2 < \cdots < t_n)$$

is expressed in terms of p(t, u, v) according to the formula

$$\mathbf{P}_u\{A\} = p(t_1, u, v_1)\, p(t_2 - t_1, v_1, v_2) \ldots p(t_n - t_{n-1}, v_{n-1}, v_n).$$

This means that if the transition functions coincide for two processes, then the probabilities of all events of the form (110) also coincide for those processes. Consequently, the probabilities of all events derived from the events (110) by means of addition (of nonoverlapping events), subtraction, and monotonic passage to the limit also coincide. It can be shown that for a class A process all events depending on the behavior of the path are so derived (correct to events of probability zero). Therefore, the difference be tween two class A processes having identical transition functions is nonessential.* Consequently, even though the class A process, as a set of probabilistic measures $\mathbf{P}_u$ on a set of paths, is not uniquely defined in terms of the transition function, for all practical purposes processes having the same transition function p(t, u, v) are indistinguishable. We will show that *the transition*

*Either of two such processes can always be derived from the other by adding to the set of paths a certain set B and subtracting from it a set C such that $P_u\{B\} = P_u\{C\} = 0$ for all $u \in E$.

function of a class A process is uniquely determined by the characteristic operator $\mathfrak{A}$.

We investigate on the continuous functions $f(u)$ $(u \in E)$ the operator R_λ defined by the relation

$$R_\lambda f(u) = \int_0^\infty e^{-\lambda t} \mathbf{M}_u f(x(t))\, dt \quad (\lambda > 0). \tag{111}$$

This operator is called the resolvent of the process x(t). (We point out that if $f \geq 0$, $R_\lambda f$ becomes the α-potential of the function f with $\alpha = e^{-\lambda}$; see Chapt. III, §8.)

In the special case when $f(u) = 1$ at a point $v \neq r$ and is equal to zero at all other points $\mathbf{M}_u f(x(t)) = p(t, u, v)$, and

$$R_\lambda f(u) = \int_0^\infty e^{-\lambda t} p(t, u, v)\, dt.$$

If $f = 1$ in all states u, $\mathbf{M}_u f(x(t)) = \mathbf{P}_u \{\zeta > t\}$, and

$$R_\lambda f(u) = \int_0^\infty e^{-\lambda t} \mathbf{P}_u \{\zeta > t\}\, dt.$$

On the right-hand side we have obtained the Laplace transform of the functions p(t, u, v) and $\mathbf{P}_u\{\zeta > t\}$. It is proved in analysis that if a function $\varphi(t)$ $(t \geq 0)$ is bounded and right continuous and if for any $\lambda > 0$

$$\int_0^\infty e^{-\lambda t} \varphi(t)\, dt = 0,$$

then φ is identically equal to zero (see, e.g., [4], p. 46). The probability $\mathbf{P}_u\{\zeta > t\}$ is right continuous in t, because as $h \downarrow 0$ the events $\{\zeta > t+h\}$ converge monotonically to the event $\{\zeta > t\}$. It follows from the right continuity of the path that the functions p(t, u, v) are also right continuous in t for $v \neq r$.* Therefore, class A

* As a matter of fact, for any pair of states $x \neq v$ the set A_h of paths starting from x and reaching v in a time $h > 0$ converges monotonically as $h \downarrow 0$ to an empty set. Inasmuch as the conditions $x(0) = x$, $x(h) = v$ imply the event A_h, it follows that $p(h, x, v) \leq$

processes having identical resolvents have equal functions p(t, u, v) ($v \neq r$) and $\mathbf{P}_u\{\zeta > t\}$. Inasmuch as

$$p(t, u, r) = \mathrm{P}_u\{\zeta > t\} - \sum_{v \neq r} p(t, u, v),$$

the functions p(t, u, r) in this case are equal, i.e., the transition functions are in complete agreement. Thus, it is sufficient to verify that the resolvent R_λ is uniquely determined by the characteristic operator $\mathfrak{A}$.

We first transform the relation (111) defining R_λ. The function $f(x(t))$ is only specified for $t < \zeta$. We denote by $\eta(t)$ a function that is equal to $f(x(t))$ for $t < \zeta$ and equal to zero for all other t. We have*

$$R_\lambda f(u) = \int_0^\infty e^{-\lambda t} \mathrm{M}_u f(x(t))\, dt = \int_0^\infty e^{-\lambda t} \mathrm{M}_u \eta(t)\, dt = \mathrm{M}_u \int_0^\infty e^{-\lambda t} \eta(t)\, dt.$$

From this we obtain

$$R_\lambda f(u) = \mathrm{M}_u \int_0^\zeta e^{-\lambda t} f(x(t))\, dt. \tag{112}$$

We now prove that the operator R_λ maps continuous

$\mathbf{P}_x\{A_h\}$, and the convergence of $\mathrm{P}_x\{A_h\}$ to zero implies that $\lim_{h \downarrow 0} p(h, x, v) = 0$ ($x \neq v$). If $v = u_n$, according to Eq. (9), $p(h, v, v) > e^{-a_n h}$, hence $\lim_{h \downarrow 0} p(h, v, v) = 1$ ($v \neq r$). Having this, the next thing is to pass to the limit as $h \downarrow 0$ in the equation

$$p(t+h, u, v) = \sum_{x \in E} p(t, u, x)\, p(h, x, v),$$

which is a direct consequence of the Markov property (passage to the limit by terms in the infinite series is legitimate, since the second factors are bounded by the number 1, whereas the series comprising the first factors is absolutely convergent).

* We have changed the order of integration over t and over the set of all paths. This is legitimate if the integrals are absolutely convergent (Fubini's theorem). The variable $\eta(t)$ is introduced in order to obtain a function defined for every t on the entire set of paths.

functions into continuous functions, i.e., that

$$\lim_{u \uparrow r} R_\lambda f(u) = R_\lambda f(r).$$

Inasmuch as the process cannot terminate before the time T of first visit of the path to the state r, we find from Eq. (112) that

$$R_\lambda f(u) = \mathbf{M}_u \int_0^T e^{-\lambda t} f(x(t))\, dt + \mathbf{M}_u \int_T^\zeta e^{-\lambda t} f(x(t))\, dt$$

$$= \mathbf{M}_u \int_0^T e^{-\lambda t} f(x(t))\, dt + \mathbf{M}_u e^{-\lambda T} \int_0^{\zeta - T} e^{-\lambda s} f(x(T+s))ds. \qquad (113)$$

Since T is a Markov time and x(T) = r, according to the strong Markov property, the process

$$y(s) \equiv x(T+s)$$

has exactly the same probability distribution as the process x(s) initiated at the point r. Here the process y(s) does not depend on the random variable T, and, clearly, $\zeta \leq T$ is the terminal time of the path y(s). Consequently, the second term in Eq. (113) is equal to

$$\mathbf{M}_u e^{-\lambda T} \mathbf{M}_r \int_0^\zeta e^{-\lambda s} f(x(s))\, ds = R_\lambda f(r) \cdot \mathbf{M}_u e^{-\lambda T},$$

and Eq. (113) assumes the form

$$R_\lambda f(u) = \mathbf{M}_u \int_0^T e^{-\lambda t} f(x(t))\, dt + R_\lambda f(r) \mathbf{M}_u e^{-\lambda T}.$$

Inasmuch as the function f is continuous, $|f(u)|$ is bounded by some constant C. Subtracting $R_\lambda f(r)$ from both sides of the latter equation and invoking the inequality $1 - e^{-x} \leq x$ $(x \geq 0)$, we obtain

$$|R_\lambda f(u) - R_\lambda f(r)| \leqslant \left| \mathbf{M}_u \int_0^T e^{-\lambda t} f(x(t))\, dt \right| +$$

$$+|R_\lambda f(r)|\,\mathbf{M}_u(1-e^{-\lambda T})\leqslant CM_uT+|R_\lambda f(r)|\cdot\lambda \mathbf{M}_uT.$$

According to Eq. (45), $\mathbf{M}_u T = S(u) - S(r) \to 0$ as $u\uparrow r$, and our assertion is thus proved.

