and $q_n(h)$ are lower semicontinuous with respect to the controls. However, in the case where x_n and a_n take only a finite number of values, everything is considerably simpler. We demonstrate this for the basic scheme of §2.2 for the functional $F_\nu(\xi)$ defined in (2.13).

Theorem 2.1 *If $f_n(\theta, h(n))$ from (2.13) are bounded from below, for each fixed n, then for any fixed $\xi \in S^N$ and for $1 \le \nu < \infty$ there exists an optimal action rule β_ν. If, in addition, Condition (A2) holds, then an optimal action rule exists for $\nu = \infty$, $\lim F_\nu(\xi) = F_\infty(\xi)$ and a topology may be introduced on the space of action rules so that any limit point of the sequence $\{\beta_\nu,\ \nu = 1, 2, \ldots\}$ will be the optimal action rule for the problem with $\nu = \infty$.*

Remark 2.7 The results of Theorem 2.1 are true for the function $W_\nu^\beta(\xi)$. This follows immediately from Remark 2.6 regarding Condition (A2). ■

Proof. It follows from the above discussion that to prove Theorem 2.1 we must introduce an appropriate topology on the space of action rules. We have already mentioned (see (2.43)) that each action rule β may be considered as a sequence of vector-valued functions $\beta = \{\pi_{n+1}(a_1 x_1 \ldots a_n x_n),\ n = 0, 1, \ldots\}$ taking values in S^m where $a_k \in \hat{S}^m$, $x_k \in \hat{S}_0^m$. The elements of this sequence for fixed n, i.e. all possible functions $\pi_{n+1}(a_1 x_1 \ldots a_n x_n)$ taking values in S^m whose arguments are from a finite set, obviously constitute a compact set with respect to the topology of pointwise convergence.

On the set of action rules we define the topology to be Tychonov's topology for the product of spaces corresponding to the different n. The set of action rules is compact in this topology by Tychonov's theorem.

We show now that for $\nu < \infty$ the functional $F_\nu^\beta(\xi)$ is lower semicontinuous with respect to the topology introduced. Indeed, $E_\xi^\beta f_n(\theta, h(n))$ may be represented as the sum of a finite number of components corresponding to the possible values of $h(n)$. For each $h(n)$ the probability of the corresponding trajectory $h := a_1 x_1 \ldots a_n x_n$ under the i^{th} hypothesis is represented as the product of a constant depending on h, the product of corresponding numbers λ_i^j and $1 - \lambda_i^j$ and the product of the corresponding coordinates of the vectors $\pi_s(a_1 x_1 \ldots a_{s-1} x_{s-1})$

for $s = 1, \ldots, n$. Thus the probabilities of the different values of $h(n)$ are continuous with respect to the action rule β. If all possible values of $f_n(\theta, h(n))$ are finite, then $E_\xi^\beta f_n(\theta, h(n))$ is continuous with respect to β, and if $f_n(\theta, h(n))$ takes the value $+\infty$, then lower semicontinuity holds, as required. ■

* * *

Finally, we prove an important property of the function $F_\nu(\xi)$, which we will need later.

Theorem 2.2

(a) If for $\nu \le \infty$ there exists ξ^0 such that $\xi_i^0 > 0$ for all $i = 1, \ldots, N$ and $F_\nu(\xi^0) < \infty$, then $F_\nu(\xi) < \infty$ for all $\xi \in S^N$.

(b) For each fixed ν, $0 \le \nu \le \infty$, the function $F_\nu(\xi)$ is convex, lower semicontinuous, continuous on the interior of the simplex S^N and its restriction to the interior of any face of any dimension is also continuous.

(c) If $\nu < \infty$ and $f_n(\theta, h(n))$ is finite for $n \le \nu$, then $F_\nu(\xi)$ is continuous on the whole simplex S^N.

Proof. To prove this we need the following lemma, which holds for arbitrary minimization problems of the type

$$\inf_{d \in \mathcal{A}} \sum_{i=1}^{N} \xi_i I_i^d, \tag{2.56}$$

where $\mathcal{A}$ is some control set, I_i^d, $i = 1, \ldots, N$, are some functionals such that $-\infty < I_i^d \le +\infty$ and $\xi \in S^N$. ■

The value of the infimum in (2.56) is denoted by $\Phi(\xi)$.

The proof of the following lemma is essentially obvious and we omit it.

