Saddle point conditions for a class of stochastic dynamical games with imperfect information

Pierre Bernhard, Annie-Laure Colomb

INRIA

UNITÉ DE RECHERCHE
INRIA-SOPHIA ANTIPOLIS

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France

Tél. (1) 39 63 55 11

Rapports de Recherche

N° 656

SADDLE POINT CONDITIONS FOR A CLASS OF STOCHASTIC DYNAMICAL GAMES WITH IMPERFECT INFORMATION

Pierre BERNHARD
Annie-Laure COLOMB

Avril 1987

Point selle pour une classe de jeux dynamiques stochastiques a information imparfaite

Pierre Bernhard et Annie-Laure Colomb

Institut National de Recherche en informatique et en Automatique
Route des Lucioles
Sophia-Antipolis
06560 Valbonne (France)

Résumé

On considère des jeux dynamiques non linéaires à deux joueurs où un des joueurs ne dispose que d'une mesure partielle, éventuellement bruitée, de l'état, tandis que l'autre à une information parfaite (causale) sur l'état, ainsi que sur la mesure et la commande de son adversaire. Nous donnons un algorithme du type programmation dynamique dans un espace étendu, qui, quand il réussit, donne un point selle en stratégies mixtes. Nous utilisons ce résultat pour donner un équilibre en stratégies prudentes pour le cas où les deux joueurs ont une information imparfaite. On examine enfin l'utilisation de la même approche pour le jeu différentiel continu linéaire quadratique gaussien, retrouvant ainsi le résultat de Behn et Ho.

Saddle point conditions for a class of stochastic dynamical games with imperfect information

Pierre Bernhard and Annie-Laure Colomb

Institut National de Recherche en informatique et en Automatique
Route des Lucioles
Sophia-Antipolis
06560 Valbonne (France)

ABSTRACT

We consider non-linear dynamical games where one player has full (causal) information on the state, the opponent's measurements and control, while the other one only has partial, possibly noise corrupted, state information. We give a dynamic programming-like algorithm in an extended space which, when successful, yields a saddle point, usually in mixed strategies. We use this to provide an equilibrium in safe strategies for the case where both players are restricted to imperfect state information. We then investigate the use of the same approach for continuous time differential games, restricting our attention to the linear quadratic gaussian case to get a usable result, thus, in effect, revisiting Behn and Ho's problem.

Introduction

It has long been recognised that the second guessing problem, which makes partial information two player dynamical games essentially untractable [1], [2], can be worked around in two extreme cases: either when one player has full information [3],[4],[5], or when one player has no information [5]. Other simple cases that avoid the second guessing altogether are when both players know the opponent's control [6], or when the information algebra does not depend on the controls [7].

Here we use the first of these situations in a nonlinear context to solve a discrete game with mixed strategies. In the last section, we show that this approach is closely related to that of Rhodes and Luenberger in [5].

1. Rabbit and Hunter

1.1 The general setup

The whole game described is in discrete time, and happens within a finite time interval $\{0,1,\ldots,T\}$. A rabbit jumps left and right along a wall. Its abscissa with respect to a specified origin will be called y, and the size of his jump u. Both are discrete, and take their values in finite sets $Y = \{1,2,\ldots,N\}$ and $U_{ad}(y) \subset \mathbf{Z}$ respectively, with $y + U_{ad}(y) \subset Y$, for all y in Y. Hence Rabbit's dynamics are discribed by the following equations:

$$y_{t+1} = y_t + u_t, \qquad y_t \in Y = \{1,2,\ldots,N\}, \quad u_t \in U_{ad}(y_t). \tag{1}$$

A hunter stands at a distance, and has a given number ν of shots available.Let v_t be its decision at time t, with $v_t = 0$ meaning that he does not shoot, (a possibility not needed if $\nu \geq T$), $v_t = \eta \in Y$ meaning that he shoots, aiming at the point of abscissa η along the wall. We shall assume perfect precision: Hunter knows y exactly and shoots where he wants. But the generalization of section 2 below would allow us to raise in part these restrictions.

Of course, Hunter would like to kill Rabbit, and he has to do so not later than time T, while Rabbit would like to survive until time T. (And then, presumably, until the next hunting party comes by).

