such an approach suffers from two disadvantages.

First, in spite of the fact that under some assumptions a one-to-one correspondence exists between the strategies and the set of action rules given for each initial point ξ (ignoring the requirement of measurability in the initial point which is required in the strict definition of strategy) to use *only* the definition of strategies is not convenient because the formulation of strategy may be considerably more complicated than that of the action rule corresponding to it and conversely.

Consider, for example, the following simple and *stationary*, i.e. independent of time, strategy in the two-armed bandit problem: fix some number ξ_0, $0 < \xi_0 < 1$ and at time n use the control a^1 if the *a posteriori probability* $\xi(n) := \xi_1(n)$ of hypothesis H_1 is less than or equal to ξ_0 and use a^2 if $\xi(n) > \xi_0$. The corresponding action rules (for each initial point) as a function of the results of observations are much more complicated and, moreover, this function depends on the initial point as argument. Now, in the same problem, consider the following action rule: the control a^1 is used initially and afterwards the control is changed after each appearance of a 0. In this case, the strategy corresponding to the action rule described (which is the same for all initial points ξ) is the more complicated.

Secondly, the use of the definition of action rule *together* with the usual definition of strategy has the advantage that the resulting problem may be presented in a special form which we will describe in the next section.

1.5 Finite Bayesian problem

In light of the equation $P_\xi^\beta = \sum_1^N \xi_i P_i^\beta$, the given problem may be presented as a particular case of an abstract optimization problem as follows:

$$\inf_{d\in\mathcal{A}} \sum_{i=1}^{N} \xi_i I_i^d \,, \tag{1.3}$$

where $\xi = (\xi_1, \ldots, \xi_N) \in S^N$, $d \in \mathcal{A}$ is some admissible control set and I_i^d is a functional defined on $\mathcal{A}$. We will call this problem simply the *finite Bayesian problem.* Designate the value of the infimum (1.3) as $\Phi(\xi)$, and the control d at which the infimum is reached, if such exists, as $d^*(\xi)$. Then it is well known (see, for example, De Groot 1970) that the function $\Phi(\xi)$ is convex and continuous inside the simplex S^N as an infimum of functions linear with respect to ξ.

The representation of the problem of the basic scheme in the form (1.3), where the rôle of $\mathcal{A}$ is played by the whole set of action rules, allows us, for example, also to obtain the continuous differentiability of the value function from the continuity of the functions $I_i^{d^*(\xi)}$ as functions of ξ (see §7.3).

Next we will describe an analogue of the basic scheme presented above in continuous time and the methods used for it, delaying a discussion of the results obtained in discrete time until later.

1.6 Transition to the continuous time case

It is known that in many situations the transition from a problem in discrete time to a problem in continuous time yields deeper and more powerful results. The basic idea of such a transition in our case consists in the replacement of Bernoulli discrete time sequences with a fixed probability of success λ observed using a fixed control with a continuous time *Poisson process* having as *intensity* the same value λ. The intensity λ defines the probability of a realized 1 (success) as $\lambda\, dt$ on a small interval $(t, t + dt)$. The realization of a 1 does not depend on the behaviour of the process up to time t (independence of process events on nonoverlapping time intervals). A Poisson process is a (unit) jump process in continuous time possessing the *stationarity property* and with *independent increments.* The expected number of realizations of a 1 in a fixed elapsed time ν is the same for the Bernoulli case in discrete time and the Poisson case in continuous time, namely $\lambda\nu$.

At first glance, it seems natural to construct a sequence of problems in discrete time depending on a time unit n^{-1} and taking limits as $n \to \infty$ to yield the problem in continuous time as follows. Divide each unit interval of observation in discrete time into n intervals $[i/n, (i+1)/n)$ and, allowing on each subinterval discrete time choice of controls, replace the probability of realizing a 1 with λ_i^j over a unit interval by the appropriate λ_i^j/n on each subinterval, so that the average number of successes observed in unit time for a *fixed* control does not change. However, the following complication arises.