Lemma 2.3

(a) From the existence of a control d such that $I_i^d < \infty$ for all $i = 1, 2, \ldots, N$, it follows that $\Phi(\xi) < \infty$ for all ξ.

(b) $\Phi(\xi)$ *is a convex lower semicontinuous function, continuous on the interior of the simplex* S^N, *and its restriction to the interior of any face of any dimension is also continuous.*

(c) If there exists $c > 0$ *such that* $I_i^d < c$ *for all* $d \in \mathcal{A}$, $i = 1, \ldots, N$, *then* $\Phi(\xi)$ *is continuous on* S^N. ■

Notice now that according to (2.5) the functional $F_\nu^\beta(\xi)$ may be represented as

$$F_\nu^\beta(\xi) = \sum_{i=1}^{N} \xi_i F_\nu^\beta(e_i^N). \tag{2.57}$$

From this and from the statement of Lemma 2.3 we derive the statement of Theorem 2.2. ■

The minimization problem of type (2.56) is naturally termed a *finite Bayesian problem*. According to (2.57), the problems of §2.1 are indeed related to this class.

Note that in §7.3 we will prove Lemma 7.2, which gives sufficient conditions for the continuous differentiability of the function $\Phi(\xi)$.

2.4 Optimality equation and optimal strategies

In the previous section we mentioned an approach allowing us, under some conditions, to establish the existence of an optimal strategy and the convergence of the optimal value function for a finite time horizon ν to the optimal value at $\nu = \infty$. However, no mention was made of how to find the optimal strategy and what properties it possesses. To answer this question we need a more detailed study of the problem, which usually involves the second approach based on the Bellman equation characterization of optimality, *without* the explicit introduction of a topology on the strategy space mentioned at the end of §2.2.

The Bellman equation is appropriate to the case in which the problem is not dependent on the parameter θ. In the second part of §2.1 it was shown that in its Bayesian formulation the basic scheme may be converted to a control problem with complete observations, i.e. where θ is absent, but to the process states the corresponding *a posteriori* probabilities of the hypotheses are added.

In the general case, with all considered spaces assumed to be Borel spaces, it is also possible in the Bayesian formulation to convert the problem to a control problem with complete observations with θ absent and a new state space at time n in the form $\mathcal{X}_n \times \mathcal{P}(\Theta)$ for $n \geq 1$. If the admissible controls do not depend on the prehistory, the strategy of nature at time n depends only on the control chosen at this moment and the cost function depends only on the last state and control, then, as in the basic scheme, $\mathcal{P}(\Theta)$ may be taken simply as the state space. The proof of these facts is accomplished similarly to that for the basic scheme, but is technically more complex and we will not present it. We only mention here that the basis of this equivalence lies in the following fact. For a fixed strategy π the measure on $\Theta \times \mathcal{H}_n$ which is constructed from the measures $\mu(d\theta)$, $p_0^\theta(\cdot)$ and the transition probabilities $p_s^\theta(\cdot|\tilde{h})$ and $\pi_s(\cdot|h)$, $s = 1, \ldots, n$, may be represented as a product (unconditional) measure on $\mathcal{H}_n$ and a transition probability $\mu_n(\cdot|h)$ from $\mathcal{H}_n$ into Θ. This transition probability is called the *a posteriori probability* of the parameter θ at time n.

Admissible states of the control problem with complete information are obtained from admissible states of the initial problem by means of the Bayesian updating formula for the *a posteriori* probabilities. Thus, in §2.1 it was shown for the basic scheme that if at time n the *a posteriori* probability equals ξ and at time $n+1$ the control e_j^m is applied, then admissible states of the process $(\Delta X(n+1), \xi(n))$ will be $(e_j^m, \Gamma^{1j}\xi)$ and $(e_0^m, \Gamma^{0j}\xi)$. In this case, as has already been stated, if the cost function depends only on the last state and control, then it suffices to consider *only* the process $\xi(n)$.

Upon transforming the Bayesian problem to the control problem with complete information, the new state space $\mathcal{X}_0'$ coincides with $\mathcal{X}_0$, and all other $\mathcal{X}_n'$ for $n \geq 1$ have the form $\mathcal{X}_n \times \mathcal{P}(\Theta)$. It is convenient for symmetry to consider the more general problem where $\mathcal{X}_0'$ also has the form $\mathcal{X}_0 \times \mathcal{P}(\Theta)$, i.e. to consider all possible *initial distributions* μ. In such a problem the condition of measurability with respect to $\mu \in \mathcal{P}(\Theta)$ must be added to the definition of strategy.