The next step is to prove the equation

$$R_\lambda - R_\mu = (\mu-\lambda)\,R_\mu R_\lambda \qquad (\lambda,\mu>0). \tag{114}$$

We note that the expression (111) for R_λ may be rewritten in the form

$$R_\lambda = \int_0^\infty e^{-\lambda t} P_t\,dt,$$

where the operator P_t is defined by the relation

$$P_t f(u) = \mathbf{M}_u f(x(t)) \qquad (t\geqslant 0).$$

It was shown in Chapt. III, §8, that

$$P_sP_t = P_{s+t} \qquad (s,\ t\geqslant 0)$$

(the reader may convince himself that only the Markov property was used in this proof, not the specific characteristics of the Wiener process). Therefore,*

$$R_\mu R_\lambda = \int_0^\infty e^{-\mu s}P_sR_\lambda ds = \int_0^\infty e^{-\mu s}P_s\int_0^\infty e^{-\lambda t}P_t dt\,ds = \int_0^\infty\int_0^\infty e^{-\mu s-\lambda t}P_{s+t}dt\,ds.$$

Going over to the new variables s and $z = s+t$, we obtain

$$R_\mu R_\lambda = \int_0^\infty e^{-\lambda z}\left(\int_0^z e^{(\lambda-\mu)s}ds\right)P_z dz = \frac{1}{\lambda-\mu}\int_0^\infty (e^{-\mu z}-e^{-\lambda z})P_z dz = \frac{R_\mu - R_\lambda}{\lambda-\mu}$$

We next show that if f is continuous and $F = R_\lambda f$, then

$$f = \lambda F - \mathfrak{A}F. \tag{115}$$

*See the footnote on pp. 203-204.

We first verify the fact that Eq. (115) is satisfied at the point r if this point is an absorbing boundary. In fact, in this case $\mathfrak{A}f(r)=0$ and

$$F(r)=\int_0^\infty e^{-\lambda t}\mathrm{M}_r f(x(t))\,dt=f(r)\int_0^\infty e^{-\lambda t}dt=\frac{1}{\lambda}f(r).$$

In all other cases $\mathrm{M}_u\tau<\infty$ for a sufficiently small neighborhood U of the point u, where τ is the time of first exit from U (it may be assumed for $u=u_n$ that $U=u_n$, whereupon $\mathrm{M}_u\tau=1/a_n$; for $u=r$ the finiteness of $\mathrm{M}_u\tau$ was established at the end of §7). Making use of the continuity of the functions f and $F=R_\lambda f$, we pick a neighborhood U of the point u within whose confines the oscillation of the function $\lambda F-f$ remains smaller than a given number $\varepsilon>0$. With Eq. (114) we represent the function $F=R_\lambda f$ in the form

$$F=R_\mu g \qquad (\mu>0),$$

where

$$g=f+(\mu-\lambda)F.$$

Inserting between 0 and ζ the time τ of first exit of the path from U and using the representation of $R_\mu g$ in the form (112), we write

$$F(u)=\mathrm{M}_u\int_0^\tau e^{-\mu t}g(x(t))\,dt+\mathrm{M}_u e^{-\mu\tau}\int_0^{\zeta-\tau}e^{-\mu s}g(x(\tau+s))\,ds \qquad (116)$$

[cf. the derivation of Eq. (113)]. According to the strong Markov property, the process $y(s)\equiv x(\tau+s)$ under the condition $x(\tau)=v$ does not depend on the random variable τ and has the same distribution as the process x(s) initiated at the point v. The time $\zeta-\tau$ acts as the terminal time for y(s). Consequently, the conditional expectation

$$\mathrm{M}_u\Big(e^{-\mu\tau}\int_0^{\zeta-\tau}e^{-\mu s}g(x(\tau+s))\,ds\,\Big|\,x(\tau)=v\Big)=$$

$$= \mathbf{M}_u(e^{-\mu\tau} \mid x(\tau) = v) = \mathbf{M}_v \int_0^z e^{-\mu s} g(x(s))\, ds$$

$$= \mathbf{M}_u(e^{-\mu\tau} \mid x(\tau) = v) R_\mu g(v) = \mathbf{M}_u(e^{-\mu\tau} R_\mu g(x(\tau)) \mid x(\tau) = v)\,.$$

Multiplying the resulting expression by $\mathbf{P}_u\{x(\tau) = v\}$ and summing over $v \in E$, we obtain

$$\mathbf{M}_u e^{-\mu\tau} \int_0^{\zeta-\tau} e^{-\mu s} g(x(\tau + s))\, ds = \mathbf{M}_u e^{-\mu\tau} R_\mu g(x(\tau)) = \mathbf{M}_u e^{-\mu\tau} F(x(\tau)).$$

Equation (116) therefore acquires the form

$$F(u) = \mathbf{M}_u \int_0^\tau e^{-\mu t} [f(x(t)) + (\mu - \lambda) F(x(t))]\, dt + \mathbf{M}_u e^{-\mu\tau} F(x(\tau))$$

and as $\mu \downarrow 0$ goes over to the equation

$$F(u) = \mathbf{M}_u \int_0^\tau [f(x(t)) - \lambda F(x(t))]\, dt + \mathbf{M}_u F(x(\tau)) \tag{117}$$

(passage to the limit in the argument of the expectation and integral is legal, since the functions f and F are bounded and $\mathbf{M}_u \tau < \infty$). From (117) we obtain

$$\frac{\mathbf{M}_u F(x(\tau)) - F(u)}{\mathbf{M}_u \tau} = \frac{1}{\mathbf{M}_u \tau} \mathbf{M}_u \int_0^\tau [\lambda F(x(t)) - f(x(t))]\, dt.$$

Subtracting from both sides the quantity

$$\lambda F(u) - f(u) = \frac{1}{\mathbf{M}_u \tau} \mathbf{M}_u \int_0^\tau [\lambda F(u) - f(u)]\, dt$$

and utilizing the fact that the oscillation of the function $\lambda F - f$ remains smaller than ε within the limits of U, we find that

$$\left|\frac{M_u F(x(\tau)) - F(u)}{M_u \tau} - [\lambda F(u) - f(u)]\right| \leqslant \frac{1}{M_u \tau} M_u \int_0^\tau \varepsilon \, dt = \varepsilon .$$

This means that

$$\lim_{U \downarrow u} \frac{M_u F(x(\tau)) - F(u)}{M_u \tau} = \lambda F(u) - f(u) ,$$

i.e., that $\mathfrak{A} F(u) = \lambda F(u) - f(u)$. Equation (115) is thus proved.

Now we are finally in a position to prove that the operator R_λ is uniquely determined by the characteristic operator $\mathfrak{A}$ and thus to establish the uniqueness of the class A process with a given characteristic operator. We have shown that the function $F = R_\lambda f$ is continuous and satisfies the equation $\lambda F - \mathfrak{A} F = f$. Consequently, it is sufficient to verify that the equation $\lambda F - \mathfrak{A} F = f$ does not have more than one continuous solution for any right-hand side thereof. If this equation were to have two different continuous solutions, their difference would be a continuous nonzero solution of the homogeneous equation

$$\lambda F - \mathfrak{A} F = 0. \qquad (118)$$

Thus, it is left for us to prove that Eq. (118) has only a zero solution in the class of continuous functions F.