We return to the symmetric 2×2 case and trace which sequences of controls we must apply to obtain an optimal strategy with n sufficiently large and initial point $\xi := (1/2) + \varepsilon$, where ε is some small positive number. According to the optimal strategy described in the introduction, at the initial time it is necessary to apply the control a^2. Since λ^1/n is close to 0, then with probability close to 1 we will observe a 0 on the interval $[0, \lambda^1/n)$ and updating the *a posteriori* probability by formula (1.1) we will get a value $\xi < 1/2$, so that at the next step we must use the control a^1. Again observing a 0, we will on updating obtain the initial point ξ and again use a^2, and so on. If the number of observations is of order n and the probability of realizing a 1 is λ^j/n, then with positive probability bounded away from 0 we will observe a sequence consisting of only 0s, and therefore will use the sequence of controls $a^1, a^2, a^1, a^2, \ldots$. It is not clear to what this sequence of alternating controls will converge as $n \to \infty$.

Randomization of the discrete time controls is not a possible solution to this problem. With randomized controls, the probability distributions on the set of basic controls serve as new controls. In our case we are given a finite number of controls $a^1, \ldots, a^m$ and therefore random controls are given by points $\alpha := (\alpha^1, \ldots, \alpha^m)$ from S^m. In control problems with a criterion of expected profit, randomization, as is well known (see, for example, De Groot 1970), does not increase the profit. This, of course, does not contradict the fact that to prove many results random controls are widely used because they allow the possibility of making the problem in some sense tractable. In the above example, the random controls $\alpha^1 := 1/2$, $\alpha^2 := 1/2$, $\alpha := (\alpha^1, \alpha^2) \in S^2$, might be tried; however, then the controls a^1 and a^2 will again be interchanged extremely irregularly.

However, randomization is not the only possible way to increase the class of admissible controls, where the controls $a^1, \ldots, a^m$ are *basic* and correspond to the set $\hat{S}^m$ of vertices of the simplex S^m, while controls from the increased class correspond to arbitrary points of the simplex. We consider another closely related method of increasing the control set for which the continuous time problem may be obtained as a "natural limit" of the prelimit problems in discrete time.

As a matter of definition consider that the choice of a vector $\alpha \in S^m$ (now this vector α is called the *control*) leads to the observation of an m-dimensional random vector with independent coordinates taking values 0 or 1 with the probability of a 1 in the j^{th} place equal to $\alpha^j \lambda_i^j$ under hypothesis H_i. (Recall that the rows of the hypothesis matrix $\{\lambda_i^j\}$ correspond to the hypotheses $H_1, \ldots, H_N$). All other conditions hold as for the basic scheme. We call this the model with *sharable resources*, interpreting α^j as a fraction of a *single* resource allocated to the j^{th} device.

In the case of the basic scheme in discrete time the situation using randomization is the same as without it; at each moment only one device is operative, so that this formulation might be termed that with *nonsharable resources.*

In the model with sharable resources, the choice $\alpha \in \hat{S}^m$, $\alpha := e_j^m$, as in the basic scheme, leads to the observation of a Bernoulli random value on the j^{th} device, while for the rest of the devices observations are absent. However, the choice of an arbitrary control in the model with sharable resources and in the basic scheme (with randomization) the observations will be different, but the expected number of realizations of 1 under the i^{th} hypothesis will be the same, i.e. $\sum_j \alpha^j \lambda_i^j$.

Since in both formulations the observations may be considered to coincide for the application of the pure controls $a^1, \ldots, a^m$, and in the basic scheme treated as a maximization problem randomization does not increase the profit function, then the optimal profit in the formulation with sharable resources is not less than that of the basic scheme.

In §3.6 a precise definition of the formulation with resource sharing is given and it is proved (Theorem 3.4) that optimal profits in both schemes coincide. We emphasize that the necessity of considering the resource sharing formulation is related to the fact that control problems in continuous time (we will preserve the term "basic scheme" for these) may be obtained by taking discrete time limits in the resource sharing model rather than in the basic scheme.