Two very different situations arise depending on how long it takes for the bullet to fly from Hunter to Rabbit.

1.2 The simple case.

The simple case is when the bullet time of flight is one step of time. Then the game can be written as a simple, complete information game. Let

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

Capture is defined by

$$x_t - y_t = 0, \tag{3}$$

and the game ends either at capture time or when $t = T$, whichever happens first.

Actually, the game also ends whenever Hunter has exhausted his amunitions, but Rabbit might not be aware of it. If there is only one shot available, or more than T, (x_t, y_t) is a state for this game. Otherwise, we introduce an extra state variable to count the shots:

$$z_{t+1} = z_t + \mathbf{1}_Y(v_t),$$

where $\mathbf{1}_Y(v) = 1$ if $v \in Y$, and 0 otherwise (i.e. if $v = 0$) and the game ends when $z_t = \nu$.

As a simple example, assume that $\nu = 1$. It is clear that this game has no pure strategy saddle point. As a matter of fact, if Rabbit had an optimal pure strategy, Hunter could compute it as well, and shoot precisely at the position this strategy makes Rabbit jump to, making it a sure win for Hunter, a contradiction.

Introduce therefore the following mixed strategies.

Let $p \in \Sigma_U$ and $q \in \Sigma_V$ be Rabbit's and Hunter's choices respectively, with

$$p_i = \text{Probability}(u = i), \quad q_j = \text{Probability}(v = j),$$

Σ_U and Σ_V being the relevant simplices.

The solution of this game in mixed strategies via dynamic programming is straightforward. Let $V(y,t)$ be the probability that Rabbit be killed before time T if it is in location y at time t. Both V and optimal strategies can be computed according to

$$V(y,t) = \min_{p \in \Sigma_U} \max_{q \in \Sigma_V} \sum_{ij} p_i V_{ij} q_j,$$

$$V_{ij} = \begin{cases} 1 & \text{if } j = y + i, \\ 0 & \text{if } j \neq y + i \text{ and } j \neq 0, \\ V(y+i, t+1) & \text{otherwise,} \end{cases}$$

and $V(y,T) = 0$.

1.3 The partial information case.

Assume now that the bullet's time of flight is two time steps. Of course, Rabbit does not know whether Hunter has shot, nor a fortiori, where he aimed at. The situation is now far more complicated, because at each instant of time, Rabbit must account for the possibility that a bullet be flying at that time, and the point Hunter is likely to have aimed at is itself a function of where Rabbit itself was at the previous time step. The problem gets only more complicated if the time of flight is more than two time steps.

Assume a time of flight of l steps. The game formulation now is (1) together with

$$x^1_{t+1} = v_t,$$
$$x^k_{t+1} = x^{k-1}_t, \qquad k = 2, \ldots, l.$$

Capture is now defined by

$$x^l_t - y_t = 0$$

But the most important new feature is that Rabbit's strategy may only be based on the knowledge of past y's (and u's), while Hunter's strategy is based on past x's, y's and v's. We may in addition notice that knowing the past y's, Hunter knows the u's actually played by Rabbit. Hence, we are looking for mixed strategies of the form

$$\begin{aligned} p_t &= \phi_t(\mathcal{Y}_t, \mathcal{U}_{t-1}), \\ q_t &= \psi_t(\mathcal{X}_t, \mathcal{Y}_t, \mathcal{U}_{t-1}, \mathcal{V}_{t-1}), \end{aligned} \tag{4}$$

where $\mathcal{X}_t = \{x_0, x_1, \ldots, x_t\}$, and likewise for $\mathcal{Y}_t$ with y's, $\mathcal{U}_t$ and $\mathcal{V}_t$ with u's and v's respectively. (It would also be possible to make the game a "noisy" duel, meaning that Rabbit knows z_t.)

It is simpler at this point to look at the more general problem of which this is a particular case. In addition we shall introduce noise in both dynamics and observations although it is not present here, only because this is hardly more complicated, and takes into account noisy "navigation" for Rabbit, as well as inprecise aiming of Hunter. The unperturbed case follows by simply assuming zero variance for the noises, and therefore ignoring the corresponding expectation symbols in the sequel. Nothing degenerates in this process, because we are in discrete time.