At some point v the continuous solution F(u) of Eq. (118) reaches its maximum value $M = F(v)$. As a matter of fact, the continuity of F(u) implies that $F(r) = \lim_{n \to \infty} F(u_n)$. If $F(u_n) \leq F(r)$ for all n, then the maximum value is reached at the point r. But if $F(u_n) > F(r)$ for some n, then, by virtue of continuity, $F(u_n) > F(u_m)$ for all m beginning with some m_0. It is apparent that in this event the largest of the numbers $F(u_0)$, $F(u_1)$, ..., $F(u_{m_0})$ is less than the maximum value of the function F. Suppose that $M > 0$. Since at the time τ of first exit from the neighborhood U of the point v

$$F(x(\tau)) \leqslant \mathbf{M} = F(v),$$

it follows that

$$\mathbf{M}_v F(x(\tau)) \leqslant F(v)$$

and, hence,

$$\mathfrak{A}F(v) = \lim_{U \downarrow v} \frac{\mathbf{M}_v F(x(\tau)) - F(v)}{\mathbf{M}_v \tau} \leqslant 0 .$$

On the other hand, it follows from Eq. (118) that

$$\mathfrak{A}F(v) = \lambda F(v) \qquad \lambda M > 0.$$

The ensuing contradiction indicates that $M \leq 0$. It is similarly verified that the smallest value m of the function F(u) is nonnegative. Consequently, F = 0, and the uniqueness is thus proved.

PROBLEMS

Average Exit Time

In Problems 1 and 2 the relations obtained in §5 for the average exit time are derived by a different method.

1. Let m_n be the average time to arrive at u_{n+1} for a particle situated at u_n. Prove the relation

$$m_n = \frac{1}{a_n} + q_n (m_{n-1} + m_n).$$

From this derive the equation

$$m_n = \frac{1}{a_0 p_0} \frac{q_1 \dots q_n}{p_1 \dots p_n} + \frac{1}{a_1 p_1} \frac{q_2 \dots q_n}{p_2 \dots p_n} + \dots + \frac{1}{a_n p_n}. \qquad (119)$$

2. From Eq. (119) obtain Eqs. (44) and (43).

Hint. In the canonical scale

$$m(u;\ b) = m(u;\ a,\ b) + \frac{b-u}{b-a}\, m(a;\ b).$$

Classification of Boundaries

We call a boundary r **finite** when $m(u; a, r) < \infty$ for all $a < u < r$ and **nonfinite** when $m(u; a, r) = \infty$ for all $a < u < r$. We call a boundary **weakly finite** when $m(u; a, r) \to \infty$ as $u \uparrow r$, and **strongly finite** when the function $m(u; a, r)$ is bounded (see [4]).

3. An attracting boundary is strongly finite for $|S(r)| < \infty$ and is nonfinite for $|S(r)| = \infty$.

4. A repelling boundary is strongly finite if $v(r) < \infty$ and the graph of S(u) has an asymptote as $u \uparrow r$, is weakly finite if $v(r) < \infty$ and S(u) does not have an asymptote, and is nonfinite if $v(r) = \infty$.

We interpose between the states u_k and u_{k+1} a **reflecting barrier**, i.e., we consider that each time, instead of going from u_k to u_{k+1}, the particle returns to u_k. We denote by $\bar{m}_k(u)$ the average time required in this case for arrival at 0 from u $(0 \le u \le u_k)$. We say that the boundary r is **passable inward** if $\bar{m}_k(u_k)$ remains bounded as $k \to \infty$ and **impassable inward** if $\bar{m}_k(u_k) \to \infty$ as $k \to \infty$ (it is evident from Problem 6 that this definition is consistent with the definition in §9).

5. The function $\bar{m}_k(u)$ satisfies Eq. (31) for all u_n situated between 0 and u_k. How must the definition of $\bar{m}_k(u)$ be augmented at the point $u = u_{k+1}$ in order for this equation to be fulfilled at the point u_k as well?

6. An attracting boundary is passable inward for $v(r) < \infty$ and is impassable inward for $v(r) = \infty$.

Hint. Find the solution of Eq. (31) for the boundary conditions $\bar{m}_k(0) = 0$ and $\bar{m}_k(u_k) = \bar{m}_k(u_{k+1})$.

7. A repelling boundary is passable inward if it is strongly finite and is impassable inward if it is weakly finite or nonfinite.

8. Strong finiteness of the boundary is equivalent to its accessibility or inward passability.

Jumps of the Path $x(t)$

In the next set of problems $x(t)$ denotes a class A process initiated at the point r, where the boundary r is assumed to be nonabsorbing. As in §§8 and 9, it is assumed that the particle hits the state -1 at the time of extinction. We recall that the paths $x(t)$ are right continuous for $t \geq 0$.

9. There exist with probability one finite limits $x(t-0)$ for all $t > 0$.

Hint. During a finite time interval $[0, t]$ the path with probability one does not enter any state u distinct from r more than a finite number of times.

We agree to say that a *jump* from a state x to a state y occurs at the time t if $x(t-0) = x$ and $x(t+0) = x(t) = y\ (y \neq x)$. It follows from Problem 9 that the path $x(t)$ with probability one does not have any discontinuities other than the jumps.

10. With probability one there are no jumps on the entire path $x(t)$ other than jumps to neighboring states and jumps from r.

Hint. Consider the first, second, and further visits of the particle to a state $x \neq r$ and invoke the strong Markov property, along with the hint to Problem 9.

Let τ_y be the time of first visit of the path to the interval of states $[-1, 0, \ldots, y)$, and let η_y be the time of last exit from r before τ_y (η_y is the upper bound of those $t \leq \tau_y$ for which $x(t-0) = r$). It is readily deduced from Problem 9 that with probability one $x(\eta_y - 0) = r$.

11. With probability $\pi(u)/\lambda(y)$ a jump occurs at the time τ_y from r to u ($u \leq y$), and with probability $\alpha(y)/\lambda(y)$ a jump occurs (at that time) to y from the nearest state to the right.

Hint. For states $u < y$ this is a consequence of Problem 10 and Eq. (69). The probability of a jump from r to the point y at the time τ_y may be calculated by adding the probabilities of a jump from r to y first at the time τ_z, first at the time of second visit from r to $[-1, z]$, etc., and making use of Eq. (82) (z is any state between y and r).

12. The distribution of the particle at the time η_y is given by the equation

$$\lambda(y)\,\mathrm{P}\{x(\eta_y)=u\}=\begin{cases}\pi(u), & u\leqslant y,\\ \pi(x)\dfrac{r-u}{r-y}, & y<u<r,\\ \dfrac{a}{r-y}, & u=r.\end{cases}$$

<u>Hint.</u> For $u \le y$ this follows from the preceding problem. In the case $y<u<r$ it is possible to choose the state z between u and r and to add the probabilities that a jump from r to u represents the time of first, second, and later visits from r to the interval $[-1, z]$. For $u=r$ it is sufficient to subtract the probabilities already determined from one.

13. If $\pi(u)=0$, the probability of a jump from r to u is equal to zero. If $\pi(u)>0$, $u\neq -1$, $\gamma=\pi(-1)>0$, and the probability of a jump from r to u is positive and less than one. If $\pi(u)>0$ and $\gamma=0$ or $u=-1$, then this probability is equal to one.

We introduce the shorthand notation $\pi(U)=\Sigma\pi(u)$, where the summation extends over states $u\in U$ [$\pi(r)$ is regarded as equal to zero]. We denote by A_U the following event: "The particle at some time jumps from the point r to U."

14. Let U be finite, and let $\pi(U)>0$. Then the probability that in the first jump from r to U the particle will jump to $u\in U$ is equal to $\pi(u)/\pi(U)$ (with regard for the conditional probability associated with the condition A_U).

<u>Hint.</u> Pick a state z lying to the right of all $u\in U$, and consider the possible jumps from r to u at the times of first, second, and later visits from r to the interval $[-1, z]$.

15. If $\pi(E)>0$, the probability is $\pi(u)/\pi(E)$ that the particle will hit the point u in the first jump from r.

<u>Hint.</u> Pass to the limit in the preceding problem as $U\uparrow E$ with fixed u.