Using the example of the basic scheme we will consider in detail how the definition of strategy changes upon transforming from a control problem with incomplete information to a control problem with complete information. For reasons of simplicity we will *assume* that

f_n depends only on Θ, $\Delta X(n)$ and $a(n)$ and that the matrix Λ has no column in which all elements coincide. In this case, as follows from §2.1, the corresponding control problem with complete data can be represented by using the terminology in §2.2 as follows.

The state of the system at time n corresponds to the *a posteriori* probabilities of hypotheses at that moment, so all spaces $\mathcal{X}_n (n \geq 0)$ coincide with the $(N-1)$-dimensional simplex S^N. As before, control consists in the choice of device which is observed, therefore all $\mathcal{A}_{n+1} (n \geq 0)$ coincide with $\widehat{S}^m$. The sets $\widetilde{\mathcal{H}}_n$, $\mathcal{H}_n$ and the transition probabilities are defined for $n \geq 1$ as follows. At each moment of time, all controls are admissible and, if at time n the control e_j^m is chosen , i.e. the j^{th} device is observed, and at time $(n-1)$ the state was ξ $(\xi \in S^m)$, then at time n the admissible states will be $\Gamma^{1j}\xi$ and $\Gamma^{0j}\xi$ and the probability of being in Γ^{1j} equals $p^j(\xi)$ and the probability of being in state $\Gamma^{0j}\xi$ equals $1 - p^j(\xi)$ (see formulae (2.16),(2.17),(2.29)). The cost functions q_n are defined according to (2.23), and the q_n corresponding to losses are of the form $\xi(n-1)(\tilde{\Lambda} - \Lambda)a^*(n)$, i.e. observing the j^{th} device at time n the loss equals

$$q^j(\xi) = \sum_{i=1}^{N} \xi_i(\lambda_i - \lambda_i^j). \tag{2.58}$$

Since a distribution on a finite number of points e_j^m $(j = 1, \ldots, m)$ is given by a vector from S^m, then in this case a strategy is defined as a sequence of functions $\pi := \{\pi_{n+1}(h),\ n = 0, 1, 2, \ldots\}$, where the measurable function $\pi_{n+1}(h)$ defining a probability distribution on $\mathcal{A}_{n+1} := \widehat{S}^m$ takes values in S^m and is given for $h \in \mathcal{H}_n$. Since for fixed $\xi_0 \in \mathcal{H}_0 := S^N$ there exist only a finite number of histories $h \in \mathcal{H}_n$, the requirement of measurability of the function $\pi_{n+1}(h)$ amounts simply to measurability with respect to ξ_0.

So, we have completely described the control problem with complete information as stated in §2.2 to which the basic scheme has been converted. However, it is more convenient to consider the *more general* problem in which the spaces of states and controls, transition functions and penalty functions are the same as before, but the constraints on admissible states are *absent*, i.e. $\widetilde{\mathcal{H}}_n'$ coincides with the *product* of the appropriate spaces. In this case a strategy is defined as a sequence of

functions

$$\pi = \{\pi_{n+1}(\xi_0, a_1, \xi_1, \ldots, a_n, \xi_n),\ n = 0, 1, \ldots\}, \tag{2.59}$$

where π_{n+1} is a function taking values in S^m and is measurable with respect to $\xi_0, \xi_1, \ldots, \xi_n$. In the problem with constraints on the set of admissible states the strategies are obtained by the projections of the functions π_{n+1} on the sets $\mathcal{H}_n$, $n = 1, 2, \ldots$.

It is easy to see that for a fixed *a priori* distribution ξ_0 the strategy (2.59) defines some action rule for the initial problem. Indeed, according to (2.18) the value of the process $\xi(n)$ may be represented as a function of ξ_0 and of $a(1), \Delta X(1), \ldots, a(n), \Delta X(n)$. Putting the corresponding functions into (2.59), we obtain an action rule which depends in a measurable way on ξ_0 and the corresponding value functions will coincide.

