Computation of equilibrium points of delayed partial information pursuit evasion games

P.BERNHARD

Institut National de Recherches en Informatique et en Automatique
Route des lucioles, 06560 Valbonne
France

Abstract.

Motivated by the "Hunter and rabbit" game [1], we study a class of discrete time, discrete state games with partial information, for which we show that the calculation of the saddlepoint at each step of the dynamic programming algorithm decouples into two trivial finite optimization problems, thus avoiding a simplex at each step of a large number of fixed point problems.

1. Introduction.

The motivation for this research is still, as in [1], [2], [3], the following Hunter and Rabbit game. A rabbit jumps back and forth along a linear wall, divided into a finite number N of discrete possible positions. It tries to avoid being killed by a Hunter, but has no information on what the latter does. At each time instant of this discrete time hunting party, the Hunter, who sees the rabbit, choses whether to shoot, and where to aim at. The important feature of this game being that the bullets take several time steps to fly from the hunter to the rabbit.

Six integer parameters specify this game. They are

- The number N of possible positions for the rabbit,
- the length l of its allowed jump, (assumed symmetrical, left and right),
- the time of flight τ of the bullet, (assumed independent of where it is aimed at),
- the number ν of shots available to the hunter,
- the lethal radius ρ of the bullets,
- the total duration T of the game.

For finite T, the only case we shall consider here, the obvious payoff is the probability that the rabbit will be killed, that the rabbit strives to minimize, and the hunter to maximize. In [2] we also considered the infinite time game, in which the payoff is the life expectation of the rabbit.

We consider it a quite intriguing fact that, even for small values of the parameter, and, say, $\tau = 2$, we do not know of a practically feasible way of computing the optimal strategies of this exceedingly simple game.

In paper [1], we gave a general theorem about a larger class of dynamical partial information games, and reported limited success in implementing the algorithm. We actually had to give up, due to the excessive computational burden of that method, and to associated problems of convergence of the fixed point algorithm when the discretization step for Q was not taken small enough. In the present paper, we report ways of drastically reducing the amount of computation required, but with no theoretical progress on the question of convergence.

It should be pointed out that the solution sought in itself, a pair of closed loop mixed strategies, can only be described by a very large number of numbers. Therefore, unless a closed form solution can be found, which is extremely unlikely, the total amount of computation required has to be large. The challenge is to keep it within feasible limits.

2. The games considered.

We shall consider games of pusuit evasion type, and as previously in discrete time and state. Let therefore time $t \in \{0, 1, \ldots, T\}$. The state x_t of the game will be assumed to separate into

$$x_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}$$

where $y_t \in \mathcal{Y}_t$ and $z_t \in \mathcal{Z}_t$ are the respective states of $\mathcal{R}$abbit ($\mathcal{R}$) and $\mathcal{H}$unter ($\mathcal{H}$). $\mathcal{Y}_t$ and $\mathcal{Z}_t$ are finite sets. The dynamics are

$$\begin{aligned} y_{t+1} &= f_t(y_t, u_t), \quad u_t \in \mathcal{U}_{ad}(y_t), \\ z_{t+1} &= g_t(z_t, v_t), \quad v_t \in \mathcal{V}_{ad}(z_t). \end{aligned}$$

However we may, with no loss of generality reformulate them as

$$\begin{aligned} y_{t+1} &= u_t, \qquad u_t \in \mathcal{U}_t(y_t), \\ z_{t+1} &= v_t, \qquad v_t \in \mathcal{V}_t(z_t). \end{aligned} \tag{1}$$

The change in notations is obvious. We shall assume that any state constraint has been included in $\mathcal{U}_t$ and $\mathcal{V}_t$.

Capture is defined via a capture set $C_t \subset \mathcal{Y}_t \times \mathcal{Z}_t$ by $x_t \in C_t$, and the game terminates at t_1 given by

$$t_1 = \min\{t \,|\, x_t \in C_t,\, T\}. \tag{2}$$

The payoff J of the game is one if $t_1 < T$, (rabbit is killed), and zero otherwise.

We shall, of course, need mixed controls:

$$\begin{aligned} p_t(u) &= \text{Probability}(u_t = u), \\ q_t(v) &= \text{Probability}(v_t = v), \end{aligned}$$

and we shall call $\mathcal{P}_t(y)$ and $\mathcal{Q}_t(z)$ the sets of admissible mixed controls, i.e. such that support $p_t \subset \mathcal{U}_t(y)$ and support $q_t \subset \mathcal{V}_t(z)$ respectively. With sequences of mixed controls $\{p_t\}$ and $\{q_t\}$ we associate the payoff $J(\{p_t\}, \{q_t\}) = \mathrm{E}J$, the mathematical expectation of J as defined above.

