LOGIC, ARITHMETIC, AND AUTOMATA(1)

By ALONZO CHURCH

This paper is a summary of recent work in the application of mathematical logic to finite automata, and especially of mathematical logic beyond propositional calculus.

To begin with a sketch of the history of the matter, let us recall that application of the "algebra of logic", i.e., elementary Boolean algebra, to the analysis of switching circuits was first suggested by Ehrenfest [A]. Nothing came of Ehrenfest's remark for many years, and it seems to have remained wholly unknown outside of Russia. Yanovskaya [G] says that details of the suggested application were worked out by Shestakoff in 1934–35. However, Shestakoff's candidate's dissertation, embodying the material, was presented to the University of Moscow in 1938, and the earliest publications by Shestakoff are [D] and [E] in 1941. Meanwhile the same idea had occurred independently to Nakasima and Hanzawa [B] and Shannon [C]. For some time the development of the idea proceeded independently in Russia, in Japan, and in the United States, the three lines of development having had at first no influence on one another.

This use of Boolean algebra is now widely familiar, and therefore requires no elaboration here. It is usually taken to be a Boolean algebra of cardinal number 2 that is used, although the character of the application would more naturally suggest propositional calculus. Use of the Boolean algebra and of propositional calculus are equivalent in a way that is well known. The choice of Boolean algebra is advantageous if algebraic methods and results are to be employed. But otherwise there is a certain artificiality in allowing only equations and inequalities to be asserted. And for further application of mathematical logic, the choice of propositional calculus provides a better basis.

Mathematical logic beyond propositional calculus is first applied to automata theory in the paper of McCulloch and Pitts [16], in which the context is biological. The authors are concerned with analyzing the behavior of a net of neurons and with the question of the existence of, and of finding, a neural net having some specified behavior. But their hypotheses about the behavior and the interaction of neurons are such that these questions become entirely similar to corresponding questions about electronic digital computing circuits. The relevance of the ideas of McCulloch and Pitts to the theory of digital computing circuits was noticed by John von Neumann, and it was evidently this that led him to suggest application of mathematical logic to the latter.

The only published reference to von Neumann's contribution to the matter is Hartree [13], where not only McCulloch and Pitts but also unpublished suggestions of both von Neumann and A. M. Turing are referred to. I am indebted to H. H. Goldstine for calling my attention to the privately cir-

(1) For support of the work represented by this paper the author is indebted to the National Science Foundation of the United States.

culated paper, von Neumann [18], in which there is some detailed discussion of the relationship between the McCulloch–Pitts neural nets and digital computing networks, but without use of mathematical logic.

The logical system employed by McCulloch and Pitts is first-order arithmetic. Both ordinary and bounded quantifiers are used, but not definition by recursion.

Hartree uses a singulary free-variable functional (or "predicate") calculus, having singulary predicates and numerical variables, but neither quantifiers nor recursions. There is a "delay operator" in the sense that not only a variable t but also $t+1$, $t+2$, etc., may appear as argument of a predicate.

The use of (in effect) such a singulary free-variable functional calculus also developed independently in Japan. The earliest paper that I have seen is Gotô [12]. But it is possible that some of the earlier papers to which Gotô refers may constitute at least a partial exception to the statement made above, that the first application of mathematical logic beyond propositional calculus was by McCulloch and Pitts.

The use of recursions in the treatment of circuits is found in Berkeley [1], and more explicitly in Burks and Wright [5]. The idea that a particular circuit or finite automaton can be completely characterized by giving a set of recursion equations—or in this context, since the functions "defined" by the recursions are propositional functions, we shall speak rather of recursion equivalences—appears first in the paper of Burks and Wright, but these authors do not have a recursion schema to which admissible recursions must conform.

The application of restricted recursive arithmetic to automata was introduced in Church [6] and [7]. This system may be described as obtained from singulary free-variable functional calculus by adding a rule of proof by mathematical induction and a rule of definition by recursion, as given below. It is therefore a specialized form of the recursive arithmetic of Skolem [19].

An abstract definition of finite automaton appears in Kleene [14], a paper of restricted circulation, and in Kleene [15], which is the published form of the same paper. The "logical nets" of Burks and Wright [5] may also be regarded as providing such a definition. But Kleene's definition is more satisfactory, and we shall here use it in the following slightly modified and generalized form.

A *finite automaton* consists of a finite number of *elements*, each of which is capable of a finite number, $\leqslant n$, of different *states*. Time is measured in discrete *instants*, $t=0,1,2,\ldots$, beginning at an *initial instant* and extending indefinitely. The elements are distinguished as *input elements*, *intermediate elements*, and *output elements*. The states of the input elements, at any instant, constitute the *input state*, at that instant; the states of the output elements, at any instant, similarly constitute the *output state*; and the states of the intermediate and output elements constitute the *internal state*. Except the input elements, whose states are imposed from outside, the state of any element at a given instant is completely determined, in some noncircular way, by (1) the states of certain other elements at the given instant, and (2) the states of certain elements, itself and others, at instants which precede the given instant by at least 1 and by not more than a certain fixed maximum *span* $h+1$.

(More fully, the foregoing statement is a definition of "discrete synchronous deterministic finite automaton with outputs." But the longer phrase is for the sake of distinction from a number of other notions with which we are not here concerned, and let us therefore say simply "finite automaton" or "automaton".)

Though there is a certain loss of structure in doing so, it is quite usual to take $n=2$, $h=0$. In this lecture we shall take $n=2$, in order to identify the two states of an element with the two truth-values, T and F. But we preserve a general value of h, $\geqslant -1$.

It is also possible to abstract from the elements of the automaton and to consider only certain numbers called input states, output states, and internal states, and the way in which the successive output states and internal states are determined by the input states. Namely there are to be a finite number each, $\geqslant 1$, of input states, output states, and internal states; the output state at any instant is to be uniquely determined by the internal state at the same instant; and the internal state at any instant is to be uniquely determined by (1) the input state at that instant and (2) the input states and the internal states at instants preceding the given instant by not more than $h+1$. In this case we have a *finite transition system.*

In either case, automaton or transition system, the sequence of input states from $t=0$ to $t=t_0$ is called an *input sequence*, and the sequence of output states from $t=0$ to $t=t_0$ is called an *output sequence*. The function which, for an input sequence from $t=0$ to $t=t_0$, as argument, has as value the output state at $t=t_0$ is the *behavior function*. The null sequence is also to be counted as an input sequence, but one for which there is no corresponding output state and hence no value of the behavior function.

Two automata, or two transition systems, are said to be *equivalent* if they have the same behavior function. Evidently, two automata are equivalent if and only if their transition systems are equivalent. In particular, two automata with the same transition system are equivalent—but they do not necessarily have the same *structure*, in an obvious sense of that word.

A class of input sequences is an *event*.

An event is *represented* by an automaton if the set of values of the behavior function for input sequences that belong to the event is disjoint from the set of values of the behavior function for other input sequences.

The foregoing (together with the definition of regularity of an event, to be given below) are the essentials of the terminology of automata theory which we shall need. It should be added that the terminology in the field is not in a very settled state, and our terminology here is somewhat eclectic. We have followed Büchi [2] in adopting a terminology that maintains the distinction between automaton and transition system.

We turn now to statement of the primitive basis of restricted recursive arithmetic. And for expository convenience we adopt a formulation which emphasizes brevity of statement rather than economy of primitives.

The primitive symbols of restricted recursive arithmetic are as follows:

i. Numerical variables $t, x, y, z, \ldots$ (having the natural numbers as their range).

ii. The symbol 0 for the number zero.

iii. The accent $'$ as a notation for successor (of a natural number).

iv. Singulary primitive predicates $i_1, i_2, i_3, \ldots$. (The primitive predicates, and other singulary predicates introduced by recursion, are to be understood as denoting propositional functions of natural numbers.)

v. Parentheses ( ) as notation for application of a singulary propositional function to its argument.

vi. The letters T and F to denote the truth-values, truth and falsehood respectively.

vii. Connectives of propositional calculus (any sufficient set), and brackets [ ] for use in connection with them.

Then a *term* consists of the symbol 0, or the symbol 0 followed by any number of accents, or a numerical variable, or a numerical variable followed by any number of accents.

An *elementary formula* consists, either of one of the letters T or F, or of a singulary predicate followed by a term between parentheses.

Further *well-formed formulas* are constructed from the elementary formulas by means of connectives of propositional calculus. (We may sometimes abbreviate "well-formed formula" as "wff"; or we may say simply "formula" if well-formedness is clear from the context.)

The primitive rules of restricted recursive arithmetic are the five following:

1. To assert any formula that is tautologous according to laws of propositional calculus.
2. From $\mathrm{A} \supset \mathrm{B}$ (as major premiss) and A (as minor premiss) to infer B.
3. From any asserted formula to infer the result of substituting any term for any numerical variable throughout.
4. From $\mathrm{S}(t) \supset \mathrm{S}(t')$ (as major premiss) and $\mathrm{S}(0)$ (as minor premiss) to infer $\mathrm{S}(t)$.
5. A rule of "definition" (or better, introduction) by recursion, allowing us at any stage to introduce new predicates in accordance with the schema of wider restricted recursion as given below.

Rules 2, 3, and 4 are, in order, *the rule of modus ponens, the rule of substitution*, and *the rule of mathematical induction.*

In Rule 4, $\mathrm{S}(t)$ is not necessarily an elementary formula, but is any assertible wff; $\mathrm{S}(t')$ is the result of substituting t' for t throughout $\mathrm{S}(t)$; and $\mathrm{S}(0)$ is the result of substituting 0 for t throughout $\mathrm{S}(t)$.

The schema of "wider restricted recursion" referred to in Rule 5 is as follows:

$$\mathrm{r}_i(j) \equiv \mathrm{P}_{ij}[\mathrm{s}_1(0), \mathrm{s}_2(0), \ldots, \mathrm{s}_k(0), \mathrm{s}_1(1), \mathrm{s}_2(1), \ldots, \mathrm{s}_k(1), \ldots, \ldots, \mathrm{s}_1(j), \mathrm{s}_2(j), \ldots, \mathrm{s}_k(j)]$$

$$\begin{aligned}\mathrm{r}_i(t+h+1) \equiv \mathrm{Q}_i[&\mathrm{s}_1(t), \mathrm{s}_2(t), \ldots, \mathrm{s}_k(t), \mathrm{s}_1(t+1), \mathrm{s}_2(t+1), \ldots, \mathrm{s}_k(t+1), \ldots, \ldots, \mathrm{s}_1(t+h+1), \mathrm{s}_2(t+h+1), \ldots, \mathrm{s}_k(t+h+1), \mathrm{r}_1(t), \mathrm{r}_2(t), \ldots, \mathrm{r}_n(t), \mathrm{r}_1(t+1),\\ &\mathrm{r}_2(t+1), \ldots, \mathrm{r}_n(t+1), \ldots, \ldots, \mathrm{r}_1(t+h), \mathrm{r}_2(t+h), \ldots, \mathrm{r}_n(t+h)]\end{aligned}$$

Here $i=1,2,\ldots,n$ and $j=0,1,\ldots,h$, so that the schema consists of $(h+1)n$ equivalences altogether. The new predicates introduced by the recursion are $\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_n$, which may be any n new letters not previously used; and $\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_k$ are given predicates, which may be either primitive predicates or predicates which were introduced in previous recursions. For convenience in indicating the form of the equivalences, $t+1$ has been written in place of what should properly be written as t', $t+2$ in place of t'', and so on; thus, e.g., $t+h$ indicates a letter t with h accents after it. Similarly, 1 has been written in place of $0'$, 2 in place of $0''$, j in place of a 0 with j accents after it. Then finally the device has been adopted of indicating the elementary parts of a wff, other than the elementary parts T and F, by writing them in brackets. Thus, e.g.,

$$P_{12}[s_1(0), s_2(0), \ldots, s_k(0), s_1(1), s_2(1), \ldots, s_k(1), s_1(2), s_2(2), \ldots, s_k(2)]$$

indicates a wff P_{12} that is built up by means of connectives of propositional calculus from some or all of the elementary formulas $s_1(0), s_2(0), \ldots, s_k(0), s_1(0'), s_2(0'), \ldots, s_k(0'), s_1(0''), s_2(0''), \ldots, s_k(0''), T, F$.