The question of connection of problems in continuous and discrete time is also discussed at the beginning of Chapter 4, but its systematic consideration is beyond the scope of this book.

1.7 Basic scheme in continuous time

A heuristic formulation of the problem in continuous time follows. Similarly to the discrete time case, the parameters of the problem are the *hypothesis matrix* $\Lambda := \{\lambda_i^j\}$, $i = 1, \dots, N$, $j = 1, \dots, m$ and the vector of *a posteriori probabilities* of the hypotheses $\xi := \{\xi_1, \dots, \xi_N\}$. The value λ_i^j corresponds to the *intensity* of realized 1s on the j^{th} device when the i^{th} hypothesis holds and the j^{th} device is used exclusively. If in a small time interval $[t, t + dt)$ the fraction of the resource used for control allocated to the j^{th} device is equal to α^j, then the corresponding intensity of realized 1s in that interval equals $\alpha^j \lambda_i^j$. On each small interval the control consists in a choice of values α^j, $\alpha^j \geq 0$, $\Sigma \alpha^j = 1$ and, similarly to the discrete time case, this choice depends *a priori* on all previous observations and controls. Since current controls are functions of past observations, *action rules* may be considered as functions β taking values in S^m and depending at time t on the history of the m-dimensional random observation process up to time t.

Such functions may be described as follows. At the initial moment a choice is made of a deterministic function $\alpha_0(s) = (\alpha_0^1(s), \dots, \alpha_0^m(s))$, measurable with respect to s, $0 \leq s \leq \nu$, which takes values in S^m and corresponds to control up to the moment of the first jump of the process (i.e. a realization of a 1). If at the random time τ_1 a jump occurs with respect to any coordinate, then depending on the value of τ_1 and the index j_1 of this coordinate, a new function $\alpha_1(s|\tau_1, j_1) = (\alpha_1^1(s), \dots, \alpha_1^m(s))$ is chosen, and so on. We will mainly consider the problem of maximization of the expected number of jumps in a fixed time and, similarly to discrete time, the corresponding value function will be designated by $V_\nu(\xi)$.

As in discrete time, the optimal control may be implemented by observing only the vector $\xi(s)$ representing the *a posteriori* distribution of the hypothesis at each moment. The rigorous derivation of the corresponding stochastic equations given in §4.2 is based on sufficiently delicate results of the theory of stochastic processes that we give here only a nonrigorous but transparent derivation of these equations.

Designate by $z_i(s|\alpha) := z_i(s)$ the probability that no jump occurs up to time s under the condition that the i^{th} hypothesis holds and the control $\alpha(s) := (\alpha^1(s), \dots, \alpha^m(s))$ is used. Since the probability

of the first jump with respect to the j^{th} coordinate of the observation process occurring on the small time interval $[s, s + ds)$ does not depend on the other coordinates and equals $\alpha^j(s)\lambda_i^j\, ds$, then $z_i(s)$ satisfies the formula

$$z_i(s) = \exp\{-\sum_{j=1}^{m} \lambda_i^j \int_0^s \alpha^j(u)\, du\}. \tag{1.4}$$

Let $z(s|\alpha) := z(s)$ be the *unconditional* probability that up to moment s no jump occurs. Obviously,

$$z(s) = \sum_{i=1}^{N} \xi_i z_i(s). \tag{1.5}$$

Denote by $p^j(\xi)\, ds$ the conditional probability of a jump with respect to the j^{th} coordinate in the small time interval $[s, s + ds)$, if $\xi(s) := \xi$ and $\alpha^j(s) := 1$. Obviously,

$$p^j(\xi) = \sum_{i=1}^{N} \xi_i \lambda_i^j. \tag{1.6}$$

Let $p_i(\alpha)\, ds$ be the conditional probability of a jump in the small time interval $[s, s + ds)$ given that the i^{th} hypothesis holds and the control $\alpha(s) := \alpha$ is used. Obviously,