Finally, we must describe the information available to each player at the time it chooses its control. We shall assume all along that both players share an exact knowledge of both initial states, although this will not be emphasized by subsequent notations. This allows us to deal with a saddle point value, instead of Cournot-Nash equilibria. In addition, $\mathcal{R}$ remembers its sequence of past states $Y_t = \{y_0, y_1, \ldots, y_t\}$, while $\mathcal{H}$ knows both Y_t and its own sequence of past states $Z_t = \{z_0, z_1, \ldots, z_t\}$. We are then looking for mixed strategies of the form

$$\begin{aligned} p_t &= \phi_t[Y_t], \\ q_t &= \psi_t[Y_t, Z_t]. \end{aligned} \tag{3}$$

Particularizing this setup to the original Rabbit and Hunter game is straightforward. Notice that we did not include a second capture set on which $J = 0$ to take into account termination of the game by exhaustion of Hunter's amunitions. As a matter of fact, we do not need terminate the game at that time, we may istead let it go, with $\mathcal{V}_t$ restricted to the single choice "do not shoot". (Although this is not computationally efficient.)

3. The general theorem.

According to the theory in [1], we shall introduce an extra state Q_t, which has the dimension of a probability law over Z_t, and represents the conditional probability law of z_t knowing Y_t if ever $\mathcal{H}$ plays according to its strategy $\hat{\psi}_t$ to be defined. The strategies sought shall depend on past values of y and z only through Q_t. This dependance will be denoted using square brackets: $\psi_t[y_t, z_t, Q_t]$ for instance. The propagation formula for Q as given in [1] simplifies here in

$$Q_{t+1}(\zeta) = \sum_z \psi_t[y_t, z_t, Q_t](\zeta) Q_t(z).$$

(We have omitted the range of variation of the summation variable z, which clearly is Z_t. We shall likewise, in the sequel, systematically omitt ranges of variation of the summation variables each time these ranges are the whole spaces these variables are in: Y_t, Z_t, U_t, V_t. It should be emphisized, though, that the last two introduce a dependance of the sum on their argument y or z.)

As it stands, there is no conditioning on Y_t in the above formula, as Y_t provides no information on what $\mathcal{H}$ has done. But actually, for games with final time unspecified, i.e. depending on the controls, there is an extra piece of information available to $\mathcal{R}$ at the time it chooses its control, and that it should use. It is "cogito ergo sum" [4], i.e. I am still alive. This rules out past behaviours of $\mathcal{H}$ that would have killed it at current time. We now set out to use this information.

Let $\Delta_t(y)$ be the section parallel to Z_t of C_t at y, i.e.

$$(y, z) \in C_t \iff z \in \Delta_t(y).$$

We need furthermore to introduce the set $D_t(y)$ of $\mathcal{H}$ states which are fatal to $\mathcal{R}$ if chooses y as its next state, whatever $\mathcal{H}$ does at current time:

$$D_t(y) = \{z | V_t(z) \subset \Delta_{t+1}(y)\}.$$

With these notations we have, using Bayes' rule:

$$Q_{t+1}(\zeta) = \left[\sum_{z \notin D_t(y_{t+1})} Q_t(z)\right]^{-1} \sum_{z \notin D_t(y_{t+1})} \frac{\psi_t[y_t, z_t, Q_t](\zeta) Q_t(z)}{\sum_{\mu \notin \Delta_{t+1}(y_{t+1})} \psi_t[y_t, z_t, Q_t](\mu)} \tag{4a}$$

if $\zeta \notin \Delta_{t+1}(y_{t+1})$,

$$Q_{t+1}(\zeta) = 0 \quad \text{if} \quad \zeta \in \Delta_{t+1}(y_{t+1}). \tag{4b}$$

This relation shall be written

$$Q_{t+1} = F(Q_t, y_t, \psi_t, y_{t+1}). \tag{4c}$$

If the first square bracket in (4a) is zero, $\mathcal{R}$ should not consider moving to that y_{t+1} if it can avoid it, and the game is essentially over if it cannot.