Now in order to represent the action of a given automaton by means of restricted recursive arithmetic, we may use a different predicate to correspond to each element of the automaton—primitive predicates $i_1, i_2, \ldots$ for the input elements, and other letters as convenient for the remaining elements. Normally we shall use $o_1, o_2, \ldots$ for the output elements.

If p is the predicate that corresponds in this way to the element e of the automaton, then $p(t)$ is to denote the state of e at time t. Thus $p(0)$ denotes T or F according as the state of the element e at time 0 is T or F, $p(1)$ denotes T or F according as the state of the element e at time 1 is T or F, and so on. And $p(t)$ may therefore be understood as asserting that the element e is in state T ("is operated," or the like) at time t.

The action of the automaton is then expressed by writing a set of recursion equivalences in the form of one or more wider restricted recursions, constituting definitions by recursion of the intermediate and output predicates (i.e., the predicates that correspond to intermediate and output elements), in terms of the input predicates $i_1, i_2, \ldots, i_\mu$. Given a particular automaton, and the specific rule by which the state of each element at any instant is determined from the states of that element at preceding instants and the states of other elements at the same and preceding instants, the corresponding set of recursion equivalences that characterizes the automaton is then uniquely determined. And conversely, given a set of recursion equivalences of the kind just described, there is then a unique corresponding automaton.

The one-to-oneness of the correspondence between automata and sets of recursion equivalences has been secured by adopting a restricted form of recursion schema. This does not mean any restriction upon the kinds of elements that are available—instead there is an assumption that elements of arbitrary kind are available, within the general definition of automaton as given above, and if we wish a limitation to elements of special kind (e.g. to "and" elements, "or" elements, "not" elements, and one-unit delay elements), it is necessary to impose further restrictions on the recursion schema. But the restrictions which are involved in the schema of wider restricted recursion, as compared to the recursion schema which would be natural if we did not have the application to automata in mind, are dictated rather by two general principles of causality which may be stated briefly as follows: (I) No direct causal action of the future upon the present. (II) No direct causal action of the remote past upon the present. The question, how far back into the past is to be regarded as the immediate rather than the remote past, has an answer in the span $h+1$, which serves as a measure of this.

Two sets of recursion equivalences of the kind here in question are called *equivalent* if the corresponding automata are equivalent. Evidently this notion may have a direct definition, in terms of the recursion equivalences themselves rather than the corresponding automata, provided that some subset of the predicates, not including any of the input predicates $i_1, i_2, \ldots, i_u$ is designated as output predicates. Hence we may state:

THE SIMPLIFICATION PROBLEM. Given a set of recursion equivalences of the kind here in question, to find one equivalent to it, which has a specified form, corresponding to a limitation to elements of special kind from which the automaton is to be constructed, and which is simplest in some suitable sense of simplicity.

The definition of simplicity adopted will evidently depend on the kinds of elements which are regarded as available to construct the automaton, and perhaps on some weighting of them. Thus there is properly not one problem but many. But little has been done with the simplification problem for automata, except in the case of combinational circuits, i.e., automata for which the span $h+1$ is 0, where only propositional calculus is required for the treatment.

In the case of transition systems it is natural to characterize simplicity as minimizing the number of internal states. This has often been followed. But it is clear that such minimization of the number of internal states will be at most a necessary condition of simplicity of a corresponding automaton, and the relevance to the simplification problem for automata is therefore uncertain.

Both the decision problem and the synthesis problem for automata involve a condition which the automaton is required to satisfy, and which we shall call the *synthesis requirement*. The synthesis requirement must be stated in some formalized language L, and we shall in this context suppose that L is either restricted recursive arithmetic itself or an extension of restricted recursive arithmetic obtained by adjoining additional notations—such as, e.g., the binary predicate $=$, the sign $+$ of addition, the quantifiers—together with rules for them. Then we may state the two problems as follows:

THE SYNTHESIS PROBLEM. Given as synthesis requirement a wff of L, containing the input predicates $i_1, i_2, \ldots, i_\mu$ and the output predicates $o_1, o_2, \ldots, o_\nu$, to find definitions by recursion of $o_1, o_2, \ldots, o_\nu$, and any number of intermediate predicates, from $i_1, i_2, \ldots, i_\mu$, in the form of one or more wider restricted recursions, such that the synthesis requirement is satisfied independently of what particular propositional functions are denoted by $i_1, i_2, \ldots, i_\mu$. Or if the given synthesis requirement is impossible in the sense that there is no automaton, and no set of recursion equivalences that satisfies it, this fact is to be ascertained.

THE DECISION PROBLEM. Given as synthesis requirement a wff of L, as before, and given a proposed solution of the synthesis problem in the form of a set of recursion equivalences (by wider restricted recursion), to determine whether the synthesis requirement is in fact a consequence of the recursion equivalences.

Both the synthesis problem and the decision problem may be divided into cases, according to the form of the synthesis requirement, or the language L in which it is expressed. Solution of any case of the synthesis problem implies, *a fortiori*, solution of the corresponding case of the decision problem, because a decision problem may always be turned into a synthesis problem by including the given recursion equivalences as a part of the synthesis requirement.

Existing results in regard to the synthesis problem and the decision problem are summarized in the following table. But it should be emphasized that the indicated synthesis algorithms and decision algorithms are by no means always of practicable length.

	Synthesis requirement expressed in restricted recursive arithmetic	Synthesis problem	Decision problem
Case 1	with one numerical variable only	Solved in [7]	Solved in [11]
Case 2	with any number of numerical variables	Solved in [7]	Solvable by the method of [11]
Case 3	plus quantifiers	Partial solution in [7]	Solved in [7]
Case 4	plus = and <	Incomplete sketch of solution in [7], properly an open problem	Solution is a consequence of [23].
Case 5	plus +, =, and quantifiers	Proved unsolvable [4]	Open

Case 1 of the decision problem was solved by Joyce Friedman [11], the earliest result. The problem of determining the equivalence of two automata which are both given by means of recursion equivalences in the form of wider restricted recursions can be formulated as a subcase of Case 1 of the decision problem, and is therefore covered by Miss Friedman's solution. Her method is also immediately applicable to solution of Case 2 of the decision problem, as pointed out by McNaughton in his review.

Case 1 of the synthesis problem was first solved by me. A modification of the synthesis algorithm, by which it is substantially shortened, was then suggested by J. B. Wright, and this is incorporated in the solution as given in [7]. A subcase of this case of the synthesis problem was also solved independently by Wang [22], without recognizing the relationship to recursive arithmetic.

The unsolvability of Case 5 of the synthesis problem is, of course, in the usual sense, that there is no uniform effective procedure which can be given once for all and by which alone every instance of this case can be solved.

A variety of cases of the synthesis problem and of the decision problem, intermediate between Case 2 and Case 5, immediately suggest themselves as being still open. We may, for example, adjoin the sign of addition, +, to restricted recursive arithmetic. Or we may adjoin the sign + and the binary predicate = to restricted recursive arithmetic; or, as would be equivalent to this, we may adjoin the ternary predicate Σ, as sign of the ternary relation of addition. Other possibilities are restricted recursive arithmetic plus + and quantifiers; or plus = and quantifiers; or plus < and quantifiers. Or, as suggested by Büchi [2], we may consider adjoining singulary predicate-variables and quantifiers for them.

If to restricted recursive arithmetic we adjoin a sign of addition and a sign of multiplication and the sign = (of equality) and quantifiers, we obtain a formalized language which includes first-order arithmetic as a part. It follows that the decision problem (*Entscheidungsproblem*), and hence *a fortiori* the synthesis problem, is then unsolvable—cf. [4].

Turning now to matters less closely related to recursive arithmetic, we introduce the definition of "regularity of an event" in the sense of Kleene, employing for this purpose the "regular-expression language" of McNaughton and Yamada [17]. We follow the original notation of McNaughton and Yamada, except that we change their $\sim$ to a $-$, out of reluctance to confuse notations of propositional calculus and of class calculus.

We suppose that there are $m+1$ input states, of an automaton under consideration, and let them be represented by the numbers $0, 1, \ldots, m$. We use (k) as a notation for the unit class of the input sequence which consists of the single input state k. Or, when convenient, and especially if $m \leqslant 9$, the parentheses enclosing the numeral k may be omitted as an abbreviation.

The notations $\cup$, $\cap$, $-$, and Λ are employed with their usual set-theoretic meanings, i.e., for union of classes, intersection of classes, complement, and null class respectively. The notation ϕ is employed for the unit class of the null sequence.

If a and b are events, then $a \cdot b$ (abbreviated when convenient as ab) is the event whose members are every sequence that is composed of the concatenation of a sequence belonging to a followed by a sequence belonging to b.

If a is an event, then the event a^* is the smallest class of sequences which contains (as members) the null sequence and all sequences belonging to a and is closed under concatenation.

The regular expressions are: (k), for any suitable numeral k; and Λ; and ϕ; and any expression built up out of these by means of $\cup$, $\cap$, $-$, $\cdot$, and $*$ (with, of course, suitable use of brackets).

The regular expressions are intended as names of events, in a way that is implicit in the foregoing explanation. And we may then define an event as being regular if, under some choice of m, there is a regular expression which denotes it. We have the important result:

KLEENE'S THEOREM. *An event can be represented by a finite automaton if and only if it is regular.*

This was first proved by Kleene in [14] and [15], under a definition of regularity equivalent to the one we have just given.

Since the behavior function of an automaton with one binary (i.e., two-state) output is uniquely characterized by giving the event for which the output is T, McNaughton and Yamada propose their regular-expression language for practical use in specifying the behavior of such an automaton, and recommend the redundancy of its notation—as compared, e.g., to the more economical regular-expression languages of [15] and [8]—as facilitating this. For example, if we assume two input states, 0 and 1, the expression

$$-[(0 \cup 1)^*000(0 \cup 1)^*] \cup (0 \cup 1)^*111-[(0 \cup 1)^*000(0 \cup 1)^*]$$

specifies an automaton as having output T at time t if and only if either there have never been three consecutive 0's in the input sequence up to and including the time t (this accounts for the part of the expression up to the end of the first square bracket) *or* (this accounts for the next symbol, $\cup$, in the expression) there have been three consecutive 1's in the input sequence since the last three consecutive 0's (this accounts for the remainder of the expression).

Kleene proposed the problem of a characterization of regular events directly in terms of their expression in a formalized language of ordinary kind, such as one of the usual formulations of first- or second-order arithmetic. This problem has since been solved by Trachtenbrot [20], and in a different way by Büchi [2] and Elgot [10].

These results of Büchi and Elgot are closely related to one another, and the substance of them was originally announced in the joint abstract [3]. Specifically, representability of an event by a finite automaton is equivalent to expressibility in any one of the three following formalized languages:

(1) *Weak second-order arithmetic* (Büchi). The primitive symbols are: numerical variables; singulary predicate-variables; the notation ( ) for application of a propositional function to its argument; the symbol 0 for the number 0; the accent $'$ as a notation for successor; T, F, and connectives of the propositional calculus; quantifiers on numerical variables; quantifiers on singulary predicate-variables. Terms are the same as in restricted recursive arithmetic. An elementary formula is T or F or a singulary predicate-variable followed by a term between ( ). Other wffs (closed and open sentences) are built up from the elementary formulas by means of connectives of the propositional calculus and both kinds of quantifiers.

(2) *Singulary second-order arithmetic with* $=$ *and* $<$ (Elgot). Differs from (1) by omission of 0 and inclusion of the two binary predicates $=$ and $<$.

(3) *First-order arithmetic with* $+$*, the binary predicate* $=$*, and a singulary predicate meaning "is a power of* 2" (Büchi).