2. The general case.

2.1 The set up.

Let a dynamical game in discrete, finite time and state be described by the dynamics

$$\begin{aligned} & & & t \in \{0, 1, \ldots, T\}, \\ & x_{t+1} = f_t(x_t, u_t, v_t, w_t), & & x_t \in X, \\ & & & u_t \in U_{ad} \subset U, \quad v_t \in V_{ad} \subset V. \end{aligned} \tag{5}$$

Here, $\{w_t\}$ is a random process, zero mean, white and of known statistics, ranging over a space W.

We shall use mixed strategies. Let therefore Σ_U and Σ_V be the relevant simplices, and

$$p_t \in \Sigma_U, \qquad p_t(i) = \Pr(u_t = i),$$
$$q_t \in \Sigma_V, \qquad q_t(j) = \Pr(v_t = j).$$

We furthermore introduce an information vector

$$y_t = h_t(x_t, z_t), \qquad y_t \in Y. \tag{6}$$

Again, $\{z_t\}$ is a random process, zero mean, white and of known statistics, assumed independant from $\{w_t\}$ for the sake of simplicity, ranging over a space Z. As previously, y_t will represent information available to player one, while player two has access to x_t, y_t, u_{t-1} and v_{t-1}. We shall also assume that both players share the same knowledge of x_0, for instance know it exactly.

Finally, let us specify the payoff. A capture set C is given in $(X \times \mathbf{N})$ that contains $(X \times t)$ for all $t \geq T$. (i.e. the game terminates before or at time T). The payoff is defined by

$$J = K(x(t_1), t_1) + \sum_{t=0}^{t_1-1} L_t(x_t, u_t, v_t), \tag{7}$$

where t_1 is final time, depending on $\{u_k\}$ and $\{v_k\}$.

Player one, $\mathcal{P}$, wants to minimize (the expectation of) this payoff, while player two, $\mathcal{E}$, wants to maximize it. They chose p_t and q_t respectively, using the information available to each, i.e. according to (4). Hence the players' choices are their closed loop strategies ϕ and ψ. We are looking for a solution concept for the expectation $\mathrm{E}J(\phi, \psi)$.

2.2 A dynamic programming solution.

We first introduce a probability law $Q_t(\cdot)$ over X that will stand for the "idea" $\mathcal{P}$ has of the state. It is not a conditional expectation, since this concept is not defined unless $\mathcal{P}$ has some information on the probability laws q_k used by $\mathcal{E}$, which we do not assume. However, with our standing assumption that $\mathcal{P}$ knows x_0 exactly, we have an obvious Q_0 as a Dirac at x_0.

Now assume for awhile that $\mathcal{P}$ knows Q_t and a strategy $\psi_t[x, Q]$ giving his opponent's choice of q_t for each (x, Q). Then he can determine Q_{t+1} as a conditional expectation. Let

$$\bar{Q}_{t+1}(\xi) = \sum_x \sum_v \mathrm{E}_w \delta\big(\xi - f_t(x, u_t, v, w)\big) \psi_t[x, Q_t](v) Q_t(x). \tag{8}$$

(δ is the discrete Dirac measure, i.e. 1 if its argument is 0, and 0 otherwise). This is just the probability $Q_t(\cdot)$ carried over by the flow f, which we use as the apriori probability for x at time $t+1$. Then use the extra information y_{t+1} according to Bayes'rule:

$$Q_{t+1}(\xi) = \frac{\mathrm{E}_z \delta\big(y_{t+1} - h_{t+1}(\xi, z)\big) \bar{Q}_{t+1}(\xi)}{\sum_x \mathrm{E}_z \delta\big(y_{t+1} - h_{t+1}(x, z)\big) \bar{Q}_{t+1}(x)}. \tag{9}$$

This defines a relation

$$Q_{t+1} = G_t(Q_t, y_{t+1}, u_t, \psi_t), \tag{10}$$

provided that ψ_t be a function of x and Q. Substituting for y_{t+1} using (6) into G, this also induces a relation

$$Q_{t+1} = g_t(Q_t, x_t, u_t, \psi_t, w_t, z_{t+1}), \tag{11}$$

but we must keep in mind that it depends on x_t, w_t and z_t only through y_{t+1} which is known to $\mathcal{P}$.

