Comput. Math. Applic. Vol. 13, No. 1-3, pp. 205-225, 1987
Printed in Great Britain.

0097-4943/87 $3.00 + 0.00

RABBIT AND HUNTER GAME: TWO DISCRETE STOCHASTIC FORMULATIONS

P. BERNHARD, A.-L. COLOMB and G. P. PAPAVASSILOPOULOS†
INRIA, Sophia-Antipolis, Route des Lucioles, 06560 Valbonne, France

Abstract—We study stationary and non-stationary versions of the same game with different information structures. In a discrete set up, we find algorithms to calculate value and saddle-point.

1. INTRODUCTION

This paper deals with different versions of the same basic game. The main difference comes with the information structure. Here, the information available to each player is, apparently, the same in all three versions of the game, but in the last version, this same piece of information is no longer the complete state of the game, because a time delay has been added. The first and second versions differ in that one is stationary, with expected capture time as payoff, the second finite time, with probability of capture before game end as the payoff. This same difference is found in the treatment of the Princess and Monster on the circle game by Foreman [1]. His derivation in [1] for the stationary game relies on the hypothesis that a finite value exists for the game. Here, in a discrete set up, we are able to show the existence of a value and saddle-point.

2. THE GENERAL SET UP

2.1. Dynamics

A rabbit R jumps back and forth along a finite wall, in a discrete world. It can therefore be in a finite number, N, of locations and is allowed to jump at each instant of time, of a limited jump size l. (We shall, for simplicity, cover mainly the cases where the jump is limited to one unit or unlimited.)

Let $x_t \in I_N = \{1, \ldots, N\}$ be the position of R. Let $u_t \in U_{ad}(x_t)$, where $x + U_{ad}(x) \subset I_N$, be its jump at time t, then the dynamics of the rabbit are simply

$$x_{t+1} = x_t + u_t. \tag{1}$$

A hunter H watches the rabbit and is trying to shoot it. We shall assume he has an arbitrarily large number of shots at its disposal. (Changing this to a given, finite number would only make the computations heavier by introducing an extra state variable, except if that number were one.)

Let $v_t \in I_N$ be the position at which H aims at time t.

In Sections 3 and 4, we shall assume that the bullet reaches the wall in one step of time. That is, a bullet hits the wall at z_t at time t, with

$$z_{t+1} = v_t. \tag{2}$$

Capture is defined by

$$t_1 = \inf\{t; x_t = z_t\}. \tag{3}$$

In the fifth section, we assume that the bullet takes several time steps to reach the wall. In practice, we shall only detail the case with two time steps. It is clear how the method we shall use

†Present address: Department of Electrical Engineering Systems, University of Southern California, Los Angeles, CA 90089-0781, U.S.A.; on leave to INRIA in January and February 1986. The work of this author was supported in part by the U.S. Air Force Office of Scientific Research under Grant AFOSR-85-0254 and by the National Science Foundation under Grant NSF-INT-8504097.

generalizes for more time steps. Using the same definition for z_t and capture, (2) is now replaced by

$$\begin{aligned} y_{t+1} &= v_t, \\ z_{t+1} &= y_t. \end{aligned} \tag{4}$$

2.2. Strategies

Although we have not yet completely described the payoff, there is no need to specialize in animal psychology to guess that the rabbit R will strive to survive as long as possible, while the hunter H will attempt to get his lunch ready. This will clearly involve mixed strategies that we introduce now.

Let a mixed strategy for R at time t be a probability distribution p_t on $U_{ad}(x)$, i.e. a vector of simplex Σ_U

$$p_t(u) = P(u_t = u), \quad p_t \in \Sigma_U. \tag{5}$$

Likewise, q_t will be a mixed strategy for H at time t, a vector of the simplex Σ_N

$$q_t(v) = P(v_t = v), \quad q_t \in \Sigma_N. \tag{6}$$

Both players have infinite memory, but while H sees R and knows where he has shot in the past, R does not know where H is shooting or has shot. (Otherwise, he would never get caught and would not need mixed strategies!)

Let therefore

$$X_t = \{x_t, x_{t-1}, \ldots, x_0\}, \quad Y_t = \{y_t, y_{t-1}, \ldots, y_0\} \text{ if appropriate.} \tag{7}$$

R must choose his mixed strategy p_t as a function of X_t; and so does H in Sections 3 and 4 [game (2)], while H has access to X_t and Y_t in Section 5 [game (4)]

$$\begin{aligned} p_t &= \Phi_t[X_t], \\ q_t &= \Psi_t[X_t, Y_t]. \end{aligned} \tag{8}$$

2.3. Payoff

Replacing u and v by mixed strategies like (5) or (6) in the game (2) [or (4)] makes x_t, y_t and z_t stochastic processes, and therefore t_1 in (3) a stopping time.

In Section 3, the payoff that H will try to minimize, while R maximizes it, will simply be R's life expectation $E(t_1)$. This game will be called the stationary game. We shall look at it only in the complete information case (2) [and therefore no Y_t in (8)].

In Sections 4 and 5, we shall assume a time T is given (and known of both players) when the game warden is going to walk by forcing the hunter to leave (did we tell you he was a trespasser?). The payoff then shall be the probability for the hunter to kill the rabbit (probability of capture) $P(t_1 \leqslant T)$ and, of course, R is seeking to minimize it while H is striving to maximize it.

3. THE STATIONARY GAME

3.1. Problem statement

The motion of the rabbit is described probabilistically as a Markovian matrix $P = p_{ij}$ defined by

$$p_{ij} \stackrel{\text{def}}{=} P(u_t = j - i / x_t = i), \quad p_{ij} \geqslant 0, \quad \sum_{j \in I_N} p_{ij} = 1, \quad i \in I_N, \quad j \in I_N.$$

The hunter, knowing x_t, chooses to shoot at any position in I_N with a certain probability. So the motion of the hunter is described probabilistically as a Markovian matrix $Q = q_{ij}$ defined by

$$q_{ij} \stackrel{\text{def}}{=} P(v_t = j/x_t = i), \quad q_{ij} \geqslant 0, \quad \sum_{j \in I_N} q_{ij} = 1, \quad i \in I_N, \quad j \in I_N.$$

Thus, if $x_t = i$, the bullet will hit position j at time $t+1$ with probability q_{ij}.

It should be pointed out that we consider in this part of paper that p_{ij} and q_{ij} are independent of t, i.e. we study stationary strategies.

Stationary strategies are motivated for our problem partly because of the infinite time horizon and partly because of their relative simplicity. We do not intend to imply that nonstationary ones can be of no relevance to the infinite time horizon problem that we study.

Since the time interval considered here is infinite, the hunter can assure that he will kill the rabbit with probability one by choosing $q_{ij} = 1/N$, for all i and j in I_N, a fact quite easy to demonstrate. Thus, in the infinite time case, what appears to be the pertinent objective of the hunter is the minimization of the average time within which the rabbit is killed. The rabbits' objective is the contrary so that the two players are engaged in a zero-sum dynamic game.

It should be notice that the average time of killing is a function of the initial position of the rabbit, x_0. Thus, if $\bar{z}_i$ denotes the average killing time if $x_0 = i$, we have a vector $\bar{z} = (\bar{z}_1, \ldots, \bar{z}_N)'$ of payoff objectives where the $'$ is the notation for the transpose.

Several questions can be posed concerning the situation described above.

First, for fixed P, what is the best Q and conversely, for fixed Q, what is the best P?

Do they exist and if yes, can one find them in a convenient manner?

If there are no restrictions on the choices of P and Q, does there exist a saddle-point solution?

If the matrix P is constrained to be of a certain form, for example $p_{ij} = 0$ if $|i-j| \geqslant l+1$ (i.e. the rabbit can move at most l positions to the right or the left of its current position $x_t = i$), does a saddle point equilibrium exist and what is it?

It should be borne in mind, that in all the questions mentioned, we are interested in the whole vector $\bar{z} = (\bar{z}_1, \ldots, \bar{z}_N)'$ and would like the optimal pairs pertaining to the questions posed above to be optimal simultaneously for each component of $\bar{z}$.

In the next sections, we study some of these questions, in the context of a simple example, and, in the later sections, we address them more generally. In the final conclusions section, we present some further questions and problems intimately related to those studied here.

3.2. *Introductory example*

Let us consider $I_N = \{1, 2\}$ and the following matrices P and Q

$$P = \begin{pmatrix} 1 & 0 \\ 1-a & a \end{pmatrix}, \quad Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}.$$

Let P be fixed and thus we have a single objective problem, i.e. choose Q as to minimize the average killing time. If $x_t = 1$, it will necessarily be that $x_{t+1} = 1$, so that the hunter obviously chooses $q_{11} = 1$, $q_{12} = 0$. Therefore, the problem of the hunter is to choose q_{21}, q_{22}. Let us consider the following two possible choices for Q

$$Q_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and study first the situation under $Q = Q_1$. Let $x_0 = 2$.

Under Q_1, the hunter shoots always at position 1 if the rabbit is at position 2 (or position 1). If z_2 denotes the time at which the rabbit is killed, given that it starts at time zero at position 2, i.e. $x_0 = 2$, it holds for the following strategy (for $t \in \{1, 2, 3, 4\}$, the rabbit chooses to go to position 2 with probability a, at time $t = 5$, it chooses to go to position 2 with probability $1-a$)

$$P(z_2 = 5/x_0 = 2, Q = Q_1) = a^4(1-a).$$

In general, we have

$$P(z_2 = t) = a^{t-1}(1-a), \quad \text{for} \quad t = 1, 2, 3, \ldots$$

and thus, the average time for killing the rabbit is

$$\bar{z}_2 = E(z_2/x_0, Q = Q_1)$$

$$= \sum_{t=1}^{t} t a^{t-1}(1-a)$$

$$= \frac{1}{1-a} \quad \text{if} \quad a < 1.$$

If $a = 1$, then the rabbit stays always at position 2, and is never killed; this is in agreement with $\lim_{a \to 1}(1-a)^{-1} = +\infty$.