In (1) and (2) the predicate-variables are to be understood as having *finite* sets of natural numbers as their values. And in the application to automata, in order to represent an input sequence, each free predicate-variable of a formula is made to correspond to a finite initial sequence of states of one input element by way of some suitable coding. The simplest convention, following Elgot, is just to take the value of each predicate-variable to be the finite set of instants at which the state of the corresponding input element is T, and then to consider the shortest input sequence which is thus represented. This results in a restriction to a proper subclass of all input sequences, the *admissible* input sequences, but it may be seen that the restriction is not very essential. Büchi uses (in effect) a method of coding that avoids this restriction to admissible input sequences.

In the case of (3), instead of using predicate-variables, each finite set of natural numbers is represented by a natural number, the finite sum $\Sigma 2^t$, where the summation is over all natural numbers t that belong to the set.

That (1) and (2) are equivalent, and that both have a redundant list of primitives, is as a matter of fact quite obvious. Definitions given by Elgot show that we may without loss omit the primitive symbol 0 from (1), or the primitive symbols $=$ and $<$ from (2).

The two notes by Trachtenbrot, [20] and [21], were unknown to me at the time this lecture was delivered, and were called to my attention by several members of the audience. The dates attached to the earlier note [20] place it as approximately simultaneous with [11] and the privately circulated first edition of [7], hence earlier than the abstract [3], later than [4] and the abstract of [11]—all of which it overlaps to some extent, though more in point of view and method than in specific content.

In [20] Trachtenbrot characterizes the behavior of an automaton with one binary output by means of the following formalized language:

(4) *Singulary second-order functional calculus with bounded quantifiers.* The primitive symbols are: numerical variables; singulary predicate variables; the notation ( ) for application of a propositional function to its argument; connectives of the propositional calculus; quantifiers on singulary predicate-variables; and bounded quantifiers on numerical variables. The elementary formulas consist of a singulary predicate-variable followed by a numerical variable between ( ), and other wffs are built up from the elementary formulas by means of connectives of the propositional calculus and quantifiers.

Trachtenbrot introduces four kinds of bounded quantifiers: $(\mathrm{a})_{\mathrm{a}<\mathrm{b}}$, $(\exists \mathrm{a})_{\mathrm{a}<\mathrm{b}}$, $(\mathrm{a})_{\mathrm{a}\leqslant\mathrm{b}}$, and $(\exists \mathrm{a})_{\mathrm{a}\leqslant\mathrm{b}}$, where a and b are any two distinct numerical variables. We shall say in each case that the variable a is bounded by the variable b. Evidently any one of these four kinds of bounded quantifiers would suffice, as the three others would be definable from it. The ordinary, unbounded quantifiers on numerical variables are not used.

In (4) the predicate-variables are to be understood as having arbitrary singulary propositional functions (or arbitrary sets) of natural numbers as their values. And in the application to automata, each free predicate-variable of a formula corresponds to one input element in the same fashion that was explained above for the predicates $i_1, i_2, \dots$ used in restricted recursive arithmetic. A *t-formula* is a wff of (4) in which the only free numerical variable is the particular variable t and in which—to state it briefly—every bound numerical variable is bounded, either by t, or by a variable that is bounded by t, or by a variable that is bounded by a variable that is bounded by t, or etc. Then we have:

TRACHTENBROT'S THEOREM. The behavior of any finite automaton with one binary output $o(t)$ can be characterized by an equivalence of the form $o(t) \equiv \mathrm{S}(t)$, where $\mathrm{S}(t)$ is a t-formula whose free predicate-variables correspond to the inputs of the automaton. And conversely every equivalence of this form characterizes the behavior of some finite automaton with one binary output.

As a characterization of the behavior of finite automata with one binary output, Trachtenbrot's theorem is evidently more direct than the results of Kleene, Büchi, and Elgot. As a characterization of events representable by a finite automaton (i.e., of regular events) it is less direct. But it can be made to yield such a characterization by taking each free predicate-variable f in $\mathrm{S}(t)$ to stand for a finite initial sequence of input states of length $t+1$, namely the sequence $\mathrm{f}(0), \mathrm{f}(1), \dots, \mathrm{f}(t)$. (This amounts to taking the formalized language in such a way that the free predicate-variables have a different range of values from the bound predicate-variables, a device which Trachtenbrot himself does not adopt explicitly in either [20] or [21].)

In order to obtain a language for practical use in specifying the behavior of automata, it may be desirable to modify Trachtenbrot's language (4) by adding the redundant primitive symbols $=$, $<$, 0, and $'$, and then to replace the class of t-formulas by the more quickly recognizable class of formulas in which the only free numerical variable is t and all bound numerical variables are bounded by t. The resulting language seems to offer less immediate facility than the regular-expression language of McNaughton and Yamada, but it may be more suitable for use in cases in which some further application of mathematical logic is to be made.

Büchi, Elgot, and Trachtenbrot are all interested, not only in applications of logic to automata theory, but also in the reverse application by which the consideration of automata is made to yield results that belong to the field of mathematical logic. And the main result announced in [21] falls under the latter head. But [21] also announces, in effect, still another characterization of regularity of events by means of expressibility in an appropriate formalized language. Namely in Büchi's characterization, weak second-order arithmetic may be replaced by another formalized language, which has the same primitive symbols and the same class of wffs as weak second-order arithmetic, but in which the values of the bound predicate-variables are taken to be arbitrary sets of natural numbers, while the values of the free predicate-variables are taken to be finite sets of natural numbers.

Finally we return to the synthesis problem, to cite some cases which belong in the present context rather than in that of restricted recursive arithmetic.

Case a. Synthesis requirement given by a regular expression, denoting an event which an automaton with one binary output is required to represent. Problem solved originally by Kleene [14], [15]. Revised treatments by Copi, Elgot, and Wright [8]; McNaughton and Yamada [17]. Elgot [10] solves a somewhat generalized form of this synthesis problem.

Case b. Synthesis requirement given by a *t*-formula that expresses the (one binary) output in terms of the inputs. Solution briefly sketched by Trachtenbrot [20].

Case c. Synthesis requirement given by a wff of the form $(\exists s_1)(\exists s_2)\ldots(\exists s_n)(t)M$, belonging to a language which is like weak second-order arithmetic except that the predicate-variables are reinterpreted as ranging over all singulary propositional functions (all sets) of natural numbers. Here $s_1, s_2, \ldots, s_n$ are predicate-variables, t is an individual variable, and the quantifier-free matrix M contains no individual variables except t. The free predicate-variables are interpreted, some as corresponding to inputs and some as corresponding to outputs, so that the synthesis requirement expressed is a condition relating the inputs and the outputs. Solved by Elgot [10].

It is clear that, as regards the form of the synthesis requirement, Case c would fit into the table that was given above in connection with restricted recursive arithmetic. But Elgot's method is different, not making use of restricted recursive arithmetic (and in fact leading in the first instance to a transition system rather than an automaton).

Case d. Synthesis requirement given by an open sentence of first-order arithmetic with $'$, $+$, and $=$, and with free singulary predicate-variables, interpreted as outputs, and as ranging over the ultimately periodic propositional functions of natural numbers. Decision problem solved, but synthesis problem unsolvable. See Elgot [10].

Case e. Synthesis requirement given by am arbitrary wff of the same formalized language that is described under Case c. The decision problem is solved by Büchi [23]. The general synthesis problem for this case is still open (as of December 1962), though unpublished results stronger than that of Case c are cited in a letter from Büchi. Evidently Case e would fit into the table that was given above, as intermediate between Cases 3–4 and Case 5.

REFERENCES

Application of propositional calculus

[A] Ehrenfest, P., Review of the Russian translation of Louis Couturat's *L'algèbre de la logique*. *Журнал Русскаго Физико-химическаго Общества*, section of physics, 42 (1910), part 2, 382–387.

[B] Nakasima, Akira & Hanzawa, Masao, A series of papers published in Japan, some by the joint authors, and some by Nakasima alone. The two earliest papers are in Japanese, by Nakasima and Hanzawa, and appeared in *J. Inst. Elect. Comm. Engrs. Japan*, no. 165 (Dec. 1936) and no. 167 (Feb. 1937). I have not seen these papers, but a condensed English translation of them which appeared in *Nippon Elect. Comm. Eng.*, no. 9 (Feb. 1938), 32–39. Later papers appeared in the latter periodical, no. 10 (Apr. 1938), 178–179; no. 12 (Sept. 1938), 310–314; no. 13 (Nov. 1938), 405–412; no. 14 (Dec. 1938), 459–466; no. 24 (Apr. 1941), 203–210. These also are English translations of papers in Japanese which preceded them.

[C] Shannon, C. E., A symbolic analysis of relay and switching circuits. *Trans. Amer. Inst. Elect. Engrs.*, 57 (1938), 713–723.

[D] Shestakoff, V. I., Алгебра двухполюсных схем, построенных исключительно из двухполюсников (Алгебра А-схем). *Автоматика и Телемеханика*, no. 2 (1941), 15–24.

[E] Shestakoff, V. I., A paper of the same title as [D]. *Журнал Технической Физики*, 11, 6 (1941), 532–549.

[F] Shestakoff, V. I., Representation of characteristic functions of propositions by expressions realizable by relay-contact circuits. (Russian with English summary.) *Bull. Acad. Sci. URSS* (sér. math., 10 (1946), 529–554.

[G]. Yanovskaya, S. A., Основания математики и математическая логика. *Математика в СССР за Тридцать Лет 1917–1947*, pp. 9–50. Moscow and Leningrad, 1948.

Application of mathematical logic beyond propositional calculus

[1]. Berkeley, E. C., *Circuit Algebra–Introduction*. New York, 1952.

[2]. Büchi, J. R., Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlagen Math.*, 6 (1960), 66–92.

[3]. Büchi, J. R. & Elgot, C. C., Decision problems of weak second order arithmetics and finite automata, Part I. Abstract in *Notices Amer. Math. Soc.*, 5, 7 (Dec. 1958), 834.

[4]. Büchi, J. R., Elgot, C. C. & Wright, J. B., The non-existence of certain algorithms of finite automata theory. Abstract in *Notices Amer. Math. Soc.*, 5, no. 1 (Feb. 1958), 98.

[5]. Burks, A. W. & Wright, J. B., Theory of logical nets. *Proc. I.R.E.*, 41 (1953), 1357–1365.

[6]. Church, Alonzo, Review of E. C. Berkeley's "The algebra of states and events". *J. Symb. Logic*, 20 (1955), 286–287.

[7]. Church, Alonzo, Application of recursive arithmetic to the problem of circuit synthesis. *Summaries of Talks Presented at the Summer Institute for Symbolic Logic, Cornell University 1957*, 2nd ed., 3–50. Princeton, 1960.

[8]. Copi, I. M., Elgot, C. C. & Wright, J. B., Realization of events by logical nets. *J. Ass. Comput. Mach.*, 5 (1958), 181–196.

[9]. Elgot, C. C., Decision problems of weak second order arithmetics and

finite automata, Part II. Abstract in *Notices Amer. Math. Soc.*, 6, 1 (Feb. 1959), 48.

[10]. ELGOT, C. C., Decision problems of finite automata design and related arithmetics. *Trans. Amer. Math. Soc.*, 98 (1961), 21–51. Errata, ibid., 103 (1962), 558–559.

[11]. FRIEDMAN, JOYCE, Some results in Church's restricted recursive arithmetic. *J. Symb. Logic*, 22, 4 (for 1957, pub. 1958), 337–342. Reviewed by Robert McNaughton in *J. Symb. Logic*, 24, 3 (for 1959, pub. 1960), 241–242. Abstract in *J. Symb. Logic*, 21 (1956), 219.

[12]. GOTÔ, MOTINORI, Application of logical mathematics to the theory of relay networks. *Japan Sci. Rev.*, 1, 3 (1950), 35–42.

[13]. HARTREE, D. R., *Calculating Instruments and Machines*. Urbana, 1949.

[14]. KLEENE, S. C., *Representation of Events in Nerve Nets and Finite Automata*. RAND memorandum, Dec. 1951.

[15]. KLEENE, S. C., Representation of events in nerve nets and finite automata. *Automata Studies*, 3–41. Princeton, 1956.

[16]. MCCULLOCH, W. S. & PITTS, WALTER, A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5 (1943), 115–133.