We wish to enphasize once more that we do not mean, in the end, to assume that $\mathcal{P}$ knows the strategy actually used by $\mathcal{E}$. Therefore, as it stands, this Q_t is not available to $\mathcal{P}$. We shall only use it in a special way which *is* available to $\mathcal{P}$.

We are now ready to state the main result of this paper.

Theorem. *Assume one can find a real function $V_t(x, Q)$, and two strategies $\hat{\phi}_t[Q]$ and $\hat{\psi}_t[x, Q]$ such that:*

$$\forall (x,t) \in C, \quad V_t(x,Q) = K(x,t), \tag{12}$$

$$V_t(x,Q) = \max_{q \in \Sigma_V} \sum_v q(v) \sum_u \hat{\phi}_t[Q](u) \left[\mathrm{E} V_{t+1}\Big(f_t(x,u,v,w), g_t(Q,x,u,\hat{\psi}_t,w,z) \Big) + L_t(x,u,v) \right], \tag{13}$$

$$\hat{\psi}_t[x,Q] \in \text{Argmax(above)}, \tag{14}$$

$$\sum_x V_t(x,Q)Q(x) = \min_{p \in \Sigma_U} \sum_u p(u) \sum_x Q(x) \sum_v \hat{\psi}_t[x,Q](v) \left[\mathrm{E} V_{t+1}\Big(f_t(x,u,v,w), g_t(Q,x,u,\hat{\psi}_t,w,z) \Big) + L_t(x,u,v) \right], \tag{15}$$

$$\hat{\phi}_t[Q] \in \text{Argmin(above)}, \tag{16}$$

then, the following pair of strategies is a saddle point: let $\{\hat{Q}_t\}$ be the sequence obtained by placing $\hat{\psi}_t[x, Q]$ in (10), and chose

$$\phi_t^*(y_t, u_{t-1}) = \hat{\phi}_t[\hat{Q}_t], \tag{17}$$

$$\psi_t^*(x_t, y_t, u_{t-1}, v_{t-1}) = \hat{\psi}_t[x_t, \hat{Q}_t]. \tag{18}$$

In addition, the saddle point value is $V_0(x_0, \delta(x_0))$.

Proof. Assume first that both players play their strategies (17) and (18). Then equation (13) may also be read

$$V_t(x,Q) = \mathrm{E}\left(V_{t+1}(x_{t+1},Q_{t+1}) + L_t(x_t,u_t,v_t)\middle| x_t = x, Q_t = Q\right),$$

where the expectation symbol extends to w and z with their natural probability laws, and to u and v with the probability laws $\hat{\phi}$ and $\hat{\psi}$. Consider the markov process $\{x_t, Q_t\}$ generated by (5), (4), and (11) where ϕ and ψ have been replaced by $\hat{\phi}$ and $\hat{\psi}$. By the classical argument of stochastic dynamic programming, (because the information algebra is increasing), it results that

$$V_0(x_0,\delta(x_0)) = \mathrm{E}\left(V_{t_1}(x_{t_1},\hat{Q}_{t_1}) + \sum_{t=0}^{t_1-1} L_t(x_t,u_t,v_t)\right),$$

and using (12) and (7),

$$J(\phi^*,\psi^*) = V_0(x_0,\delta(x_0)). \tag{19}$$

Assume now that P plays according to ϕ^*, (i.e. using $\hat{\psi}$ to compute Q), but that $\mathcal{E}$ uses an arbitrary stochastic process $\{q_t(\omega)\}$ of mixed strategies, adapted to the algebra of past events. Because we have left $\hat{\psi}_t$, and not q, as an argument of g_t in the r.h.s. of (13), this r.h.s., without the "max" operator is the expectation of $V(x_{t+1},Q_{t+1})$ given that $x_t = x$ and $Q_t = Q$ when P uses ϕ^* and for an arbitrary q as q_t. We therefore have, with the probability laws $\hat{\phi}_t[Q_t]$ and $q_t(\omega)$ for u and v,

$$V_t(x,Q) \geq \mathrm{E}\left(V_{t+1}(x_{t+1},Q_{t+1}) + L_t(x_t,u_t,v_t)\middle| x_t = x, Q_t = Q\right).$$