Let us now study the situation under $Q = Q_2$ and $x_0 = 2$. There are now only two possible trajectories for the rabbit:

if the rabbit starts at $x_0 = 2$ and goes to position 2, it will be killed at time 1 and this happens with probability a,

if the rabbit starts at $x_0 = 2$ and goes to position 1, it will be killed at time 2 and this happens with probability $1 - a$.

Thus

$$P(z_2 = 1/x_0 = 2, Q = Q_2) = a,$$

$$P(z_2 = 2/x_0 = 2, Q = Q_2) = 1 - a,$$

$$P(z_2 = t/x_0 = 2, Q = Q_2) = 0 \quad \text{if} \quad t \geqslant 3,$$

and

$$\bar{\bar{z}}_2 = E(z_2/x_0 = 2, Q = Q_2) = a + 2(1-a) = 2 - a.$$

One can draw $\bar{z}_2$ and $\bar{\bar{z}}_2$ as a function of a and the two curves intersect at $\bar{a} = (3 - \sqrt{5})/2$. It is clear that

if a belongs to $[0, \bar{a}]$, $\bar{z}_2 < \bar{\bar{z}}_2$ and Q_1 is preferred over Q_2,

if a belongs to $[\bar{a}, 1]$, $\bar{\bar{z}}_2 < \bar{z}_2$ and Q_2 is preferred over Q_1.

For $a = \bar{a}$, both Q_1 and Q_2 result in the same average killing time $\bar{z}_2(\bar{a}) = \bar{\bar{z}}_2(\bar{a}) = (1 + \sqrt{5})/2$.

There are several interesting facts revealed by this simple example. One is that, although Q_2 guarantees that the rabbit will be killed no later than time 2, whereas Q_1 allows the rabbit to be alive after an arbitrarily large time, Q_1 is preferable if a belongs to $[0, \bar{a}]$.

In the context of the example considered here with $I_N = \{1, 2\}$, the reader can easily persuade himself that a zero sum equilibrium cannot be formed by a pair of matrices P^* and Q^* which have only zeros and ones, since, if, for example, the hunter shoots always at the position i when the rabbit is at position j, the rabbit will always go from i to $k \neq j$. And so, it will never be killed. Analogously can do the hunter and always kill the rabbit in the next instant of time, if P^* is composed of zeros and ones; thus, an equilibrium pair P^*, Q^* with zeros and ones cannot exist.

Using the results of the next sections, one can show that if the choices of P and Q are arbitrary, there exists a unique zero sum equilibrium pair P^*, Q^* with

$$P^* = Q^* = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix},$$

and the resulting average killing time $\bar{z}_i$, if the rabbit starts at time $t = 0$ at $x_0 = i$, $i = 1, 2$, is $\bar{z}_1 = \bar{z}_2 = 2$.

Actually, this pair (P^*, Q^*), constitutes a zero-sum equilibrium for either one of the costs $E(z_1)$ or $E(z_2)$.

Although it is reasonable to assume that Q is chosen arbitrarily by the hunter, i.e. that he can shoot anywhere he wants, it might not be so for P, i.e. the rabbit might be restricted as to where it can go within one instant of time, due for example to its finite speed. Thus, one may be interested in investigating zero-sum equilibria subject to the constraint that P is of a certain form.

Let us examine a situation of this type. Let it be that P is to be chosen of the form

$$\begin{pmatrix} 1 & 0 \\ 1-a & a \end{pmatrix},$$

i.e. the rabbit can choose only a.

The hunter suffers no restriction as to his choice of Q, but it is obvious that he will choose $q_{11} = 1$, $q_{12} = 0$. Thus, the hunter chooses

$$\begin{pmatrix} 1 & 0 \\ 1-q & q \end{pmatrix}.$$

The average capture times can be calculated directly, or by using the more general results of the next section to be

$$\bar{z}_1 = 1, \quad \bar{z}_2 = \frac{1+q(1-a)}{1+a(1-q)},$$

and it is easy to see that

$$\bar{a} = \frac{3-\sqrt{5}}{2} \quad \text{and} \quad \bar{q} = \frac{-1+\sqrt{5}}{2}$$

constitute the zero-sum equilibrium with resulting value for $\bar{z}_2$, $(1+\sqrt{5})/2$, (i.e. the intersection point of the two curves $[\bar{a}, \bar{z}_2(\bar{a})]$ appears, as may be expected).

Thus, we see that it is possible to have zero sum equilibria in cases where P is restricted, although not every restriction of P will allow such an existence. The issue of study of zero-sum equilibria under some restrictions on the choice of P is undertaken in paragraph 3.5.2.

3.3. *Calculation of the average capture time*

For a given pair of two $N \times N$ Markovian matrices P and Q, let z be the random variable that the rabbit is killed at some time; it obviously depends on P and Q as well as on the initial value of x_0. It holds

$$P(z = t+1/x_0 = i) = \sum_{j=1}^{N} P(z = t/x_0 = j)p_{ij}(1-q_{ij}), \tag{9}$$

i.e. the probability that the rabbit is killed at time $t+1$, given that it started at $x_0 = i$, equals the probability that it is not killed in going from $x_0 = i$ to some $x_1 = j$ multiplied by the probability that it is killed at time $t+1$ if it started at $x_1 = j$ at time $t = 1$; the fact that $P(z = t+m/x_m = j) = P(z = t/x_0 = j)$, which is due to the stationarity of P and Q, is also used with $m = 1$ in deriving (9). Let

$$y_{t+1} = \begin{bmatrix} P(z = t+1/x_0 = 1) \\ P(z = t+1/x_0 = 2) \\ \vdots \\ P(z = t+1/x_0 = N) \end{bmatrix}, \quad e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbf{R}^N,$$

and

$$P * Q \overset{\text{def}}{=} \begin{bmatrix} p_{11}q_{11} & p_{12}q_{12} \cdots p_{1N}q_{1N} \\ p_{21}q_{21} & p_{22}q_{22} \cdots p_{2N}q_{2N} \\ \vdots & \vdots \quad\quad \vdots \\ p_{N1}q_{N1} & p_{N2}q_{N2} \cdots p_{NN}q_{NN} \end{bmatrix}, \quad L = P - P * Q.$$

(9) can be written as

$$y_{t+1} = Ly_t, \quad t = 1, 2, 3, \ldots, \quad y_1 = (P * Q)e.$$

If some $p_{i,j} = 0$, i.e. the rabbit does not ever go from position i to position j, the hunter can obviously choose $q_{ij} = 0$, since a shot at position j will be an obvious waste. Thus, without loss of generality, we assume that

$$p_{ij} = 0 \Rightarrow q_{ij} = 0. \tag{10}$$

Gregorskin's theorem, applied to the matrix L, yields that all the eigenvalues λ of L lie in the disc

$$|\lambda| \leq \max_i \sum_{j=1}^{N} p_{ij}(1 - q_{ij}) = 1 - \min_i \sum_{j=1}^{N} p_{ij}q_{ij}.$$

Under assumption (10), it holds that

$$|\lambda| \leq 1 - \min_{i,j} \{p_{ij}; p_{ij} > 0\} < 1,$$

and thus, the matrix L has all its eigenvalues strictly within the unit disc of the complex plane. This guarantees that y_t tends to $0 . e$ as t tends to $+\infty$. So, the matrix inverses, series forms and infinite series differentiations that will be used next, are valid. It holds

$$\begin{aligned} P(z < +\infty) &= y_1 + y_2 + y_3 + \cdots \\ &= y_1 + Ly_1 + L^2 y_1 + \cdots \\ &= (I - L)^{-1} y_1 \\ &= (I - P + P * Q)^{-1}(P * Q)e. \end{aligned}$$

Since $Pe = e$, it holds

$$(I - P + P * Q)e = (P * Q)e \quad \text{i.e.} \quad (I - L)e = y_1,$$

and

$$P(z < +\infty) = e. \tag{11}$$

Since $P(z = +\infty) = 0$, we can calculate the average capture time $\bar{z}$ by

$$\bar{z} = \sum_{t=1}^{+\infty} t y_t = (I - L)^{-1} e,$$

where (11) may be used in the last step. Thus

$$\bar{z} = (I - L)^{-1} e = (I - P + P * Q)^{-1} e. \tag{12}$$

Formula (12) will be used repeatedly in the sequel.

3.4. *The hunter's problem*

The hunter's problem is to minimize $\bar{z}$ with respect to Q, P being fixed.
Here, we consider the problem

$$\min_Q \bar{z} = \min_Q (I - P + P * Q)^{-1} e. \tag{13}$$

Notice that we are interested in Qs that minimize all the components of $\bar{z}$ simultaneously.

One way of going about this problem is the following. It is known that the inverse of an $N \times N$ matrix A, assuming it exists, has (i, j)th elements $(-1)^{i+j}|A_{ji}|/|A|$, where |A| is the determinant of A and $|A|_{ji}$ is the determinant of the minor of the (j, i) element of A. Thus, it is easy to see that although each component, say $\bar{z}_1$, of $\bar{z}$ is a quotient of nonlinear functions of the q_{ij}s, these nonlinear functions are multilinear in the sense that they are linear in each q_{ij}, the rest of the q_{ij}s considered fixed. Thus, the extremal values of $\bar{z}_1$ can be achieved at a Q, the elements of which are zeros and ones. Consequently, one may minimize $\bar{z}_1$ by checking which ones of these (N^2 in multitude) Qs results in the smallest value. If one is interested in minimizing all the components of $\bar{z}$ simultaneously, one may check whether this is possible by calculating $\bar{z}$ for all such Qs and find whether such solution exists and what it is. This procedure is quite cumbersome and as it stands quite uninformative. If one considers in addition that such Qs cannot serve as pairs of zero-sum equilibrium, one is bound to search for a different method for handling (13).

Let us consider that a given choice Q results in $\bar{z}$ and that another choice $\hat{Q}$ results in $\hat{\bar{z}}$. Thus, $(I - P + P * Q)\bar{z} = e$ and $(I - P + P * \hat{Q})\hat{\bar{z}} = e$. We assume that (10) holds for both Q and $\hat{Q}$.