[17]. MCNAUGHTON, R. & YAMADA, H., Regular expressions and state graphs for automata. *I.R.E. Trans.*, EC-9, no. 1 (1960), 39–47.

[18]. VON NEUMANN, JOHN, *First Draft of a Report on the EDVAC*. Moore School of Electrical Engineering, University of Pennsylvania, June 30, 1945.

[19]. SKOLEM, THORALF, *Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich*. Skrifter utgit av Videnskapsselskapet i Kristiania, I. matematisk-naturvidenskabelig klasse, no. 6 (1923).

[20]. TRACHTENBROT, B. A., Синтез логических сетей, операторы которых описаны средствами исчисления одноместных предикатов. *Доклады Академии Наук СССР*, 118, 4 (1958), 646–649. Reviewed by J. C. Shepherdson in *Zbl. Math.*, 84 (1960), 11–12.

[21]. TRACHTENBROT, B. A., Некоторые построения в логике одноместных предикатов. *Доклады Академии Наук СССР*, 138, 2 (1961), pp. 320–321. English translation in *Soviet Math.*, 2, 3 (1961), 623–625.

[22]. WANG, HAO, Circuit synthesis by solving sequential Boolean equations. *Z. Math. Logik Grundlagen Math.*, 5 (1959), 291–322. Reviewed by C. C. Elgot in *J. Symb. Logic*, 25, 4 (for 1960, pub. 1962), 373–375.

(Added in proof):

[23]. BÜCHI, J. R., On a decision method in restricted second-order arithmetic. *Logic, Methodology, and Philosophy of Science, Proc. of the 1960 Internat. Congress*, Stanford 1962, 1–11.

МАРКОВСКИЕ ПРОЦЕССЫ И ЗАДАЧИ АНАЛИЗА

Е. Б. ДЫНКИН

Связи между марковскими процессами и дифференциальными уравнениями известны уже давно. Еще в начале тридцатых годов они были объектом глубоких исследований Колмогорова, Петровского, Хинчина. За последние годы были обнаружены новые замечательные связи между марковскими процессами и классическим анализом. Эти связи плодотворны не только для теории вероятностей, но и для анализа. Современная теория марковских процессов становится все более и более не только разделом теории вероятностей, но и важной областью анализа, оказывающей влияние на развитие таких классических областей как теория эллиптических и параболических дифференциальных уравнений, теория потенциала и т. п.

О широком интересе к этой области свидетельствует место, которое уделено ей в программе нашего съезда, так же, как и в программе предыдущего международного съезда в Эдинбурге. Сыгравшие выдающуюся роль в новейшем развитии теории марковских процессов исследования Феллера были освещены автором в часовом докладе на Эдинбургском конгрессе. Важные направления в теории марковских процессов связаны с именами Ито и Ханта, которые выступают с обзорными докладами на нашем съезде.

Это облегчает мою задачу. Я не буду пытаться охватить все аспекты современного развития теории марковских процессов (задача, неразрешимая в рамках часового доклада), а сосредоточу внимание на нескольких направлениях, связанных с моей собственной работой и работой группы молодых советских математиков, находящихся со мной в постоянном научном контакте.

Речь будет идти о понятии характеристического оператора марковского процесса, позволяющем дать вероятностную формулировку и вероятностное решение ряда задач теории дифференциальных уравнений. Значительное место будет уделено понятию аддитивного функционала от марковского процесса и различным преобразованиям марковских процессов, связанным с аддитивными функционалами. Будет показано, что с помощью таких преобразований можно получить некоторое представление о строении наиболее общего строго марковского процесса с непрерывными траекториями на топологическом многообразии E. Мы введем определенные классы функций, которые естественно назвать гармоническими и супергармоническими функциями, связанными с марковским процессом, и остановимся на некоторых задачах, возникающих в связи с изучением множества всех неотрицательных гармонических функций.

§ 1. Современное определение марковского процесса

1.1. В 1828 году английский ботаник Броун заметил, что мелкие частицы, погруженные в жидкость, хаотически движутся, беспрестанно меняя направление движения. В первых работах, посвященных математическому описанию этого явления, изучалась так называемая переходная функция $P(t, x, \Gamma)$, выражающая вероятность того, что частица, вышедшая из точки x, через время t попадает в множество Γ.

Более развитая теория имеет своим объектом траекторию движения x_t. Случайный характер движения выражается математически предположением, что $x_t = x_t(\omega)$, где ω принадлежит некоторому множеству Ω, на котором задан набор вероятностных мер P_x (это множество называется пространством элементарных событий). Множества A, для которых определены значения $P_x(A)$, называются событиями, и значение $P_x(A)$ истолковывается как вероятность события A в предположении, что движение начинается из точки x. Переходная функция определяется формулой $P(t,x,\Gamma) = P_x\{x_t \in \Gamma\}$.

1.2. Мы приходим таким образом к современному определению марковского процесса как пары (x_t, P_x), где $x_t = x_t(\omega)$ — функция, заданная при $t \geqslant 0$ и $\omega \in \Omega$, и P_x — набор вероятностных мер в пространстве Ω. *Фазовое пространство*, в котором принимает значения функция x_t, является в случае броуновского движения некоторой областью трехмерного евклидова пространства. Вообще же это произвольное множество E, в котором выделена некоторая система «измеримых подмножеств». Основное условие, связывающее функцию x_t и меры P_x, — это марковский принцип независимости будущего от прошлого при известном настоящем. Точнее, при известном значении x_t прогноз дальнейшего движения частицы не зависит от характера движения до момента t (¹).

Разобьем всю траекторию процесса на две части: до момента τ первого достижения некоторого множества Γ и после этого момента. Предположим, что нам известно x_τ. Существенно ли знание движения до момента τ для прогноза движения после момента τ? Физическая интуиция требует отрицательного ответа. Однако такой ответ вовсе не вытекает из определения марковского процесса, в котором участвует фиксированный момент t, а не случайный момент τ. Марковские процессы, для которых условие независимости будущего от прошлого при известном настоящем выполняется не только для постоянного, но и для определенного класса случайных моментов τ, называются *строго марковскими процессами* (²).

(¹) Строгая формулировка опирается на понятие условной вероятности и требует, чтобы для любого измеримого множества Γ и любых $t, h \geqslant 0$ $P_x\{x_{t+h} \in \Gamma / x_s, s \leqslant t\} = P(h, x_t, \Gamma)$ при почти всех ω.

(²) Изучение строго марковских процессов как самостоятельного класса марковских процессов было начато в 1955–1956 годах в работах Дынкина [5], [6], [8], Дынкина и Юшкевича [18] и Рэя [38]. В работе [18] впервые дано общее определение строго марковских процессов, строятся примеры марковских процессов, не являющихся строго марковскими, и выводятся условия, достаточные для того, чтобы марковский процесс являлся строго марковским.

1. 3. До сих пор мы считали, что x_t определено для всех $t \geqslant 0$. Однако, многие естественные конструкции приводят к процессам, для которых $x_t(\omega)$ определено лишь в некотором интервале $(0, \zeta(\omega))$. Современная теория марковских процессов охватывает и такие обрывающиеся процессы.

Определение марковского процесса не требует, чтобы в фазовом пространстве была задана какая-нибудь топология. Однако в развитой теории широко используются условия топологического характера и ту или иную топологию в фазовом пространстве приходится вводить. Среди различных топологий особое место занимает так называемая естественная топология, в которой открытые множества выделяются следующим условием: траектория, начинающаяся в произвольной точке такого множества, с вероятностью единица не выходит из этого множества в течении положительного времени (см. Дуб [2], Хант [28], Дынкин [10]). Для широкого класса процессов доказано, что измеримая функция $f(x)$ является непрерывной в естественной топологии тогда и только тогда, когда она непрерывна справа вдоль почти всех траекторий процесса (¹).

§ 2. Операторы сдвига функций. Характеристические и инфинитезимальные операторы

2. 1. Общая схема, связывающая марковские процессы с анализом, основана на понятии сдвига функции, заданной в фазовом пространстве E. Рассмотрим какую-нибудь неотрицательную измеримую функцию $\tau(\omega)$. Пусть $f(x)$ — измеримая функция в пространстве E. Тогда $f(x_\tau)$ является функцией в пространстве Ω. Интеграл этой функции по мере P_x (если он имеет смысл) и есть значение сдвинутой функции в точке x. В виде формулы это записывается так

$$T_\tau f(x) = M_x f(x_\tau)\ (^2).$$

Рассмотрим внимательнее случай, когда $\tau = t$ не зависит от ω. Соответствующий оператор сдвига выражается через переходную функцию по формуле

$$T_t f(x) = \int_E P(t, x, dy) f(y).$$

Из марковского принципа вытекает, что $T_s T_t = T_{s+t} (s, t \geqslant 0)$, т. е. операторы T_t образуют полугруппу. Оператор

$$A f(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t}$$

Дальнейшие указания относительно литературы по строго марковским процессам можно найти в [11].

(¹) Этот результат принадлежит Гирсанову [23], опиравшемуся на некоторые идеи Дуба [2].

(²) Через $M_x \xi$ обозначается интеграл функции ξ по мере P_x, распространенный на всю область определения функции $\xi(\omega)$.

называется *инфинитезимальным оператором* марковского процесса. При очень широких предположениях инфинитезимальный оператор определяет однозначно переходную функцию процесса (см. [7], [17]).

Другой важной характеристикой процесса является характеристический оператор, который определяется формулой

$$\mathfrak{A} f(x) = \lim_{U \downarrow x} \frac{T_{\tau(U)} f(x) - f(x)}{M_x \tau(U)}.$$

Здесь U — окрестность точки x, $\tau(U)$ — момент первого выхода из U, предельный переход производится, когда U стягивается к x. (Предполагается, что в фазовом пространстве задана некоторая топология.)

Между характеристическим и инфинитезимальным оператором марковского процесса существует тесная связь. Именно, если в некоторой топологии траектории процесса непрерывны справа, а операторы сдвига T_t оставляют инвариантным множество всех ограниченных непрерывных функций, то характеристический оператор $\mathfrak{A}$ является расширением инфинитезимального оператора A (1).

Из формулы, определяющей оператор $\mathfrak{A}$, видно, что если $\mathfrak{A} f(x)$ определено и функция f достигает в точке x неотрицательного максимума, то $\mathfrak{A} f(x) \leqslant 0$. Это свойство часто называют *принципом максимума*.

Начиная с этого места мы сосредоточим все внимание на процессах, траектории которых непрерывны. Для таких процессов в момент первого выхода из множества U частица попадает на границу U. Поэтому значение $\mathfrak{A} f(x)$ определяется по значениям функции f в сколь угодно малой окрестности точки x. Оператор $\mathfrak{A}$ является *локальным* линейным оператором.

§ 3. Диффузионные процессы. Вероятностное решение дифференциальных уравнений

3. 1. Строго марковский процесс с непрерывными траекториями условимся называть *диффузионным процессом*, если $\mathfrak{A} f(x)$ определено для каждой функции f, дважды непрерывно дифференцируемой в окрестности точки x. (Предполагается, что в фазовом пространстве введена структура гладкого многообразия.) Доказывается, что для диффузионного процесса характеристический оператор совпадает на дважды непрерывно дифференцируемых функциях с некоторым дифференциальным оператором второго порядка

$$L f(x) = \sum a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum b_i(x) \frac{\partial f(x)}{\partial x_i} - c(x) f(x).$$

Из принципа максимума вытекает, что $c(x) \geqslant 0$ и выполняется ослабленное условие эллиптичности: $\Sigma a_{ij}(x) \lambda_i \lambda_j \geqslant 0$ для любых вещественных $\lambda_1, \lambda_2, \ldots$. Оператор L называется *производящим дифференциальным*

(1) Понятие характеристического оператора и связь между характеристическими и инфинитезимальными операторами были впервые изучены в работах [5], [8].

оператором диффузионного процесса. (Он является сужением характеристического оператора.) Величины a_{ij}, b_i и c называются, соответственно, коэффициентами диффузии, коэффициентами сноса и коэффициентом обрыва (если процесс не обрывается, то $c(x)=0$).

