Information And Strategies in Dynamic Games

P.BERNHARD

Institut National de Recherche en Informatique et en Automatique

Sophia Antipolis

Summary

We extend to the setting of stochastic dynamic games with incomplete information a theorem of Kuhn, and use it to prove the existence of a saddle point in a suitable class of strategies. We then particularize this result to the situation where one of the players has full information to show existence of a saddle point in another class of strategies exhibiting a constant dimension sufficient statistic. A dynamic programming-like algorithm is naturally associated with this class of strategies, and was proposed in a previous paper in a sufficient condition setting. For this same class of games, we give an example of another use of the main theorem, leading to a different dynamic programming-like algorithm.

Keywords. Games, saddle point, mixed strategies, dynamic programming.

1. Introduction

In an historical paper in 1953 [5], Kuhn introduced the modern concept of game in extensive form, extending and simplifying the concept introduced by Von Neuman and Morgenstern [11]. In the same paper, he proved that all such games *with complete memory* have a saddle point in behavioural strategies.

We shall consider the equivalent property in the setting of dynamic, and more specifically multistage, games. When all variables range over finite sets, the latter are a special case of games in extensive form. However, our approach allows us to deal with the case where the decision variables range over infinite sets (we shall restrict them to compact sets for technical reasons), and with noisy information. Notice also that the property of *perfect memory*, which was somewhat technical in Kuhn's set up, becomes extremely natural and simple in the set up of dynamical systems.

For the sake of simplicity, we restrict our attention to two-player zero-sum games. It is clear that our form of Kuhn's theorem extends, as well as the original one, to many-person games. (The simplest way to see that is, following Aumann [1], to lump into "player two" the actions of all the other players). Again for simplicity, we shall first derive it for deterministic games, and extend it to stochastic games after. We could also, of course, as in refs [8],[9] discussed below, assume compacity of the control space of one player only, and use the non symmetric version of Sion's theorem.

In his paper [1], Aumann proposed an extension of Kuhn's theorem to infinite games. He states (p 628) "A mixed strategy can be thought of as a probability distribution, i.e. a measure, on the set of all pure strategies". This is is exactly what we do here. He prefers not to place a measurable structure on the set of pure strategies. (We use Radon measures, with various topologies). As a result, he has to use a rather restrictive class of mixed srategies, which are not really the direct generalization of Kuhn's mixed strategies, but some superset of our, and his, behavioural strategies. But the difference between mixed and behavioral then appears somewhat artificial, since his definition of behavioral is that, in our notations, for $i > j$, the random variables u_i and u_j should be independant. However, the probability law of u_i is $p_i = \varphi_i(u^{t-1}, y^t)$, *explicitly depending on the realization* u_j. (In the notations of [1], our u_i is $y_i = b_i(x_i, \omega)$, where x_i plays the role of our r_i, and encodes, through the function u_i^j, the information on $y_j = u_i^j(x_i)$.)

In the past few years, several papers have dealt with this type of games, see e.g. [8] and [9]. These two references and the related literature use a similar set up to ours, and their strategies are our behvioral strategies. The link with what we call mixed strategies is not made there. They do not either consider deterministic games (which are topologically more difficult to handle), nor, more importantly, partial or noisy state information. They deal with infinite time games. The main obstacle to doing so here is that the second part of the paper would not carry over in a simple way.

Starting with section 7, we examine in more detail the case where one of the players has full (causal) information, where it is known that the second guessing problem simplifies. See [2] and the bibliography therein. A dynamic programming approach lets us achieve two things. On the one hand, it allows us to show that there exists a saddle point in the class of strategies used in [2], using finite dimensional sufficient statistics, which turns the sufficient condition of that paper into a necessary and sufficient condition. On the other hand, it may be more effective for short duration games to stay with the space of behavioural strategies, and we shall derive a different dynamic programming-like algorithm for the particular case of the rabbit and hunter game.

Notations. We shall study only *discrete time* games, so that we shall have to deal with finite sequences of objects. We shall adopt the following conventions. Let $a = \{a_1, a_2, \dots, a_T\}$ be a finite sequence, where $a_t \in A_t$. A subscript will refer to a particular element of the sequence, while a superscript will refer to the restriction of the sequence to its first elements: $a^t = \{a_1, a_2, \dots, a_t\} \in A_1 \times A_2 \times \cdots \times A_t = A^t$. The notation A^t will therefore mean the cartesian power of A *only if* $A_1 = A_2 = \cdots = A_t = A$. Likewise, if α is a function ranging over A^T, α_t will be its component in A_t, and α^t its first t components. Finally, if $a \in A^t$ et $b \in A_{t+1}$, then $a \cdot b$ stands for the element of A^{t+1} obtained by concatenating a and b.

Let us also agree that for a topological space A, we shall call $\pi(A)$ the set of all (Radon) probability measures over A.

2. Multistage game.

A deterministic two-player zero-sum multistage game is given by

- An integer T called the *horizon* of the game. Let $\mathbf{T} = \{1, 2, \ldots, T\}$ and $t \in \mathbf{T}$ is called the time.
- A sequence of *state spaces* X_t. We shall use $x_t \in X_t$, the *state at time* t.
- An *initial state* $x_1 \in X_1$, which is assumed to be part of the common knowledge of both players.
- Two sequences of *output spaces* Y_t and Z_t. $y_t \in Y_t$ and $z_t \in Z_t$ are the *measurements* of player **1** and **2** at time t.
- Two *control sets* $\mathcal{U}$ and $\mathcal{V}$ and two point to set maps *admissible controls* $y_t \mapsto U(y_t) \subset \mathcal{U}$ and $z_t \mapsto V(z_t) \subset \mathcal{V}$. We shall oftentimes write U_t and V_t instead of $U(x_t)$ and $V(x_t)$ when what is meant is clear. $u_t \in U_t$ and $v_t \in V_t$ are the *controls* of player **1** and **2** respectively at time t.
- A sequence of functions *dynamics* $f_t : X_t \times U_t \times V_t \to X_{t+1}$.

$$x_{t+1} = f_t(x_t, u_t, v_t). \tag{1}$$

- Two sequences of *output functions* $h_t : X_t \to Y_t$ and $k_t : X_t \to Z_t$, defining

$$y_t = h_t(x_t), \tag{2a}$$

$$z_t = k_t(x_t), \tag{2b}$$

(Actually, h_t and k_t need only be defined for $t \geq 2$)

- A *Capture set* $C \subset X^T \times \mathbf{T}$ defining the *final time* t_1 through

$$t_1 = \begin{cases} \min\{t \mid (x_t, t) \in C\} & \text{if } capture \text{ happens,} \\ T & \text{otherwise.} \end{cases} \tag{3}$$