Let $\hat{\bar{z}} = \bar{z} + \delta, \delta \in R^N$. It holds

$$\delta = (I - P + P * \hat{Q})^{-1}(P * Q - P * \hat{Q})\bar{z}.$$

It is obvious that all the elements of $P - P * Q$ and of $P - P * \hat{Q}$ are nonnegative and since it holds

$$(I - P + P * \hat{Q})^{-1} = (I - (P - P * \hat{Q}))^{-1} = I + (P - P * \hat{Q}) + (P - P * \hat{Q})^2 + \dots \quad (14)$$

and similarly for $(I - P + P * Q)^{-1}$.

Thus, if we want $\hat{Q}$ to be preferable over Q, it must be $\delta \leqslant 0$, so we have

$$(P * Q - P * \hat{Q})\bar{z} \leqslant 0. \quad (15)$$

If, by a choice of $\hat{Q}$, we make the first component of $(P * Q - P * \hat{Q})\bar{z}$ negative and the other components less or equal than zero, we have guaranteed that all the components of δ will be less or equal than zero. In addition, the first component of δ, δ_1 will be strictly negative since the unit matrix in the right-hand side of (14) guarantees that the first element of the first raw of $(I - P + P * \hat{Q})$ is strictly positive. For example, if

$$p_{1l}z_l = \max\{p_{11}z_1, p_{12}z_2, \dots, p_{1N}z_N\},$$

we can choose the first raw of $\hat{Q}$ by

$$\hat{q}_{11} = \dots = \hat{q}_{1(l-1)} = \hat{q}_{1(l+1)} = \dots = \hat{q}_{1N} = 0, \quad \hat{q}_{1l} = 1,$$

and the other raws of $\hat{Q}$ to be the same as those of Q, which results in

$$(P * Q - P * \hat{Q})\bar{z} = \begin{bmatrix} -p_{1l}\bar{z}_l + \sum_{j \in I_N} p_{1j}q_{1j}\bar{z}_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leqslant 0.$$

No reduction of value in going from $(Q, \bar{z})$ to some $(\hat{Q}, \hat{\bar{z}})$ is possible by changing only the first raw of Q if

$$p_{11}q_{11}\bar{z}_1 + \dots + p_{1N}q_{1N}\bar{z}_N \geqslant p_{11}\bar{z}_1, \dots, p_{1N}\bar{z}_N,$$

which is equivalent to

$$p_{1j}\bar{z}_j = p_{1l}\bar{z}_l, \quad \forall j, l \quad \text{with} \quad p_{1j} \neq 0 \quad \text{and} \quad p_{1l} \neq 0.$$

The proof of the following theorem is a straightforward application of the ideas delineated above.

Theorem 1

(i) There exists a Q that minimizes simultaneously all the components $\bar{z}_1, \dots, \bar{z}_N$ of $\bar{z}$.

(ii) A Q is optimum if and only if

$$p_{ij}\bar{z}_j = p_{il}\bar{z}_l, \quad \forall i, j, l \quad \text{with} \quad p_{ij} \neq 0, \quad p_{il} \neq 0, \quad (16)$$

where the $\bar{z}_j$s are the solution of (13) for the aforementioned Q.

An algorithm for finding all the optimal Qs is the following

Step 1: Choose a $Q = Q_1$, so that condition (10) is satisfied. Calculate

$$\bar{z}_1 = (I - P + P * Q_1)^{-1}e, \quad \bar{z}_1 = (\bar{z}_{11}, \dots, \bar{z}_{N1})^T.$$

Step 2: Calculate $p_{ij}\bar{z}_j$ for $p_{ij} \neq 0$.

Step 3: Find for each i, the l for which

$$p_{il}\bar{z}_l = \max(p_{i1}\bar{z}_{11}, p_{i2}\bar{z}_{21}, \dots, p_{iN}\bar{z}_{N1}).$$

Step 4: Choose Q_2 such that $q_{il} = 1$ for each i, where l is the l found for this i in the step 3.

Step 5: Set $Q = Q_2$.

This algorithm will converge in a finite number of steps. It operates essentially on the extreme points of the set of the $N \times N$ Markovian Qs and will converge much faster than the primitive algorithm suggested in the paragraph 5. which checks N^2 possible Qs since it reduces simultaneously all the components of $\bar{z}$.

It is also clear that, as soon as an optimum has been found, all the other optima Q can be generated as follows if Q is optimal consider for each i, the ls for which $p_{il}\bar{z}_l = \max(p_{i1}\bar{z}_{11}, \ldots, p_{1N}\bar{z}_{N1})$ where $\bar{z} = (\bar{z}_{11}, \ldots, \bar{z}_{N1})'$ corresponds to the optimal Q^* that has been found. Any Q, which has at the ith raw $q_{ij} = 1$ for any j for which the maximum of the $p_{ij}\bar{z}_j$s for $j = 1$ to N, is achieved, is also optimal.

The set of convex combinations of all these Qs is the solution set of problem (13). Finally, by construction of the algorithm $(I - P + P * Q_k)^{-1}$ exists at each step k, i.e. (10) will be automatically satisfied throughout the operation of the algorithm.

Example

Let

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/4 & 1/4 & 1/2 \\ 1/5 & 2/5 & 2/5 \end{pmatrix}.$$

Step 1: Let

$$Q_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then

$$\bar{z}_1 = (I - P + P * Q_1)^{-1} e = \begin{pmatrix} 3 + 3/17 \\ 3 + 6/34 \\ 2 + 16/17 \end{pmatrix}.$$

Using the criterion of the step 3, we choose

$$Q_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{which yields} \quad \bar{z}_2 = \begin{pmatrix} 1 + 1/2 \\ 1 + 5/6 \\ 2 + 5/18 \end{pmatrix}.$$

Notice that $\bar{z}_2$ is better than $\bar{z}_1$ componentwise.

Using again the criterion of the step 3, we choose

$$Q_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{which yields} \quad \bar{z}_3 = \begin{pmatrix} 1 + 1/2 \\ 1 + 5/6 \\ 2 + 1/30 \end{pmatrix}.$$

Use of the criterion 3 shows that this is the optimal $\bar{z}$. It is worth noticing that in this example, q_{ij} equals one at the position of the raw-maxima of the p_{ij}s.

This is not in general true as other examples can demonstrate. As a matter of fact, a simple continuity argument can show that Q_3 remains optimum if we perturb the last raw of P into $(1/5, 2/5 + \epsilon, 2/5 - \epsilon)$ where $\epsilon > 0$, ϵ small, so that the optimal Q_3 for this new P will not have $q_{32} = 1$, whereas $p_{32} > p_{31}, p_{32}$. It is nonetheless reasonable to expect that large p_{ij}s deserve large q_{ij}s, so that a good initial choice of Q for starting the algorithm is obviously to choose $q_{ij} = 1$ for $p_{ij} = \max\{p_{ik}; k = 1, \ldots, N\}$.

Before leaving this section, it is worth pointing out an intuitive justification of the first part of the theorem.

First, a simple continuity and compactness argument shows that there exist Qs which minimize $\bar{z}_i$.

Secondly, let us assume that Q_i minimizes $\bar{z}_i$, $i = 1, \ldots, N$ at time t, the hunter shoots according to the jth raw of Q_i. If he fails to kill the rabbit, which now at time $t+1$ is at $x_{t+1} = k$, he uses the kth raw of Q_i, whereas common sense suggests that the kth raw of Q_j is pertinent now.

The fact that there exists a Q that minimizes $\bar{z}_1, \ldots, \bar{z}_N$ can also be justified by the fact that the control of the hunter can be expected to be of the feedback type by which at position $x_t = i$, the ith raw of Q gives the optimal control.

3.5. *The rabbit's problem*

The rabbit's problem is to maximize $\bar{z}$ with respect to P, Q being fixed.

Here we consider the problem

$$\max_P \bar{z} = \max_P (I - P + P * Q)^{-1} e. \tag{17}$$

Let P_1 correspond to $\bar{z}_1$ and P_2 correspond to $\bar{z}_2$. It holds

$$(I - P_1 + P_1 * Q)\bar{z}_1 = (I - P_2 + P_2 * Q)\bar{z}_2 = e.$$

Let $\delta = \bar{z}_2 - \bar{z}_1$. So, we have

$$\delta = (I - P_2 + P_2 * Q)^{-1}((P_2 - P_1) - (P_2 - P_1) * Q)\bar{z}_1,$$

where we assume that (10) holds for both P_1 and P_2. Here, we are interested in increasing $\bar{z}$, i.e. we would like to move from $\bar{z}_1$ to $\bar{z}_2$ with $\delta \geqslant 0$. We cannot increase δ if

$$\bar{z}_j(1 - q_{ij}) \leqslant \sum_{j=1}^{N} \bar{z}_j p_{ij}(1 - q_{ij}), \quad j = 1, \ldots, N, \tag{18}$$

or equivalently if

$$\bar{z}_j(1 - q_{ij}) = \bar{z}_k(1 - q_{ik}), \quad \forall i, j, k \quad \text{with} \quad p_{ij} \neq 0 \quad \text{and} \quad p_{ik} \neq 0. \tag{19}$$

For example, if $p_{11}, \ldots, p_{1k}$ are not zero and $p_{1(k-1)}, \ldots, p_{1N}$ are zero, it should hold

$$z_1(1 - q_{11}) = \cdots = z_k(1 - q_{1k}). \tag{20}$$

The whole development of the previous section can also be carried out here in a completely analogous fashion but we omit it for the sake of brevity. The only difference is that if P satisfies (10), and (18) does not hold for some (i, j) and we update P accordingly to some P_3 which yields a $\bar{z}_3$ greater, there is no guarantee that P_3 satisfies also (10), so that some of the components of $\bar{z}_3$ may be infinite.

3.5.1. *The zero-sum case with unrestricted P.* If $p_{ij} = 1/N$, for i and j in I_N, (13) yields that $\bar{z} = Ne$ for any Q. Similarly, if $q_{ij} = 1/N$, for i and j in I_N, (13) yields that $\bar{z} = Ne$ for any P.