3. 2. Важно научиться строить диффузионные процессы, отправляясь от дифференциальных операторов L. Диффузионный процесс в евклидовом пространстве, отвечающий оператору Лапласа Δ, был построен еще в 1923 году Винером. Мы будем называть его *винеровским процессом*. Конструкцию Винера можно распространить на любой эллиптический оператор L, для которого удается построить фундаментальное решение параболического дифференциального уравнения

$$\frac{\partial u_t}{\partial t} = L u_t.$$

В случае евклидова пространства для этого достаточно, чтобы коэффициенты оператора L были ограничены и удовлетворяли условию Гельдера и чтобы существовала положительная нижняя граница для собственных значений матрицы коэффициентов при старших производных ([1]).

Более простой, вероятностный метод построения необрывающихся диффузионных процессов был предложен Ито [30], [31].

Пусть (x_t, P_x) — одномерный винеровский процесс и $(\tilde{x}_t, \tilde{P}_x)$ — диффузионный процесс с производящим дифференциальным оператором

$$L = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + m(x)\frac{d}{dx}.$$

Ито показал, что оба процесса можно реализовать на одном и том же пространстве элементарных событий Ω так, чтобы $\tilde{P}_x = P_x$ и траектории $\tilde{x}_t$ и x_t были связаны соотношением

$$\tilde{x}_t = \tilde{x}_0 + \int_0^t m(x_u)\,du + \int_0^t \sigma(x_u)\,dx_u \tag{1}$$

(справа стоит так называемый стохастический интеграл, общее определение которого было дано Ито). Формулу (1) можно рассматривать как интегральное уравнение, позволяющее выразить $\tilde{x}_t$ через x_t. Аналогичное уравнение пишется и для многомерных процессов.

Уравнение (1) можно записать в дифференциальной форме

$$d\tilde{x}_t = m(x_t)\,dt + \sigma(x_u)\,dx_u.$$

Последнее соотношение представляет собой стохастический аналог обыкновенного дифференциального уравнения и обладает многими свойствами таких уравнений: его можно решать методом последовательных приближений или методом ломанных Эйлера, его исследование нисколько не

([1]) См. Погожельский [36], [37].

усложняется с увеличением числа измерений пространства. Еще одним замечательным преимуществом метода Ито является его нечувствительность к вырождению исходного дифференциального оператора.

3. 3. Рассмотрим несколько типичных задач теории дифференциальных уравнений, связанных с оператором L:

А. (Задача Дирихле.) $Lf(x)=0$ для $x \in G, f(x)=\varphi(x)$ для $x \in G'$.

Б. $Lf(x)=-g(x)$ для $x \in G, f(x)=0$ для $x \in G'$.

В. $\dfrac{\partial u_t(x)}{\partial t}=Lu_t(x)$ для $t>0, x \in G; u_0(x)=v(x)$ для $x \in G; u_t(x)=0$ для $x \in G'$.

Здесь G означает некоторую область в евклидовом пространстве, G' — ее границу, φ, g, v — известные функции, f и u_t — неизвестные функции, которые надо найти.

Заменив в формулировках задач А, Б, В оператор L на характеристический оператор $\mathfrak{A}$, мы получим новые задачи А′, Б′, В′. Поскольку $\mathfrak{A} \supseteq L$, то всякое решение какой-нибудь из задач А, Б, В является в то же время решением соответствующей задачи А′, Б′, В′. Естественно назвать решения задач А′, Б′, В′ *обобщенными решениями* задач А, Б, В.

Из принципа максимума для оператора $\mathfrak{A}$ вытекает, что при широких предположениях задачи А′, Б′, В′ имеют не более одного решения. Если это решение дважды непрерывно дифференцируемо, то она является решением соответствующей задачи А, Б, В. Таким образом, в этом случае обобщенное решение задачи А, Б, В является «классическим решением».

Замечательно, что решение задач А′, Б′, В′ всегда выражается явными вероятностными формулами. Именно, обозначая через β момент первого выхода области G, имеем:

для А′: $f(x)=M_x \varphi(x_\beta)$

для Б′: $f(x)=M_x \int_0^\beta g(x_t)\,dt$

для В′: $u_t(x)=M_x v(x_t)\,\chi_{\beta<t}$ $\quad \left(\chi_A(\omega)=\begin{cases}1 \text{ при } \omega \in A\\ 0 \text{ при } \omega \notin A\end{cases}\right)$

Эти формулы позволяют исследовать решение с качественной стороны, в частности, исследовать его зависимость от начальных и граничных данных и правых частей (см., например, Хасьминский [24]–[27], Фрейдлин [21], [22]). Кроме того они могут служить основой для вычислений по методу Монте-Карло.

Аналогичные формулы можно дать и для более сложных уравнений. Например, обобщенное решение дифференциального уравнения

$$Lf - Vf = -g$$

при нулевых граничных условиях выражается формулой

$$f(x) = M_x \int_0^\beta \exp\left(- \int_0^t V(x_u)\, du \right) g(x_t)\, dt.$$

3.4. В задаче В мы рассматривали простейшее, нулевое граничное условие для параболического дифференциального уравнения $\partial u_t / \partial t = L u_t$. Известно, что для этого уравнения можно ставить и другие граничные условия, например, можно потребовать, чтобы обращалась в нуль нормальная производная $\partial u_t / \partial n$. Различным граничным условиям соответствуют различные типы поведения движущейся частицы на границе области. Например, нулевым граничным условиям соответствует исчесновение в момент первого достижения границы; условию $\partial u_t / \partial n = 0$ соответствует отражение по направлению нормали.

Задача определения всех возможных граничных условий для диффузионных процессов была изучена в одномерном случае Феллером [19], [20] и в многомерном случае Вентцелем [47], [48].

Задачу можно уточнить следующим образом. Пусть G — область с гладкой границей G' и L — дифференциальный оператор в G, коэффициенты которого продолжаются непрерывно на $G \cup G'$. Требуется описать все диффузионные процессы в замкнутой области $G \cup G'$, для которых производящий дифференциальный оператор во всех внутренних точках совпадает с L. Пока еще не было определено, что значит, что процесс является диффузионным в граничной точке x. Так же, как и для внутренней точки это понятие определяется через характеристический оператор. Требуется, чтобы значение $\mathfrak{A} f(x)$ было определено либо (а) для всех гладких функций, либо (б) для всех гладких функций, подчиненных одному линейному условию (под гладкими функциями понимаются здесь функции, дважды непрерывно дифференцируемые в некоторой окрестности точки x).

Вентцель показал (¹), что в случае (а) на гладких функциях

$$\mathfrak{A} f(x) = -qf + \gamma \frac{\partial f}{\partial n} + L' f,$$

где L' — дифференциальный оператор второго порядка на G'(²); в случае (б) линейное условие, задающее область определения имеет вид

$$0 = -qf + \gamma \frac{\partial f}{\partial n} + L' f.$$

Если функция $\mathfrak{A} f$ непрерывна, то в первой из написанных формул можно заменить $\mathfrak{A}$ на L, и обе формулы записываются в виде

$$-\sigma L f - qf + \gamma \frac{\partial f}{\partial n} + L' f = 0.$$

(¹) Результаты Вентцеля излагаются в переработанном виде.

(²) Мы считаем, что траектории процесса непрерывны. Феллер и Вентцель допускали возможность разрывов на границе. Поэтому у них появляются дополнительные интегральные члены.

Итак, для того, чтобы задать граничное условие, надо определить на границе функции σ, q, γ и дифференциальный оператор L'. При $L' \equiv 0$ получаем граничные условия, хорошо известные в теории дифференциальных уравнений. Теоремы существования и единственности для общего случая разбирались Вентцелем и Уено [44]. Однако, полученные в этом направлении результаты еще никак нельзя считать окончательными.

С наглядной точки зрения сформулированный выше результат означает, что единственно возможными типами поведения диффундирующей частицы на границе области являются остановка, исчезновение, отражение, диффузия по границе и их различные комбинации. Слово «комбинация» означает просто линейную комбинацию соответствующих граничных условий, но вероятностный смысл такого комбинирования совсем не прост.

Каждому типу граничных условий соответствует определенный процесс, происходящий на границе. Он определен на случайном множестве моментов, в которые частица находится на границе. Изучение граничных процессов можно поэтому рассматривать как одну из задач, еще не построенной общей теории марковских процессов со случайной областью определения. Построение такой теории представляется весьма интересной проблемой. (Интересные соображения по этому поводу были высказаны в одном из докладов Колмогорова на секции теории вероятностей и математической статистики Московского математического общества.)

§ 4. Аддитивные функционалы

4.1. За последние годы одним из важнейших орудий исследования и применения марковских процессов становятся аддитивные функционалы от этих процессов.

Аддитивный функционал от марковского процесса (x_t, P_x) представляет собой функцию, сопоставляющую каждому интервалу времени $[s,t]$ случайную величину φ_t^s зависящую лишь от течения процесса за время $[s,t]$ (1). При этом требуется, чтобы $\varphi_t^s + \varphi_u^t = \varphi_u^s$ для любых $s \leqslant t \leqslant u$. Мы будем рассматривать только аддитивные функционалы, удовлетворяющие следующему условию однородности: при замене траектории x_t на x_{t+h} значение φ_t^s заменяется на φ_{t+h}^{s+h} (2).

Примером аддитивного функционала является

$$\varphi_t^s = \int_s^t V(x_u)\, du,$$

где $V(x)$ — какая-нибудь измеримая функция в фазовом пространстве. Если функция V неотрицательна, то функционал φ_t^s принимает только неотрицательные значения. Такие функционалы представляют особый интерес для применений. В 1958 г. Волконский [45] (см. также [46]) описал все непрерывные неотрицательные аддитивные функционалы от

(1) Точнее, функция $\varphi_t^s(\omega)$ измерима относительно σ-алгебры, порожденной функциями $x_u(\omega)$ при $u \in [s,t]$.

(2) Точнее, если для некоторых ω_1 и ω_2 из Ω и $h > 0$ $x_t(\omega_1) = x_{t+h}(\omega_2)$ при всех t, то $\varphi_t^s(\omega_1) = \varphi_{t+h}^{s+h}(\omega_2)$ при всех $s \leqslant t$.

одномерного винеровского процесса. Оказалось, что эти функционалы находятся во взаимно однозначном соответствии с мерами на прямой, принимающими конечные значения на каждом конечном отрезке. Это соответствие непрерывно, если рассматривать для мер слабую сходимость, а для функционалов сходимость по вероятности. Функционал, соответствующий мере μ, записывается в виде

$$\varphi_t^s = \int_s^t \frac{d\mu}{dx}(x_u)\, du. \tag{2}$$

При этом $d\mu/dx$ означает плотность меры μ относительно меры Лебега. Если такой плотности не существует, то запись носит символический характер и расшифровывается посредством естественного предельного перехода.

Для многомерного винеровского процесса задача была решена в 1960 г. Вентцелем [49] и Маккином и Танака [33]. Так же, как и в одномерном случае, имеет место однозначное соответствие между неотрицательными аддитивными функционалами и некоторым классом мер в фазовом пространстве. Однако, этот класс мер описывается сложнее. (Он содержит, в частности, все меры, обладающие ограниченной плотностью относительно меры Лебега, а также все меры с ограниченным ньютоновым потенциалом.) Непрерывность соответствия между мерами и функционалами больше не имеет места, однако интегралу, стоящему в правой части формулы (2) все же можно придать смысл с помощью более осторожного предельного перехода.

Не исключено, что теорему о непрерывности соответствия между мерами и функционалами можно спасти, если при определении слабой сходимости мер рассматривать не обычную топологию евклидова пространства, а естественную топологию, связанную с винеровским процессом. Было бы интересно выяснить, верно ли это предположение.