16. For $0<\pi(E)<\infty$, with probability one there exists a first jump from r, while for $\pi(E)=\infty$, with probability one there does not exist a first jump from r.

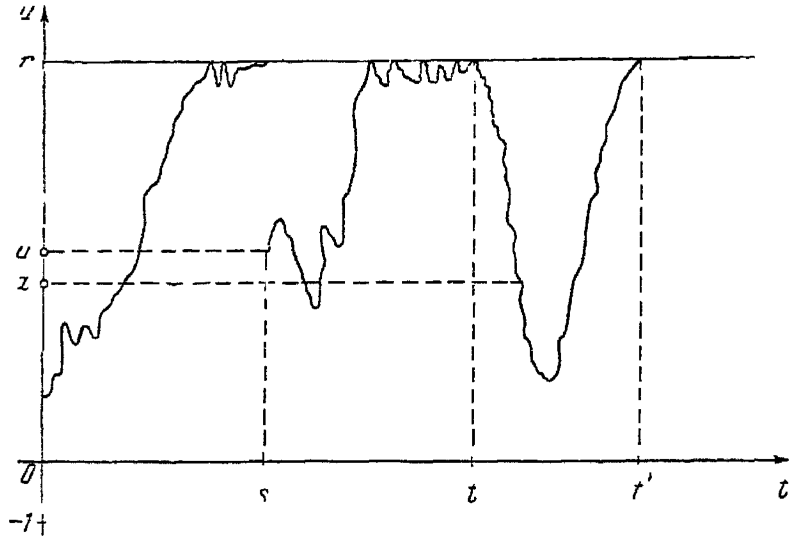

Fig. 50

17. If $\pi(E) = \infty$, then with probability one infinitely many other jumps from r occur prior to a jump from r to a fixed state u.

Reflection of the Path x(t) from the Boundary r

If a jump from r occurs at time s, then $x(s-0) = r$, $x(s) = x(s+0) \neq r$. We now assume that $x(t-0) = x(t+0) = r$. It may happen that the particle returns to r in any interval $(t, t+\delta)$ $(\delta > 0)$. If this is not the case, then we say that t is the *reflection time*, and we refer to the maximum interval (t, t') in which the particle exists outside r as the reflection interval (see Fig. 50, in which the step-function broken line is replaced by a continuous curve for better visualization). If in the course of this interval the particle sojourns in the state x, we say that an *x-reflection* has taken place.

18. Let x by any state distinct from r. With probability one no more than a finite number of x-reflections occur on any finite time interval.

It follows from Problem 18 that with probability one either there exists a first x-reflection or there exists no x-reflection. We agree to denote the time of the first x-reflection by δ_x (setting $\delta_x = +\infty$ if x-reflections do not exist).

19. If $\alpha = 0$, then

$$\mathbf{P}_r\{\delta_x = +\infty \text{ for all } x\} = 1$$

(with probability one no reflection occurs).

Hint. Use the embedding

$$\{\delta_r < \tau_x\} \subseteq \{x(\eta_x) = r\}$$

(see Problem 12).

20. If $\alpha > 0$, then

$$\mathbf{P}_r\{\delta_x < \infty \quad \text{for all} \quad x\} > 0$$

(for $\alpha > 0$ and $\gamma = 0$ this probability is equal to one).

Hint. See Problem 12.

21. For $\alpha > 0$ and $0 \le x \le y < r$

$$\mathbf{P}_r\{\delta_y = \delta_x | \delta_y < \infty\} = \frac{r-y}{r-x}.$$

22. For $\alpha > 0$, with probability one a first reflection does not exist.

Hint. If δ is the time of first reflection, then for some state x

$$\delta_x = \delta < \infty.$$

It is clear that then $\delta_y = \delta_x < \infty$ for all $x < y < r$, and, by Problem 21,

$$\mathbf{P}_r\{\delta_x = \delta < \infty\} \leqslant \frac{r-y}{r-x}.$$

23. If $\alpha > 0$ and $\pi(E) = \infty$, then with probability one there does not exist an x-reflection that would not be preceded by some kind of a jump from r.

Hint. Let $p_y = \mathbf{P}_r\{$ Before the time $\delta_x < \infty$ there was no jump from r to the interval $[-1, y]\}$. With the help of Problem 12, for $y > x$, we obtain the equation

$$p_y = \frac{\alpha}{(r-y)\lambda(y)}\left(\frac{r-y}{r-x} + \frac{y-x}{r-x}p_y\right),$$

which, combined with Eq. (72), yields the estimate

$$p_y \leqslant \frac{\alpha}{(r-x)\,\pi([-1,\ y])}.$$

24. If $\alpha > 0$, we have in the notation of Problem 12

$$\lim_{y \uparrow r} \mathbf{P}_r \{\eta_y = \delta_y\} = 1$$

(for y close to r the probability of a first visit to the interval [−1, y] by a y-reflection, rather than a jump, is close to one).

Hint. It follows from the convergence of the series $\Sigma\pi(u)\,\overline{(r-u)}$ that

$$\lim_{y \uparrow r} (r-y) \sum_{u \leqslant y} \pi(u) = 0.$$

Therefore $\lambda(y)\,(r-y) \to \alpha$ as $y \uparrow r$.

25. If $\alpha > 0$, then with probability one there does not exist a jump from r that would not be preceded by some kind of reflection.

Hint. Take advantage of Problem 24.

Local Time at the Boundary Point r

In Problems 26–31 the additional assumption was made that $\beta > 0$ and $\gamma = 0$ (the average time of sojourn at the point r is positive, and extinction is impossible).* The function s(t) denotes the time spent by the particle at the point r up until the time t (a graph of this function is shown in Fig. 51).

26. The function s(t) is continuous, is nondecreasing, and with probability one tends to ∞ as $t \to \infty$.

Hint. In order to arrive at the equation

$$\mathbf{P}_r \{\lim_{t \to \infty} s(t) = \infty\} = 1$$

---

* The local time may also be introduced in the case $\beta = 0$. In this respect, for example, see Ito and McKean [8], Chapt. 2. The ideas behind this set of problems trace their origin to Lèvy [30].

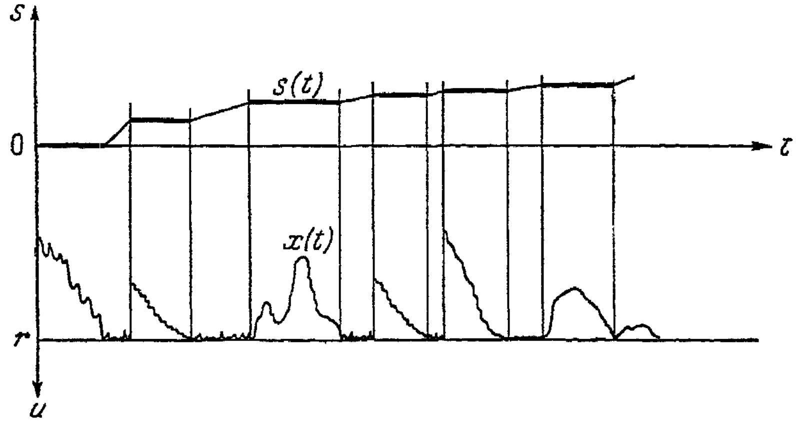

Fig. 51

for a nonabsorbing boundary r (the remaining assertions of the problem are quite simple to prove), we note that the particle with probability one hits 0 from r and returns thereto infinitely often (see the end of §7). Let s_n be the sojourn time at r during the nth such cycle. Then

$$\lim_{t \to \infty} s(t) = \sum_{n=1}^{\infty} s_n,$$

where the random variables s_n are mutually independent, but are distributed, are nonnegative, and with positive probability are positive [see Eq. (96)].