$$p_i(\alpha) = \sum_{j=1}^{m} \alpha^j \lambda_i^j. \tag{1.7}$$

Denote by $p(\xi, \alpha)\, ds$ the unconditional probability of a jump with respect to at least one coordinate of the vector process in the small time interval $[s, s + ds)$, if $\xi(s) := \xi$ and the control $\alpha(s) := \alpha$ is used. Clearly,

$$p(\xi, \alpha) = \sum_{i=1}^{N} \xi_i p_i(\alpha) = \sum_{j=1}^{m} \alpha^j p^j(\xi) = \sum_{i=1}^{N} \sum_{j=1}^{m} \xi_i \alpha^j \lambda_i^j. \tag{1.8}$$

By the Bayesian formula for the *a posteriori* distribution of hypotheses $\xi(s) = (\xi_1(s), \dots, \xi_N(s))$ under the condition that no jump has occurred up to time s,

$$\xi_i(s) = \xi_i z_i(s)/z(s), \qquad i = 1, \dots, N. \tag{1.9}$$

From (1.5) and (1.4) using (1.9) we have

$$\dot{z}(s) = -z(s)p(\xi(s), \alpha).$$

Differentiating (1.9) and (1.4) and using the last relation we have that the differential equations describing the evolution of $\xi(s)$ on the time interval up to the first jump of the process are given by

$$\dot{\xi}_i(s) = \xi_i(s)[p(\xi(s), \alpha) - p_i(\alpha)] := f_i(\xi(s), \alpha), \quad i = 1, \ldots, N. \quad (1.10)$$

This equation can be rewritten in the following vector form

$$\dot{\xi} = \alpha\Lambda^*(\xi^*\xi - \text{diag } \xi),$$

where * denotes transpose and $\text{diag}(\xi)$ is the diagonal matrix formed from the entries of ξ.

Obviously, the same equation holds on the time interval between any two successive jumps.

Let $\Gamma^j\xi$ denote the value of the vector of *a posteriori* probabilities of hypotheses given that a jump occurred with respect to the j^{th} coordinate and previous to the jump this value was given by ξ. Again applying Bayes' formula to the probability of a jump on a small time interval it is not difficult to see that

$$(\Gamma^j\xi)_i = \xi_i\lambda_i^j / \sum_{k=1}^{N} \xi_k\lambda_k^j, \quad i = 1, \ldots, N. \quad (1.11)$$

The original formulation of the basic scheme in discrete time was reduced to a control problem for a Markov chain whose states are the *a posteriori* probabilities of hypotheses. Similarly, it may be shown that the basic scheme in continuous time can be reduced as follows. At the initial moment choose a control which is a deterministic vector-valued function $\alpha_1(s)$ taking values in S^m. The system is described by a point (s, ξ) in the $(N+1)$-dimensional state space "time—*a posteriori probabilities*," $0 \le s \le \nu$, $\xi \in S^m$, whose motion from the initial point $(0, \xi)$ is determined by the differential equation (1.10). At some random moment τ_1, the trajectory is transformed by a jump from the state $\xi(\tau_1)$ to the state $\Gamma^j(\xi(\tau_1))$ if a jump of type j occurs $(j = 1, \ldots, m)$. In the original formulation this corresponds to the moment of realization of a 1 with respect to the j^{th} coordinate of the observation process. The probability density that the transition occurs to the state $\Gamma^j\xi(s)$ on the small time interval $[s, s + ds)$ equals $\alpha_1^j(s)p^j(\xi(s))$. After the first jump the control $\alpha_2(s)$ is chosen, which in general depends upon when and in what state the first jump occurred, and so on. The sequence of such successive controls $\{\alpha_1(s), \alpha_2(s), \ldots\}$ is called a *strategy*. It is required to find the strategy maximizing the *expected number of jumps* over the horizon ν. The optimal value of this problem will also be denoted by $V_\nu(\xi)$.