Again, considering the process $\{x_t, Q_t\}$ generated by ϕ^*, $\{q_t\}$, and using the increasing algebra property, this results in

$$V_0(x_0,\delta(x_0)) \geq \mathrm{E}\left(V_{t_1}(x_{t_1},\hat{Q}_{t_1}) + \sum_{t=0}^{t_1-1} L_t(x_t,u_t,v_t)\right) = J(\phi^*,\{q_t\}). \tag{20}$$

Notice that it is so although P has used the wrong q to compute Q which therefore is not a conditional probability law for x, a property we did not use.

Assume finally that P plays an arbitrary stochastic process $\{p_t(\omega)\}$, adapted to the algebra generated by y_t and u_{t-1}, while the maximizer uses ψ^*. Now, $\hat{Q}_t$ is indeed the conditional probability law of x_t given y_t and u_{t-1}. Therefore, we can write

$$\bar{V}_t(\hat{Q}_t) = \sum_x V_t(x,\hat{Q}_t)\hat{Q}_t(x) = \mathrm{E}\left(V_t(x_t,\hat{Q}_t)\middle| y_t, u_{t-1}\right). \tag{21}$$

By the optimality principle as applied to dynamical games, (see [11]), the problem faced by P at each instant of time depends on the past only through x_t. Therefore Q_t is a sufficient statistics, and (15)(16) imply that, with the probability laws $p_t(\omega)$, $\hat{\psi}_t[x_t, Q_t]$ for u and v,

$$(22) \quad \mathrm{E}\left(V_t(x_t, Q)\middle|y_t, u_{t-1}\right) \leq \mathrm{E}\left(V_{t+1}(x_{t+1}, \hat{Q}_{t+1}) + L_t(x_t, u_t, v_t)\middle|y_t, u_{t-1}, \hat{Q}_t = Q\right).$$

The rest of the proof is almost classical, and similar to the previous one. Let us nevertheless give it in more detail since it is slightly more subtle, because the expectation sign appears on both sides of the inequality. Let us consider again the markov process $\{x_t, \hat{Q}_t\}$ generated by $\{p_t\}$, $\{\hat{\psi}_t\}$, and consider

$$\mathrm{E}\left(V_{t+1}(x_{t+1}, \hat{Q}_{t+1}) + L_t(x_t, u_t, v_t) + L_{t-1}(x_{t-1}, u_{t-1}, v_{t-1})\middle|y_{t-1}, u_{t-2}\right).$$

(Notice that $\hat{Q}_t$ is measurable on y_t, u_{t-1}). Because the information algebra is increasing, we have, droping the arguments in V and L:

$$\mathrm{E}\left(V_{t+1} + L_t + L_{t-1}\middle|y_{t-1}, u_{t-2}\right) = \mathrm{E}\left(\mathrm{E}\left(V_{t+1} + L_t + L_{t-1}\middle|y_t, u_{t-1}\right)\middle|y_{t-1}, u_{t-2}\right).$$

Use the linearity to expand the r.h.s. above:

$$\mathrm{E}\left(V_{t+1} + L_t + L_{t-1}\middle|y_{t-1}, u_{t-2}\right) = \mathrm{E}\left(\mathrm{E}\left(V_{t+1} + L_t\middle|y_t, u_{t-1}\right) + \mathrm{E}\left(L_{t-1}\middle|y_t, u_{t-1}\right)\middle|y_{t-1}, u_{t-2}\right).$$

Since inequality (22) holds for each Q, it holds for the conditional expectation. We have thus, regrouping sums of expectations,

$$\mathrm{E}\left(V_{t+1} + L_t + L_{t+1}\middle|y_{t-1}, u_{t-2}\right) \geq \mathrm{E}\left(\mathrm{E}\left(V_t + L_{t-1}\middle|y_t, u_{t-1}\right)\middle|y_{t-1}, u_{t-2}\right).$$

Use again the property of nested algebras in the other direction, and finally inequality (22) one step of time earlier to obtain

$$\mathrm{E}\left(V_{t+1} + L_t + L_{t-1}\middle|y_{t-1}, u_{t-2}\right) \geq \mathrm{E}\left(V_{t-1}\middle|y_{t-1}, u_{t-2}\right).$$

By induction, and using the fact that $\hat{Q}_0 = \delta(x_0)$, we end up with

$$(23) \qquad J(\{p_t\}, \psi^*) \geq V_0\big(x_0, \delta(x_0)\big).$$

Finally, (19), (20) and (23) together prove the theorem. ∎

Comments. **1.** We apparently did not use the equality in (15), but only the fact that $\hat{\phi}$ provides the min of the r.h.s. As a matter of fact, the equality is a consequence of that in (13).