Thus the pair (P^*, Q^*), such that for all i and j in I_N, $p^*_{ij} = q^*_{ij} = 1/N$, is a zero-sum equilibrium, for each component of the vector $\bar{z}$. The averaging capture time at the equilibrium is N units of time, i.e. it equals the dimension of P (and Q). The remaining question is whether there exists another zero-sum equilibrium $(\hat{P}, \hat{Q})$. If it does, it will hold for all Q and all P

$$J(\hat{P}, Q) \geqslant J(\hat{P}, \hat{Q}) = J(\hat{P}, Q^*) = Ne = J(P^*, \hat{Q}) \geqslant J(P, \hat{Q}), \tag{21}$$

where $J(P, Q)$ denotes $\bar{z} = (I - P + P * Q)^{-1}e$. The left-hand side of (21), in conjunction with the condition (15) for optimality of $Q = \hat{Q}$ yields $(\hat{P} * Q^* - \hat{P} * \tilde{Q})\bar{z} \geqslant 0$ for any $\tilde{Q}$, i.e.

$$\hat{p}_{i1}(q^*_{i1} - \tilde{q}_{i1})N + \ldots + \hat{p}_{iN}(q^*_{iN} - \tilde{q}_{iN})N \geqslant 0. \tag{22}$$

But $q^*_{ij} = 1/N$ and $\Sigma_{j=1}^{N} \hat{p}_{ij} = 1$, so (22) yields $1/N \geqslant \hat{p}_{i1}\tilde{q}_{i1} + \cdots + \hat{p}_{iN}\tilde{q}_{iN}$, for any $\tilde{Q}$ and thus, for any i, j, $1/N \geqslant \hat{p}_{ij}$ which implies $\hat{p}_{ij} = 1/N = p^*_{ij}$, i.e. $P^* = \tilde{P}$. Since $\hat{Q}$ is an optimal response to P^*, it has to satisfy (20) with $\bar{z}_2 = N$ and thus $Q = Q^*$. We have thus proven

Theorem 2

If there is no restriction on P or Q, there exists a unique zero-sum equilibrium which is

$$P^* = Q^* = \begin{pmatrix} 1/N \dots 1/N \\ \vdots \quad \ddots \quad \vdots \\ 1/N \dots 1/N \end{pmatrix}, \tag{23a}$$

with optimal value for $\bar{z}$

$$z^* = \begin{pmatrix} N \\ \vdots \\ N \end{pmatrix}. \tag{23b}$$

3.5.2. The zero-sum solution for a class of constrained Ps. In this section, we assume that P is restricted so as to reflect the fact that the rabbit, due to its finite speed, cannot move further than l positions, i.e. we assume that $p_{ij} = 0$ if $|j - i| \geqslant (l + 1)$, $l \geqslant 1$, so we have

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1(l+1)} & & & & \\ p_{21} & p_{22} & \cdots & p_{2(l+1)} & p_{2(l+2)} & & (0) & \\ \vdots & \vdots & & \ddots & & \ddots & & \\ p_{(l+1)1} & p_{(l+1)2} & \ddots & & \ddots & & \ddots & \\ & p_{(l+2)2} & & \ddots & & \ddots & & p_{(N-l)N} \\ & & \ddots & & \ddots & & & \vdots \\ & (0) & & \ddots & & & & p_{(N-1)N} \\ & & & & p_{Nl} & \cdots & p_{N(N-1)} & p_{NN} \end{bmatrix}. \tag{24}$$

Obviously, if a zero-sum equilibrium P^*, Q^* exists, it will be $q_{ij}^* = 0$ if $|i - j| \geqslant (l + 1)$. Let us examine whether such an equilibrium with $p_{ij}^* \neq 0$, for $|i - j| < (l + 1)$ exists. If it does, and the associated optimal value is $z^* = (z_1^*, \dots, z_N^*)$, it must hold

$$p_{ij}^* z_j^* = p_{ik}^* z_k^*, \quad |i - j| < (l + 1), \quad |i - k| < (l + 1),$$

and thus

$$p_{ij}^* = 1 \Bigg/ \left(z_j^* \sum_{k = \max[1, i - l]}^{\min[N, i + l]} \frac{1}{z_k^*} \right). \tag{25a}$$

Also using (19), we obtain that

$$q_{ij}^* = 1 - (\min [N, i + l] - \max [1, i - l] - 1) p_{ij}^*, \tag{25b}$$

where p_{ij}^* is defined by (25a).

We can write P^* and Q^* in a more compact form by introducing the following notation. Let E be an $N \times N$ matrix with each (i, j) element equal to 1 if $|i - j| \leqslant l$ and 0 otherwise. E has the same structure as P in (24) but with one's in place of the p_{ij}s.

Let $e = (1 \dots 1)'$ in $\mathbf{R}^N$ and

$$\begin{pmatrix} \Sigma_1 \\ \vdots \\ \Sigma_N \end{pmatrix} = E \begin{pmatrix} 1/z_1^* \\ \vdots \\ 1/z_N^* \end{pmatrix}, \quad \rho = Ee - e.$$

Then

$$P^* = \begin{pmatrix} 1/\Sigma_1 & & (0) \\ & \ddots & \\ (0) & & 1/\Sigma_N \end{pmatrix} E \begin{pmatrix} 1/z_1^* & & (0) \\ & \ddots & \\ 0 & & 1/z_N^* \end{pmatrix}. \tag{26a}$$

$$Q^* = E - \begin{pmatrix} \rho_1/\Sigma_1 & & (0) \\ & \ddots & \\ (0) & & \rho_N/\Sigma_N \end{pmatrix} E \begin{pmatrix} 1/z_1^* & & (0) \\ & \ddots & \\ 0 & & 1/z_N^* \end{pmatrix}. \tag{26b}$$

For the P^*, Q^*, z^* to exist [since the choice Q as in (23a) is also admissible, we know that if P^* and Q^* are optimal, (10) will be satisfied since the resulting z^* has to be finite] and satisfy (13) and (25), it must hold

$$(I - P^* + P^* * Q^*)z^* = e. \tag{27}$$

Substituting P^* and Q^* from (26) into (27) yields

$$z^* - \begin{pmatrix} 1/\Sigma_1 & & (0) \\ & \ddots & \\ (0) & & 1/\Sigma_N \end{pmatrix} Ee + \begin{pmatrix} 1/\Sigma_1 \\ \vdots \\ 1/\Sigma_N \end{pmatrix} = e. \tag{28}$$

Introducing for any i in I_N,

$$a_i = 1/z_i^*, \tag{29}$$

denoted by (29), we can write (28) in the following equivalent form: let

$$E\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \Sigma_1 \\ \vdots \\ \Sigma_N \end{pmatrix} = \Sigma(a) \quad \text{and} \quad Ee - e = \rho = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_N \end{pmatrix}$$

$$a_i = \frac{\Sigma_i}{\Sigma_i + \rho_i}, \quad \forall i \in I_N. \tag{30}$$

what we need to show is the existence of a nonzero solution of (30). To prove existence, we introduce the function f from $\mathbf{R}_+^N$ to $\mathbf{R}^N$ such that, for any x in $\mathbf{R}_+^N$, we have $f(x) = y$ with

$$y_i = \max\left[\epsilon, \frac{\hat{\Sigma}_i}{\hat{\Sigma}_i + \rho_i}\right], \quad \forall i \in I_N,$$

where $\hat{\Sigma}$ is defined by the first equality of (20) and ϵ in [0, 1] will be specified shortly. First, notice that f is continuous and $f([\epsilon, 1]^N) \subseteq [\epsilon, 1]^N$ by construction. Thus, by Brouwer's fixed point theorem, f has a fixed point. But for our purpose, i.e. existence of a nonzero solution of (30), the fixed point of f would be worthless if ϵ is such that, for some of the components of the fixed point of f, $x = y$, we have $x_i = y_i = \epsilon > [\hat{\Sigma}_i(x)]/(\hat{\Sigma}_i + \rho_i)$. To exclude that this holds for the fixed point of f, for any x_i, we work as follows. If it holds for x_i, it will be

$$x_i = y_i = \epsilon > \frac{\hat{\Sigma}_i(x)}{\rho_i + \hat{\Sigma}_i(x)} \Rightarrow \frac{\epsilon\rho_i}{1-\epsilon} > \hat{\Sigma}_i(x).$$

But

$$\hat{\Sigma}(x) = Ex \geqslant E\begin{pmatrix} \epsilon \\ \vdots \\ \epsilon \end{pmatrix} = \epsilon Ee = \epsilon(\rho + e) = \epsilon\begin{pmatrix} \rho_1 + 1 \\ \vdots \\ \rho_N + N \end{pmatrix},$$

and thus

$$\frac{\epsilon\rho_i}{1-\epsilon} > \epsilon(\rho_i + 1) \Rightarrow \epsilon > \frac{1}{\rho_i + 1}.$$

Thus, if $\epsilon < 1/\rho_i + 1$, it cannot be that the fixed point $x = y$ of f satisfies

$$x_i = y_i = \epsilon \geqslant \frac{\hat{\Sigma}(x)}{\rho_i + \hat{\Sigma}(x)}.$$

To exclude that this happens for any component of the fixed point, it suffices to choose

$$\epsilon < \frac{1}{1 + \max(\rho_1, \ldots, \rho_N)} = \bar{\epsilon}. \tag{31}$$

With such an ϵ, the fixed point of f serves also as a nonzero solution of (30).

Since any solution of (30) creates through (26), (29) a solution to the zero-sum game at hand and since the value of this game is uniquely determined, we immediately conclude that (30) not only has a solution, but a unique one. Also, this solution satisfies $a_i \geqslant \epsilon$ for any $\epsilon < \bar{\epsilon}$ and thus $z_i^* = 1/a_i \leqslant 1/\epsilon$, i.e. the average capture time at the zero-sum equilibrium is less or equal than $2l + 1$, which is a much better bound than the already mentioned bound N, especially if N is much larger than l.

Remark: notice that $\rho = (l, l+1, \ldots, 2l-1, 2l, 2l, \ldots, 2l, 2l, 2l-1, \ldots, l)'$ if $N \geqslant 2l+1$, $\rho = (l, l+1, \ldots, N-1, N-1, \ldots, l)'$ if $N \leqslant 2l+1$ and thus $\bar{\epsilon} = (1+2l)^{-1}$, correspondingly.