The dynamic programming procedure of [1] uses the relations

$$V_t(y,z,Q) = \max_{q\in\mathcal{Q}_t(z)} \sum_v \sum_u V_{t+1}(u,v,F(Q,y,\hat{\psi}_t,u))\hat{\phi}_t[y,Q](u)q(v), \quad (5)$$

$$\hat{\psi}_t[y,z,Q] \in \text{Argmax(above)}, \quad (6)$$

$$\sum_z V_t(y,z,q)Q(z) = \quad (7)$$

$$\min_{p\in\mathcal{P}_t} \sum_u \sum_z \sum_v V_{t+1}(u,v,F(Q,y,\hat{\psi}_t,u))\hat{\psi}_t[y,z,Q](v)Q(z)p(u),$$

$$\hat{\phi}_t[y,Q] \in \text{Argmax(above)}, \quad (8)$$

initialized at t_1 by

$$\forall(Q,t), \quad \forall(y,z) \in C_t, \quad V_t(y,z,Q) = 1, \quad (9)$$

$$\forall Q, \quad \forall(y,z) \notin C_T, \quad V_T(y,z,Q) = 0. \quad (10)$$

The theorem states:

Theorem. *If there exists a real function V and strategies $\hat{\phi}$ and $\hat{\psi}$ satisfying (5) to (8), with F given by (4), for all t,y,z,Q, and (9), (10), then the strategies obtained by placing $\hat{\psi}$ in (4) and using the resulting $\hat{Q}$ in $\hat{\phi}$ and $\hat{\psi}$ constitute a saddle point for the game considered.*

4. Separation of max and min problems

Introduce the notation

$$\bar{V}_{t+1}(y,Q) = \sum_z V_t(y,z,Q)Q(z).$$

This will be considered as $\mathcal{R}$'s value function, although Q may not be a true conditional probability. As a matter of fact, since the strategy $\hat{\psi}$ is the worst possible against $\hat{\phi}$, and since then, Q generated by (4) *is* the conditonal probability for z, this actually is the best guaranteed payoff to $\mathcal{R}$.

Using condition (9) in (7), the latter may be expanded into

$$\bar{V}_t(y,Q) = \min_p \sum_u \sum_z p(u) \Bigg[\sum_{v \in \Delta_{t+1}(u)} \hat{\psi}_t[y,z,Q](v)Q(z) + \sum_{v \notin \Delta_{t+1}(u)} V_{t+1}(u,v,F(Q,y,\hat{\psi}_t,u))\hat{\psi}_t[y,z,Q](v)Q(z) \Bigg].$$

Now, for each u, for the z's that are in $D_t(u)$ all v's are in $\Delta_{t+1}(u)$. Then, the $\hat{\psi}_t[\cdots](v)$ sum up to one. So that the first summation above can again be expanded into two terms, yielding

$$\bar{V}_t(y,Q) = \min_p \sum_u p(u) \Bigg[\sum_{z \in D_t(u)} Q(z) + \sum_{z \notin D_t(u)} \sum_{v \in \Delta_{t+1}(u)} \hat{\psi}_t[y,z,Q](v)Q(z) + \sum_{z \notin D_t(u)} \sum_{v \notin \Delta_{t+1}(u)} V_{t+1}(u,v,F(Q,y,\hat{\psi}_t,u))\hat{\psi}_t[y,z,Q](v)Q(z) \Bigg].$$

The Rabbit and Hunter game of the introduction is a typical instance of a *delayed* game, i.e. one where the Hunter's choice of action at current time t does not influence capture at next step $t+1$, but only at a later stage, specifically at $t+\tau$. We shall therefore introduce th following:

Definition. *A game of the form (1)(2) will be called a delayed game for the second player if,* $\forall t$, $\forall y \in \mathcal{Y}_{t+1}$, *and* $\forall z \in \mathcal{Z}_t$, *either* $\{y, V_{t+1}(z)\} \subset C_{t+1}$, *or* $\{y, V_{t+1}(z)\} \cap C_{t+1} = \emptyset$