On the other hand, as was mentioned regarding (2.43), a fixed action rule may be considered as a sequence of functions

$$\beta = \{\pi_{n+1}(a(1), \Delta X(1), \ldots, a(n), \Delta X(n)),\ n = 0, 1, \ldots\}\,. \tag{2.60}$$

But $\Delta X(r)$ may, according to (2.18), be represented as a function of $\xi(r-1), a(r), \xi(r)$. Putting the corresponding functions into (2.60), we obtain functions π_{n+1} defined only on $\mathcal{H}_n$. Consider a strategy π of type (2.59) which on $\mathcal{H}_n$ coincides with the obtained functions and outside $\mathcal{H}_n$ is defined arbitrarily. Because the set $\mathcal{H}_n$ is measurable, such a strategy always exists. In the following sense the obtained strategy does *not* depend on ξ_0.

If $h^{(k)} = (\xi^{(k)}(0),\ a(1),\ \xi^{(k)}(1), \ldots, a(n),\ \xi^{(k)}(n))$, $k = 1, 2$, are two histories from $\mathcal{H}_n$ with *different* ξ_0 but corresponding to the same set $(a(1), \Delta X(1), \ldots, a(n), \Delta X(n))$, i.e. such that for any s $(1 \leq s \leq n)$ from $\xi^{(1)}(s) = \Gamma^{1j}\xi^{(1)}(s-1)$ it follows that $\xi^{(2)}(s) = \Gamma^{1j}\xi^{(1)}(s-1)$, then $\pi(h^{(1)}) = \pi(h^{(2)})$.

* * *

The Bellman optimality equation connects the optimal value functions for problems started at different times. It has the most simple form in the Markov case, which we now formulate.

A special case of the model is called a *Markov* control problem with partial observations when the following conditions hold:

(1) Transition probabilities defining the strategy of nature depend only on the last state and control, i.e.

$$p^{\theta}_{n+1}(\cdot|x_0a_1\dots x_na_{n+1}) := p^{\theta}_{n+1}(\cdot|x_na_{n+1}).$$

(2) The sections of the set $\widetilde{\mathcal{H}}_{n+1}$ for points $h \in \mathcal{H}_n$ depend only on x_n and the sections of $\mathcal{H}_n$ for points $\tilde{h} \in \widetilde{\mathcal{H}}_n$ depend only on x_{n-1} and a_n.

(3) Cost functions depend only on the last control and the previous state, i.e.

$$q^{\theta}_n(x_0a_1x_1\dots x_{n-1}a_nx_n) := q^{\theta}_n(x_{n-1}a_n).$$

Condition 2 means that for each point the set of admissible controls does not depend on the path by which we come to this point, the set of admissible states depends only on the last control and the previous state. In Condition 3 it might be assumed that q_n also depends on x_n, but this is not necessary by Remark 2.3. Obviously, Conditions 1 and 2 are satisfied for the basic scheme, and Condition 3 is satisfied for the maximization of the number of successes and the minimization of loss problems.

Formally, the Markov control problem is a *special* case of the general problem of sequential control. However, it is known that the *general* case may be written as a Markov problem. It suffices to take the $\mathcal{H}_n$ as the new state spaces and the $\widetilde{\mathcal{H}}_{n+1}$ as the new control spaces. But it is not always convenient to convert to a Markov problem, because this may lead to an unreasonable expansion of state and control spaces.

Essentially, upon the transformation of the Bayesian Markov control problem with incomplete observation to a control problem with complete observation, we again have a Markov problem. For the basic scheme this follows directly from the form of the transition probability and cost functions.

In the remainder of this section we consider a Markov control problem with *complete observation.*

In the study of Markov problems, an important rôle is played by *randomized Markov strategies*, i.e. strategies of the form

$$\pi_{n+1}(\cdot|x_0a_1\dots a_nx_n) := \pi_{n+1}(\cdot|x_n), \quad n = 0, 1, \dots .$$

It may be shown that for *any* strategy π in the Markov problem a *Markov* strategy $\overline{\pi}$ such that $w^{\overline{\pi}} \le w^{\pi}$ can be found, but we will not discuss this result further.

A basic approach to investigation of any dynamical optimization problem is the (Bellman) *optimality principle*, the heuristic formulation of which is as follows: *an optimal strategy possesses the property that if by applying it we arrive at some intermediate time at a point x, then it remains optimal for the problem over the remaining interval of time for which the point x serves as the initial point.*

A rigorous formulation of this principle requires the definition of the optimization problem at a fixed intermediate point. This may be done by two methods.