It is also clear that one can slightly modify the definition of f as to define it only for x symmetrics, i.e. $x_1 = x_N, x_2 = x_{N-1}, \ldots$, which will guarantee that the fixed point is symmetric and thus $z_1^* = z_N^*$, $z_2^* = z_{N-1}^*$, and so on, which will guarantee that the ith row of P^* (or Q^*) is the mirror image of its $N - i + 1$ row.

Finally, since any solution of (30) provides the value vector of the zero-sum game at hand, which is uniquely determined one concludes that (30) has unique nonzero solution. The only thing missing is an algorithm for finding the solution of (30), since the obviously suggested iteration $x_{k+1} = f(x_k)$ is not guaranteed to produce the solution even as one of its cluster points. In this paragraph, we are going to remedy this weakness. Let us consider the function g from $\mathbf{R}_+^N$ to $\mathbf{R}^N$ defined by $g(x) = y$ with

$$y_i = \frac{\hat{\Sigma}_i(x)}{\hat{\Sigma}_i(x) + \rho_i}, \quad \forall i \in I_N. \tag{32}$$

It holds

$$\nabla g(x) = E' \begin{pmatrix} \rho_1/(\rho_1 + \hat{\Sigma}_1(x))^2 & & (0) \\ & \ddots & \\ (0) & & \rho_N/(\rho_N + \hat{\Sigma}_N(x))^2 \end{pmatrix}.$$

Let us consider the iteration $x_{k+1} = g(x_k)$. It holds

$$x_{k+1}^i - x_k^i = (\nabla g_i(\tilde{x}_k^i))'(x_k - x_{k-1}). \tag{33a}$$

where $\tilde{x}_k^i$ is some vector in $\mathbf{R}_+^N$, for $i = 1, \ldots, N$. Since $\nabla g(x)$ has nonnegative elements, if $x_k \geqslant x_{k-1}$, it will be $x_{k+1} \geqslant x_k$. Thus, if we can find an initial point x_0 to start the iteration, with $g(x_0) \geqslant x_0$ and $x_0 \neq 0$ we are guaranteed to create an increasing sequence of vectors $\{x_k\}$ which is obviously bounded in $[0, 1]^N$ [see (32)] and thus we have guaranteed convergence of x_k to some $x^* \neq 0$ which solves $x^* = g(x^*)$. We claim that any $x_0 = \epsilon e$, denoted by (33b), where $0 < \epsilon < \bar{\epsilon} = 1/(1 + \max(\rho_1, \ldots, \rho_N))$ performs this task. It holds

$$E(\epsilon e) = \epsilon E e = \epsilon(\rho + e) = \epsilon \begin{pmatrix} \rho_1 + 1 \\ \vdots \\ \rho_N + 1 \end{pmatrix} \tag{33b}$$

and thus

$$g(\epsilon e) = y \quad \text{with} \quad y_i(\epsilon e) = \frac{\epsilon(\rho_i + 1)}{\rho_i + \epsilon(\rho_i + 1)}.$$

It suffices

$$\frac{\epsilon(\rho_i + 1)}{\rho_i + \epsilon(\rho_i + 1)} \geqslant \epsilon \quad \text{or} \quad \epsilon < \frac{1}{1 + \rho_i}. \tag{34}$$

Thus, if $\epsilon < \bar{\epsilon}$ and we start the iteration $x_{k+1} = g(x_k)$ with any initial condition $x_0 = \epsilon e, \epsilon < \bar{\epsilon}, \epsilon \neq 0$, we will create an increasing sequence converging to a solution of (30).

Similarly, if we wish to have a decreasing sequence of x_ks, it suffices that, at the first step, $g(x_0) \leqslant x_0$. For example, if $x_0 = a = \mu e$ for some $\mu > 0$, in order to have $g(x_0) \leqslant x_0$, it suffices that $\Sigma_i(x_0)/(\rho_i + \Sigma_i(x_0)) \leqslant \mu$ or $\Sigma_i(x_0) \leqslant \rho_i \mu/(1-\mu)$ or $\mu \geqslant (1+\rho_i)^{-1}$ for all is which is equivalent to $\mu \geqslant (1+l)^{-1}$. If we do not wish to choose $x_0 = \mu e$, since it holds that $\Sigma_i/(\Sigma_i + \rho_i) \leqslant (\rho_i + 1)/(2\rho_i + 1)(\Sigma/\Sigma + \rho)$ is an increasing function of Σ), it suffices to choose the ith component of x_0 greater than $(1+\rho_i)/(1+2\rho_i)$. In conclusion, if

$$x_0 = \mu e, \quad 1 \geqslant \mu > \frac{1}{(1+l)} \quad \text{or} \quad e \geqslant x_0 \geqslant \begin{pmatrix} (1-\rho_1)/(1+2\rho_1) \\ \vdots \\ (1+\rho_N)/(1+2\rho_N) \end{pmatrix}, \tag{35}$$

the iteration $x_{k+1} = g(x_k)$ creates a decreasing sequence and thus $\{x_k\}$ converges. It is conceivable that since $\{x_k\}$ is decreasing, it might converge to zero, in which case this iteration would not provide a nonzero solution of (30). But, this can be excluded by showing that it is not possible to have $g(x) \leqslant x$ for x sufficiently close to zero. The proof is the following.

For x close to 0, it holds $g_i(x) = g_i(0) + (\nabla g_i(\tilde{x}^i))'x$ or

$$g(x_k) = \begin{pmatrix} \rho_1/(\rho_1 + \hat{\Sigma}_1(\tilde{x}^1))^2 & & (0) \\ & \ddots & \\ (0) & & \rho_N/(\rho_N + \hat{\Sigma}_N(\tilde{x}^N))^2 \end{pmatrix} Ex,$$

where $\tilde{x}^i$ is in $[0, x]$, for $i = 1, \ldots, N$. If the algorithm with decreasing x_ks converges to zero, it will be

$$g(x_k) = \begin{pmatrix} \rho_1/(\rho_1 + \hat{\Sigma}_1(\tilde{x}_k^1))^2 & & (0) \\ & \ddots & \\ (0) & & \rho_N/(\rho_N + \hat{\Sigma}_N(\tilde{x}_k^N))^2 \end{pmatrix} Ex_k \leqslant x_k. \tag{36}$$

It is clear from the form of g (32), that if $x_k = 0$ then $x_{k-1} = 0$ and thus, as long as $x_0 \neq 0$, it will be $x_k \neq 0$ for every k. Let $\delta_k = x/\|x_k\|$ and divide both sides of (36) by $\|x_k\|$ to get

$$\begin{pmatrix} \rho_1/(\rho_1 + \hat{\Sigma}_1(\tilde{x}_k^1))^2 & & (0) \\ & \ddots & \\ (0) & & \rho_N/(\rho_N + \hat{\Sigma}_N(\tilde{x}_k^N))^2 \end{pmatrix} E\delta_k \leqslant \delta_k. \tag{37}$$

Since $\|\delta_k\| = 1$, there is a subsequence of $\{\delta_k\}$ which converges to some δ, $\|\delta\| = 1$, $\delta \geqslant 0$. For this subsequence, the corresponding subsequences of $\tilde{x}_k^i$s go to zero for $i = 1, \ldots, N$, and thus taking limits with respect to this subsequence in (37) yields

$$\begin{pmatrix} 1/\rho_1 & & (0) \\ & \ddots & \\ (0) & & 1/\rho_N \end{pmatrix} E\delta \leqslant \delta \quad \text{or} \quad E\delta \leqslant \begin{pmatrix} \rho_1 & & (0) \\ & \ddots & \\ (0) & & \rho_N \end{pmatrix} \delta.$$

Multiplying both sides with e' and using (30) yields

$$(1+\rho_1, \ldots, 1+\rho_N)\delta \leqslant (\rho_1, \ldots, \rho_N)\delta \quad \text{or} \quad \delta_1 + \cdots + \delta_N \leqslant 0.$$

But this cannot be for $\delta = (\delta_1, \ldots, \delta_N)' \geqslant 0$ and $\|\delta\| = 1$. Thus, we conclude that any sequence $x_{k+1} = g(x_k)$ with $x_{k+1} \leqslant x_k$ cannot converge to zero.

We have thus established two algorithms, the one increasing, if x_0 is as in (33b) and the other decreasing, if x_0 is as in (34) which provide in the limit the solution of (30). One can carry out the first steps of these algorithms to create upper and lower bounds for the z_is.

Thus, starting with $x_0 = (1+2l)^{-1}e$, we calculate $g(x_0) \geqslant x_0$ and a_i is greater or equal to the ith component of $g(x_0)$; starting with $\bar{x}_0 = (1+l)^{-1}$, we calculate $g(\bar{x}_0) \geqslant \bar{x}_0$ and a_i is less or equal to the ith component of $g(\bar{x}_0)$. It turns out

$$1 + \frac{\rho_i}{1+\rho_i}(1+2l) \geqslant z_i \geqslant 1 + \frac{\rho_i}{1+\rho_i}(1+l).$$

For z_1, this means

$$1 + 2l - \frac{l}{1+l} \geqslant z_1 \geqslant 1 + l. \tag{38}$$

For z_m somewhere in the middle, where $\rho_m = 2l$, we have

$$1 + 2l \geqslant z_m \geqslant 1 + 2l + \frac{l}{2l+1}. \tag{39}$$

The bounds (38) and (39) are in agreement with the fact that we expect the average capture times for i close to 1 and N, to be smaller than those corresponding to is far from 1 to N, since the closer the rabbit starts to the barrier (i.e. $i = 1$ or N) the more restricted its moves are.

Having established the existence of a solution of (30), it is trivial to show that the P^*, Q^* constructed as in (26) provide a zero-sum solution to the game. By arguments similar to those used in the part 3.5.1., one can show that it is unique.

Let us formally state the results of this part in the following theorem.