We mention that, just as in discrete time (see §1.4), it is sometimes more convenient to make a change of variables from ξ to η (see formula (1.2)) where, for example, defining $\eta_i := \ln(\xi_i/\xi_N)$, the differential equation (1.10) takes the simple form

$$\dot{\eta}_i = -\sum_{j=1}^{m} (\lambda_i^j - \lambda_N^j)\alpha^j, \quad (1.12)$$

and formula (1.11) takes the form

$$\tilde{\Gamma}^j\eta = \eta + \gamma^j, \quad \gamma^j := (\gamma_1^j, \ldots, \gamma_N^j), \quad \gamma_i^j := \ln(\lambda_i^j/\lambda_N^j), \quad (1.13)$$

i.e. the value of the jump in the new process depends exclusively on the jump coordinate j of the old.

Two different approaches exist to solve the problems described above for the basic continuous time scheme. The first, traditional, approach is connected with the application of the (Bellman) *optimality equation*, i.e. with the equation satisfied by the value function given by the supremum of the criterion functional (in our case the expected number of jumps) maximized with respect to all admissible strategies. The second approach is related to the (*Pontryagin*) *maximum principle* (see §1.9 and Chapter 6).

With both approaches it is sometimes convenient to *embed* the initial formulation, that is, to consider the problem not as a single problem but as a set of problems with a *fixed* time of *final* observation (horizon) ν, but with *initial* time t running through the different values, $t \le \nu$. (Formally, it is possible to stay in the framework of a single process by considering the conditional values of the criterion functional at given intermediate points, see Gikhman & Skorohod (1977).)

To solve problems of the basic scheme in continuous time the optimality equation may be rewritten in two forms: namely, in terms of the values of the value function at *successive jumps* of the observation process or on a *small time interval.* In the second case, the optimality equation is called *local.*

1.8 Local optimality equation

We give a heuristic derivation of the local optimality equation for the problem of maximization of the expected number of jumps (successes). Let the final observation time ν be fixed ($\nu \leq \infty$). By $V(t,\xi)$ denote the value of the value function for problems on the time interval $[t,\nu)$ with *a priori* distribution ξ, so that $V(t,\xi) := V_{\nu-t}(\xi)$. We shall consider a small interval $[t,t+dt)$ and write the relation connecting $V(t,\xi)$ and $V(t+dt,\xi+d\xi)$ and the control on this interval. Without loss of generality we may take the control $\alpha(s)$ to be constant on the interval, since, similarly to the situation for deterministic optimal control problems in continuous time (see Boltyanski 1969), it is sufficient to replace the measurable functions $\alpha(s)$ by piecewise continuous functions. Let $V^\alpha(t,\xi)$ designate the expected value of the number of jumps for the strategy defined by using the control α on the interval $[t,t+dt)$ and using the strategy optimal for the initial point $(t+dt,\xi(t+dt))$ on the remaining interval $[t+dt,\nu)$. With probability $p^j(\xi)\alpha^j dt + o(dt)$ the system will be in the state $(t+dt,\Gamma^j\xi + o(dt))$, $j = 1,\dots,m$, for $t+dt$ and with probability $1 - \sum_j \alpha^j p^j(\xi)dt + o(dt)$ in the state $(t+dt,\xi+d^\alpha\xi)$, where $d^\alpha\xi$ is defined in terms of evolution according to (1.10). Therefore, we have that up to terms of $o(dt)$

$$V^\alpha(t,\xi) = \sum_{j=1}^m \alpha^j p^j(\xi)dt[1 + V(t+dt,\Gamma^j\xi + o(dt))] + (1 - \sum_{j=1}^m \alpha^j p^j(\xi)dt)\, V(t+dt,\ \xi+d^\alpha\xi)\,. \tag{1.14}$$