4.2. Примером непрерывного аддитивного функционала, принимающего значения разных знаков, может служить

$$\varphi_t^s = f(x_t) - f(x_s), \tag{3}$$

где f — произвольная непрерывная функция. Дынкин [13] показал, что непрерывные аддитивные функционалы от винеровского процесса можно получить с помощью стохастических интегралов по формуле

$$\varphi_t^s = \int_s^t b(x_u) dx_u. \tag{4}$$

Опираясь на некоторые идеи Скорохода [42] Вентцель [50], [51] доказал, что всякий непрерывный аддитивный функционал от одномерного винеровского процесса представляется в виде суммы функционалов вида (3) и (4). Аналогичный результат получен им и для многомерного случая, но его формулировка несколько сложнее.

§ 5. Супергармонические и гармонические функции

5.1. Супергармонические и гармонические функции играют важную роль в классическом анализе.

Одно из возможных определений супергармонической функции требует, чтобы функция была непрерывна сверху и чтобы среднее ее значение по любой сфере не превосходило ее значения в центре сферы.

Пусть τ — момент первого выхода траектории винеровского процесса из сферы S с центром в точке x. Очевидно, точка x_τ равномерно распределена по сфере S, и поэтому среднее значение функции f по S равно $T_\tau f(x) = M_x f(x_\tau)$. Таким образом, основное неравенство для супергармонических функций можно записать в виде $T_\tau f(x) \leqslant f(x)$. Доказывается, что это неравенство сохраняет силу, если сфера с центром в x заменяется произвольной компактной окрестностью точки x. Доказывается также, что супергармонические функции непрерывны в естественной топологии и ограничены снизу на каждом компакте. Более того, оказывается, что перечисленные требования (непрерывность в естественной топологии, ограниченность снизу на каждом компакте и неравенство $T_\tau f(x) \leqslant f(x)$ для каждой компактной окрестности точки x) равносильны супергармоничности функции f. Мы приходим к вероятностному определению супергармонических функций.

Это определение сохраняет смысл при замене винеровского процесса на любой марковский процесс в локально компактном фазовом пространстве. Таким образом, для каждого марковского процесса можно ввести класс связанных с ним супергармонических функций (¹). Для процессов с непрерывными траекториями это определение естественно локализуется, так что можно говорить о функциях, супергармонических в некоторой области G.

5.2. В классическом анализе заметное место занимает задача о представлении супергармонических функций в виде потенциалов. Ньютонов потенциал меры μ в трехмерном евклидовом пространстве E представляет собой функцию $f(x)$, определяемую формулой

$$f(x) = \int_E \frac{\mu(dy)}{|y-x|}$$

($|y-x|$ означает длину вектора $y-x$). Эта функция неотрицательна и супергармонична.

Нетрудно доказать, что если φ_t^s — аддитивный функционал от винеровского процесса, отвечающий мере μ, то потенциал дается формулой

$$f(x) = M_x \varphi_\infty^0. \tag{5}$$

Естественно поставить общую задачу о представлении неотрицательных супергармонических функций, связанных с произвольным марковским процессом, в виде (5). Для широкого класса процессов эта задача решена

(¹) Приведенное определение принадлежит докладчику [10], [17]. Как показано в [10] класс всех неотрицательных супергамонических функций совпадает с классом эксцессивных функций, введенных Хантом [28].

в работах Волконского [46], Вентцеля [49], Мейера [35], Шура [39], [40]. Доказано, что представление (5) возможно для всех ограниченных неотрицательных супергармонических функций. Получены критерии представимости для неограниченных функций.

5.3. Функция f называется *гармонической*, если функции f и $-f$ являются супергармоническими. Гармонические функции являются решениями уравнения $\mathfrak{A}f(x)=0$, где $\mathfrak{A}$ — характеристический оператор процесса. (В случае винеровского процесса они являются решениями уравнения Лапласа $\Delta f(x)=0$.)

Пусть (x_t, P_x) — необрывающийся непрерывный строго марковский процесс на n-мерном топологическом многообразии E. Условимся говорить, что функции $f_1, \ldots, f_n$ представляют собой *гармонические координаты* вблизи x, если эти функции являются гармоническими в некоторой окрестности U точки x и определяют топологическое отображение U на некоторую область евклидова пространства. Если кроме того всякая функция, гармоническая в U, является функцией класса C^k (1) от $f_1, \ldots, f_n$, то мы будем говорить, что функции $f_1, \ldots, f_n$ суть гармонические координаты класса C^k.

Нетрудно видеть, что если вблизи каждой точки x из E можно ввести гармонические координаты класса C^k, то набор всех таких координат определяет в E структуру гладкого многообразия класса C^k. Мы назовем ее *естественной гладкой структурой*.

Условимся говорить, что марковский процесс *регулярен в точке* x, если вблизи x можно ввести гармонические координаты. В одномерном случае это условие равносильно требованию, чтобы существовал интервал $(a,b)\supset x$ со следующим свойством: траектория, начинающаяся из любой точки $y \in (a,b)$ попадает с положительной вероятностью в каждую из точек a,b в момент первого выхода из (a,b). Интуитивно ясно, что в многомерном случае условие регулярности имеет аналогичный смысл. Было бы интересно найти соответствующую точную формулировку.

Очень интересно было бы выяснить, насколько широк класс процессов, допускающих естественную гладкую структуру класса C^k. Этим свойством во всяком случае обладают (при $k=\omega$) все регулярные одномерные процессы, а также все диффузионные процессы, для которых производящий дифференциальный оператор является эллиптическим оператором с аналитическими коэффициентами.

Понятие регулярности процесса и понятие естественной гладкой структуры были введены нами для необрывающихся процессов. Если процесс может обрываться, эти определения необходимо видоизменить. Для одномерного случая корректное определение естественной гладкой структуры дано в [17]. В многомерном случае такое определение еще должно быть найдено.

---

(1) Через C^k ($k=0, 1, 2, \ldots, \infty$) обозначается класс всех k раз непрерывно дифференцируемых функций. Через C^ω обозначается класс всех вещественных аналитических функций.

§ 6. Преобразования марковских процессов, связанные с аддитивными функционалами

6. 1. Знание аддитивных функционалов от марковского процесса позволяет строить по этому процессу ряд новых марковских процессов.

Мы рассмотрим несколько операций над марковскими процессами. Одна из них — построение подпроцесса (¹). Эта операция состоит в том, что траектория процесса x_t обрывается в некоторый момент $\tilde{\zeta}$. При заданной траектории момент $\tilde{\zeta}$ остается случайным, и для того, чтобы определить подпроцесс, надо задать соответствующее условное распределение вероятностей. Оказывается, что все эти условные распределения связаны с аддитивными функционалами от исходного процесса. Именно, если φ_t^s — аддитивный функционал, неотрицательный и непрерывный относительно t, то можно построить подпроцесс, для которого условная вероятность события $\tilde{\zeta}>t$ равна $e^{-\varphi_t^0}$.

Рассмотрим простейший случай, когда функционал φ_t^s задается формулой

$$\varphi_t^s=\int_s^t V(x_u)\,du \tag{6}$$

($V(x)$ — непрерывная неотрицательная функция.) Легко подсчитать, что в этом случае условная вероятность события $t\leqslant\tilde{\zeta}<t+dt$ равна $V(x_t)dt+o(dt)$ и, следовательно, траектория, находящаяся в некоторый момент t в точке x, обрывается за время dt с вероятностью $V(x)dt+o(dt)$ независимо от характера движения до момента t. Если оборвать таким образом диффузионный процесс с производящим дифференциальным оператором L, то получится диффузионный процесс с производящим оператором

$$\tilde{L}f=Lf-Vf\ (^2).$$

Функцию $V(x)$ естественно назвать *плотностью вероятности обрыва*. Для широкого класса процессов эта плотность может быть определена для каждого подпроцесса, только, вообще говоря, она является обобщенной функцией.

Поясним это утверждение для случая винеровского процесса. Для такого процесса каждый неотрицательный аддитивный функционал характеризуется некоторой мерой μ. Следовательно, каждый подпроцесс винеровского процесса также характеризуется некоторой мерой μ. Производная $d\mu/dx$ меры μ относительно меры Лебега является обобщенной функцией, и эта функция представляет собой естественное обобщение введенной выше плотности вероятности обрыва. Меру μ мы назовем *убивающей мерой*: если для некоторой области G $\mu(G)=0$, то траектория процесса с вероятностью 1 не обрывается в области G; на-

(¹) См. [11], [17].

(²) Этот результат легко следует из общей теоремы устанавливающей аналогичную связь инфинитезимальными операторами процесса и подпроцесса (см. [4]).

против, если значение $\mu(G)$ велико, то частица имеет мало шансов пройти область G, не погибнув.

6.2. Рассмотрим далее преобразование марковского процесса (x_t, P_x), при котором траектории не меняются, а меры преобразуются по формуле

$$\tilde{P}_x(d\omega) = \xi(\omega) P_x(d\omega).$$

Чтобы пара $(x_t, \tilde{P}_x)$ задавала марковский процесс, достаточно положить

$$\xi = e^{-\varphi_\infty^0},$$

где φ_t^s $(0 \leqslant s \leqslant t \leqslant \infty)$ — аддитивный функционал, удовлетворяющий условию

$$M_x e^{-\varphi_\infty^0} = 1.$$

Для винеровского процесса функционалы φ, удовлетворяющие последнему условию, можно строить по формуле

$$\varphi_t^s = \int_s^t b(x_u)\, dx_u + \frac{1}{2}\int_s^t b(x_u)^2\, du.$$

(первый член представляет собой стохастический интеграл, функция $b(x)$ принимает векторные значения и $b(x)^2$ означает ее скалярный квадрат). При преобразовании мер, отвечающем этому функционалу, из винеровского процесса получается диффузионный процесс с производящим дифференциальным оператором

$$\tilde{L}f = \Delta f - \sum b_i(x) \frac{\partial f}{\partial x_i}$$

($b_i(x)$ — декартовы координаты вектора $b(x)$).

6.3. В каждом из рассмотренных нами случаев построение нового марковского процесса происходит с помощью некоторого аддитивного функционала φ_t^s. Простая выкладка показывает, что переходная функция преобразованного процесса задается каждый раз одной и той же формулой

$$\tilde{P}(t, x, \Gamma) = \int_{x_t \in \Gamma} e^{-\varphi_t^0} P_x(d\omega).$$

В настоящее время известен обширный класс преобразований, обладающих тем же свойством (см. [14], [17]). При каждом из этих преобразований обрывается траектория процесса и в то же время определенным образом преобразуются меры P_x. К числу наиболее важных примеров таких преобразований относится преобразование, связанное с аддитивным функционалом

$$\varphi_t^s = -\log \frac{f(x_t)}{f(x_0)},$$

где f — положительная супергармоническая функция (1). При этом преобразовании переходная функция преобразуется по формуле

$$\tilde{P}(t, x, \Gamma) = \frac{1}{f(x)} \int_{\Gamma} P(t, x, dy) f(y).$$

6. 4. Весьма важное место занимает в теории марковских процессов *случайная замена времени*, впервые систематически изученная Волконским [45], [46] (см. также Ито и Маккин [32]). Эта операция состоит в том, что вдоль каждой траектории $x_t(\omega)$ вводится новое время $\tau_t(\omega)$, другими словами, вместо процесса (x_t, P_x) рассматривается пара (x_{τ_t}, P_x). Функция $\tau_t(\omega)$ должна быть, конечно, непрерывной и возрастающей функцией от t. Возникает вопрос: как выбрать среди таких функций все функции, для которых пара (x_{τ_t}, P_x) определяет марковский процесс. Оказывается, для этого достаточно рассмотреть всевозможные аддитивные функционалы $\varphi_t^s(\omega)$ от процесса (x_t, P_x), непрерывные относительно t и положительные при любых $t > s \geqslant 0$, и построить для каждого такого функционала функцию $\tau_t(\omega)$, обратную для функции $\varphi_t^0(\omega)$. Простейшим примером аддитивного функционала, удовлетворящего всем необходимым условиям, является функционал (6) при положительной функции $V(x)$. Значение τ_t находится из соотношения

$$\int_0^{\tau_t} V(x_u) du = t.$$

Ясно, что $d\tau_t/dt = V(x_{\tau_t})^{-1}$ и, следовательно, замена времени сводится к тому, что в каждой точке x движение ускоряется (или замедляется) с коэффициентом $V(x)$.