It follows from Problem 26 that with probability one there exists, inverse to s(t), an increasing left continuous function $t(s) = \min\{t : s(t) = s\}$, specified on the entire semiaxis $0 \leq s < \infty$. Here $t(0) = 0$, and t(s) is a Markov time for any $s \geq 0$. Making use of Problem 9, we can exclude from the discussion the set of paths x(t) along which the limits $x(t-0)$ do not always exist. Then, by the actual definition of t(s), for every $s > 0$ we have the equation $x(t(s) - 0) = r$. It can be proved that for any fixed $s > 0$ the path x(t) with probability one is left continuous at the time t(s), so that therefore,*

$$\mathbf{P}_r\{x(t(s)) = r\} = 1 \qquad (s \geqslant 0).$$

* A process x(t) of the class A is left quasi-continuous (see [4], Theorem 3.13), i.e., if the τ_n are Markov times and $\mathbf{P}_x\{\tau_n \uparrow \tau < \infty\} = 1$, then $\mathbf{P}_x\{x(\tau - 0) = x(\tau)\} = 1$. Here it is permissible to let $t(s - 1/n)$ represent τ_n.

27. For an initial state r the time ξ_y spent at the point r until the first visit to the interval [0, y] has an exponential distribution with a mean value of $\beta/\lambda(y)$.

Hint. Analyzing the Markov times t(s), deduce the relation

$$\mathbf{P}_r\{\xi_y \geqslant s_1 + s_2\} = \mathbf{P}_r\{\xi_y \geqslant s_1\} \cdot \mathbf{P}_r\{\xi_y \geqslant s_2\}$$
$$(s_1,\ s_2 > 0)$$

and use §3 of the Appendix, along with Eq. (96).

On the local time axis s for the state r we mark off the points $s_1, s_2, s_3, \ldots$ corresponding to the visits of the particle to the interval [0, y]. It follows from the well-known properties of an exponential distribution that the points $\{s_i\}$ form a Poisson flow with parameter $\beta/\lambda(y)$. This means that the number of points falling on an interval of length s has a Poisson distribution with parameter $[\lambda(y)/\beta]s$ and that for nonoverlapping time intervals the number of points falling on them are mutually independent.

28. Let $\{s_i\}$ be a Poisson flow with parameter $1/\mu$, and let each of the points s_i be labeled with a star independently of the preceding points with probability p. Then the starred points form a Poisson flow with parameter $1/\mu p$.

Hint. The time ξ to the first starred point satisfies the relation

$$\mathbf{P}\{\xi \geqslant a + b\} = \mathbf{P}\{\xi \geqslant a\}\, \mathbf{P}\{\xi \geqslant b\}$$
$$(a,\ b > 0).$$

The expectation of the index of the first starred point s_i is equal to 1/p.

29. The time σ_n spent at the point r up to the first jump from r to u has an exponential distribution with mean value $\beta/\pi(u)$.

Hint. See Problems 12 and 28.

30. The time ρ_x spent at the point r up to the first x-reflection has an exponential distribution with mean value $\beta(r - x)/\alpha$.

31. We modify Problem 28, labeling a point s_i, independently of the preceding ones, with probability p_1 with a star and with probability p_2 with a cross ($p_1 + p_2 \leq 1$). Then for any time interval the number of points labeled with a star does not depend on the number of points labeled with a cross.

Hint. Let s be the length of the interval, m_1 the number of points labeled with a star, and m_2 the number of points labeled with a cross on that interval. According to Problem 28,

$$\mathbf{P}\{m_1 = k\} = \frac{(p_1 s\mu)^k}{k!} e^{-p_1 s\mu},$$

$$\mathbf{P}\{m_2 = l\} = \frac{(p_2 s\mu)^l}{l!} e^{-p_2 s\mu},$$

$$\mathbf{P}\{m_1 + m_2 = k + l\} = \frac{[(p_1 + p_2)s\mu]^{k+l}}{(k+l)!} e^{-(p_1 - p_2)s\mu}.$$

No matter where the labeled points are situated, the number of starred points among the successively chosen $k + l$ labeled points has a binomial distribution with $p = p_1/(p_1 + p_2)$ (the stars and crosses alternate, like the hits and misses in the Bernoulli scheme). Consequently,

$$\mathbf{P}\{m_1 = k,\ m_2 = l\} = \mathbf{P}\{m_1 + m_2 = k + l\} \cdot \mathbf{P}\{m_1$$

$$= k \mid m_1 + m_2 = k + l\} = \frac{[(p_1 + p_2)s\mu]^{k+l}}{(k+l)!} e^{-(p_1 + p_2)s\mu}$$

$$\times \frac{(k+l)!}{k!\,l!} \left(\frac{p_1}{p_1 + p_2}\right)^k \left(\frac{p_2}{p_1 + p_2}\right)^l = \mathbf{P}\{m_1 = k\} \cdot \mathbf{P}\{m_2 = l\}.$$

Problem 31 is easily generalized to the case of n probabilities

$$p_1, p_2, \ldots, p_n \qquad (p_1 + p_2 + \ldots + p_n \leqslant 1).$$

Among the points $\{s_i\}$ (see the text leading up to Problem 28) we set apart the points corresponding to jumps to u_0; then the points corresponding to jumps to u_1; ...; then the points corresponding to jumps to $u_n = y$; finally, the points corresponding to x-reflections (x is a fixed state on the interval [0, y]). It follows from Problem 31 that they form $n + 2$ mutually independent Poisson flows (the parameters of these flows are found in Problems 29 and 30). Inasmuch as n can be made as large as we please, the times of the jumps from r to any different states also form independent Poisson flows on the s axis. The times of the x-reflections (x fixed) form a Poisson flow, which is independent of all the jumps from r. The times of x-reflections and y-reflections, of course, are mutually independent. But the reflections can also be split into independent flows if we investigate x-reflections which are not u-reflections for any $u < x$ (they may be called x-reflections in the narrow sense). It is inferred from Problems 28 and 31, as in the jump case, that x-reflections in the narrow sense for distinct x form independent Poisson flows on the local time axis s, and these flows do not depend on the flows of jumps from r. All this permits a more graphic visualization of the behavior of the particle at the point r.

Appendix

§1. Estimation of the Function g(x, y)

We wish to prove, following Duffin [31], the asymptotic estimate given in Chapt. I, §3, for the function g(x, y). As in the derivation of the recurrence criterion, we confine ourselves to the case of three dimensions.

Inasmuch as g(x, y) depends only on the difference x − y, it is sufficient to analyze g(x, 0). We introduce the abbreviated notation

$$\theta = \{\theta_1, \theta_2, \theta_3\}, \qquad \|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2},$$
$$d\theta = d\theta_1\, d\theta_2\, d\theta_3, \qquad \rho = \sqrt{\theta_1^2 + \theta_2^2 + \theta_3^2},$$
$$\theta x = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

and, as before, denote by Q the cube $|\theta_1| \le \pi$, $|\theta_2| \le \pi$, $|\theta_3| \le \pi$. Then

$$g(x, 0) = \frac{3}{(2\pi)^3} \int_Q F(\theta)\, e^{i\theta x}\, d\theta, \tag{1}$$

where

$$F(\theta) = \frac{1}{3 - \cos\theta_1 - \cos\theta_2 - \cos\theta_3}. \tag{2}$$

We will show that

$$\lim_{\|x\| \to \infty} \|x\| g(x, 0) = \frac{3}{2\pi}. \tag{3}$$

Equation (1) means that g(x, 0), correct to a constant, is the Fourier coefficient with index $x = \{x_1, x_2, x_3\}$ (x_1, x_2, x_3 are integers) for the function $F(\theta)$.