Theorem 2

The zero-sum game, with P restricted as in (24), admits a unique solution given by (26) where the z_i^*s are found by solving (29) and (30). The solution can be found by finding the a_is that solve (30), by using the iteration $x_{k+1} = g(x_k)$, where g is given in (32) and x_0 is as in (35) or $x_0 = \epsilon e$, where $0 < \epsilon < \bar{\epsilon}$, and $\bar{\epsilon}$ is given in (31).

3.6. Interpretation of the solution

Having derived the optimal strategies, let us elaborate on their meaning. Before doing that, let us find out where the rabbit spends most of its time. Since P^* is clearly composed of a single ergodic class, it holds

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n} P^k = e\mu',$$

where $\mu' = \mu' P$ is a probability vector. It can be verified that

$$\mu = \frac{1}{\theta} \begin{pmatrix} a_1 \Sigma_1 \\ \vdots \\ a_N \Sigma_N \end{pmatrix}, \quad \theta = \sum_{i=1}^{N} a_i \Sigma_i = \sum_{i=1}^{N} a_i.$$

Let $\lambda = Q'\mu$; μ_i denotes the probability with which the rabbit will be at position i, after the lapse of a lot of time, assuming it is still alive, and λ_i denotes the probability that the hunter will shoot at position i, i.e. λ gives the distribution of the bullets as time goes to infinity. Let us proceed now with some intuitive interpretations of what happens, by employing an example. Example: let $N = 3$, $l = 1$. Calculating the a_is, z_is, P, Q, μ, λ at the optimum yields

$$a_1 = a_3 = 0.453, \quad a_2 = 0.375,$$

$$z_1 = z_3 = 2.207, \quad z_2 = 2.666,$$

$$\Sigma_1 = \Sigma_3 = 0.824. \quad \Sigma_2 = 1.281,$$

$$P = \begin{pmatrix} 0.55 & 0.45 & 0 \\ 0.355 & 0.29 & 0.355 \\ 0 & 0.45 & 0.55 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.45 & 0.55 & 0 \\ 0.29 & 0.42 & 0.29 \\ 0 & 0.55 & 0.45 \end{pmatrix},$$

$$\mu = \frac{1}{\theta} \begin{pmatrix} 0.305 \\ 0.390 \\ 0.305 \end{pmatrix}, \quad \lambda = \begin{pmatrix} 0.250 \\ 0.499 \\ 0.250 \end{pmatrix}, \quad \theta = (1, 0, 0).$$

Thus, the average expected time for the rabbit to leave, increases, the further, the rabbit starts at $t = 0$, from the boundary ($z_2 > z_1$). The rabbit has a tendency to move to the boundary whereas the hunter prefers to shoot more towards the middle. One can say, intuitively, that the hunter exhibiting a tendency to shoot more towards the middle, forces the rabbit towards the boundary, where the restricted moves of the rabbit make easier the hunters' task. Nonetheless, things are such that the rabbit ends up spending more time around the middle (since $\mu_2 > \mu_1 = \mu_3$) where his life expectation is higher ($z_2 > z_1 = z_3$) and actually that is where most of the bullets fall ($\lambda_2 = 0.499 > \lambda_1 = \lambda_3 = 0.25$). Thus, two different tendencies appear. At each instant of time (short time horizon), the rabbit moves towards the boundary, forced by the hunter's tendency to shoot more in the middle. But in the long term horizon, the rabbit frequents more the middle where his life expectation is higher and similarly the hunter ends up most of his bullets there.

The stationary strategies applied to the finite horizon game.

It is worthy finding out in what situations, the stationary strategies are good for finite time problems. This obviously concerns the magnitute of the time horizon.

It holds $y_{t+1} = (P - P * Q)y_t$, $y_1 = (P * Q)e$ and

$$P(z \leqslant te) = y_1 + \cdots + y_t = e - \left[\begin{pmatrix} \rho_1 & & (0) \\ & \ddots & \\ (0) & & \rho_N \end{pmatrix} (P * P) \right]^t e.$$

Thus

$$\begin{pmatrix} \rho_1 & & (0) \\ & \ddots & \\ (0) & & \rho_N \end{pmatrix} (P * P)e = \begin{bmatrix} \rho_1 \dfrac{\Delta_1}{\Sigma_1^2} \\ \vdots \\ \rho_N \dfrac{\Delta_N}{\Sigma_N^2} \end{bmatrix} \leqslant \max\left(\frac{\rho_i \Delta_i}{\Sigma_i^2}, i = 1, \ldots, N \right) e = \theta e$$

with

$$\Delta = E \begin{pmatrix} a_1^2 \\ \vdots \\ a_N^2 \end{pmatrix}.$$

Thus, if

$$\left[\begin{pmatrix} \rho_1 & & (0) \\ & \ddots & \\ (0) & & \rho_N \end{pmatrix} (P * P) \right]^t e \leqslant \theta^t e$$

and if $\theta < 1$, θ gives a rate of convergence of $\lim_{t \to +\infty} P(z \leqslant te) = e$. This can now be used as follows. The stationary strategy applied to the finite time horizon problem with time horizon t_f, will give a very good strategy for the hunter, who will kill the rabbit fast in average times $\bar{z}_1, \ldots, \bar{z}_N$, and the killing will take place with probability $99\% = 1 - \epsilon$, $\epsilon = 10^{-2}$, if $\theta^{t_f} < 10^{-2}$, i.e.

$$t_f > \frac{2}{|\log_{10} \theta|}. \tag{40}$$

Let us show that $\theta < 1$, by calculating explicitly a $\bar{\theta}$ with $\theta \leqslant \bar{\theta} < 1$. We will need the following fact

$$\text{if} \quad \epsilon_1 \leqslant x_i \leqslant \epsilon_2, i = 1, \ldots, N, \epsilon_1, \epsilon_2 > 0, \quad \text{then} \quad \phi(x) \frac{x_1^2 + \cdots + x_N^2}{(x_1 + \cdots + x_N)^2} \leqslant \frac{\epsilon_2^2 + (N-1)\epsilon_1^2}{(\epsilon_2 + (N-1)\epsilon_1)^2}.$$

The proof of this fact is as follows: if $\phi(x)$ achieves its maximum in the interior of the constraint set, it will be $x_1 = \cdots = x_N$ and the value of ϕ will be $1/N$. Checking now the values of ϕ at the boundary, it is easy to show that ϕ achieves its maximum by taking $(N-1)$ components of x to equal the minimum value ϵ_1 and only one component of x to equal the maximum value ϵ_2.

Using this fact and taking $\epsilon_1 = 1/(2l+1)$, $\epsilon_2 = 1/(l+1)$, we can show that

$$\frac{\Delta_i}{\Sigma_i^2} \leqslant \frac{\epsilon_2^2 + \rho_i \epsilon_1^2}{(\epsilon_2 + \rho_i \epsilon_1)^2}.$$

Thus

$$\rho_i \frac{\Delta_i}{\Sigma_i^2} \leqslant \rho_i \frac{\epsilon_2^2 + \rho_i \epsilon_1^2}{(\epsilon_2 + \rho_i \epsilon_1)^2}, \quad \text{with} \quad \epsilon_1 = \frac{1}{2l+1}, \epsilon_2 = \frac{1}{l+1}.$$

We can now use the fact that, if $0 < \lambda_1 < \lambda_2$ then

$$\lambda_1 \frac{\epsilon_2^2 + \lambda_1 \epsilon_1^2}{(\epsilon_2 + \lambda_1 \epsilon_1)^2} < \lambda_2 \frac{\epsilon_2^2 + \lambda_2 \epsilon_1^2}{(\epsilon_2 + \lambda_2 \epsilon_1)^2},$$

to show that

$$\rho_i \frac{\Delta_i}{\Sigma_i^2} \leqslant 2l \frac{\epsilon_2^2 + 2l\epsilon_1^2}{(\epsilon_2 + 2l\epsilon_1)^2} = \bar{c}.$$

Letting $\epsilon_1 = 1/(2l+1)$, $\epsilon_2 = 1/(l+1)$ in the preceding equality, we find

$$\max\left(\rho_i \frac{\Delta_i}{\Sigma_i^2}\right) = \theta \leqslant \bar{\theta} = \frac{4l^4 + 16l^3 + 12l^2 + 2l}{4l^4 + 16l^3 + 12l^2 + 2l + 8l^2 + 6l + 1} < 1.$$

Thus

$$\bar{\theta} = \frac{1}{1 + \dfrac{8l^2 + 6l + 1}{4l^4 + 16l^3 + 12l^2 + 2l}}$$

is a number smaller than 1, (which is independent of N) and can be used in (40), in place of θ to provide lower bounds for the duration of the game, in order that the stationary strategies are "good" for the finite time case. Notice, that for l large

$$\bar{\theta} \cong \frac{1}{1 + \dfrac{2}{l^2}} \cong 1 - \frac{2}{l^2},$$

and thus

$$t_f \geqslant \frac{2}{\log_{10}\left(1 - \dfrac{2}{l^2}\right)} \cong \frac{l^2}{2} \ln \epsilon.$$

So that for large l, we have $t_f \cong l^2/2 \ln(1 - \bar{p})$ where $\bar{p}$ is the desired probability of killing.

4. FIRST VERSION OF THE NON-STATIONARY GAME

Let Φ and Ψ be given strategies. Let the payoff $J(\Phi, \Psi)$ be the probability that R be killed at time T or before knowing the initial state. Let the stopping time

$$t_f = \begin{cases} \inf\{t; t \in \{1, \ldots, T\} & \text{and} \quad x_t = z_t\} \\ T \quad \text{if} \quad \forall\, t \in \{1, \ldots, T\} x_t \neq z_t. \end{cases}$$

We use the same payoff and stopping time in Sections 4 and 5.

We use dynamic programming to solve this game.