- A *criterion* (or *cost function*) G that player **1**, chosing the control, $u_t \in U_t$ strives to minimize, and **2**, chosing the controls $v_t \in V_t$ to maximize. G is defined via two sequences of functions: $L_t : X_t \times \mathcal{U} \times \mathcal{V} \to \mathbf{R}$ and $K_t : X_t \to \mathbf{R}$ by

$$G = \sum_{t=1}^{t_1 - 1} L_t(x_t, u_t, v_t) + K_{t_1}(x_{t_1}). \tag{4}$$

Remark. The above set up is that of dynamical systems, and is by now classical. The important point for us is that it defines functions

$$y_t = \tilde{h}_t(u^{t-1}, v^{t-1}), \tag{5a}$$

$$z_t = \tilde{k}_t(u^{t-1}, v^{t-1}). \tag{5b}$$

We shall also make use of

$$y^t = h^t(x^t) = \tilde{h}^t(u^{t-1}, v^{t-1}), \tag{6a}$$

$$z^t = k^t(x^t) = \tilde{k}^t(u^{t-1}, v^{t-1}). \tag{6b}$$

And also

$$G = \tilde{G}_{t_1}(u^{t_1-1}, v^{t_1-1}). \tag{7}$$

Let us introduce a last notation. We shall be interested in games with complete memory. By this we mean that the *information* available to the players to make up their choice of control values at each instant of time is the whole sequence of their own past controls, and their past and present measurements. We shall therefore write:

$$r^t = (u^{t-1}, y^t) \quad \text{player } \mathbf{1}\text{'s information}, \tag{8a}$$

and

$$s^t = (v^{t-1}, z^t) \quad \text{player } \mathbf{2}\text{'s information}. \tag{8b}$$

We shall also use the following definition

Definition. If the sets $\mathcal{U}$ and $\mathcal{V}$, (and for stochastic games, $\mathcal{W}$), are all finite, the game shall be called *finite*.

Notice that for a finite game, the sets X_t, Y_t, and Z_t may, with no loss of generality be restricted to finite sets.

For infinite games, we shall use the following *topological hypothesis*

Hypothesis. The sets U_t, V_t, X_t, Y_t, and Z_t are all topological spaces, the first two *compact*. The functions f_t, h_t, k_t, L_t, and K_t are all continuous, (making G in (4) a continuous function of (x^T, u^{T-1}, v^{T-1}), or $\tilde{G}$ in (7) a continuous function of (u^{T-1}, v^{T-1})).

3. Strategies

To make precise the definition of the game, we must now specify the strategy sets, and what are the quantities to be minimized or maximized.

Definition 2. We call *pure strategy* of player **1** a *non anticipatory* measurable map

$$\alpha : Y^{T-1} \to U^{T-1} : (y_1, \ldots, y_{T-1}) \mapsto (u_1, \ldots, u_{T-1}) \tag{9a}$$

and we call A the set of all such non anticipatory maps, i.e. such that, for a and b in Y^{T-1}

$$a^t = b^t \quad \Longrightarrow \quad \alpha_t(a) = \alpha_t(b).$$

Likewise, pure strategies of the second player are non anticipatory measurable maps

$$\beta \in B, \quad \beta : Z^{T-1} \to V^{T-1} : (z_1, \ldots, z_{T-1}) \mapsto (v_1, \ldots, v_{T-1}). \tag{9b}$$

Any pair of pure strategies $(\alpha, \beta) \in A \times B$ generates a well defined game history by (1), (2), and

$$u_t = \alpha_t(y_1, \ldots, y_t, \eta), \tag{10a}$$

$$v_t = \beta_t(z_1, \ldots, z_t, \zeta), \tag{10b}$$

where η and ζ are arbitrary sequences that do not affect the resulting values of u_t and v_t, by the hypothesis of nonanticipativity. As a consequence, we shall omit them in the future. There corresponds to it a well defined cost

$$G = \hat{G}(\alpha, \beta).$$

We can state the following fact:

Lemma. *Under the topological hypothesis, the sets A and B of pure strategies are compact in the topology of pointwise convergence.*

Proof. The set of *all* functions from Y^{T-1} into U^{T-1} is isomorphic to the power set

$$\left(U^{T-1}\right)^{Y^{T-1}}.$$

Since the U_t are all assumed compact, by Tychonov's theorem so is U^{T-1}, and therefore also the above power set, in the product topology, which coincides with the topology of pointwise convergence. Now, the property of non anticipativity is clearly preserved in the pointwise limit, so that the sets of pure strategies are closed subsets of compact sets in that topology. Q.E.D.

In the case where $\hat{G}$ has no saddle point over $A \times B$, it is natural to introduce mixed strategies. Assuming we have chosen a topology on A and B, we may give the following definition :

Definition. We call *mixed strategies* probability measures λ and μ over A and B respectively:

$$\lambda \in \pi(A), \tag{11a}$$

$$\mu \in \pi(B). \tag{11b}$$

We then wish to take as a new criterion

$$J(\lambda,\mu) = \mathrm{E}(G) = \int_{A\times B} \hat{G}(\alpha,\beta)\, d\lambda(\alpha)\, d\mu(\beta). \tag{12}$$

The existence of the above integral is not guaranteed a priori, since $\hat{G}$ is in general not continuous in α and β, and there is no guarantee that it be measurable. Three alternate sets of hypotheses are provided here as examples of setups where this integral exists, and that preserve the compactness of the set of allowed pure strategies. The first set is rather trivial:

Hypothesis 1. The game is finite.

Then the integral in (12) is mearly a finite sum. The game is just a matrix game, (possibly with a very large matrix !), and we are in the classical setting of Von Neuman and Morgenstern.

The second set allows us to avoid any finiteness hypothesis, but imposes rather strong restrictions on the allowed pure strategies:

Hypothesis 2. The sets U_t, V_t, Y_t, and Z_t, are metric compact spaces, the pure strategies are restricted to be Lipshitz continuous with a prescribed Lipshitz modulus.

We then have the following fact

Lemma. *Under hypothesis 2, the sets A and B are compact in the topology of uniform convergence, and the sequences u^{T-1}, v^{T-1}, and x^T depend continuously on (α, β).*

Proof. The first claim is a direct consequence of the Arzela Ascoli theorem. The second claim derives from the fact that the x_t's are then continuous functions of the strategies, as is easily seen by induction on t.

Finally, it is also possible to avoid regularity assumptions on the pure strategies, still keeping infinite control sets, by assuming finite observation sets. Notice that it follows from the standing topological hypothesis that the reachable game space is bounded, so that any quantization of the measurement with a finite mesh will produce a finite measurement set. This type of hypothesis was first proposed by Levine [6].

Hypothesis 3. The sets Y_t and Z_t are finite. The functions h_t and k_t can therefore not be continuous. Assume that the sets $h^{-1}(y_k) \cap k^{-1}(z_l)$ have non void interiors, the union of their boundaries is made of the finite union of sets on which h and k are constant, that satisfy the same hypothesis in the relative topology of this boundary, and so on recursively, the whole construction defining a finite partition of X^T.

Lemma. *Under hypothesis 3, the set $A \times B$ can be partitioned in a finite union of Borel sets (in the topology of pointwise convergence), the state trajectory x^T depending continuously on (α, β) over each of them, and thus also the control sequences u^{T-1}, v^{T-1}.*

Proof. Let, for simplicity, $A \times B = C$, $(\alpha, \beta) = \gamma$, and, for this proof, y stand for (y, z). Let also $P_k, k = 1, \ldots, K$ be the *interior* of the sets in X such that $y = const$, $P = \cup P_k$, and

$$C_t = \{\gamma \mid x_s \in P, \forall s \leq t\}.$$

It is easy to see by induction on t that C_t is open in C. As a matter of fact, let $\gamma^{(n)}$ be a sequence of strategies converging to $\gamma \in C_t$. Since y_1 is fixed, $u_1^{(n)} = \gamma_1^{(n)}(y_1) \to u_1$. By continuity of f_1, $x_2^{(n)} \to x_2$. Since x_2 is interior to $h_2^{-1}(y_2)$, for n large enough, $y_2^{(n)} = y_2$. Therefore, we can use the convergence of $\gamma_2^{(n)}$ and the continuity of f_2 to conclude that $x_3^{(n)} \to x_3$, and so on. Therefore $x_t^{(n)} \to x_t$, and for n large enough, $x_t^{(n)}$ is also in the interior of its set P_k, thus $\gamma^{(n)}$ is in C_t.