The first consists in considering the minimization with respect to all possible strategies of the functional

$$\sum_{r=s+1}^{\nu} E^{\pi}(q_r(x_{r-1}a_r)|x_s = x) = \Phi^{\pi}_{s\nu}(x), \quad 0 \le s < \nu \le \infty. \tag{2.61}$$

Since $\Phi^{\pi}_{s\nu}(x)$ is uniquely defined for each π only up to a set of P^{π} measure 0 in Ω, then in the case of general state and control spaces the questions arise of how to define $\Phi^{\pi}_{s\nu}(x)$ independently of the strategy used in the interval $[0, s)$ and how to define $\inf_{\pi} \Phi^{\pi}_{s\nu}(x)$ so that it is measurable. These questions do not arise in the case in which the initial distribution concentrates on a countable number of points and, for each point $x \in \mathcal{X}_s$, $s = 0, 1, \dots$, the union of the sets of admissible states in $\mathcal{X}_{s+1}$ with respect to all admissible controls is countable. If the last of these conditions holds, then we will say that we have a case of *discrete transitions*, but the space itself need not necessarily be discrete. The basic scheme considered in terms of *a posteriori* probabilities falls exactly in this case.

We will use here another method of settling the optimality question; that based on the fact that in a Markov problem after each transition we arrive in the situation of a new Markov problem. This method is connected with the introduction of *remaining* Markov models, i.e. problems where the initial state space is not $\mathcal{X}_0$ but some $\mathcal{X}_s$ for $s \ge 1$, and further evolution may be described similarly to that of the initial problem.

In the intermediate Markov problem at time s, admissible controls and states, and also transition and loss functions, are defined *only* for those points in $\mathcal{X}_s$ to which an admissible trajectory leads. It is, however, convenient for consideration of arbitrary models that these sets and functions are defined for *all* $x \in \mathcal{X}_s$. Therefore, to study the Markov problem we introduce the following changes in the general scheme.

We will consider that for all $s = 0, 1, \ldots$ measurable sets $\tilde{F}_{s+1}$ and F_{s+1} are given such that $\tilde{F}_{s+1} \subset \mathcal{X}_s \times \mathcal{A}_{s+1}$, the projection of $\tilde{F}_{s+1}$ on $\mathcal{X}_s$ coincides with $\mathcal{X}_s$, $F_{s+1} \subset \tilde{F}_{s+1} \times \mathcal{X}_{s+1}$ and the projection of F_{s+1} on $\mathcal{X}_s \times \mathcal{A}_{s+1}$ coincides with $\tilde{F}_{s+1}$. The set F_{s+1} defines the admissible controls and states for points from $\mathcal{X}_s$. Now $h := x_0 a_1, \ldots, a_{s+1} x_{s+1} \in \mathcal{H}_{s+1}$ if, and only if, for all $i \leq s$, $x_i a_{i+1} x_{i+1} \in F_{i+1}$ and the measures $p_{i+1}(\cdot | x_i a_{i+1})$ and the functions q_i are defined on $\tilde{F}_{i+1}$. Further, we will assume that the distribution $p_0(dx)$ in the Markov problem (2.31) is *not* fixed, and consider the set of problems for *all* possible distributions $p_0(dx)$ which are concentrated on a single point, termed the *initial (state) point.*

We call such an object a *Markov model* and designate it by Z_0 or simply Z. Designate by Z_s ($s \geq 1$), and call the s^{th} *remaining model*, the Markov model given by the corresponding set (2.31), where n runs through the values s, $s+1, \ldots$. In such a model the strategies are defined on histories of the type $x_s a_{s+1} x_{s+1} \ldots$. The set of strategies for the s^{th} remaining model will be designated by Π_s.

As before, the strategy $\pi \in \Pi_s$ and the initial point $x \in \mathcal{X}_s$ define a measure P_x^π on the corresponding $\mathcal{H}$. Expectation with respect to this measure is denoted by E_x^π. Previously the subscript on the expectation symbol corresponded to the *a priori* parameter distribution, but with the transition to the problem with complete information the *a priori* distribution of parameters corresponds to the initial point, and therefore the notation introduced here agrees with the notation of §2.2. The *value of strategy* $\pi \in \Pi_s$ on the time interval $[s, \nu)$, $\nu \leq \infty$, for the initial point $x \in \mathcal{X}_s$ is given by

$$w_{s\nu}^\pi(x) := \sum_{r=s+1}^{\nu} E_x^\pi q_r(x_{r-1} a_r) ,$$

and the (optimal) *value function*, i.e. $\inf_\pi w_{s\nu}^\pi(x)$, is denoted by $w_{s\nu}(x)$.