According to the Bellman optimality principle we must have that

$$V(t,\xi) = \sup_{\alpha\in S^m} V^\alpha(t,\xi). \tag{1.15}$$

Assuming that the value function $V(t,\xi)$ is continuously differentiable, using equalities $V(t+dt,\Gamma^j\xi + o(dt)) = V(t,\Gamma^j\xi) + 0(dt)$ and $V(t+dt,\xi+d^\alpha\xi) = V(t,\xi) + [\frac{\partial}{\partial t}V(t,\xi)]dt + \sum_{i=1}^N [\frac{\partial}{\partial\xi_i}V(t,\xi)]f_i(\xi,\alpha)dt + o(dt)$, neglecting second order terms, dividing by dt and taking limits as $dt \to 0$, yields

$$-\frac{\partial}{\partial t}V(t,\xi) = \sup_{\alpha\in S^m}\Big\{\sum_{j=1}^m \alpha^j p^j(\xi)V(t,\Gamma^j\xi) + \sum_{i=1}^N [\frac{\partial}{\partial\xi_i}V(t,\xi)]f_i(\xi,\alpha) - (V(t,\xi)-1)p(\xi,\alpha)\Big\}$$
$$:= \sup_{\alpha\in S^m}\sum_{j=1}^m \alpha^j T^j(t,\xi), \tag{1.16}$$

where $f_i(\xi,\alpha)$ is defined in (1.10).

The approach to the optimality equation is based on a theorem which states that if there exists a continuously differentiable solution $V^*(t,\xi)$ of equation (1.16) satisfying the boundary condition $V^*(\nu,\xi) = 0$ for $\nu < \infty$, or its analogue for $\nu = \infty$, and a function $\alpha^*(t,\xi)$ making (1.16) an identity and such that equation (1.10) has a unique solution with $\alpha^*(t,\xi(s))$ replacing α, then α^* defines the optimal strategy and $V^*(t,\xi)$ coincides with the value function $V(t,\xi)$.

The function $\alpha^*(t,\xi)$ is called an optimal *synthesis* and it gives the optimal strategy as a function of time and state as in the usual theory of optimal control.

This theorem will be proved in §4.5 and follows a proof scheme similar to analogous theorems for other types of controlled random processes (for example, for controlled processes of diffusion type (see Krylov 1977).

Note that for such a *verification theorem* it is *not* necessary to give a proof that the value function is sufficiently smooth and satisfies equation (1.16) (i.e. a proof of the *necessity* of the local optimality equation which is here *sufficient*).

1.9 Reduction to the Pontryagin problem

The second approach to the solution of basic scheme problems in continuous time involves reduction to optimal control problems in which the motion of the system (object) is described by the differential equation

$$\dot{x} = f(t, x, \alpha), \tag{1.17}$$

and it is necessary to maximize (minimize) the integral functional

$$F^{\alpha} = \int_t^{\nu} f^0(s, x(s), \alpha(s))ds \tag{1.18}$$

over a class of piecewise continuous (or measurable) functions $\alpha(\cdot)$ with values in some *control set*. The method, called the *Pontryagin maximum principle*, used to solve such problems (and some more general problems) is well known and is related to the introduction of the *Hamiltonian* and *conjugate* variables. The specific characteristics of basic scheme problems reduced to problems of type (1.17),(1.18) yield a sequence of interesting properties.

To obtain the system dynamics for the basic scheme, the differential equation (1.10) is augmented by the differential equation for $z(s)$ (see the equation following (1.9)). Thus we get a system of differential equations for the state variables as

$$\dot{z}(s) = -z(s)p(\xi(s), \alpha) \tag{1.19}$$

$$\dot{\xi}_i(s) = f_i(\xi(s), \alpha) = \sum_{j=1}^{m} \alpha^j f_i^j(\xi(s)), \qquad i = 1, \ldots, N. \tag{1.20}$$

The solution of these equations with initial (at time t) state $z(t) = 1$, $\xi(t) = \xi$ and with control $\alpha := (\alpha(s), s \geq t)$ will be denoted by $z(s|t, \xi, \alpha)$, $\xi(s|t, \xi, \alpha)$.