Therefore the complement D_t of C_t in C_{t-1} is also a Borel set. This is the set of strategies γ such that $x_s \in P$, for $s = 1, \dots t-1$, $x_t \in \partial P$. Now, assume again that $x_t \in Q_i$ where Q_i is the relative interior of one of the subsets of constant y in ∂P. The same type of argument will show that the subset E_l of D_t such that the next x_s, up to $s = l$, are in P is again open in D_t.

We finally have a finite number of subsets of C, depending on the subsets of the partition of X in which each of the x_t lie. All are Borel sets. And since we have shown the convergence of the $x_t^{(n)}$ in each case, $\hat{G}$ is continuous in the relative interior of all of them. This proves the lemma.

As a consequence, $\hat{G}$ is measurable, and J in (12) is well defined.

So we now have at least three cases where the following *existence hypothesis* is satisfied:

Hypothesis. For a suitable topology on A and B, these sets are compact, and the state trajectory x^T, as well as the control histories u^{T-1}, v^{T-1}, are measurable functions of the pure strategies α and β.

As a consequence of this hypothesis, we have the following two facts.

Proposition 0. *J in (12) is well defined, since G as defined in (4) is a continuous function of (x^T, u^{T-1}, v^{T-1}).*

Theorem 0. *Under the existence hypothesis, and if furthermore, $\hat{G}$ is continuous, the game has a saddle point in mixed strategies.*

Proof. See Ekeland [4], p 25. The proof makes use of Sion's theorem (see [10]) together with the vague topology on the set of measures.

The idea behind mixed strategies is that the players choose a pure strategy at random, according to the probability laws λ and μ respectively, once for all at the begining of the game and for its whole duration. The spaces of mixed strategies are very large and complicated sets, and this type of behavior may not appear very natural. We shall therefore introduce another concept of strategies. But we first need a preliminary definition.

Definition. We call *mixed control* of the first player at time t a probability law p_t over U_t, and likewise for the second player.

Let thus

$$p_t \in P_t = \pi(U_t), \tag{13a}$$

$$q_t \in Q_t = \pi(V_t). \tag{13b}$$

We now define the new class of strategies:

Definition. We call *behavioral strategy* of the first player a sequence of measurable maps

$$\varphi_t \in \Phi_t, \quad \varphi_t : U^{t-1} \times Y^t \to P_t : (u^{t-1}, y^t) \mapsto p_t = \varphi_t(u^{t-1}, y^t) = \varphi_t(r^t), \tag{14a}$$

and similarly for the second player:

$$\psi_t \in \Psi_t, \quad \psi_t : V^{t-1} \times Z^t \to Q_t : (v^{t-1}, z^t) \mapsto q_t = \psi_t(v^{t-1}, z^t) = \psi_t(s^t). \tag{14b}$$

The game is then extended by considering $\{x_t\}$, $\{u_t\}$, $\{v_t\}$, $\{y_t\}$, and $\{z_t\}$ as stochastic processes, generated by (1) and (2), u_t and v_t being stochastic variables with probability distributions p_t and q_t respectively, given by (14). Then the sequences u^{T-1} and v^{T-1} are stochastic variables, and we define the payoff of the game as

$$J = \mathrm{E}(G) = \mathrm{E}\left(\tilde{G}(u^{T-1}, v^{T-1})\right). \tag{15}$$

This is well defined, since $\tilde{G}$ is continuous, and (14) clearly defines a Radon probability over $U^{T-1} \times V^{T-1}$. (By a finite, elementary, version of the Ionescu Tulcea theorem.)

The idea behind behavioral strategies is that at each instant of time, the players choose their controls at random, according to a probability distribution function of their information.

One might believe that this new set of strategies is richer than the previous mixed strategies, since it involves many random choices instead of a single one. A very simple example will teach us that this is not so.

4. Example.

The following example is the smallest possible version of the Hunter and Rabbit game. Let $T = 4$. (i.e. the game has three time steps.) The state is $x = (y, w, z) \in \{1, 2\} \times \{0, 1, 2\}^2$ and $U = V = \{1, 2\}$. The dynamics are

$$y_{t+1} = u_t, \quad y_1 = 1,$$

$$w_{t+1} = v_t, \quad w_1 = 0,$$

$$z_{t+1} = w_t, \quad z_1 = 0.$$

Player **1** has no other information than the sequence u^{t-1} at time t. Player **2** knows the whole state in addition to the past controls. The capture set is $y_t - z_t = 0$. The payoff to player **2** is 1 if $y_{t_1} = z_{t_1}$, i.e. if *capture* has occured, and 0 otherwise. Hence, $J = \mathrm{E}(G)$ is the capture probability. Recall that we always assume that the initial state is part of the rule of the game, and is therefore common knowledge.

Player **1** actually plays open loop, and therefore chooses among $2^3 = 8$ pure strategies.

Let us look at the possible strategies of player **2**. The pure strategies are specified by the decision rules at time 1 and 2, since actions taken at time 3 have no effect on the outcome of the game. At time 1, no information is available beyond the rules of the game. The only two possibilities are either $v_1 = 1$ or $v_1 = 2$. At time 2, a measurement $y_2 = u_1$ is available, which can be equal to 1 or 2. Since v_2 can also be chosen as either 1 or 2, there are 4 possible decision rules for $\beta_2(y_2)$, i.e. $v_2 = 1\ \forall y$, that we denote by **1**, or, with the same convention, **2**, or $v_2 = y_2$, or finally $v_2 = 3 - y_2$. We thus have the following list of 8 possible pure strategies:

strategy	1	2	3	4	5	6	7	8
time								
$t = 1$	1	1	1	1	2	2	2	2
$t = 2$	**1**	**2**	y_2	$3 - y_2$	**1**	**2**	y_2	$3 - y_2$

The mixed strategies are therefore given by 8 probabilities $(\mu_1, \ldots, \mu_8)$ whose sum is one, thus seven degrees of freedom.

Let us now look at the behavioral strategies, still for player **2**. They are defined by the probability ψ_1 of playing $v_1 = 1$, (and $1 - \psi_1$ of playing $v_1 = 2$), and for time $t = 2$

the four probabilities $\psi_2(s^2)$ of playing $v_2 = 1$ according to the four possible values of $s^2 = (v_1, y_2)$. Thus, these strategies are defined by five independent probabilities.

We also see on this example that one can make a mixed strategy correspond to a unique behavioral strategy, its *behavior*, by seeing the latter as a conditional marginal probability. Here, for instance, assuming that the pure strategies are numbered according to the above table, we have

$$\psi_1 = \Pr(v_1 = 1) = \mu_1 + \mu_2 + \mu_3 + \mu_4,$$

$$\psi_2(1,1) = \Pr(v_2 = 1|v_1 = 1, y_2 = 1) = \frac{\mu_1 + \mu_3}{\mu_1 + \mu_2 + \mu_3 + \mu_4}$$

$$\psi_2(1,2) = \Pr(v_2 = 1|v_1 = 1, y_2 = 2) = \frac{\mu_1 + \mu_4}{\mu_1 + \mu_2 + \mu_3 + \mu_4}$$

Likewise, the conditioning on $v_1 = 2$ would leed to a denominator $\mu_5 + \mu_6 + \mu_7 + \mu_8$, and so on. It is easy to verify that this map is onto, but cannot be one to one. An infinity of different mixed strategies lead to the same behavioral strategy.