In this section we will *assume* that for any strategy π and any x, $x \in X_s$, $s = 0, 1, \ldots$,

$$\sum_{r=s+1}^{\infty} E_x^\pi q_r^-(x_{r-1} a_r) < \infty , \tag{2.62}$$

where $q^- := \max(-q, 0)$. This implies that

$$w_{s\infty}^\pi(x) := \sum_{r=s+1}^{\infty} E_x^\pi q_r = E_x^\pi \sum_{r=s+1}^{\infty} q_r > -\infty.$$

The *optimality equation* connects the values $w_{s-1,\nu}(x)$ and $w_{s\nu}(y)$. First, we write the equation which, for a fixed strategy $\pi \in \Pi_{s-1}$, $\pi := \{\pi_{s+r}(\cdot|h),\ r = 0, 1, \ldots\}$, connects the values $w_{s-1,\nu}^\pi(x)$ and $w_{s\nu}^{\pi'}(y)$, where $\pi' := \pi'_{xa}$ is the strategy in the model Z_s which is the *continuation* of the strategy π, so that $\pi'_{xa}(\cdot|h') := \pi(\cdot|xah')$, where $h' = x_s a_{s+1} x_{s+1} \ldots$ is a history in the model Z_s. This relation (the *fundamental equation*) for $s < \nu$ is given by

$$w_{s-1,\nu}^\pi(x) = \int_{\mathcal{A}_s} \pi_s(da|x)[q_s(xa) + \int_{\mathcal{X}_s} w_{s\nu}^{\pi'_{xa}}(y) p_s(dy|xa)]. \tag{2.63}$$

This formula follows directly from (2.33) and may be written more concisely as follows. If a is an admissible control for the point $x \in \mathcal{X}_{s-1}$ and the function $f(ay)$ is defined on the product $\mathcal{A}_s \times \mathcal{X}_s$, then let

$$\begin{aligned} (M_s^a f)(x) &:= M_s^a f(x) := \int_{\mathcal{X}_s} f(ay) p_s(dy|xa) , \\ (T_s^a f)(x) &:= T_s^a f(x) := q_s(xa) + M_s^a f(x), \quad s = 1, 2, \ldots . \end{aligned} \tag{2.64}$$

It is obvious that M_s^a and T_s^a have the following properties:

$$\begin{aligned} &(1)\ M_s^a(f+g)(x) = M_s^a f(x) + M_s^a g(x), \\ &(2)\ T_s^a(f+g)(x) = T_s^a f(x) + M_s^a g(x). \end{aligned} \tag{2.65}$$

If μ is a distribution on $\mathcal{A}_s$, then we define

$$T_s^\mu f(x) := \int_{\mathcal{A}_s} \mu(da) T_s^a f(x) . \tag{2.66}$$

In this notation the fundamental equation may be written as

$$w^{\pi}_{s-1,\nu}(x) = T^{\pi_s}_s w^{\pi'}_{s\nu}(x). \tag{2.67}$$

If $\nu < \infty$, then (2.63) holds true also for $s = \nu$ if by definition we let

$$w^{\pi}_{\nu\nu} := 0. \tag{2.68}$$

We note, also, that if the transition function $\pi_s(\cdot|x)$ is fixed, then $T^{\pi_s(\cdot|x)}_s$ is an operator transforming functions given on the admissible subset of the product space $\mathcal{A}_s \times \mathcal{X}_s$ into functions given on $\mathcal{X}_{s-1}$. If for each $x \in \mathcal{X}_{s-1}$ the set of admissible controls coincides with $\mathcal{A}_s$, then for the sake of simplicity we will assume further that T^a_s and M^a_s for fixed a may be considered as operators transforming functions given on $\mathcal{X}_s$ into functions given on $\mathcal{X}_{s-1}$.

We now define the operator T_s by the formula

$$T_s f(x) := \inf_{a \in \mathcal{A}_s} T^a_s f(x), \quad s = 1, 2, \ldots . \tag{2.69}$$

If the transition function $\nu(\cdot|x)$ from $\mathcal{X}_{s-1}$ into $\mathcal{A}_s$ satisfies the relation

$$T^{\nu(\cdot|x)}_s f(x) = \inf_{a \in \mathcal{A}_s} T^a_s f(x), \tag{2.70}$$

i.e. for each x the measure $\nu(\cdot|x)$ is concentrated on the control a which achieves the infimum in (2.69), then we say that $\nu(\cdot|x)$ *realizes the operator T for a function f*.