We note that if a periodic function $H(\theta)$ (period 2π with respect to each argument) has continuous second derivatives, its Fourier coefficients

$$h(x) = \int_Q H(\theta) e^{i\theta x} d\theta \tag{4}$$

satisfy the condition

$$h(x) = O\left(\frac{1}{\|x\|^2}\right) \tag{5}$$

[here and elsewhere $O(\alpha)$ denotes a quantity that does not exceed the product of α multiplied by a certain constant]. Thus, let Δ be the Laplace operator in the space of the variables θ. According to Green's formula

$$\int_Q H \cdot \Delta e^{i\theta x} d\theta = \int_Q \Delta H \cdot e^{i\theta x} d\theta, \tag{6}$$

because, owing to the periodicity of the integrated functions, the surface integrals over opposite faces of the cube Q cancel one another. Since $\Delta e^{i\theta x} = -\|x\|^2 e^{i\theta x}$, it follows from (6) that

$$|h(x)| = \frac{1}{\|x\|^2} \left| \int_Q \Delta H \cdot e^{i\theta x} d\theta \right| \leqslant \frac{1}{\|x\|^2} \int_Q |\Delta H| \, d\theta, \tag{7}$$

and we arrive at the estimate (5).

The estimate (5) remains valid in the case when the derivatives of the function H have a singularity of not too high order at zero (and the function H is twice continuously differentiable at all other points of the cube Q). Specifically, it is sufficient to demand that the function H be bounded, its first partial derivatives equal to $O(1/\rho)$, and its second partial derivatives $\partial^2 H/\partial\theta_1^2$, $\partial^2 H/\partial\theta_2^2$, $\partial^2 H/\partial\theta_3^2$ equal to $O(1/\rho^2)$. In fact, we apply Green's formula to the domain $Q \setminus K$, where K is a small cube enclosing the point 0; the integral over its surface approaches zero by virtue of the estimate for the derivatives $\partial H/\partial\theta_1$, $\partial H/\partial\theta_2$, and $\partial H/\partial\theta_3$, and in the limit we obtain Eq. (6). As a result of the estimate for the second derivatives, the integrals in (7) converge, and we again arrive at Eq. (5).

The function in which we are interested, $F(\theta)$, has a higher-order singularity at zero. Differentiating Eq. (2) as many times as necessary and writing out the first two or three terms of the expansion of the sine or cosine in a Taylor series, we obtain for small ρ

$$\left.\begin{aligned} F(\theta) &= \frac{2}{\rho^2 + O(\rho^4)}, \\ \frac{\partial F}{\partial\theta_i} &= \frac{-4\theta_i + O(\rho^3)}{\rho^4 + O(\rho^6)}, \\ \frac{\partial^2 F}{\partial\theta_i^2} &= \frac{16\theta_i^2 - 4\rho^2 + O(\rho^4)}{\rho^6 + O(\rho^8)}. \end{aligned}\right\} \tag{8}$$

This singularity may be weakened by subtracting the function $2/\rho^2$ from $F(\theta)$, as the former behaves similarly to the latter near zero. It is readily deduced from Eq. (8) that the function $F(\theta) - (2/\rho^2)$ already meets the restrictions imposed on $H(\theta)$ in the preceding paragraph. For example,

$$\frac{\partial}{\partial\theta_i}\left(F - \frac{2}{\rho^2}\right) = \frac{-4\theta_i + O(\rho^3)}{\rho^4 + O(\rho^6)} + \frac{4\theta_i}{\rho^4} = \frac{O(\rho^7)}{\rho^8 + O(\rho^{10})} = O\left(\frac{1}{\rho}\right).$$

It is still impossible, however, to use the estimate (5), because the subtracted function $2/\rho^2$, if continued periodically beyond the limits of the cube Q, will not have continuous first and second derivatives on the face of this cube. In order to remove this obstacle, we multiply $2/\rho^2$ by a nonincreasing twice continuously differentiable function $S(\rho)$ equal to one for $0 < \rho \leq 1/2$ and equal to zero for $1 \leq \rho < \infty$. It is clear that the function $2S(\rho)/\rho^2$ will, as before, "extinguish" the singularity of the function $F(\theta)$ at zero, and in the integration over the cube Q this function may be regarded as periodic with period 2π, without disturbing its smoothness. Now, therefore, the estimate (5) is applicable to the function

$$H(\theta)=F(\theta)-\frac{2S(\rho)}{\rho^2}$$

and we find that the Fourier coefficients of the functions $F(\theta)$ and $2S(\rho)/\rho^2$ differ from one another by an amount $O(1/\|x\|^2)$. Thus,

$$g(x,0)=\frac{6}{(2\pi)^3}\int\limits_Q \frac{S(\rho)\,e^{i\theta x}}{\rho^2}\,d\theta+O\left(\frac{1}{\|x\|^2}\right). \tag{9}$$

We proceed now with the computation of the integral in Eq. (9). Inasmuch as the function S is equal to zero outside the cube Q, we are in a position to replace integration over the domain Q by integration over the entire space R^3. After this we rotate the coordinates axes θ_1, θ_2, and θ_3 so that the θ_1 axis will pass through the point $x=\{x_1, x_2, x_3\}$. The quantities ρ, $S(\rho)$, and $d\theta$ remain unchanged when this is done, and the scalar product $\theta x=\theta_1 x_1+\theta_2 x_2+\theta_3 x_3$ goes over to $\theta_1\|x\|$, since the vector x in the new system has the coordinates $\{\|x\|, 0, 0\}$. Further, we replace $e^{i\theta_1\|x\|}$ by $\cos\theta_1\|x\|+i\sin\theta_1\|x\|$; inasmuch as $A(\rho)/\rho^2$ is an even function of the argument θ_1, the integral containing the sine will be equal to zero. It turns out, therefore, that

$$\int\limits_Q \frac{S(\rho)}{\rho^2}e^{i\theta x}\,d\theta=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\frac{S(\rho)\cos\theta_1\|x\|}{\rho^2}\,d\theta_1\,d\theta_2\,d\theta_3.$$

In the latter integral we transform to spherical coordinates according to the relations

$$\theta_1=\rho\cos\psi,\quad \theta_2=\rho\sin\psi\cos\varphi,\quad \theta_3=\rho\sin\psi\sin\varphi.$$

Recognizing that the Jacobian of the transformation is equal to

$\rho^2 \sin \psi$, we obtain

$$\int_Q \frac{S(\rho)\, e^{i\theta x}}{\rho^2}\, d\theta = \int_0^\infty d\rho \int_0^{2\pi} d\varphi \int_0^\pi S(\rho) \cos(\|x\| \rho \cos \psi) \sin \psi \, d\psi$$

$$= \frac{4\pi}{\|x\|} \int_0^\infty \frac{S(\rho) \sin(\|x\| \rho)}{\rho}\, d\rho = \frac{4\pi}{\|x\|} \int_0^\infty \frac{S\left(\frac{\lambda}{\|x\|}\right) \sin \lambda}{\lambda}\, d\lambda. \tag{10}$$

Since the integral $\int_0^\infty \frac{\sin \lambda}{\lambda} d\lambda$ converges and the function $S(\lambda/\|x\|)$ is monotonic in λ and bounded for all x by the same number, the integral obtained in Eq. (10) converges uniformly in x (see [19], p. 477). It is permissible, therefore, to pass to the limit in the integrand, and we obtain

$$\lim_{\|x\| \to \infty} \int_0^\infty \frac{S\left(\frac{\lambda}{\|x\|}\right) \sin \lambda}{\lambda}\, d\lambda = \int_0^\infty \frac{\sin \lambda}{\lambda}\, d\lambda = \frac{\pi}{2}.$$

Returning to Eqs. (9) and (10), we find

$$\lim_{\|x\| \to \infty} \|x\| g(x, 0) = \frac{6}{(2\pi)^3} \cdot 4\pi \cdot \frac{\pi}{2} = \frac{3}{2\pi}.$$