The aim of the next section of this paper is to show that it is enough to consider behavioral strategies, for the outcome of the game only depends on the behavior of the strategies used, as rigorously defined hereafter.

Let us remark, before we close this example, that the solution of this game is almost obvious: each player should play either 1 or 2 with probability 1/2.

5. Kuhn's theorem.

We must first make precise the relationship between a mixed strategy and its associated behavioral strategy.

Definition. We define the map *behavior* γ from the set $\pi(A)$ of mixed strategies to the set Φ of behavioral strategies in the following way:

$\varphi_t(u^{t-1}, y^t)$ is the marginal law on $\alpha_t(y^t)$ knowing $\alpha^{t-1}(y^{t-1}) = u^{t-1}$.

In the finite case, for instance, this can be made explicit in the following way. Let y^t and u^{t-1} be fixed. Let

$$A_1 = \left\{\alpha \in A | \alpha^{t-1}(y^{t-1}) = u^{t-1}\right\},$$

$$A_2 = \left\{\alpha \in A | \alpha^t(y^t) = u^{t-1} \cdot \bar{u}\right\} \subset A_1.$$

Then, $\gamma(\lambda) = \varphi$ with, if $A_1 \neq \emptyset$

$$\varphi_t[u^{t-1}, y^t](\bar{u}) = \left(\sum_{\alpha \in A_2} \lambda(\alpha)\right) \left(\sum_{\alpha \in A_1} \lambda(\alpha)\right)^{-1}.$$

If $A_1 = \emptyset$, $\varphi_t[u^{t-1}, y^t]$ may be arbitrarily specified, and we shall always assume that $\varphi = \gamma(\alpha)$ has been so extended to all values of its arguments.

We shall also call γ the behavior map of the second player, from $\pi(B)$ into Ψ.

We now state the main theorem, which is an extension of Kuhn [5]:

Theorem 1. *There exists a function $\bar{J}$ from $\Phi \times \Psi$ into $\mathbf{R}$ such that,*

$$\forall (\lambda, \mu) \in \pi(A) \times \pi(B), \quad J(\lambda, \mu) = \bar{J}\big(\gamma(\lambda), \gamma(\mu)\big)$$

i.e., the criterion $J(\lambda, \mu)$ of the game only depends on the behaviors $\gamma(\lambda)$ and $\gamma(\mu)$.

Proof. Under the existence hypothesis, (u^{T-1}, v^{T-1}) is a measurable function of (α, β). Thus a pair of mixed strategies (λ, μ) generates a probability distribution Π^{T-1} over the set $U^{T-1} \times V^{T-1}$, and one has

$$J(\lambda, \mu) = \int_{U^{T-1} \times V^{T-1}} \tilde{G}(u^{T-1}, v^{T-1})\, d\Pi^{T-1}(u^{T-1}, v^{T-1}),$$

or, in the discrete case

$$J(\lambda, \mu) = \sum_{u^{T-1}, v^{T-1}} \tilde{G}(u^{T-1}, v^{T-1}) \Pi^{T-1}(u^{T-1}, v^{T-1}).$$

Therefore, J only depends on the probability law Π^{T-1}. Notice that, for fixed $\bar{u}^{T-1}, \bar{v}^{T-1}$, the event $(u^{T-1}, v^{T-1}) = (\bar{u}^{T-1}, \bar{v}^{T-1})$ can be writen

$$(u^{T-2}, v^{T-2}) = (\bar{u}^{T-2}, \bar{v}^{T-2}) \quad \text{and} \quad (u_{T-1}, v_{T-1}) = (\bar{u}_{T-1}, \bar{v}_{T-1}).$$

Furthermore, it follows from the definition (12) that the strategies λ and μ are chosen independently by the two players, so that, for $u^{T-2} = \bar{u}^{T-2}$ and $v^{T-2} = \bar{v}^{T-2}$ fixed, the random variables u_{T-1} and v_{T-1} are independent.

Let therefore λ and μ be fixed, $\varphi = \gamma(\lambda)$ and $\psi = \gamma(\mu)$ their behaviors,

$$\tilde{h}^{T-1}(\bar{u}^{T-2}, \bar{v}^{T-2}) = \bar{y}^{T-1}, \qquad \tilde{k}^{T-1}(\bar{u}^{T-2}, \bar{v}^{T-2}) = \bar{z}^{T-1},$$

$$\varphi_{T-1}(\bar{u}^{T-2}, \bar{y}^{T-1}) = \bar{p}_{T-1}, \qquad \psi_{T-1}(\bar{v}^{T-2}, \bar{z}^{T-1}) = \bar{q}_{T-1}.$$

One has the equality

$$d\Pi^{T-1}(\bar{u}^{T-1}, \bar{v}^{T-1}) = d\Pi^{T-2}(\bar{u}^{T-2}, \bar{v}^{T-2})\, d\bar{p}_{T-1}(\bar{u}^{T-1})\, d\bar{q}_{T-1}(\bar{v}^{T-1}).$$

(See, for instance, the second part of proposition V.1.1 in [7].) Or in the discrete case,

$$\Pi^{T-1}(\bar{u}^{T-1}, \bar{v}^{T-1}) = \Pi^{T-2}(\bar{u}^{T-2}, \bar{v}^{T-2})\bar{p}_{T-1}(\bar{u}_{T-1})\bar{q}_{T-1}(\bar{v}_{T-1}),$$

by Bayes rule.

One then iterates this process to write Π^{T-2} in terms of Π^{T-3} and of φ_{T-2} and ψ_{T-2}, and so on. This proves the theorem.

This allows us to carry over results obtained by topological means on games in normal form, such as theorem 0 above, to dynamic games in behavioral strategies. We thus have, for instance, the obvious following fact :

Corollary 1. *Under the existence hypothesis, and if furthermore $\hat{G}$ is continuous, a dynamic game with perfect memory admits a saddle point in behavioral strategies over the sets $\Phi = \gamma(A)$, $\Psi = \gamma(B)$.*

The difficult task, however, is to characterize the sets $\gamma(A)$ and $\gamma(B)$. This is trivial for finite games, where all behavioral strategies will be included. We shall see that the question is easy for non degenerate stochastic games. For deterministic continuous games, let us look, for instance, at the set up defined by hypothesis 2 of section 3. A is the set of causal functions from Y^{T-1} into the compact U^{T-1}, Lipshitz continuous with Lipshitz modulus ℓ.

Proposition. *The behavioral strategies of $\gamma(A)$ have the following property : let $\tau : U_t \to$ **R** be a function with Lipshitz modulus m, then*

$$E^{r^t}\tau = \int_{U_t} \tau(u) d\varphi_t(r^t)(u)$$

is a Lipshitz continuous function of y^t with Lipshitz modulus ℓm.

The proof is elementary. This includes the obvious fact that if the pure strategies are restricted to open loop controls, so are the corresponding behavioral strategies, since this is the case $\ell = 0$. We conjecture that this is a complete characterization of the set $\gamma(A)$, but this is not sure. We did not investigate this point in detail, since we do not need it in the sequel.