As well as the problem of maximization of the expected number of jumps over a fixed time interval, we may also consider the problem of maximization of the probability of the event that at least k jumps occur in the time interval to ν. We term this problem B_k. Let us first consider the relations between the problems B_k, $k = 1, 2, \ldots$. The value of the criterion functional in the problem B_k for the strategy $\beta := (\alpha_1(s|t, \xi), \alpha_2(s|\tau_1, \xi(\tau_1), t, \xi), \ldots)$ and the initial point (t, ξ) will

be designated by $F_k^{\beta}(t, \xi)$ and that of the value function by $F_k(t, \xi) := \sup_{\beta} F_k^{\beta}(t, \xi)$. The probability of the first jump after time t occurring in the small time interval $[s, s + ds)$ and being a jump of the j^{th} type is equal to

$$z(s|t, \xi, \alpha_1)\alpha_1^j(s)p^j(\xi(s|t, \xi, \alpha_1))ds. \tag{1.21}$$

By the complete probability formula we have

$$F_k^{\beta}(t, \xi) = \int_t^{\nu} z(s|t, \xi, \alpha_1) \sum_{j=1}^{m} \alpha_1^j(s)p^j(\xi(s|t, \xi, \alpha_1))F_{k-1}^{\tilde{\beta}}(s, \Gamma^j\xi(s|t, \xi, \alpha_1))ds. \tag{1.22}$$

Here $\alpha_1^j(s) := \alpha_1^j(s|t, \xi)$ and $\tilde{\beta} := (\alpha_2(s|\tau_1, \xi(\tau_1), t, \xi), \alpha_3(\cdot), \ldots)$, i.e. the *continuation* of the strategy β after the first jump. According to the Bellman optimality principle

$$F_k(t, \xi) = \sup_{\alpha} \int_t^{\nu} z(s|t, \xi, \alpha) \sum_{j=1}^{m} \alpha^j(s)p^j(\xi(s|t, \xi, \alpha))F_{k-1}(s, \Gamma^j\xi(s|t, \xi, \alpha))ds. \tag{1.23}$$

Here $\alpha := \alpha(s|t, \xi)$ is an arbitrary piecewise continuous (measurable) vector function with values in S^m defined on the interval $[t, \nu)$.

If the function $F_{k-1}(s, \xi)$ is known and continuously differentiable with respect to its arguments, then taking into account also the smoothness of the transformation $\Gamma^j\xi$, we obtain a nonautonomous optimal control problem with fixed time, fixed left-hand end point and free right-hand end point, state variables (z, ξ), $0 < z \leq 1$, $\xi \in S^N$ satisfying the system (1.19),(1.20) and with integral functional (1.22). For simplicity we will also call this problem B_k.

Since $F_0(s, \xi) :\equiv 1$, the problem B_1 may in principle be solved and the values of the function $F_1(s, \xi)$ found. If this function is continuously differentiable, then problem B_2 can thereby be solved, and so on.

From the optimal controls $\alpha_1^*(s|t, \xi)$, $\alpha_2^*(s|t, \xi)$ obtained from the solution of problem B_1, from problem B_2, and so on, the optimal strategy in the problem B_k, $\beta := (\alpha_k^*(s|t, \xi), \alpha_{k-1}^*(s|\tau_1, \xi(\tau_1)), \ldots,$

$\alpha_1^*(s|\tau_{k-1}, \xi(\tau_{k-1})))$, is constructed. Supposing the function $F_{k-1}(t,\xi)$ to be continuously differentiable, we may write the *Hamiltonian* $\widetilde{\mathcal{H}}_k$ of the appropriate Pontryagin maximum principle in the form

$$\begin{aligned}\widetilde{\mathcal{H}}_k &= z\,\mathcal{H}_k \\ &= z\left\{\sum_{j=1}^{m}\alpha^j[p^j(\xi)F_{k-1}(\Gamma^j\xi)+\sum_{i=1}^{N}\psi_i f_i^j(\xi)-\phi p^j(\xi)]\right\} \\ &:= z\sum_{j=1}^{m}\alpha^j L_k^j(\xi,\phi,\psi). \qquad (1.24)\end{aligned}$$

Here $\psi_i := \tilde{\psi}_i/z$, $i = 1,...,N$, where $\tilde{\psi}_i$, ϕ are the *conjugate* variables for the problem B_k.