2. Assume the solution of (13) to (16) is not unique, and let $(\hat{\phi}_1, \hat{\psi}_1)$ and $(\hat{\phi}_2, \hat{\psi}_2)$ be two different solutions. Then ϕ_1^* and ψ_1^* are constructed with $\hat{Q}_1$ obtained by placing $\hat{\psi}_1$ in (10), while ϕ_2^* and ψ_2^* are obtained likewise with $\hat{Q}_2$. What happens if $\mathcal{P}$ plays ϕ_1^* while $\mathcal{E}$ plays ψ_2^*? Then $\hat{Q}_1$ used by $\mathcal{P}$ to construct his strategy is the wrong one. But not more so than against any other control. And since both (ϕ_1^*, ψ_1^*) and (ϕ_2^*, ψ_2^*) are saddle points, so are (ϕ_1^*, ψ_2^*) and (ϕ_2^*, ψ_1^*), with the same value. Thus nonunicity does not preclude use of this theory.

3. We assumed that both players know exactly x_0, so that $Q_0 = \delta(x_0)$. If $\mathcal{P}$ only had an apriori probability distribution Q_0, we could carry out the same theory, but we would end up with a Nash point. As a matter of fact, although there is only one performance index, since the players optimize expectations conditioned by different apriori information, it is in effect a non zero sum game. Then we recover all the classical difficulties associated with the Cournot-Nash equilibrium, and the above comment does not hold any more.

4. The max in (13) is equivalent to that obtained by multiplying both sides by $Q(x)$ and summing over x. (Since q is allowed to depend on x, the maximum of this positively weighted sum is obtained by maximizing each term). Let therefore r be the vector of $\mathbf{R}^{N \times V}$ of all the values of $q(x)$ for $x = 1, \ldots, N$. Let, for a fixed Q,

$$A_{ij}(x,r) = \mathrm{E}_{wz}\left[V_t\big(f(x,i,j,w), g(Q,x,i,r,w,z)\big) + L_t(x,i,j)\right],$$

and

$$B(r) = \left[A(1,r)\ A(2,r)\ \ldots\ A(N,r)\right].$$

For a given $\bar{r}$, the bilinear form

$$p' B(\bar{r}) r$$

(where ' means transposed), has a saddle point $(\hat{p}(\bar{r}), \hat{q}(\bar{r}))$ over the relevant sets for p (the simplex of $\mathbf{R}^U$) and r (the product of N simplices of $\mathbf{R}^V$). Given $B(\bar{r})$, this saddle point can be computed by a linear progam in an almost classical way. See [12] for more details. Equations (13) to (16) can be interpreted as: for each Q, find r^* such that $\hat{r}(r^*) = r^*$. This is therefore a fixed point problem. Because B is generally not continuous in r, we were not able to prove the existence of this fixed point.

5. As a consequence of the above remark, the present theory is not an existence result, but only a sufficiency condition. Notice however that Lévine [9] has proved the existence of a value to such a game. If it were possible to prove that this value is measurable on (x, Q), this would provide the existence result, since then (13) to (16) become necessary conditions.

6. We have programmed this algorithm for the Rabbit and Hunter game, with $l = 2$ and $\nu \geq T$. Great care must be taken in doing so, among others in the discretization and approximation process for probability laws, in the linear programming, in the fixed point algorithm, etc. Due to the great size of the state space, we could only check the overall algorithm for very small values of N and T. We did get existence of the fixed point for several of the cases checked. (See [12] for more details). But more numerical experience is still needed.