To derive the optimality equation from (2.67) the following operations, which require some justification, are needed. We must take the infimum with respect to $\pi \in \Pi_{s-1}$ on both sides of equation (2.67). In the right-hand side the infimum is divided into an outside infimum with respect to the transition function $\pi(\cdot|x)$ from $\mathcal{X}_{s-1}$ into $\mathcal{A}_s$ and an inside infimum with respect to $\pi'_{xa} \in \Pi_s$. We must show that we may interchange the latter with the operator T^a_s and replace it with infimum with respect to $\pi \in \Pi_s$. As will be seen below, to justify this operation we need, for example, the measurability of the function $w_{s\nu}(x)$ and the existence of uniformly ε-optimal strategies, which are defined in the following way.

By the definition of $w_{s\nu}(x)$, for any $\varepsilon > 0$ and any $x \in \mathcal{X}_s$ there exists an *ε-optimal strategy*, i.e. $\pi(x) \in \Pi_s$ such that

$$w^{\pi(x)}_{s\nu}(x) \le w_{s\nu}(x) + \varepsilon.$$

If in this inequality $\pi \in \Pi_s$ may be chosen independently of x, then such a strategy is called *uniformly ε-optimal* and, if $\varepsilon = 0$, *uniformly optimal*.

Now we will give Theorem 2.3 which states that $w_{s\nu}(x)$ satisfies the optimality equation. This theorem holds under general assumptions, for example, for Borel spaces of states and controls. However, to avoid a discussion of measurability, we prove it here only for the cases of interest to us, when at least one of the two following *assumptions* hold:

(1) Z_s is a model with discrete transitions (defined above).

(2) In the model Z_s there exists a uniformly ε-optimal strategy for any $\varepsilon > 0$ and it is known that the function $w_{s\nu}(x)$ is measurable.

Theorem 2.3 *Let at least one of the two assumptions (1) and (2) hold. Then:*

(a) The functions $w_{s\nu}(x)$, $0 \le s < \nu \le \infty$ satisfy the following relations (optimality equation):

$$w_{s-1,\nu}(x) = T_s w_{s\nu}(x). \tag{2.71}$$

(b) If $\pi = (\pi_s(\cdot|x), \pi_{s+1}(\cdot|x a_{s+1} s_{s+1}), \ldots)$ is a uniformly optimal strategy in the model Z_{s-1}, then the distribution $\pi_s(\cdot|x)$ realizes the operator T_s for function $w_{s\nu}(x)$.

Proof. Fix some $x \in \mathcal{X}_{s-1}$. From (2.67) it follows for any strategy $\pi \in \Pi_{s-1}$ that

$$\begin{aligned} w^{\pi}_{s-1,\nu}(x) &\ge \int_{a \in \mathcal{A}_s} \pi_s(da|x) T^a_s w_{s\nu}(x) \\ &:= T^{\pi_s(\cdot|x)} w_{s\nu}(x) \ge T_s w_{s\nu}(x). \end{aligned} \tag{2.72}$$

If Assumption 1 holds, then the operator T_s^a may be applied to the function $w_{s\nu}(x)$, since for all possible controls transitions from the point x to at most a countable number of admissible points $y \in \mathcal{X}_s$ is made. If it is known that $w_{s\nu}(x)$ is measurable, then the question of the applicability of T_s^a does not arise.

We choose an arbitrary $\varepsilon > 0$ and, in the case of discrete transitions, for each point y of the countable set in $\mathcal{X}_s$ mentioned above we take a strategy $\pi(y) \in \Pi_s$ such that $w_{s\nu}^{\pi(y)}(y) < w_{s\nu}(y) + \varepsilon$. Then we will take as $\pi_\varepsilon \in \Pi_s$ the strategy which, for each y from the countable set mentioned above, coincides with $\pi(y)$ for all histories h in the model Z_s. For the case when it is known that in the model Z_s an ε-uniformly optimal strategy exists, then this strategy is taken as π_ε.