6. Stochastic games.

Up to here, we have dealt with games with deterministic dynamics and measurements. The information available to the players is incomplete, but not noisy. We now extend all the previous theory to the case of stochastic games.

A stochastic multistage game is defined as in section 2, except that (1) and (2) involve an extra stochastic process $\{w_t\}$, with values in sets $\mathcal{W}_t$, usually multidimensional. It will always be assumed to be white, with probability distributions W_t known to both players. Thus (1) and (2) are replaced respectively by

$$x_{t+1} = f_t(x_t, u_t, v_t, w_t), \tag{17}$$

$$y_t = h_t(x_t, w_{t-1}), \tag{18a}$$

$$z_t = k_t(x_t, w_{t-1}). \tag{18b}$$

The one time step shift in the argument w of h_t and k_t makes sense since x_1 being known to both players, the first relevant measurement is y_2, z_2. Moreover, with that convention, $\tilde{h}_t$ and $\tilde{k}_t$, as well as r^t and s^t, all depend on u^{t-1}, v^{t-1}, and w^{t-1}, greatly simplifying the sequel. Of course, the hypothesis that w is a white process makes this much more than a pure notational trick. One case where this is not restrictive, and equivalent to the more classical approach, is when w enters in the dynamics and the measurements through distinct independent components.

In this setup, the sequences x, u, v, become stochastic processes, even with pure strategies, and (4) is replaced with a mathematical expectation:

$$G = \mathrm{E}\left[\sum_t L_t(x_t, u_t, v_t) + K_{t_1}(x_{t_1})\right]. \tag{19}$$

For simplicity, we shall assume that the sets U_t and V_t do not depend on the current state.

The definitions of the strategies are unchanged. Notice that the three sets of hypotheses that have been proposed as alternate setups that insure satisfaction of the existence hypothesis still stand here. Hypothesis 1 leads to a standard matrix game the entries of which are the expected cost incurred. Hypothesis 2 needs no modification either. Convergence of the trajectories is insured for each value of w^{T-1}, and thus the expected values converge. Hypothesis 3 must be extended to hold on the $X \times W$ space. It is just a bit cumbersome to state and deal with. A simple case is when $w = (\xi, \eta)$ is made of two independent components, ξ entering in the dynamics, and $h_t(x, w) = \hat{h}_t(x + \eta)$, and likewise for k_t. However, everything becomes much simpler if we assume that for all x_t, u_t, v_t, the transition probability induced by f_t is absolutely continuous with respect to the Lebesgue measure. (The so called *non degenerate* case). Then, the existence hypothesis and continuity of $\hat{G}$ for pointwise convergence of the strategies is just a consequence of the Ionescu Tulcea theorem.

Theorem 1 is still valid in this context, its proof is slightly modified as follows.

Proof. (Theorem 1 for stochastic games.) As previously, let Π^{T-1} be the distribution law of $(u^{T-1}, v^{T-1}, w^{T-1})$ generated by a given pair of mixed strategies (λ, μ). One has

$$J(\lambda, \mu) = \int_{U^{T-1} \times V^{T-1} \times W^{T-1}} \tilde{G}(u^{T-1}, v^{T-1}, w^{T-1})\, d\Pi^{T-1}(u^{T-1}, v^{T-1}, w^{T-1}).$$

Moreover, as previously

$$d\Pi^{T-1}(u^{T-1}, v^{T-1}, w^{T-1}) = d\Pi^{T-2}(u^{T-2}, v^{T-2}, w^{T-2})\, d\Pi^c(u_{T-1}, v_{T-1}, w_{T-1}),$$

where Π^c is the conditional law of $(u_{T-1}, v_{T-1}, w_{T-1})$ knowing $(u^{T-2}, v^{T-2}, w^{T-2})$. Using the notation

$$p_{T-1} = \varphi_{T-1}\big(u^{T-2}, \tilde{h}^{T-1}(u^{T-2}, v^{T-2}, w^{T-2})\big)$$
$$q_{T-1} = \psi_{T-1}\big(v^{T-2}, \tilde{k}^{T-1}(u^{T-2}, v^{T-2}, w^{T-2})\big)$$

for the behaviors, which are, by definition, the conditional laws of u_{T-1}, v_{T-1} for a given r^{T-2} and s^{T-2}, and remembering that w_{T-1} is by hypothesis independent of the past, we derive, still as previously

$$d\Pi^{T-1}(u^{T-1}, v^{T-1}, w^{T-1}) = \\ d\Pi^{T-2}(u^{T-2}, v^{T-2}, w^{T-2})\, dp_{T-1}(u_{T-1})\, dq_{T-1}(v_{T-1})\, dR_{T-1}(w_{T-1}).$$

We may anew iterate the process, to conclude the proof.

7. Semicomplete information in finite games.

One of the superiorities of behavioral strategies over mixed strategies is that due to their sequential nature, they lend themselves to dynamic programming. We shall exploit this fact in the case of *finite games* with *semicomplete information.*

By this we mean that one of the players, say **2**, has full knowledge of the relevant variables of the game at each instant of time. More specifically, we shall assume that at time t, **2** knows x_t, y_t, and also u_{t-1}. This is not beyond the scope of our previous theory, since we may always augment the state with the variables η_t and ξ_t, with $\eta_{t+1} = w_t$, and $\xi_{t+1} = u_t$, so that knowing the full state also yields $y_t = h_t(x_t, \eta_t)$, and $u_{t-1} = \xi_t$. Of course the players are still assumed to have perfect memory, so that they also remember past values of their measurements.

Let ν_t be a probability distribution over $X^t \times V^{t-1}$. We think of ν_t as being player **1**'s conditional probability on (x^t, v^{t-1}). Assume u^t given, as well as a behavioral strategy ψ^t of the second player. Then, using the dynamics (17), we can propagate ν_t into a probability $\bar\nu_{t+1}$ over $X^{t+1} \times V^t$, in the following way. Let $a \in X^{t+1}$ and $b \in V^t$,

$$\bar\nu_{t+1}(a,b) = \nu_t(a^t, b^{t-1})\psi_t[a^t, u^{t-1}, b^{t-1}](b_t) \sum_w \delta\big(a_{t+1} - f_t(a_t, u_t, b_t, w)\big) W_t(w).$$

Then, when the measurement y_{t+1} comes in, one may compute the new conditional probability ν_{t+1} on (x^{t+1}, v^t). We shall give explicit formulas only for the simple example of the next section. Anyhow, this defines a filter of the form

$$\nu_{t+1} = F_t(\nu_t, u^t, y_{t+1}, \psi_t), \tag{20}$$

and also a function

$$\nu_t = N_t(u^{t-1}, y^t, \psi^{t-1}) = N_t(r^t, \psi^{t-1}). \tag{21}$$

By summation over the component subspaces, we can project ν_t on X^t alone, let ρ^t be that law on x^t. We can further project on the component X_t, yielding a law

$$\rho_t = R_t(r^t, \psi^{t-1}). \tag{22}$$

We can now state a theorem of dynamic programming.