For a delayed game, formula (4a) simplifies into

$$Q_{t+1}(\varsigma) = \left[\sum_{z \notin D_t(y_{t+1})} Q_t(z) \right]^{-1} \sum_{z \notin D_t(y_{t+1})} \psi_t[y_t, z_t, Q_t](\varsigma)Q_t(z),$$

since the denominator in (4a) is identically one. Moreover, the middle summation sum in the last formula for $\bar{V}_{t+1}$ disappears, and the last one simplifies, using the new form of (4). Introduce the natural notations

$$\sum_{z \in D_t(u)} Q(z) = Q(D_t(u))$$

and $F(Q, y, \hat{\psi}_t, u) = \hat{Q}_{t+1}$. Formula (7) now reads

$$\bar{V}_t(y, Q) = \min_{p \in P_t} \sum_u p(u) \left[Q(D_t(u)) + (1 - Q(D_t(u))) \bar{V}_{t+1}(u, \hat{Q}_{t+1}) \right]$$

Finally, this itself is known to be equivalent to

$$\bar{V}_t(y, Q) = \min_u \left[Q(D_t(u)) + (1 - Q(D_t(u))) \bar{V}_{t+1}(u, \hat{Q}_{t+1}) \right]. \tag{11}$$

Any p weighting only the minimal elements provides the required minimum in p.

Equation (11) could have been derived directly from probabilistic considerations. It is nevertheless important to check that it is nothing else than the original equation (7) particularized to the delayed game. The important point, of course, is that in (11), $\hat{\psi}_t$ enters only through $\hat{Q}$, an entry which is anyway the subject of a fixed point problem.

We still have a dynamic algorithm in (y, z, Q) space, where each step involves the solution of a fixed point problem. At (t, y, Q) fixed, let r denote the vector made of all the values of the $\psi_t[y, z, Q]$ vectors for z ranging over Z_t. For a fixed r, we can compute an "optimal" p, denoted $\hat{p}(r)$, in (11), and placing r for $\hat{\psi}_t$ and $\hat{p}(r)$ for $\hat{\phi}_t[z, Q]$ in (5), we get a $\hat{q}(r)$ as the argmax. The fixed point problem to solve is: find $\hat{r}$ such that $\hat{q}(\hat{r}) = \hat{r}$. If a solution can be found, this $\hat{r}$ will stand for $\hat{\psi}_t[y, \cdot, Q]$, and $\hat{p}(\hat{r})$ for $\hat{\phi}_t[y, Q]$.

Whatever the solution technique used for this fixed point problem is, it involves far less computation than the original algorithm, since the two finite extremalization problems of this new one replace a large matrix saddle point calculation, i.e. a large simplex algorithm.

5. Simplifications.

Further simplifications are possible, at the expense of optimality. Let us first notice that we may of course multiply borh sides of (5) by $Q(z)$, since this number is nonnegative, and sum over z. Let again r denote the set of values of $\psi[y,z,Q]$ as z ranges over Z_t. The function F depends on ψ only through r. The above weighted sum process yields

$$\bar{V}_{t+1}(y,Q) = \max_r \sum_u \hat{\phi}_t[y,Q](u)\Big[Q(D_t(u))+ \\ (1-Q(D_t(u)))\sum_v V_t(u,v,F[\hat{r}])F[r](v)\Big]. \quad (12)$$

Here, of course, $F[r]$ stands for $F(Q,y,r,u)$.

Equations (11) and (12) may be the basis of various suboptimal schemes, involving a drastically reduced amount of computation.

The simplest case is when we are mainly interested in a strategy for $\mathcal{R}$. We can decide on an arbitrary, but sensible, probability distribution for z. Buid the corresponding r and use this to compute $\hat{Q}$ (which still depends on u) in (11). Then we get a suboptimal strategy ϕ via a simple one-player dynamic programming in (y,Q) space.

Of course, the resulting ϕ and the same r may be used together with (12) to compute a suboptimal strategy for $\mathcal{H}$, still via a dynamic programming in the (comparatively) reduced (y,Q) space, although this yields a strategy ψ actually depending on z also, since a whole r is computed at each step.

More elaborate schemes can be devised, that progress towards the solution of the full fixed point problem, but stop before full convergence has been obtained. Such schemes are currently being investigated.

Bibliography.

[1] 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**, 1987.

[2] P. Bernhard, A. L. Colomb and G. Papavassilopoulos, "Rabbit and Hunter game: two discrete stochastic formulations." *Comput. Math. Applic.* **13**, No 1–3, pp 205–225, 1987.

[3] A. L. Colomb, "Étude de jeux à deux joueurs en information incomplète: le jeu du chasseur et du lapin." Thèse, Université de Provence, Marseille, France, 1986.

[4] R. Descartes, "discours de la méthode," Leyden, 1637.