Вообще при случайной замене времени «карта траекторий» процесса не меняется. Поэтому не меняются вероятности, характеризующие положение частицы в момент первого выхода из любого открытого множества G (мы называем их для краткости *вероятностями выхода*). Естественно спросить, вытекает ли из совпадения вероятностей выхода для двух марковских процессов, что каждый из них может быть получен из другого посредством случайной замены времени. В случае, когда один из процессов — винеровский, положительный ответ на этот вопрос дается в работе Маккина и Танака [33]. В работе Блюменталя, Гетура и Маккина [1] этот результат распространяется на широкий класс процессов.

Пусть $\mathfrak{A}$ — характеристический оператор некоторого процесса, и $\tilde{\mathfrak{A}}$ — характеристический оператор процесса, полученного посредством случайной замены времени. В формуле, определяющей $\mathfrak{A}f(x)$ числитель выражается через f и вероятности выхода, а знаменатель не зависит от f. Поэтому для всех функций f, входящих одновременно в области определения $\mathfrak{A}$ и $\tilde{\mathfrak{A}}$, отношение $\mathfrak{A}f(x)/\tilde{\mathfrak{A}}f(x)$ принимает одно и то же значение. Обозначая его через $V(x)$, имеем формулу

$$\tilde{\mathfrak{A}} f(x) = V(x)^{-1} \mathfrak{A} f(x).$$

(1) Это преобразование строится при дополнительном предположении, что f представимо в виде (5).

Опираясь на эту формулу, нетрудно доказать, что диффузионный процесс с производящим дифференциальным оператором $\tilde{L}$ можно получить посредством случайной замены времени из диффузионного процесса с производящими дифференциальным оператором L тогда и только тогда, когда $\tilde{L}=V^{-1}L$, где $V(x)$ — некоторая положительная функция. При этом замена времени может быть осуществлена с помощью аддитивного функционала (6).

§ 7. Обобщенное броуновское движение

7. 1. Условимся называть *обобщенным броуновским движением* всякий марковский процесс, который можно получить из винеровского процесса посредством операций образования подпроцесса и случайной замены времени. Любое число таких операций можно заменить одной операцией образования подпроцесса и одной операцией случайной замены времени. Как уже говорилось, каждая из этих двух операций определяется аддитивным функционалом. Эти функционалы находятся во взаимно однозначном соответствии с определенными классами мер. Таким образом, всякое обобщенное броуновское движение характеризуется двумя мерами: первая — *«убивающая мера»* μ задает, каким образом винеровский процесс обрывается; вторая — *«мера скорости»* ν определяет случайную замену времени.

Нетрудно видеть, что вероятности выхода обобщенного броуновского движения мажорируются вероятностями выхода винеровского процесса. Как показал недавно Шур, для весьма широкого класса процессов это требование не только необходимо, но и достаточно для того, чтобы процесс являлся обобщенным броуновским движением.

7. 2. Рассмотрим броуновское движение с убивающей мерой μ и мерой скорости ν. Если существуют непрерывные производные $d\mu/d\nu$ и $d\nu/dx>0$, то этот процесс является диффузионным и его производящий дифференциальный оператор равен

$$Lf=\frac{1}{\frac{d\nu}{dx}}\Delta f-\frac{d\mu}{d\nu}f. \tag{7}$$

В общем случае равенству (7) можно придать смысл, если переписать его в виде

$$\frac{d\nu}{dx}Lf(x)=\Delta f(x)-f(x)\frac{d\mu}{dx}$$

и понимать $d\nu/dx$, Δf и $d\mu/dx$ как обобщенные функции ([1]). Доказывается,

---

([1]) Произведение $fd\mu/dx=g$ определяется как функционал на финитных бесконечно дифференцируемых функциях, заданный формулой

$$g(\varphi)=\int\varphi f d\mu.$$

что если Lf определено этим равенством и функции f и Lf непрерывны в естественной топологии, то f принадлежит области определения характеристического оператора $\mathfrak{A}$ и $\mathfrak{A}f = Lf$. Отправляясь от этого описания оператора $\mathfrak{A}$ можно описать и инфинитезимальный оператор A (1).

Оператор L можно определить также следующим образом. Пусть f — функция точки и ψ — мера со знаком. Условимся писать $\psi = \Psi f$, если $d\psi/dx = -\Delta f$ (левые и правые части понимаются как обобщенные функции). Далее, пусть μ — некоторая мера. Будем писать $\psi = \mu f$, если $d\psi/d\mu = f$. Тогда

$$L = \frac{d}{dv}[\Psi + \mu].$$

Для винеровского процесса

$$L = \frac{d}{dx}\Psi.$$

7. 3. Из аналитических задач, связанных с обобщенным броуновским движением, мы рассмотрим две задачи: 1) изучение гармонических функций; 2) изучение переходной функции.

Доказывается, что функция f является гармонической для броуновского движения с убивающей мерой μ, тогда и только тогда, когда она непрерывна в естественной топологии и удовлетворяет уравнению

$$\Delta f = \frac{d\mu}{dx} f \tag{8}$$

(правая и левая части понимаются как обобщенные функции). Из общей теории гармонических функций, связанных с марковскими процессами, вытекают простые условия существования и единственности решения задачи Дирихле для уравнения (8) и явные вероятностные формулы, выражающие это решение. Пусть, в частности, мера μ сосредоточена на некоторой гладкой поверхности S и имеет там непрерывную плотность ϱ Тогда уравнение (8) сводится к уравнению Лапласа вне S и требованию, чтобы функция f испытывала при переходе через S скачок нормальной производной, равный $f\varrho$ (2).

Относительно переходной функции обобщенного броуновского движения известно пока еще немного. В одномерном случае $P(t,x,\Gamma)$ как функция от Γ имеет относительно меры скорости v плотность $p(t,x,y)$, симметричную относительно x и y. (Для процессов с нулевой убивающей мерой это было впервые доказано Вентцелем [52].) Верно ли аналогичное утверждение для многомерного случая, остается неизвестным.

(1) В случае, когда убивающая мера равна нулю, инфинитезимальные операторы описаны Маккином и Танака [33]. Общий случай рассмотрен в [16], [17].

(2) Это доказано Молчановым.

§ 8. Как устроен общий непрерывный строго марковский процесс?

8. 1. Одной из наиболее актуальных задач теории марковских процессов является выяснение строения наиболее общего непрерывного строго марковского процесса на топологическом многообразии E. Одномерный случай в настоящее время исследован полностью. Имеются некоторые подходы и к многомерному случаю.

Начнем с процессов, заданных на открытом интервале (r_1, r_2). Доказывается, что если из каждой точки интервала траектория может попасть с положительной вероятностью в любую другую точку, то процесс можно получить посредством монотонного непрерывного преобразования фазового пространства из некоторого обобщенного броуновского движения. Другими словами, его можно получить из винеровского процесса, последовательно выполняя три операции: образование подпроцесса, случайную замену времени и преобразование координат. Если процесс необрывающийся, то из этих трех операций нужны только две последние.

Сходный результат можно получить и относительно необрывающихся непрерывных строго марковских процессов на топологических многообразиях любого числа измерений. Однако, в общем случае исходным материалом для построения служит не только винеровский процесс, но и целый класс процессов, близкий к классу диффузионных процессов (мы называем эти процессы квазидиффузионными). К тому же сведение к квазидиффузионным процессам удается провести только локально и при некоторых ограничениях, которые нуждаются в дальнейшем анализе.

8. 2. Напомним, что непрерывный строго марковский процесс называется диффузионным, если его характеристический оператор определен на всех дважды непрерывно дифференцируемых функциях. Чтобы получить определение *квазидиффузионного* процесса, достаточно заменить здесь характеристический оператор на *квазихарактеристический* оператор, который определяется следующим образом. Пусть f и F — функции, заданные в некоторой окрестности точки x и пусть для каждой достаточно малой окрестности U этой точки

$$T_{\tau(U)} f(x) - f(x) = M_x \int_0^{\tau(U)} F(x_s) ds.$$

($\tau(U)$ — момент первого выхода из U.) Тогда мы говорим, что f принадлежит области определения квазихарактеристического оператора $\hat{\mathfrak{A}}$ в точке x и $\hat{\mathfrak{A}} f(x) = F(x)$. Если функция F непрерывна в точке x, то $\mathfrak{A} f(x)$ также равно $F(x)$, так что квазихарактеристический оператор совпадает с характеристическим (¹).

(¹) Аналогичным образом можно видоизменить определение инфинитезимального оператора. Условимся говорить, что функция f принадлежит области определения квазиинфинитезимального оператора $\hat{A}$ и $\hat{A} f = F$, если

$$T_t f - f = \int_0^t T_s F ds, \quad (t \geqslant 0).$$

8. 3. Рассмотрим теперь произвольный необрывающийся непрерывный строго марковский процесс (x_t, P_x) на многообразии E. Построение, сводящее такой процесс к квазидиффузионным процессам, основано на понятии гармонической функции и аддитивного функционала. Основная идея построения принадлежит Скороходу [43]. Скороход показал, что каждой паре гармонических функций f, g можно сопоставить непрерывный аддитивный функционал $\varphi_t^s(f,g)$, который определяется следующим образом ([1]). Разобьем отрезок $[s,t]$ на мелкие отрезки $[u_k, u_{k+1}]$, составим произведение приращений функций $f(x_u)$ и $g(x_u)$ на каждом из мелких отрезков. Сумма этих произведений при $\max|u_{k+1}-u_k|\to 0$ сходится в среднем квадратическом к $\varphi_t^s(f,g)$.

Пусть $f_1, \ldots, f_n$ — гармонические координаты вблизи точки x. Положим

$$\varphi_t^s = \sum_i \varphi_t^s(f_i, f_i).$$

Опираясь на теорему Радона-Никодима, можно показать, что

$$\varphi_t^s(f_i, f_j) = \int_s^t a_{ij}(x_u)\, d\varphi_u^0.$$

Сделаем случайную замену времени, отвечающую функционалу φ_t^s ([2]). Мы получим квазидиффузионный процесс с производящим дифференциальным оператором

$$\sum a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}.$$

Если коэффициенты $a_{ij}(x)$ непрерывны, то этот процесс является диффузионным.

Условие существования гармонической системы координат является, как мы уже говорили, некоторым условием регулярности процесса. Таким образом исследование локального строения любых регулярных необрывающихся непрерывных строго марковских процессов сводится к изучению квазидиффузионных процессов.

Доказывается, что квазихарактеристический оператор является расширением оператора $\hat{A}$.

([1]) Строго говоря, результат Скорохода относится к гармоническимфункциям, подчиненным дополнительному условию: при $t\downarrow 0$

$$\sup_x P_x\{|f(x_t)-f(x)|>\varepsilon\} = o(t)$$

для любого $\varepsilon>0$. Эту оговорку надо иметь в виду и в дальнейшем.

([2]) Функционалы $\varphi_t^s(f_i, f_j)$ и φ_t^s определены до момента τ первого выхода из окрестности, в которой определены гармонические координаты. Случайная замена времени проводится после того, как исходный процесс оборван в момент τ.

§ 9. Неотрицательные гармонические функции и предельное поведение траекторий марковского процесса.

9. 1. В 1941 г. Мартин [34] исследовал множество всех неотрицательных решений уравнения Лапласа в произвольной области G евклидова пространства. Значение конструкций Мартина для теории вероятностей было выяснено в работах Дуба [3] и Ханта [29]. Дуб и Хант рассматривали гармонические функции, связанные с дискретными цепями Маркова. Случай диффузионных процессов был изучен недавно Шуром [41].

Мы начнем со следующей конкретной задачи. Пусть эллипсоид с закрепленным центром и фиксированным объемом случайно изменяется (поворачивается и растягивается). Предположим, что этот процесс является марковским, непрерывным и инвариантным относительно группы аффинных преобразований, не меняющих объема. Что будет происходить с эллипсоидом при $t\to\infty$?