Theorem 2. *Let a stochastic dynamic game be given by (17) to (19), and (φ^*, ψ^*) be a saddle point in behavioral strategies (which exists according to corollary 1). There exists a sequence of functions V_t from $X^t \times U^{t-1} \times V^{t-1}$ to $\mathbf{R}$ such that for all (x^t, u^{t-1}, v^{t-1}) reached with a non zero probability while playing according to (φ^*, ψ^*), one has*

$$(23) \qquad V_t(x^t, u^{t-1}, v^{t-1}) = \max_{v \in V_t} \sum_{w \in \mathcal{W}^t} \sum_{u \in U_t} \left[V_{t+1}\big(x^t \cdot f_t(x_t, u, v, w_t), u^{t-1} \cdot u, v^{t-1} \cdot v\big) + L_t(x_t, u, v) \right] p_t^*(u) W^t(w)$$
$$= \sum_{v \in V_t} \sum_{w \in \mathcal{W}^t} \sum_{u \in U_t} \left[V_{t+1} + L_t \right] p_t^*(u) W^t(w) q_t^*(v).$$

(the arguments in V_{t+1} and L_t in the third term are of course the same as in the second) and

$$(24) \qquad \sum_{x^t v^{t-1}} V_t(x^t, u^{t-1}, v^{t-1}) \nu_t^*(x^t, v^{t-1}) =$$
$$\min_{u \in U_t} \sum_{w \in \mathcal{W}^t} \sum_{v \in V_t} \sum_{x^t v^{t-1}} \Big[V_{t+1}\big(x^t \cdot f_t(x_t, u, v, w_t), u^{t-1} \cdot u, v^{t-1} \cdot v\big) + L_t(x_t, u, v) \Big] \nu_t^*(x^t, v^{t-1}) q_t^*(v) W^t(w)$$
$$= \sum_{u \in U_t} \sum_{w \in \mathcal{W}^t} \sum_{v \in V_t} \sum_{x^t v^{t-1}} \left[V_{t+1} + L_t \right] \nu_t^*(x^t, v^{t-1}) q_t^*(v) W^t(w) p_t^*(u),$$

where p_t^* and q_t^* stand for $\varphi_t^*[r^t]$ and $\psi_t^*[x^t, u^{t-1}, v^{t-1}]$ respectively, ν_t^* for $N_t(r^t, \psi^*)$, and y^t in r^t for $\tilde{h}^t(x^t, w^{t-1})$, and

$$(25) \qquad \forall (x, \tau) : (x_\tau, \tau) \in C,\ \forall (u, v),\ V_\tau(x^\tau, u^{\tau-1}, v^{\tau-1}) = K_\tau(x_\tau),$$

Conversely, if a sequence of functions V_t together with a pair of behavioral strategies φ^*, ψ^* satisfy equations (23) to (25), these strategies constitute a saddle point of the game, and the value of the game is $V_1(x_1)$.

Proof. The sufficiency part of the claim is a direct adaptation of the theorem in [2] and shall not be repeated in detail. The proof amounts to using (23)(25) to show that, for an arbitrary sequence v^T, one has $G(\varphi^*, v^T) \le V_1(x_1)$, and using (24)(25) to derive the other inequality of the saddle point. This second part uses the fact that when calculating

$G(u^T, \psi^*)$, since **2** plays ψ^*, $N_t(r^t, \psi^*)$ is actually a conditional probability. This fact is not true, but not needed either, in the first calculation.

Let us now look at necessity.

Notice first that the second equalities in (23) and (24) amount to the fact that q^* has its support contained in the set of v's that provide the maximum in (23), and likewise for p^* with the minimum in (24).

Let (x^t, u^{t-1}, v^{t-1}) be a state of the game reached with a nonzero probability while playing (φ^*, ψ^*). For each w^{t-1}, there corresponds a y^{t-1}, and we can describe the game history from there on under the strategies φ^*, ψ^*. Let

$$V_t(x^t, u^{t-1}, v^{t-1}) = \mathbf{E}\left[\sum_{i=t}^{t_1-1} L_i(x_i, u_i, v_i) + K(x_{t_1})\right].$$

It is clear that V_t thus defined satisfies the second equality of (23), and thus of (24) by summing both sides of (23). Assume now that there exists $\hat{v} \in V_t$ that gives to the second term of (23) a value larger than V_t. Consider the strategy $\hat{\psi}$ that coincides with ψ^* everywhere, except at (x^t, u^{t-1}, v^{t-1}) where it is a dirac distribution at $\hat{v}$. Let, for simplicity, $L_{t_1} = K(x_{t_1})$, and write

$$\mathbf{E}\tilde{G}(u^{t_1-1}, v^{t_1-1}) = \mathbf{E}\sum_{i=1}^{t-1} L_i + \mathbf{E}\sum_{i=t}^{t_1} L_i.$$

Using the fact that the information algebra is increasing, write the second expectation above as

$$\mathbf{E}\sum_{i=t}^{t_1} L_i = \mathbf{E}\left[\mathbf{E}\left(\sum_{i=t}^{t_1} L_i \;\middle|\; x^t, u^{t-1}, v^{t-1}\right)\right].$$

The inner expectation is larger for $(\varphi^*, \hat{\psi})$ than for (φ^*, ψ^*) by hypothesis. From all other possible states at time t, this expectation coincides for the two strategy pairs, since ψ^* and $\hat{\psi}$ coincide. However, this particular state is reached by hypothesis with a nonzero probability. Therefore the outer expectation is larger with the strategy $\hat{\psi}$, which contradicts the definition of the saddle point.

We do likewise with φ and (24), noticing as in the sufficiency proof that, when player **2** does play ψ^*, the quantity minimized in (24) actually is the expectation of V_t for player **1**. We thus contradict the other inequality of the saddle point. The theorem is proved.

Notice that as in [2], this may be viewed as a fixed point theorem: multiply equation (23) by $\nu_t^*(x^t, v^{t-1})$ on both sides, and sum over all (x^t, v^{t-1}). Then (23)(24) together express the fact that φ_t^*, ψ_t^* provide a saddle point (over a product of simplices for ψ_t^*) of the matrix made up of the blocks $\big[V_{t+1} + L_t\big]$ weighted by $\nu_t^* = N_t(r^t, \psi^{*t-1})$. So the problem is to find a ψ^* that gives rise to a ν^* for which this ψ^* is an argument of this sequence of saddle points.

In itself, this theorem is of little use. We shall see in the next section a case where it simplifies to the point where it can be used to compute the saddle point of the game. At this time, we show a theoretical consequence of interest.