4.1. Set up

Let Φ, Ψ be given strategies. Let $W(x, t)$ be the probability that R be killed at time T or before when $x_t = x$. We have

$$W(x, t) = \sum_{u \in U_{ad}(x)} \left(p(u)q(u + x) + \sum_{v \in I_N; v \neq x + u} p(u)q(v)W(x + u, t + 1) \right).$$

R wants to minimize this probability and H to maximize. Isaacs' optimality principle gives us the optimal value

$$V(x, t) = \min_{p \in \Sigma_U} \max_{q \in \Sigma_N} \left(\sum_{u \in U_{ad}(x)} p(u)q(u + x) + \sum_{v \in I_N; v \neq x + y} p(u)q(v)V(x + u, t + 1) \right),$$

$$= \min_{p \in \Sigma_U} \max_{q \in \Sigma_N} p' B_{t+1}(x) q.$$

where $B_{t+1}(x)$ is a matrix of dimension less or equal to $N \times (N + 1)$.

Therefore, we have to solve a matrix game at each stage of the dynamic programming algorithm. The equivalence between solving such a problem and solving a linear programming problem gives us the existence of a mixed saddle point for this game.

4.2. Results and some properties

We show some properties of symmetry with respect to x, so we can study the game for x in $\{1, \ldots, N/2\}$ if N is even or x in $\{1, \ldots, (N+1)/2\}$ if N is odd.

Typical results are as follows for $l = 1$.

Let $a = V(x-1, t+1)$, $b = V(x, t+1)$ and $c = V(x+1, t+1)$. For the game value, we have

$$V(1, t) = \frac{b(1-c) + (1-b)}{(1-c) + (1-b)},$$

$$V(x, t) = \frac{1 + \dfrac{a}{(1-a)} + \dfrac{b}{(1-b)} + \dfrac{c}{(1-c)}}{\dfrac{1}{(1-a)} + \dfrac{1}{(1-b)} + \dfrac{1}{(1-c)}} \quad \text{for} \quad x \neq 1$$

and

$$1/3 \leqslant V(x, T-1) \leqslant 1/2 \qquad \forall x,$$

$$1/2 \leqslant V(x, t) \leqslant 1 \qquad \forall x \quad \text{and} \quad t \neq T-1,$$

$$V(x+1, t) \leqslant V(x, t) \leqslant V(x+1, t-1) \leqslant V(x, t-1) \qquad \forall (x, t).$$

So that the table of Vs against x and t can be easily computed. We give an example for $N = 12$:

	1	2	3	4	5	6
$T-1$	1/2	1/3	1/3	1/3	1/3	1/3
$T-2$	0.7140	0.6000	5/9	5/9	5/9	5/9
$T-3$	0.8330	0.7570	0.7140	0.7040	0.7040	0.7040
$T-4$	0.9000	0.8500	0.8200	0.8050	0.8050	0.8050
$T-5$	0.9400	0.9100	0.8800	0.8700	0.8688	0.8683
$T-6$	0.9640	0.9450	0.9260	0.9160	0.9127	0.9123

For the optimal strategies, we have

For R

$$\Phi_t^*[1](0) = \frac{1}{1 + \dfrac{1-b}{1-c}},$$

$$\Phi_t^*[1](1) = \frac{1}{1 + \dfrac{1-c}{1-b}}.$$

For H

$$\Psi_t^*[1](v) = \begin{cases} \Phi_t^*[1](1) & \text{if} \quad v = 1, \\ \Phi_t^*[1](0) & \text{if} \quad v = 2, \\ 0 & \text{otherwise.} \end{cases}$$

For R and $x \neq 1$

$$\Phi_t^*[x](-1) = \frac{1}{1 + \dfrac{1-a}{1-b} + \dfrac{1-a}{1-c}},$$

$$\Phi_t^*[x](0) = \frac{1}{1 + \dfrac{1-b}{1-a} + \dfrac{1-b}{1-c}},$$

$$\Phi_t^*[x](1) = \frac{1}{1 + \dfrac{1-c}{1-a} + \dfrac{1-c}{1-b}}.$$

For H and $x \neq 1$

$$\Psi_t^*[x](v) = \begin{cases} 1 - 2\Phi_t^*[x](v-x) & \text{if} \quad v = x-1, x, x+1, \\ 0 & \text{otherwise.} \end{cases}$$

At x and t fixed, the probability that R jumps in $x+1$ is less than the probability that R stays in x which is less than the probability that R jumps in $x-1$.

$$\Phi_t^*[x](1) \leqslant \Phi_t^*[x](0) \leqslant \Phi_t^*[x](-1).$$

At x fixed, $x \neq 1$, the probability that R jumps in $x-1$ decreases and the probability that R jumps in $x+1$ increases when t increases.

$$\Phi_t^*[x](-1) \leqslant \Phi_{t-1}^*[x](-1)$$

$$\Phi_t^*[x](1) \geqslant \Phi_{t-1}^*[x](1).$$

The properties of $(\Psi_t^*[x])_{(x,t)}$ are derived from the properties of $(\Phi_t^*[x])_{(x,t)}$.

5. SECOND VERSION OF THE NON-STATIONARY GAME

5.1. Definitions

Let now the bullet take two time steps to reach the wall. The game is described by (1), (3) and (4).

Let $U_t = (u_0, \ldots, u_t)$ and $V_t = (v_0, \ldots, v_t)$. We can remark that $Y_t = V_{t-1}$ and $X_t = (x_0, U_{t-1})$. The players' information are given by

—for H, (X_t, Y_t) or (x_0, U_{t-1}, V_{t-1}) for all t in $\{0, \ldots, T\}$,
—for R, (x_0, y_0, z_0) for $t = 0$ and X_t or (x_0, U_{t-1}) for all t in $\{1, \ldots, T\}$.

R knows exactly x at each time t, so we can introduce a distribution law Q_t on the space I_N for y. This law depends on

—the strategy of H at $t-1$, denoted by Ψ_{t-1},
—Q_{t-1},
—the information of R at $t-1$, x_{t-1}.

$Q_t(y)$ is the *a posteriori* probability that y_t equals y. Let E be defined by

$$E = \left\{ Q = (Q(0), \ldots, Q(N)) \in [0,1]^{N+1}; \quad \sum_{y=0}^{N} Q(y) = 1, Q(0) = 0 \quad \text{or} \quad Q(0) = 1 \right\}.$$

At time t, the strategies of the two players are defined by

—R's strategy depends on its state, x in I_N and Q a distribution law in I_N

$$\Phi_t[x, Q] = p_t,$$

—H's strategy depends on R's state, x in I_N, the control V_{t-1} that he has chosen at time $t-1$ or y in I_N and Q a distribution law in I_N

$$\Psi_t[x, y, Q] = q_t.$$

Then, we can write explicitly the dependance between Q_t and Q_{t-1}

$$Q_t(y) = \sum_{j \in I_N} Q_{t-1}(j) \Psi_{t-1}[x, j, Q_{t-1}](y), \quad \text{for} \quad t \geqslant 2,$$

and

$$Q_1(y) = \Psi_0[x_0, 0, Q_0](y) \quad \text{with} \quad Q_0 \in E; \quad Q_0(0) = 1,$$

denoted by $Q_t(y) = F(Q_{t-1}, \Psi_{t-1})(y)$.

Let Φ and Ψ be given strategies. Let $W_0^{\Phi\Psi}(x, y, z, Q, t)$ be the probability for H that R be killed at time T or before when $x_t = x$, $y_t = y$, $z_t = z$, $Q_t = Q$. Now, we have

$$W_0^{\Phi\Psi}(x, y, z, Q, t) = \begin{cases} \sum_{u \in U_{ad}(x)} \sum_{v \in I_N} \Phi_t[x, Q](u)\Psi_t[x, y, Q](v) W_0^{\Phi\Psi}(x + u, v, y, F(Q, \Psi_t), t + 1) \\ \qquad \text{if} \quad x \neq z \quad \text{and} \quad t < t_f, \\ 0 \qquad \text{if} \quad x \neq z \quad \text{and} \quad t = t_f \quad (\text{and then, } t_f = T), \\ 1 \qquad \text{if} \quad x = z \quad (\text{and then } t_f = t). \end{cases}$$

We remark that the first term of the right-hand side of the equality does not depend on z, so we denote it by $W_1^{\Phi\Psi}(z, y, Q, t)$, furthermore by convention, we set $W_1^{\Phi\Psi}(x, y, Q, t_f) = 0$.

Let $W_2^{\Phi\Psi}(x, Q, t)$ be the probability for R to be killed at time T or before when $x_t = x$, $Q_t = Q$. We have

$$W_2^{\Phi\Psi}(x, Q, t) = \sum_{y \in I_N} Q(y) W_1^{\Phi\Psi}(x, y, Q, t) \quad \text{for} \quad t \geqslant 1,$$

and

$$W_2^{\Phi\Psi}(x_0, Q_0, 0) = W_1^{\Phi\Psi}(x_0, 0, Q_0, 0) \quad \text{for} \quad t = 0.$$

We define

—the H cost function by

$$V(x, y, Q, t) = \max_{q \in \Sigma_N} \sum_{u \in U_{ad}(x)} \sum_{v \in I_N} q(v)\hat{\Phi}_t[x, Q](u) V_y(x + u, v, F(Q, \hat{\Psi}_t), t + 1), \tag{41}$$

where

$\hat{\Psi}_t[x, y, Q] = q^*$ belongs to the set of arguments of the maximum sought, the function V_z is defined by

V_z: $I_N^2 \times E \times \{0, \ldots, t_f\} \to B$, B bounded

$$(x, y, Q, t) \to V_z(x, y, Q, t) = \begin{cases} V(x, y, Q, t) & \text{if} \quad x \neq z \quad \text{and} \quad t < t_f, \\ 0 & \text{if} \quad x \neq z \quad \text{and} \quad t = t_f, \\ 1 & \text{if} \quad x = z, \end{cases}$$

—the R cost function by

$$\begin{aligned} \bar{V}(x, Q, t) &= \sum_{y \in I_N} Q(y) V(x, y, Q, t) \\ &= \min_{p \in \Sigma_U} \sum_{y \in I_N} Q(y) \sum_{u \in U_{ad}(x)} \sum_{v \in I_N} p(u)\hat{\Psi}_t[x, y, Q](v) V_y(x + u, v, F(Q, \hat{\Psi}_t), t + 1), \end{aligned} \tag{42}$$

where $\hat{\Phi}_t[x, Q] = p^*$ belongs to the set of arguments of the minimum sought.