§2. Certain Properties of Concave Functions

A function $f(x)$, $x \in [a, b]$, is called concave on the indicated interval if any chord joining two points of the graph of f lies entirely on or below this graph (Fig. 52). Analytically, for any values of $x_1 < x_2$ on the interval $[a, b]$ and any numbers p and q satisfying the conditions $p \geq 0$, $q \geq 0$, $p + q = 1$, the following inequality is fulfilled:

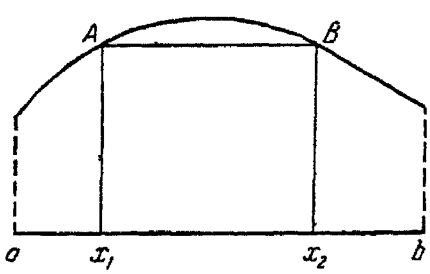

Fig. 52

$$f(px_1 + qx_2) \geqslant pf(x_1) + qf(x_2). \tag{11}$$

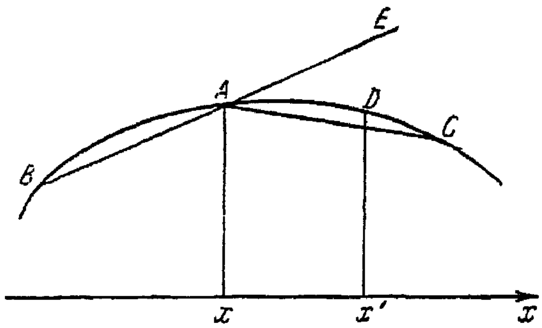

Fig. 53

The following properties of concave functions were used in Chapter III and now need to be proved.

I. The function f is continuous at all interior points of the interval $[a, b]$ and has finite limits as $x \downarrow a$ and $x \uparrow b$, where $f(a+0) \geq f(a)$, $f(b-0) \geq f(b)$.

First let x be an interior point of the interval, and let A be the corresponding point of the graph (Fig. 53). On the graph of f we pick points B and C to the left and right of A and investigate on the graph a variable point D with abscissa x' tending on the right to x. We draw the chord AC and the half-line AE representing the continuation of the chord BA. The point D cannot go above the line AE; otherwise, the chord BD would pass above the point A. On the other hand, after D goes to the left of C, it cannot drop below the chord AC. Hence, as $x' \downarrow x$ the point D will not emerge from the angle EAC, and its ordinate will tend to the ordinate of the point A. The function f is therefore right continuous at the point x. It is demonstrated analogously that the function f is left continuous at the point x.

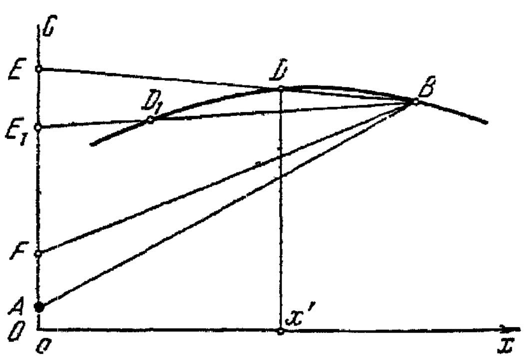

Fig. 54

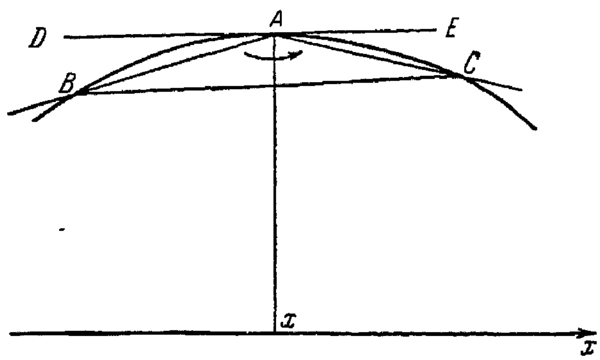

Fig. 55

We now investigate the left end point A of the graph of the function (the case of the right end point is analyzed analogously). On the graph we pick a point B distinct from A and draw the chord AB and vertical half-line AC (Fig. 54). Let D be a variable point on the graph, its abscissa x' tending on the right to a. We continue the chord DB until it intersects with the line AC at the point E. By the same arguments as in the preceding paragraph, a point D_1 situated to the left of D cannot lie above the segment ED. Therefore as $x' \downarrow a$ the point E moves along the line AC monotonically downward, without passing the point A. In the limit the point E occupies some position F, where OF ≥ OA. Inasmuch as the segments FE and ED shrink to zero as $x' \downarrow a$, the ordinate of the point D tends to the ordinate of the point F, hence $f(a+0) = OF \geq OA = f(a)$.

II. For any interior point x it is possible to choose a linear function $\bar{f}$ that coincides with f at the point x and is greater than or equal to f at all other points.

On the graph of the function f we pick variable points B and C to the left and right of a fixed interior point A (Fig. 55). Arguing as before, we are readily convinced that the half-line AB majorizes the graph of the function to the left of the point B, the line AC doing the same to the right of the point C, and that as B and C tend to A, these lines rise monotonically upward. Since the chord BC cannot pass above the point A, the angle BAC never exceeds 180° (the angles at the point A are measured counterclockwise). In the limit, therefore, the lines AB and AC occupy some positions AD and AE, the angle DAE again never exceeding 180°. If this angle is equal to 180°, the line DE is then the graph of the function $\bar{f}$ we are looking for. If, on the other hand, the angle DAE is smaller than 180°, any line passing through the point A outside the angle DAE will serve as the graph of $\bar{f}$.

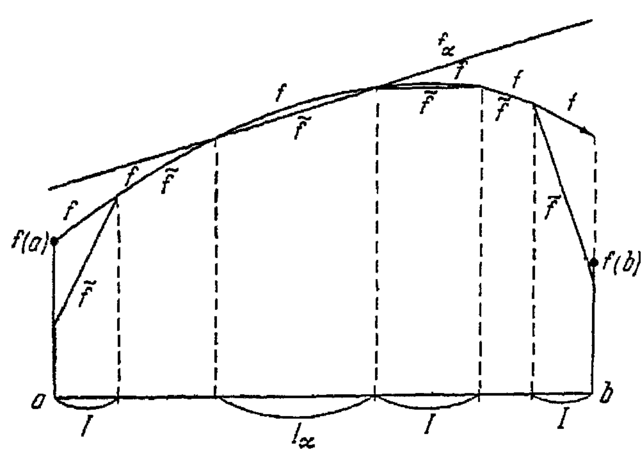

Fig. 56

III. Let us choose an arbitrary system of nonoverlapping segments I_α belonging to the interval $[a, b]$. On every segment I_α we replace the function f by a linear function f_α that coincides with f at the end points of the segment I_α, except that if an end point of I_α coincides with the point a [or point b], the function f_α can be either equal to or smaller than $f(a)$ [or $f(b)$] at the point a [or b]. At all other points we leave the function f unchanged. Then the resulting function $\tilde f$ is again concave on the interval $[a, b]$ (Fig. 56).

It follows from the foregoing considerations that $f_\alpha \geq f$ outside the segment I_α. Therefore, if $x \in I_\alpha$, then $f_\beta(x) \geq f(x) \geq f_\alpha(x) = \tilde f(x)$ for all $\beta \neq \alpha$, and if x does not belong to any of the segments I_α, then $f_\alpha(x) \geq f(x) \geq \tilde f(x)$ for all α. Hence, the function $\tilde f$ is a lower bound of the functions f and f_α (α spans all possible values). Inasmuch as the functions f and f_α are concave, all that is left to prove is that the lower bound $\tilde f$ of any family $\{f_\alpha\}$ of concave functions is also a concave function. For this it is sufficient to invoke the analytic condition of concavity of the functions (11) and note that for any α

$$f_\alpha(px_1 + qx_2) \geqslant pf_\alpha(x_1) + qf_\alpha(x_2) \geqslant p\tilde f(x_1) + q\tilde f(x_2).$$

§3. Solution of the Equation $p(s)p(t) = p(s + t)$

We need to show that any bounded solution of the functional equation

$$p(s)\,p(t) = p(s+t) \qquad (s,\ t > 0), \tag{12}$$

which was investigated in Chapt. IV, §2, has the form

$$p(t) = e^{-at}, \tag{13}$$

where $0 \leq a \leq +\infty$ (considering $e^{-\infty} = 0$).