Corollary 2. *Let $\rho_t^* = R_t(r^t, \psi^*)$. The game admits a saddle point in behavioral strategies of the form $\varphi_t^*[r^t] = \hat{\varphi}_t[\rho_t^*]$, $\psi_t^*[x^t, u^{t-1}, v^{t-1}] = \hat{\psi}_t[x_t, \rho_t^*]$. We can define a filter $\rho_{t+1} = g_t(\rho_t, u_t, y_{t+1}, \psi_t)$, and there exists a sequence of functions $V_t(x_t, \rho_t)$ such that for all (x_t, ρ_t) that are reached with a non zero probability while playing optimally,*

$$\text{(26)} \qquad V_t(x_t, \rho_t) = \max_{v \in V_t} \sum_{w \in \mathcal{W}_t} \sum_{u \in U_t} \left[V_{t+1}\left(f_t(x_t, u, v, w), g_t(\rho_t, u, y_{t+1}, \hat{q}_t)\right) + L_t(x_t, u, v)\right] \hat{p}_t(u) W_t(w)$$
$$= \sum_{v \in V_t} \sum_{w \in \mathcal{W}_t} \sum_{u \in U_t} \left[V_{t+1} + L_t\right] \hat{p}_t(u) W_t(w) \hat{q}_t(v)$$

and

$$\text{(27)} \qquad \sum_{x_t \in X_t} V_t(x_t, \rho_t)\rho_t(x) = \min_{u \in U_t} \sum_{x_t} \sum_{w \in \mathcal{W}_t} \sum_{v \in V_t} \left[V_{t+1}\left(f_t(x_t, u, v, w), g_t(\rho_t, u, y_{t+1}, \hat{q}_t)\right) + L_t(x_t, u, v)\right] \hat{q}_t(v) W_t(w) \rho_t(x_t)$$
$$= \sum_{u \in U_t} \sum_{x_t} \sum_{w \in \mathcal{W}_t} \sum_{v \in V_t} \left[V_{t+1} + L_t\right] \hat{q}_t(v) W_t(w) \rho_t(x_t) \hat{p}(u).$$

where $\hat{p}_t$ stands for $\hat{\varphi}_t[\rho_t]$, $\hat{q}_t$ for $\hat{\psi}_t[x_t, \rho_t]$, and y_{t+1} for $h_{t+1}\big(f_t(x_t, u, v, w), w\big)$, and

$$\text{(28)} \qquad \forall (x_\tau, \tau) \in C,\ \forall \rho,\ V_\tau(x_\tau, \rho) = K_\tau(x_\tau).$$

Conversely, if a sequence of functions $\hat{\varphi}_t$, $\hat{\psi}_t$, and V_t satisfy these equations, they provide a saddle point of the game.

Proof. The fact that with strategies of this form, there exists such a filter, and the sufficiency part of the proof is exactly the main theorem of [2]. Let us look at the necessity.

Notice that, for each value of (x^t, v^{t-1}) in $X^t \times V^{t-1}$, we may multiply all terms of equation (23) by $\nu_t^*(x^t, v^{t-1})$, which is nonnegative, and sum over all such terms still preserving the inequalities. Conversely, writing the resulting summed inequality (implicit in the max operation) is *equivalent* to the separate inequalities, provided it is specified that q^* is allowed to depend on (x^t, v^{t-1}). (Therefore the combined maximizing variable ranges over a product of simplices).

Therefore, the mean value $\bar{V}(r^t) = \sum V_t \nu_t^*$ appears as the saddle point of the kernel $\sum_w \sum_{(x^t,v^{t-1})} [V_{t+1} + L_t] \nu_t^*(x^t, v^{t-1}) W^t(w)$, over a simplex for the minimizing variable, and a product of simplices for the maximizing variable. Now look at this definition for $t = T-1$. Then, $V_{t+1} = K_T(x_T)$, and the kernel depends on past values of the various variables only through ν_{T-1}^*. Moreover, since the variables x^{T-2}, v^{T-2}, and w^{T-2} do not appear in V_T and L_{T-1}, we may first sum over these variables, ending up with the kernel $\sum_w \sum_x [V_T + L_{T-1}] \rho_{T-1}(x) W_{T-1}(w)$. Therefore, the value of the saddle point depends only on ρ_{T-1}. And using the sufficiency argument, we can replace φ_{T-1}^* and ψ_{T-1}^* by strategies of the form proposed for $\hat{\varphi}$ and $\hat{\psi}$. So V_{T-1} also depends only on x_{T-1} and ρ_{T-1}.

Finally, using the propagation of ρ_t as stated in the theorem, we can iterate this process for time $T-2$, and so on down to 1. This proves the theorem.

This theorem shows that (x_t, ρ_t) constitutes a *sufficient statistic* of constant dimension for the decision problem at hand. The dynamic programming algorithm that one would like to derive from this theory remains quite cumbersome for two reasons. On the one hand, one must work in a space with a *continuous* component, while the game is discrete, and even finite. On the other hand, as was pointed out in [2], each step involves the solution of a difficult fixed point problem, since $(\hat{\varphi}_t, \hat{\psi}_t)$ must be the saddle point of a kernel that itself depends on $\hat{\psi}_t$ through its appearance in g_t. The above theory proves that this fixed point exists, a result which could not be obtained through the classical topoligical techniques, since the dependance of the kernel on $\hat{\psi}$ is not continuous. But computing it may remain a formidable task.

So we turn now to an example where one may prefer to stick with the full behavioral strategies.

8. Rabbit and Hunter game.

This game is an extension of the example of section 4. A hunter tries to shoot a rabbit that moves in a finite space made of N positions, assumed for this example to lie on a straight line. The game is specified by six integers:

- T, the horizon of the game. As usual, $\mathbf{T} = \{1, \ldots, T\}$.
- N, the number of possible positions of rabbit. The game space is therefore $\mathbf{N} = \{1, \ldots, N\}$
- a, the amplitude of rabbit's jumps, (or a^-, a^+, the left and right amplitudes).
- b, the number of bullets available to hunter,
- c, the capture radius (or lethal radius) of a bullet,
- d, the delay or time taken by the bullets to fly from hunter to rabbit.

The state x_t s composed of the following scalar variables:

- $y_t \in \mathbf{N}$, the position of rabbit,
- $w_t^k \in \mathbf{N} \cup \{0\}$, $k = 1, \ldots, d-1$, the position the bullet shot k time steps earlier is flying to,
- $z_t \in \mathbf{N} \cup \{0\}$, the position where the bullet shot d time steps earlier is arriving. (It will be convenient to use this notation rather than w^d).
- κ_t, the counter of expended bullets.

The players controls are

- $u_t \in U_t = [y_t - a, y_t + a] \cap \mathbf{N}$, (or $U^t = [y_t - a^-, y_t + a^+] \cap \mathbf{N}$), the next position of rabbit,
- $v_t \in V_t$, the position hunter aims at. $v_t = 0$ means that he does not shoot, since 0 is not a possible position of rabbit. To take into account the budget constraint, we set

$$V_t = \begin{cases} \mathbf{N} \cup \{0\} & \text{if } \kappa_t < b, \\ \{0\} & \text{if } \kappa_t = b. \end{cases}$$

The dynamics are

$$y_{t+1} = u_t, \qquad y_1 \text{ given},$$

$$
\begin{aligned}
w^1_{t+1} &= v_t, \qquad w^1_1 = 0,\\
w^{k+1}_{t+1} &= w^k_t \qquad w^k_1 = 0,\\
z_{t+1} &= w^{d-1}_t \quad z_1 = 0,\\
\kappa_{t+1} &= \kappa_t + 1 - \delta(v_t), \quad \kappa_1 = 0.
\end{aligned}
$$

Having included the budget constraint into V_t, we may take for the capture set $\{|z_t - y_t| \leq c\}$, pretending that the games goes on, even if hunter cannot shoot anymore.

In fact, to simplify the calculations below, we shall from now on assume that

$$b \leq N,$$

so that hunter shoots at all time steps, and we may ignore κ.

We shall need the following notations:

$$C(w) = \{u \mid |u - w^{d-1}| \leq c\}$$

and

$$\chi_w(u) = \begin{cases} 1 & \text{if } u \in C(w), \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, with a slight abuse of notations, we shall write either $u \in C(w)$ or $w \in C(u)$, meaning $(u, w) \in C$.