Remark: The equality (42) is equivalent to this equality:

$$\bar{V}(x, Q, t) = \min_{p \in \Sigma_U} \max_{q \in \Pi_{i=1}^N \Sigma_N} \sum_{y \in I_N} Q(y) \sum_{u \in U_{ad}(x)} \sum_{v \in I_N} p(u) q(y)(v) V_y(x + u, v, F(Q, \hat{\Psi}_t), t + 1).$$

Remark: Thus, we have $V_z(x, y, Q, t) = W_0^{\hat{\Phi}\hat{\Psi}}(x, y, z, Q, t)$, $V = W_1^{\hat{\Phi}\hat{\Psi}}$ and $\bar{V} = W_2^{\hat{\Phi}\hat{\Psi}}$.

Let Q^* be a solution of (43):

$$\begin{aligned} &Q_{t+1}^*(\cdot) = \sum_{y \in I_N} Q_t^*(y)\hat{\Psi}_t[x, y, Q_t^*](\cdot) \quad \text{for} \quad t \geqslant 1, \\ &Q_1^*(\cdot) = \hat{\Psi}_0[x_0, 0, Q_0^*](\cdot) \quad \text{with} \quad Q_0^* \quad \text{such that} \quad Q_0^*(0) = 1. \end{aligned} \tag{43}$$

Let $\Phi_t^*(x_0, U_{t-1}) = \hat{\Phi}_t[x_t, Q_t^*]$ and $\Psi_t^*(x_0, U_{t-1}, V_{t-1}) = \hat{\Psi}_t[x_t, y_t, Q_t^*]$ for all x_t, y_t in I_N.

5.2. Theorem

Theorem 3

If it exists Q^*, Φ^*, Ψ^*, V from $I_N^2 \times E \times \{0, \dots, t_f\}$ to B_0, B_0 bounded, $\bar{V}$ form $I_N \times E \times \{0, \dots, t_f\}$ to B_1, B_1 bounded verifying the equalities (41)–(43) then

(i) $J(\Phi^*, \Psi^*) = V(x_0, 0, Q_0^*, 0) = \bar{V}(x_0, Q_0^*, 0)$,

(ii) For all admissible strategies Φ and Ψ, we have the following inequalities

$$J(\Phi^*, \Psi) \leqslant J(\Phi^*, \Psi^*) \leqslant J(\Phi, \Psi^*).$$

Proof. First, notice that (41) implies that $V = W_1^{\Phi^*\Psi^*}$ and $\bar{V} = W_2^{\Phi^*\Psi^*}$. It follows the equality (i) of the theorem.

Take now an arbitrary $\bar{\Psi}$. $\hat{\Phi}$ and $\bar{\Psi}$ together generate trajectories (depending on w). Let $\{\bar{x}_t\}$, $\{\bar{y}_t\}$, $\{\bar{q}_t\}$ be such a trajectory. Also, let $\bar{Q}_t$ be generated by (43) along this trajectory (i.e. placing $\hat{\Psi}$ and not $\bar{\Psi}$ in the equation for Q_{t+1}). Let t such that $t < t_f$.

We have

$$E^{\hat{\Phi}\bar{\Psi}}(W_0^{\hat{\Phi}\hat{\Psi}}(x_{t+1}, y_{t+1}, z_{t+1}, \bar{Q}_{t+1}, t+1)/X_t, Y_t)$$

$$= \begin{cases} \sum\limits_{u \in U_{ad}(x)} \sum\limits_{v \in I_N} \bar{q}_t(v)\hat{\Phi}_t[x_t, \bar{Q}_t](u)\, W_0^{\hat{\Phi}\hat{\Psi}}(x_t + u, v, y_t, F(\bar{Q}_t, \hat{\Psi}_t), t+1) & \\ \qquad \text{if} \quad x_{t+1} \neq z_{t+1} \quad \text{and} \quad t < t_f, & \\ 0 \qquad \text{if} \quad x_{t+1} \neq z_{t+1} \quad \text{and} \quad t = t_f, & \\ 1 \qquad \text{if} \quad x_{t+1} = z_{t+1}. & \end{cases}$$

$$\leqslant W_0^{\hat{\Phi}\hat{\Psi}}(x_t, y_t, z_t, \bar{Q}_t, t) \text{ by definition of } \hat{\Psi}.$$

Then

$$\begin{aligned} J(\hat{\Phi}, \bar{\Psi}) &= E^{\hat{\Phi}\bar{\Psi}}(W_0^{\hat{\Phi}\hat{\Psi}}(x_{t_f}, y_{t_f}, z_{t_f}, \bar{Q}_{t_f}, t_f)/x_0), \\ &= E^{\hat{\Phi}\bar{\Psi}}(\dots E^{\hat{\Phi}\bar{\Psi}}(W_0^{\hat{\Phi}\hat{\Psi}}(x_{t_f}, y_{t_f}, z_{t_f}, \bar{Q}_{t_f}, t_f)/X_{t_f-1}, Y_{t_f-1}) \dots /x_0), \\ &\quad \text{with the increasing algebra property,} \\ &\quad \leqslant W_0^{\hat{\Phi}\hat{\Psi}}(x_0, 0, 0, \bar{Q}_0, 0) = W_1^{\hat{\Phi}\hat{\Psi}}(x_0, 0, \bar{Q}_0, 0), \\ &= J(\hat{\Phi}, \hat{\Psi}). \end{aligned}$$

So, we have $J(\Phi^*, \bar{\Psi}) \leqslant J(\Phi^*, \Psi^*)$ for all admissible strategy $\bar{\Psi}$.

Take now an arbitrary $\bar{\Phi}$. $\bar{\Phi}$ and $\hat{\Psi}$ together shall generate trajectories. Let $\{\bar{x}_t\}$, $\{\bar{p}_t\}$ be such a trajectory. Let $\bar{Q}_t$ be generated by (43) along this trajectory. Then, $(\bar{Q}_t)$ is the conditional distribution law of y_t knowing X_t. Let t such that $t < t_f$. By definition of $W_2^{\Phi\Psi}$, we have

$$W_2^{\Phi\Psi}(x_{t+1}, Q_{t+1}, t+1) = E(W_1^{\Phi\Psi}(x_{t+1}, y_{t+1}, Q_{t+1}, t+1)/X_{t+1}).$$

Then

$$\begin{aligned} E(W_2^{\Phi\Psi}(x_{t+1}, Q_{t+1}, t+1)/X_t) &= E(E(W_1^{\Phi\Psi}(x_{t+1}, y_{t+1}, Q_{t+1}, t+1)/X_{t+1})/X_t), \\ &= E(W_1^{\Phi\Psi}(x_{t+1}, y_{t+1}, Q_{t+1}, t+1)/X_t) \quad \text{since} \\ &\qquad \sigma(X_t) \subset \sigma(X_{t+1}). \end{aligned}$$

So, we have

$$\begin{aligned} &E^{\bar{\Phi}\hat{\Psi}}(W_2^{\hat{\Phi}\hat{\Psi}}(x_{t+1}, \bar{Q}_{t+1}, t+1)/X_t) \\ &\quad = E^{\bar{\Phi}\hat{\Psi}}(W_1^{\hat{\Phi}\hat{\Psi}}(x_{t+1}, y_{t+1}, \bar{Q}_{t+1}, t+1)/X_t), \\ &\quad = \sum_{y \in I_N} \bar{Q}_t(y) \sum_{u \in U_{ad}(x_t)} \sum_{v \in I_N} \bar{p}_t(u)\hat{\Psi}_t[x_t, y, \bar{Q}_t](v)\, W_0^{\hat{\Phi}\hat{\Psi}}(x_t + u, v, y, F(\bar{Q}_t, \hat{\Psi}_t), t+1), \end{aligned}$$

$$\geqslant \sum_{y \in I_N} \bar{Q}_t(y) \sum_{u \in U_{ad}(x_t)} \sum_{v \in I_N} \hat{\Phi}_t[x_t, \bar{Q}_t](u) \hat{\Psi}_t[x_t, y, \bar{Q}_t](v)$$

$$\times W_0^{\hat{\Phi}\hat{\Psi}}(x_t + u, v, y, F(\bar{Q}_t, \hat{\Psi}_t), t+1), \text{ by definition of } \hat{\Phi},$$

$$= \sum_{y \in I_N} \bar{Q}_t(y) W_1^{\hat{\Phi}\hat{\Psi}}(x_t, y, \bar{Q}_t, t),$$

$$= W_2^{\hat{\Phi}\hat{\Psi}}(x_t, \bar{Q}_t, t).$$

Then

$$J(\bar{\Phi}, \hat{\Psi}) = E^{\bar{\Phi}\hat{\Psi}}(W_2^{\hat{\Phi}\hat{\Psi}}(x_{t_f}, \bar{Q}_{tf}, t_f)/x_0),$$

$$= E^{\bar{\Phi}\hat{\Psi}}(\ldots E^{\bar{\Phi}\hat{\Psi}}(W_2^{\hat{\Phi}\hat{\Psi}}(x_t, \bar{Q}_{t_f}, t_f)/X_{t_f-1}) \ldots /x_0) \text{ with the increasing algebra property,}$$

$$\geqslant E(W_2^{\hat{\Phi}\hat{\Psi}}(x_{t_f}, \bar{Q}_{t_f}, t_f)/x_0)$$

$$= J(\hat{\Phi}, \hat{\Psi}).$$

So, we have $J(\Phi^*, \Psi^*) \leqslant J(\bar{\Phi}, \Psi^*)$ for all admissible strategy $\bar{\Phi}$.

The set of inequations (41) and (42) is a saddle-point in p, q, with $\hat{\Psi}$ fixed in V, which must coincide with the optimal q in the saddle-point. The algorithm amounts therefore to solving what is essentially a fixed point problem $\hat{\Psi} = q^*(\hat{\Psi})$ at each point in the extended state space (x, Q, t). This is a formidable computational problem for large values of N, but can be tackled for small values. This is currently being attempted.

REFERENCE

1. J. Foreman, The Princess and the Monster game on the circle. *SIAM J. Control Optim.* **15**, 841–856 (1977).