We point out that if p(t) goes to zero at some point $t_0 > 0$, then, according to (12), p(t) = 0 for all $t \geq t_0$. Moreover, it follows from the relation

$$p\left(\frac{t}{2}\right)^2 = p(t) \tag{14}$$

that $p(t_0/2) = 0$, hence that p(t) = 0 for all $t \geq t_0/2$. Repeating this argument, we obtain p(t) = 0 for all t > 0, and Eq. (13) is valid with $a = +\infty$.

It now remains for us to consider the case when $p(t) \neq 0$ for all t > 0. Equation (14) implies that p(t) > 0 in this case, and we are entitled to set

$$f(t) = \ln p(t).$$

Now Eq. (12) goes over to the equation

$$f(s) + f(t) = f(s+t) \qquad (s,\ t > 0), \tag{15}$$

and the problem is reduced to one of finding all solutions of this equation that are bounded above.

It is readily deduced from Eq. (15) by induction that for any natural n

$$f(nt) = nf(t). \tag{16}$$

Picking the number a on the basis of the condition

$$f(t_1) = -at_1,$$

where t_1 is a fixed positive number, we obtain by means of Eq. (16)

$$f\left(\frac{t_1}{n}\right) = \frac{f(t_1)}{n} = -a\frac{t_1}{n}$$

Applying Eq. (16) once again, we find that for any natural numbers m and n

$$f\left(\frac{m}{n}t_1\right) = mf\left(\frac{t_1}{n}\right) = -a\frac{m}{n}t_1.$$

Consequently, for all $t > 0$ commensurable with t_1 we have

$$f(t) = -at. \tag{17}$$

If it turned out for some $t_2 > 0$ that $f(t_2) \neq -at_2$, then, determining the number b from the condition $f(t_2) = -bt_2$, we would have found in fully analogous fashion that

$$f(t) = -bt$$

for all $t > 0$ commensurable with t_2, where $b \neq a$. Let $b > a$ for definiteness. If s is commensurable with t_2, and $s + t$ with t_1, then

$$f(t) = f(s+t) - f(s) = -a(s+t) + bs = (b-a)s - at. \tag{18}$$

Inasmuch as numbers commensurable with a given number are densely distributed everywhere, s can be made as large as we like and $s + t$ as close to s as we like in the above equation. In this case t is small, and Eq. (18) gives arbitrarily large values for $f(t)$. We have thus arrived at a contradiction with the upper-boundedness requirement on $f(t)$. This means that Eq. (17) for the function $f(t)$ is valid for all $t > 0$. Since $f(t)$ is bounded above and t can be as large a number as we like, in this equation $a \geq 0$.*

Returning to the function $p(t) = e^{f(t)}$, we obtain the representation (13) for it.

---

*We note that Eq. (18) makes it possible to obtain arbitrarily large values for $f(t)$ when t varies over any predetermined interval. For the derivation of Eq. (17) from (15), therefore, it is sufficient to require that the function $f(t)$ be bounded above in some interval of variation of t (the number a in this case can be of any sign).

Literature Cited

The following books are recommended primarily for a familiarization with the more classical aspects of the theory of Markov processes:

1. W. Feller, An Introduction to Probability Theory and Its Applications, Wiley, New York (1957).
2. J. G. Kemeny and J. L. Snell, Finite Markov Chains, Van Nostrand, Princeton (1960).

The following monographs are devoted to new developments in this theory:

3. E. B. Dynkin, Theory of Markov Processes, Pergamon, New York (1960).
4. E. B. Dynkin, Markov Processes, Springer–Verlag, Berlin (1965).
5. G. A. Hunt, "Markov processes and potentials (I-III)," Illinois J. Math., 1:44-93 (1957); 1:316-369 (1957); 2:151-213 (1958).
6. K. L. Chung, Markov Chains with Stationary Transition Probabilities, 2nd ed., Springer–Verlag, Berlin (1967).
7. F. Spitzer, Principles of Random Walk, Van Nostrand, Princeton (1964).
8. K. Ito and H. P. McKean, Diffusion Processes and Their Sample Paths, Springer–Verlag, Berlin (1965).

(In this group, [6] represents a slight departure from the theme of the present book)

The following books also devote considerable attention to some related problems:

9. K. Ito, Probabilistic Processes, Nos. I and II [Russian translation], IL, Moscow (1960 and 1963).
10. M. M. Loève, Probability Theory, 3rd ed., Van Nostrand, Princeton (1963).
11. I. I. Gikhman and A. V. Skorokhod, Introduction to the Theory of Random Processes, Nauka, Moscow (1965).

The following publications are cited in the main text of the present work:

12. K. Ito and H. P. McKean, "Potentials and the random walk," Illinois J. Math., 4:119-132 (1960).
13. E. B. Dynkin, "The optimum choice of the instant for stopping a Markov process," Dokl. Akad. Nauk SSSR, 150:238-240 (1963).

14. W. Feller, "The birth and death processes as diffusion processes," J. Math. Pure Appl., 38:301-345 (1959).
15. A. A. Yushkevich, "Remarks on the boundary conditions for birth and death processes," Trans. Fourth Prague Conf. Information Theory, Statistical Decision, and Random Processes, Prague (1967), pp. 381-387.
16. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol. 2: Partial Differential Equations, Wiley, New York (1962).
17. R. S. Bucy, "Recurrent sets," Ann. Math. Statistics, 36(2):535-545 (1965).
18. M. A. Naimark, Normed Rings, Gostekhizdat, Moscow (1956).
19. G. M. Fikhtengol'ts, Course in Differential and Integral Calculus, Gostekhizdat, Moscow– Leningrad (1948).
20. A. Cayley, "Mathematical questions and their solutions, Problem No. 4528," The Educational Times, 27:189 (1874).
21. A. Cayley, "Mathematical questions and their solutions, solution to Problem No. 4528," The Educational Times, 27:237 (1875).
22. Gusein-zade, "The sampling problem and optimal stopping rule for a sequence of independent trials," Teoriya Veroyatnostei i Ee Primeniya, 11(3):534-537 (1966).
23. L. Moser, "On a problem of Cayley," Scripta Math., 22:289-292 (1956).
24. J. L. Doob, "Discrete potential theory and boundaries," J. Math. Mech., 8:433-458 (1959).
25. P. R. Halmos, Measure Theory, Van Nostrand, New York (1950).
26. L. A. Lyusternik and V. I. Sobolev, Elements of Functional Analysis, Gostekhizdat, Moscow–Leningrad (1951).
27. G. Choquet and P. A. Meyer, "Existence et unicité des representations intégrales dans les convexes compacts quelconques [Existence and uniqueness of integral representations on convex compact sets]," Ann. Inst. Fourier, Grenoble, 13:139-154 (1963).
28. E. B. Dynkin and M. B. Malyutov, "Random walk on groups with a finite number of generators," Dokl. Akad. Nauk SSSR, 137:1042-1045 (1961).
29. Wang Tzŭ-k'uang, Classification of all birth and death processes," Nauchnye Doklady Vysshei Shkoly, Fiz. Mat. Nauki, No. 4, pp. 19-25 (1958).
30. P. Lèvy, Processus stochastiques et mouvement Brownien [Stochastic Processes and Brownian Motion], Villars, Paris (1948).
31. R. J. Duffin, "Discrete potential theory," Duke Math. J., 20:233-251 (1953).

Index