2.3 Safe strategies in the all imperfect information case

We can now use an idea of Kumar and van Schuppen [4] to propose an equilibrium concept in the case where both players have different, imperfect information on the state, and no knowledge of the opponent's control. Assume therefore that $\mathcal{P}$ has again an information of the form (6), and $\mathcal{E}$ has a symmetric information, say

$$\eta_t = k_t(x_t, \varsigma_t)$$

where ς_t is a white sequence of known statistics.

The concept of "safe" strategies calls for both players to behave in a worst case hypothesis. Thus, $\mathcal{P}$'s problem would be

$$\min_{\phi} \max_{\{q_t\}} \mathrm{E}(J|\mathcal{Y})$$

where $\{q_t\}$ ranges over all possible stochastic processes of mixed strategies for $\mathcal{E}$. Of course, $\mathcal{E}$'s problem would be the symmetric one. Letting the process $\{q_t(\omega)\}$ be arbitrary in the max above is equivalent to assuming that $\mathcal{E}$ has all the information he wishes to make his choice of control, i.e. precisely the information structure assumed in the previous section. Therefore, the solution of this problem is that given above for $\mathcal{P}$, and the symmetric one for $\mathcal{E}$. And as pointed out in [4], by playing that way, both players have a lower bound on how well they will do.

3. The continuous time case

3.1. A preliminary lemma.

We turn now to the continuous time case. Reference [9] provides all the necessary technical tools to write the general nonlinear theory. We see, however, little incentive to do so, since the resulting theory is so involved that we do not see any use to it. We shall therefore restrict our attention to the linear-quadratic-gaussian case.

However, it turns out to be simpler to proove first a lemma in general nonlinear terms. But this lemma is so specialized that the L.Q.G. problem is the only one we know of that fits into it.

Let a stochastic two-player dynamic system in $\mathbf{R}^n$ be given by a diffusion

$$dx = f(x,u,v,t)dt + db_1, \tag{24}$$

an observation process in $\mathbf{R}^p$ be given by

$$dy = h(x,t)dt + db_2, \tag{25}$$

where b_1 and b_2 are two independant vector brownian motions of covariance coefficient matrices $W_1(t)$ and $W_2(t)$ respectively, and a performance index be given as

$$J = \mathrm{E}\left(K(t_1) + \int_{t_0}^{t_1} L(x,u,v,t)dt\right), \tag{26}$$

where t_0 and t_1 are given time instants.

Standard hypotheses are assumed on f, h, and the admissible processes $u(\cdot)$ and $v(\cdot)$ for $x(\cdot)$ and $y(\cdot)$ to be well defined, and K and L are supposed to be globally C^1.

We again consider the game where the minimizer only knows $y(t)$, with perfect memory, while the maximizer knows $x(t)$, $y(t)$ and $u(t)$. To stay with saddle points, as opposed to Nash points, we assume that in addition, the minimizer knows x_0.

We call *strategies* functions $\phi(\xi,t)$ from $\mathbf{R}^n \times \mathbf{R}$ into $\mathbf{R}^m$ and $\psi(x,\xi,t)$ from $\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}$ into $\mathbf{R}^{m'}$ such that, if $u(t)$ and $v(t)$ are replaced by $\phi(\hat{x}(t),t)$ and $\psi(x(t),\hat{x}(t),t)$ in equations (24) and (27) below, the processes x and $\hat{x}$ are well defined in the Ito sense.

Lemma. *If there exist a process $\hat{x}$, a C^2 function $V(x,\hat{x},t)$ from $\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}$ into $\mathbf{R}$, strategies $\hat{\phi}$ and $\hat{\psi}$, such that*

i) *the process $\hat{x}$ obeys a diffusion law of the form*

$$d\hat{x} = G_1(\hat{x},u,t)dy + G_2(\hat{x},u,t)dt, \tag{27}$$

and thus

$$d\hat{x} = g(x,\hat{x},u,t)dt + k(x,\hat{x},u,t)db_2,$$

and is such that, whenever $v(t) = \hat{\psi}(x(t),\hat{x}(t),t)$, the conditional law of x given the information algebra $\mathcal{Y}_t$ generated by $\{y(s), s \le t\}$, is of the form $\pi_{\hat{x}}(x) = \pi(x-\hat{x},t)$, where π is a fixed zero-mean law, i.e. independant of the control process $u(\cdot)$, (it is this condition which is never met in nonlinear problems),

ii) $\forall (x,\hat{x},t) \in \mathbf{R}^n \times \mathbf{R}^n \times [t_0,t_1]$, *we have (omitting the arguments x, $\hat{x}$, t in $\frac{\partial V}{\partial x}$, ϕ and ψ)*