Consider now the strategy $\pi^{x\varepsilon} \in \Pi_{s-1}$ such that for $x = x_{s-1}$ all transition functions $\pi_{s+r}^{x\varepsilon}(\cdot|x_{s-1}a_s \dots x_{s+r-1})$ for $r \geq 1$ coincide with the corresponding functions defining the strategy π_ε. The distribution of $\pi_s(\cdot|x)$ is concentrated on some control $\bar{a}$ for which $T_s^{\bar{a}} w_{s\nu}(x) < T_s w_{s\nu}(x) + \varepsilon$ and for the remaining x the strategy may be chosen arbitrarily. Then by (2.66) and by the choice of $\bar{a}$ and π_ε

$$w_{s-1,\nu}^{\pi^{x\varepsilon}}(x) = T_s^{\bar{a}} w_{s\nu}^{\pi_\varepsilon}(x) < T_s^{\bar{a}}(w_{s\nu}(x) + \varepsilon) \leq T_s w_{s\nu}(x) + 2\varepsilon.$$

Since ε is arbitrary, from this and (2.72) we obtain that (2.71) holds.

If π is a uniformly optimal strategy in the model Z_{s-1}, then for this model the first and the last terms in (2.72) coincide with $w_{s-1,\nu}(x)$ and this equality, by definition, means that $\pi_s(\cdot|x)$ realizes equation (2.71). ■

Remark 2.8 The statement of Theorem 2.3 holds for models in which $w_{s\nu}(x)$ can be represented as the solution of a finite Bayesian problem with $\mathcal{X}_n := S^N$ and $\Pi_s := \mathcal{A}$. Assumption 2 holds for such models. Indeed, measurability of $w_{s\nu}(x)$ follows from the convexity proved in Lemma 2.3. Existence of a uniformly ε-optimal strategy follows from the results of Lemma 2.3 and from the fact that for any fixed π which is not dependent on x the function $w_{s\nu}^{\pi}(x)$ is linear with respect to x. ■

* * *

It is useful to have a condition under which the solutions of the equations $f_{s-1} = T_s f_s$ coincide with the value functions $w_{s-1,\nu}(x)$ and randomized Markov strategies composed from the distributions realizing the corresponding operators are optimal.

We now formulate the appropriate theorems separately for the cases of finite and infinite horizon, since only the latter case is non-trivial.

Theorem 2.4 *Let the sequence of functions $f_s(x)$, $0 \leq s \leq \nu < \infty$, and the randomized Markov strategy $\pi^0 := \{\pi_{r+1}(\cdot|x_r),\ r = 0,1,\dots\}$ be such that:*

(1) the functions $f_s(x)$ satisfy the equations

$$f_{s-1}(x) = T_s f_s(x), \tag{2.73}$$

(2) the distributions $\pi_s(\cdot|x_{s-1})$ realize the operators T_s for the functions $f_s(x)$,

(3) $f_\nu(x) \equiv 0$.

Then $f_s(x) = w_{s\nu}(x)$ and the Markov strategy $\pi^s = \{\pi_{r+1}(\cdot|x_r)$, $r = s, s+1, \dots\}$ is uniformly optimal for the model Z_s.

Proof. We give the proof by induction, beginning from the time at the horizon ν. By definition $w_{\nu\nu}(x) = f_\nu(x)$ (see (2.68) and Condition 3 of the theorem). Let the statement of the theorem be proved for all $s = n, \dots, \nu$, so that the operator T_n is applicable to $w_{n\nu}$. For any strategy $\tilde{\pi}$, using the fundamental equation (2.63), the induction assumption and equation (2.73) we obtain

$$\begin{aligned} w_{n-1,\nu}^{\tilde{\pi}}(x) &= T_n^{\tilde{\pi}_n} w_{n\nu}^{\tilde{\pi}'_{x^n}}(x) \geq T_n w_{n\nu}^{\tilde{\pi}'_{x^n}}(x) \geq T_n w_{n\nu}(x) \\ &= T_n f_n(x) = f_{n-1}(x). \end{aligned} \tag{2.74}$$

Consider now the strategy π^{n-1} in model Z_{n-1}. Applying sequentially the fundamental equation (2.63), the induction assumption, Condition 2 of the theorem and the equalities (2.73), we obtain

$$\begin{aligned} w_{n-1,\nu}^{\pi^{n-1}}(x) &= \int_{\mathcal{A}_n} \pi_n(da|x)\, T_n^a w_{n\nu}^{\pi^n}(x) \\ &= \int_{\mathcal{A}_n} \pi_n(da|x)\, T_n^a f_n(x) = T_n f_n(x) = f_{n-1}(x). \end{aligned}$$