Let the optimal control be given by some synthesis $\alpha_k^*(t,\xi)$ and let $\xi(t)$ be the corresponding trajectories. Then there exist the functions $\psi(t,\xi)$, $\phi(t,\xi)$, $L_k^j(t,\xi)$ such that $\psi(t) := \psi(t,\xi(t))$, $\phi(t) := \phi(t,\xi(t))$, $L_k^j(\xi(t), \psi(t), \psi(t)) := L_k^j(t,\xi(t))$. Under suitable assumptions it can be proven that $\phi(t,\xi) = F_k(t,\xi)$, $\psi_i(t,\xi) = \partial F_k(t,\xi)/\partial\xi_i$, and under these assumptions $\mathcal{H}_k$ in (1.24) coincides with the right-hand side of an equation for the value function similar to (1.16).

A specific feature of the basic scheme problem is that the derivative of the function L_k^j along optimal trajectories at a point (t,ξ) is expressed in terms of the value of the functions L_{k-1}^r, $r = 1,\ldots,m$, and the optimal synthesis for the problem B_{k-1} at the points $(t,\Gamma^j\xi)$.

If the problem B_{k-1} has been solved, then the functions L_{k-1}^r can be constructed and the relations presented above used to construct the functions L_k^j as necessary conditions. In particular, such a method is used in §6.5 to solve the B_k problems for the symmetric case.

Moreover, the identity mentioned above between the Hamiltonian and the right-hand side of the local optimality equations makes it possible to use the formula for L_k^j in the construction of the optimal synthesis for the Bellman equation.

The solution is more complicated for the basic formulation problem requiring the maximization of the expected number of jumps. Formulated in an analogous way to the above, the Pontryagin maximum principle leads to maximization of an integral functional of the following type:

$$\begin{aligned}F(t,\xi) = \sup_{\alpha}\int_t^{\nu} z(s|t,\xi,\alpha)\sum_{j=1}^{m}\alpha^j(s)p^j(\xi(s|t,\xi,\alpha)) \\ \times\,(1+F(s,\Gamma^j\xi(s|t,\xi,\alpha)))ds, \qquad (1.25)\end{aligned}$$

where an unknown function appears in the left- and right-hand sides of the equation.

One possible way of solving this problem is the following. Suppose that the value function $F(t,\xi)$ is known and has the required smoothness. Then the maximum principle can be written in terms of the corresponding Hamiltonian $\widetilde{\mathcal{H}} := z\sum_{j=1}^{m}\alpha^j L^j$. For the derivative of the function L^j we can obtain a formula analogous to the formula for the derivative of L_k^j. Using this formula as a necessary condition, a synthesis can be constructed and its optimality proved.

1.10 Linear control problems with Poisson jumps

Above we have indicated that basic scheme problems can be converted to particular cases of the optimal control of a system whose evolution consists of deterministic motion defined by differential equations linear in the controls with m types of trajectory jumps occurring at random times whose distributions also depend on the controls. We can apply the Pontryagin maximum principle approach described above to problems whose differential equations have a more general character than (1.20), but with all other specifications remaining the same.

At each moment the control is represented by an m–dimensional vector $\alpha(s) = (\alpha^1(s),\ldots,\alpha^m(s))$ with values in S^m. As before, $\alpha^j(s)$ is interpreted as the "intensity" of use of the j^{th} device relative to the m others at time s. The state of the system is described by the vector $x = (x_1,\ldots,x_N)$. Now the coordinates x_i do not correspond to *a posteriori* probabilities and thus, in considering this problem, we diverge from the Bayesian approach. The motion of the system in the intervals between jumps is described by differential equations linear with respect to α, *viz.*

$$\dot{x}_i = \sum_{j=1}^{m}\alpha^j(s)a_i^j(x). \qquad (1.26)$$