Доказывается (см. [15]), что с вероятностью 1 главные направления эллипсоида стремятся к определенным предельным положениям и эллипсоид сплющивается; точнее, если $\varrho_1\geqslant\varrho_2\geqslant\varrho_3$ — длины его полуосей, то с вероятностью 1

$$\log\varrho_1\to\infty,\qquad \frac{\log\varrho_2}{\log\varrho_1}\to 0,\qquad \frac{\log\varrho_3}{\log\varrho_1}\to -1.$$

Множество всех эллипсоидов с закрепленным центром и фиксированным объемом представляет собой некоторое гладкое многообразие, и случайное движение, которое мы рассмотрели, является диффузионным процессом на этом многообразии. Пусть вообще на некотором гладком многообразии E задан диффузионный процесс с производящим дифференциальным оператором L. Оказывается, что вопрос о предельном поведении траектории при $t\to\infty$ тесно связан с вопросом о строении множества всех отвечающих процессу неотрицательных гармонических функций, т. е. решений уравнения $Lf=0$. Обозначим это множество через K и назовем функцию $f\in K$ минимальной, если ее нельзя разложить на сумму двух не пропорциональных друг другу функций из K.

В случае, когда L — оператор Лапласа в единичном шаре, то минимальные функции описываются формулой

$$f_\varrho(x)=\frac{1-|x|^2}{|x-\varrho|^3},$$

где $|\varrho|=1$ (через $|x|$ обозначается евклидова длина вектора x). Таким образом, минимальные функции находятся во взаимно однозначном соответствии с точками границы исходного шара. Оказывается, что и в общем случае можно погрузить фазовое пространство E в некоторый компакт таким образом, что минимальные функции оказываются во взаимно однозначном соответствии с некоторой частью границы E

(эта граница называется границей Мартина). Доказывается, что всякая неотрицательная гармоническая функция f записывается в виде интеграла

$$\int f_\varrho(x)\,\mu(d\varrho),$$

причем мера μ однозначно определяется по f. В частности, в предположении, что $L1=0$ (¹), имеем

$$1=\int f_\varrho(x)\,\mu_0(d\varrho).$$

Доказывается, что траектория процесса входит с вероятностью 1 в определенную точку границы Мартина и вероятность того, что эта точка принадлежит множеству Γ, дается формулой

$$\int_\Gamma f_\varrho(x)\,\mu_0(d\varrho).$$

Таким образом, $f_\varrho(x)$ представляет собой плотность вероятности вхождения в точку ϱ для частицы, отправляющейся из точки x.

Сформулированный выше результат о предельном поведении диффундирующего эллипсоида получен именно таким путем. Выбирая какой-нибудь ортонормированный базис в евклидовом пространстве, можно описывать эллипсоиды положительно определенными матрицами. Оказывается, что минимальные функции представляют собой произведения степеней угловых миноров этой матрицы. Если мы хотим получить плотность вероятности того, что в пределе главные направления эллипсоида определяются ортонормированным базисом e_1, e_2, e_3, мы должны построить, отправляясь от базиса e_1, e_2, e_3, матрицу, соответствующую начальному положению эллипсоида, и взять произведение угловых миноров этой матрицы в степени -2. (Все сказанное распространяется на эллипсоиды в комплексном евклидовом пространстве любого числа измерений (см. [15]). Аналогичные результаты можно получить и для других симметрических пространств отрицательной кривизны.

Множество минимальных неотрицательных решений эллиптического дифференциального уравнения $Lf=0$ еще мало исследовано и здесь имеется много интересных нерешенных задач. Мы сформулируем две такие задачи.

Первая задача: при каких условиях на дифференциальный оператор L можно ввести в множество минимальных функций структуру гладкого многообразия так, чтобы $f_\varrho(x)$ являлась гладкой функцией от ϱ и x?

Вторая задача касается связей между локальной геометрией полного риманова многообразия и строением границы Мартина этого многообразия. (В качестве оператора L рассматривается дифференциальный оператор Бельтрами, определяемый римановой метрикой.) Во всех примерах полных односвязных многообразий отрицательной кривизны,

---

(¹) Это предположение равносильно требованию, чтобы траектория процесса выходила, не обрываясь, из любого компакта.

для которых подсчитана размерность множества минимальных функций, эта размерность отличается на единицу от размерности многообразия. Верно ли, что и в общем случае отрицательность кривизны и односвязность влекут за собой определенное богатство множества минимальных функций?

ЛИТЕРАТУРА

[1]. Blumenthal, R. M., Getoor, R. K., & McKean, H. P., Jr., Markov processes with identical hitting distributions. *Illinois J. Math.*, 8 (1962), 402–421.

[2]. Doob, J. L., Semimartingales and subharmonic functions. *Trans. Amer. Math. Soc.*, 77 (1954), 86–121.

[3]. —— Discrete potential theory and boundaries. *J. Math. Mech.*, 8 (1959), 433–458.

[4]. Дынкин, Е. Б., Функционалы от траекторий марковских случайных процессов. *Доклады АН СССР*, 104 (1955), 691–694.

[5]. —— Бесконечно малые операторы марковских случайных процессов. *Доклады АН СССР*, 105 (1955), 206–209.

[6]. —— Непрерывные одномерные марковские процессы. *Доклады АН СССР*, 105 (1955), 405–408.

[7]. —— Марковские процессы и полугруппы операторов. *Теор. вер. и ее прим.*, 1 (1956), 25–37.

[8]. —— Инфинитезимальные операторы марковских процессов. *Теор. вер. и ее прим.*, 1 (1956), 38–60.

[9]. —— Одномерные непрерывные строго марковские процессы. *Теор. вер. и ее прим.*, 4 (1959), 3–54.

[10]. —— Естественная топология и эксцессивные функции, связанные с марковским процессом. *Доклады АН СССР*, 127 (1959), 17–19.

[11]. —— *Основания теории марковских процессов*. Москва, 1959.

[12]. —— Марковские процессы и связанные с ними задачи анализа. *Успехи матем. наук*, 15: 2 (92) (1960), 3–24.

[13]. —— Аддитивные функционалы от винеровского процесса, определяемые стохастическими интегралами. *Теор. вер. и ее прим,*. 5 (1960), 441–452.

[14]. Dynkin, E. B., Transformations of Markov processes connected with additive functionals. *Proc. Fourth Berkeley Symp. on Math. Stat. and Prob.*, 2, 117–142. Berkeley, 1961.

[15]. —— Неотрицательные собственные функции оператора Лапласа-Бельтрами и броуновское движение в некоторых симметрических пространствах. *Доклады АН СССР*, 141 (1961), 288–291.

[16]. —— Броуновское движение с убивающей мерой μ и мерой скорости ν. *Доклады АН СССР*, 144 (1962), 483–486.

[17]. —— *Марковские процессы*. Москва, 1963.

[18]. Дынкин, Е. Б., и Юшкевич, А. А., Строго марковские процессы. *Теор. вер. и ее прим .*, 1 (1956), 149–155.

[19]. Feller, W., The parabolic differential equations and the semi-groups of transformations. *Ann. Math.*, 55 (1952), 468–519.

[20]. —— Generalized second order differential operators and their lateral conditions. *Illinois J. Math.*, 1 (1957), 459–504.

[21]. Фрейдлин, М. И., Диффузионные процессы с отражением и задача с

косой производной на многообразии с краем. *Теор. вер. и ее прим.*, 8 (1963), 80–88.

[22]. —— Задача Дирихле для уравнения с малым параметром и разрывными коэффициентами. *Доклады АН СССР*, 144 (1962), 501–504.

[23]. Гирсанов, И. В., О некоторых топологиях, связанных с марковским процессом. *Доклады АН СССР*, 129 (1959), 488–491.

[24]. Хасьминский, Р. З., Диффузионные процессы и эллиптические дифференциальные уравнения, вырождающиеся на границе области. *Теор. вер. и ее прим.*, 3 (1958), 430–451.

[25]. —— О положительных решениях уравнения $\mathfrak{A}u + Vu = 0$. *Теор. вер. и ее прим.* 4 (1959), 332–341.

[26]. —— Эргодические свойства возвратных диффузионных процессов и стабилизация решений задачи Коши для параболических уравнений. *Теор. вер. и ее прим.*, 5 (1960), 196–214.

[27]. —— О принципе усреднения для параболических и эллиптических дифференциальных уравнений и марковских процессов с малой диффузией. *Теор. вер. и ее прим.*, 8 (1963), 3–25.

[28]. Hunt, G. A., Markoff processes and potentials. *Illinois J. Math.*, 1 (1957), 44–93; 1 (1957), 316–369; 2 (1958), 151–213.

[29]. —— Markoff chains and Martin boundaries. *Illinois J. Math.*, 4 (1960), 313–340.

[30]. Itô, K., Stochastic differential equations in a differentiable manifold. I. *Nagoya Math. J.*, 1 (1950), 35–47; II. *Mem. Coll. Sci., Univ. Kyoto, Ser. A, 28, Mathematics*, 1 (1953), 81–85.

[31]. —— On stochastic differential equations. *Mem. Amer. Math. Soc.*, 4 (1951).

[32]. Itô, K., & McKean, H. P., Jr. *Diffusion theory.* (To appear.)

[33]. McKean, H. P., Jr., & Tanaka, H., Additive functionals of the Brownian path. *Mem. Coll. Sci., Univ. Kyoto, Ser. A, Mathematics*, 33 (1961), 479–506.

[34]. Martin, R. S., Minimal positive harmonic functions. *Trans. Amer. Math. Soc.*, 49 (1941), 137–172.

[35]. Meyer, P. A., Fonctionelles multiplicatives et additives de Markov. *Ann. Inst. Fourier, Grenoble*, 12 (1962), 125–230.

[36]. Pogorzelski, W., Étude de la solution fondamentale de l'équation parabolique. *Ricerche di Mat., Napoli*, 5 (1956), 25–37.

[37]. Погожельский, В., Исследование интегралов параболического уравнения и краевых задач в неограниченной области. *Матем. сборн.* 47 (1959), 397–430.

[38]. Ray, D., Stationary Markov processes with continuous paths. *Trans. Amer. Math. Soc.*, 82 (1956), 452–493.

[39]. Шур, М. Г., Непрерывные аддитивные функционалы от марковских процессов и эксцессивные функции. *Доклады АН СССР*, 137 (1961), 800–803.

[40]. —— Эксцессивные функции и аддитивные функционалы от марковских процессов. *Доклады АН СССР*, 143 (1962), 293–296.

[41]. —— Граница Мартина для линейного эллиптического оператора второго порядка. *Известия АН СССР, сер. матем.*, 27 (1963), 45–60.

[42]. Скороход, А. В., Аддитивные функционалы от процесса брауновского движения. *Теор. вер. и ее прим.*, 6 (1961), 430–439.

[43]. —— Об однородных непрерывных марковских процессах, являющихся мартингалами. *Теор. вер. и ее прим.*, 8 (1963).

[44]. Ueno, T., The diffusion satisfying Wentzell's boundary conditions and the Markov processes on the boundary. *Proc. Japan Acad.*, 36 (1960), 533–538, 625–629.

[45]. Волконский, В. А., Случайная замена времени в строго марковских процессах. *Теор. вер. и ее прим.*, 3 (1958), 332–350.

[46]. —— Аддитивные функционалы от марковских процессов. *Труды Москов. матем. общества*, 9 (1960), 143–189.

[47]. Вентцель, А. Д., О граничных условиях для многомерных диффузионных процессов. *Теор. вер. и ее прим.*, 4 (1959), 172–185.

[48]. —— Общие граничные задачи, связанные с диффузионными процессами. *Успехи матем. наук*, 15 (1960), 202–204.

[49]. —— Неотрицательные аддитивные функционалы от марковских процессов. *Доклады АН СССР*, 137 (1961), 17–20.

[50]. —— Аддитивные функционалы от многомерного винеровского процесса. *Доклады АН СССР*, 139 (1961), 13–16.

[51]. —— О непрерывных аддитивных функционалах от многомерного винеровского процесса. *Доклады АН СССР*, 142 (1962), 1223–1226.

[52]. —— Об абсолютной непрерывности переходных вероятностей одномерного диффузионного процесса. *Теор. вер. и ее прим.*, 6 (1961), 439–446.