Introduce finally the following shift operator operating on the vector w:

$$\sigma(v)\cdot w = \begin{pmatrix} v \\ w^1 \\ \vdots \\ w^{d-2} \end{pmatrix}$$

In the sequel, let

$$
\begin{aligned}
p^*_t &= \varphi^*_t[y^t],\\
q^*_t &= \psi^*_t[x^t].
\end{aligned}
$$

The filter must take into account that the rabbit computes ρ_{t+1} only if it is alive, which gives extra information on the past. We get (see [3])

(29) $$\rho_{t+1}(w) = \sum_{\omega \notin C(y_{t+1})} \delta\big(w - \sigma(w^1)\cdot\omega\big) q^*_t(w^1)\rho_t(\omega)\Big[\sum_{\omega \notin C(y_{t+1})} \rho_t(\omega)\Big]^{-1}.$$

The dynamic programming equations (23)and (25) now read

$$V_t(y^t, w^t, z^t) = \max_{v \in V_t} \sum_{u \in U_t} V_{t+1}(y^t \cdot u, w^t \cdot \sigma(v) \cdot w, z^t \cdot w^{d-1}) p_t^*(u)$$

if $y_t \notin C(z_t)$, i.e. $(y_t, z_t) \notin C$,

$$V_t(y^t, w^t, z^t) = 1 \qquad \text{if } (y_t, z_t) \in C.$$

We may simplify this expression, and more importantly reduce the size of the space to be scanned, in the following manner. Introduce a function $W_t(y^t, w_t)$ which will be related to the function V_t according to the following definition:

$$V_t(y^t, w^t, z^t) = \begin{cases} W_t(y^t, w_t) & \text{if } (y_t, z_t) \notin C, \\ 1 & \text{if } (y_t, z_t) \in C. \end{cases} \tag{30}$$

This function satisfies the following dynamic programming equations, that show that it only depends on the indicated variables, exactly in the same way as we proved corollary 2 above:

$$W_t(y^t, w_t) = \max_{v \in V_t} \sum_{u \in U_t} \Big[\chi_{w_t}(u) + \big(1 - \chi_{w_t}(u)\big) W_{t+1}(y^t \cdot u, \sigma(v) \cdot w_t) \Big] p_t^*(u),$$

or equivalently

$$W_t(y^t, w_t) = \max_{v \in V_t} \Big[\sum_{u \in C(w_t)} p_t^*(u) + \sum_{u \notin C(w_t)} W_{t+1}(y^t \cdot u, \sigma(v) \cdot w_t) p_t^*(u) \Big]. \tag{31}$$

The second dynamic programming equation, equation (24), becomes now

$$\sum_w W_t(y^t, w) \rho_t(w) = \min_{u \in U_t} \sum_w \Big[\chi_w(u) + \big(1 - \chi_w(u)\big) \sum_v W_{t+1}(y^t u, \sigma(v) w) q_t^*(v) \Big] \rho_t(w). \tag{32}$$

Equations (31) and (32) may be used as the basis for a numerical algorithm, provided that the horizon T be short enough, (and a and N small enough) so that the state space remain of a tractable size. The algorithm again involves a fixed point search: for a given set of ρ_t's at each point of the space, compute $\hat{\varphi}$ and $\hat{\psi}$ rearwards in time, then solve for the fixed point $N_t(y^t, \hat{\psi}) = \rho_t$. This can be done for instance via a successive approximation scheme, using subrelaxation as necessary. No convergence proof is available at this time, however.

Notice finally that as in [3], the mean value

$$\sum_{w} W_t(y^t, w)\rho_t(w) = \bar{W}_t(y^t)$$

can be computed in a faster way, avoiding the separate computations according to the values of w. Taking ρ_{t+1} in (29), and for a fixed $y_{t+1} = u$, we have

$$\sum_{\omega} W_{t+1}(y^t \cdot u, \omega)\rho_{t+1}(\omega) =$$

$$\left[1 - \rho_t\big(C_t(u)\big)\right]^{-1} \sum_{\omega} \sum_{w \notin C(u)} \sum_{v} W_{t+1}(y^t \cdot u, \omega)\delta\big(\omega - \sigma(v) \cdot w\big) q_t^*(v)\rho_t(w).$$

Take the summation in ω inside the other two, and use the fact that then, the δ selects the only value $\omega = \sigma(v) \cdot w$, to get

$$\bar{W}_{t+1}(y^t \cdot u) = \left[1 - \rho_t\big(C_t(u)\big)\right]^{-1} \sum_{w \notin C(u)} \sum_{v} W_{t+1}(y^t \cdot u, \sigma(v) \cdot w) q_t^*(v)\rho_t(w),$$

or equivalently

$$\left[1 - \rho_t\big(C_t(u)\big)\right] \bar{W}_{t+1}(y^t \cdot u) = \sum_{w} \big(1 - \chi_w(u)\big) \sum_{v} W_{t+1}(y^t \cdot u, \sigma(v) \cdot w) q_t^*(v)\rho_t(w).$$

We recognize the right hand side here as being the second term in the right hand side of (32) above. Substituting into it we get a recurrent equation for $\bar{W}_t$:

$$\bar{W}_t(y^t) = \min_{u} \Big[\rho_t\big(C(u)\big) + \left[1 - \rho_t\big(C_t(u)\big)\right] \bar{W}_{t+1}(y^t \cdot u)\Big].$$

Aknowledgement.

This work ows much to fruitful discussions with Tamer Başar, of the University of Illinois, then on leave at INRIA Sophia Antipolis, who among other things, suggested the example of section 4 which led to the understanding of the nature of mixed vs behavioral strategies in dynamic games.

Bibliography.

[1] R.J. Aumann, *Mixed and behavior strategies in infinite extensive games*, Advances in Game Theory, Annals of Mathematical Studies **52**, 1964, pp 627–650.

[2] P. Bernhard and A. L. Colomb, *Saddle point condition for a class of stochastic dynamical games with imperfect information*, IEEE trans. on Automatic Control, **AC 10**-23, pp 98–101, 1987.

[3] P. Bernhard, *Computation of equilibrium points of delayed partial information games*, IEEE Conference on Decision and Control, Los Angeles, 1987.

[4] I. Ekeland, *La théorie des jeux et ses applications a l'économie mathématique*, P.U.F., Paris, 1974.

[5] H. W. Kuhn, *Extensive Games and the Problem of Information*, in Annals of Mathematical Studies **28**, Princeton, pp 193–216, 1953.

[6] J. Lévine, *Incomplete Information and Optimality, a Dynamic Programming Approach to Necessary and Sufficient Conditions*, internal report A/93, Centre d'Automatique et Informatique of Ecole Nationale Superieure des Mines de Paris, Fontainebleau, France, 1980.

[7] J. Neveu, *Bases mathématiques du calcul des probabilités*, Masson, Paris, 1964.

[8] A. S. Nowak, *Existence of Optimal Strategies in Zero Sum Nonstationary Stochastic Games with Lack of Information on Both Sides*, SIAM Jal on Control and Optimization, vol. **27**, pp 289–295, 1989.

[9] M. Schäl, *Stochastic Nonstationary Two Person Zero Sum Games*, Z. Angew. Math. Mech., vol. **61**, pp 352–353, 1981.

[10] M. Sion, *On General Minimax Theorems*, Pacific Journal of Mathematics, vol. **8**, pp 171–176, 1958.

[11] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behaviour*, 2nd ed., Princeton, 1947.