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(x,\hat{\phi},\hat{\psi},t) + \frac{\partial V}{\partial \hat{x}} g(x,\hat{x},\hat{\phi},t) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 V}{\partial x^2} W_1\right) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 V}{\partial \hat{x}^2} k W_2 k'\right) + L(x,\hat{\phi},\hat{\psi},t) = 0, \tag{28}$$

$$\forall (x,\hat{x}) \in \mathbf{R}^n \times \mathbf{R}^n, \quad V(x,\hat{x},t_1) = K(x). \tag{29}$$

iii) $\forall (x,\hat{x},t) \in \mathbf{R}^n \times \mathbf{R}^n \times [t_0,t_1]$ *and* $\forall v \in \mathbf{R}^{m'}$,

$$\frac{\partial V}{\partial x} f(x,\hat{\phi},\hat{\psi},t) + L(x,\hat{\phi},\hat{\psi},t) \ge \frac{\partial V}{\partial x} f(x,\hat{\phi},v,t) + L(x,\hat{\phi},v,t). \tag{30}$$

iv) $\forall (\hat{x},t) \in \mathbf{R}^n \times [t_0,t_1]$ *and* $\forall u \in \mathbf{R}^m$,

$$\int_{\mathbf{R}^n} \left[\frac{\partial V}{\partial x} f(x,\hat{\phi},\hat{\psi},t) + \frac{\partial V}{\partial \hat{x}} g(x,\hat{x},\hat{\phi},t) + L(x,\hat{\phi},\hat{\psi},t)\right] d\pi_{\hat{x}}(x) \le \int_{\mathbf{R}^n} \left[\frac{\partial V}{\partial x} f(x,u,\hat{\psi},t) + \frac{\partial V}{\partial \hat{x}} g(x,\hat{x},u,t) + L(x,u,\hat{\psi},t)\right] d\pi_{\hat{x}}(x). \tag{31}$$

then, $(\hat{\phi}(\hat{x},t),\hat{\psi}(x,\hat{x},t))$ *is a saddle point strategy pair, and the value of the game is* $V(x_0,x_0,t_0)$.

Proof. This is the exact continuous time counterpart of the discrete time case seen earlier. We only made it simple by the very special assumption that the conditional law of x_t depends only on its mean $\hat{x}$. (And we deleted equal terms on both sides of inequalities (30) and (31)). There is no need to detail it again. ■

3.2. The Linear Quadratic Gaussian case.

We shall now apply the above theory to the L.Q.G. problem, thus in effect revisiting Behn and Ho's problem [3], from a less specialized viewpoint. We do not know at this time how the nonlinear theory could be extended to the situation described in [4] which generalizes this result.

Let equations (24), (25), (26) be respectively (all matrices are time varying, piecewise continuous)

$$dx = (Fx + Gu + Ev)dt + db_1, \tag{32}$$

$$dy = Hx\,dt + db_2, \tag{33}$$

$$J = \mathrm{E}\left[x'(t_1)Ax(t_1) + \int_{t_0}^{t_1}\Big(x'(t)Qx(t) + u'(t)Ru(t) - v'(t)Bv(t)\Big)dt\right]. \tag{34}$$

Then, we can state the following fact:

Theorem. (Behn and Ho) *If equations (38), (39) and (40) have a solution over* $[t_0,t_1]$, *then the above game, with the partial information specified, has a saddle point, given by*

$$d\hat{x} = \big[(F + EB^{-1}E'P)\hat{x} + Gu - \Sigma H'W_2^{-1}H\hat{x}\big]dt + K\,dy, \tag{35}$$

$$\hat{\phi}(\hat{x},t) = -R^{-1}G'P\hat{x}, \tag{36}$$

$$\hat{\psi}(x,\hat{x},t) = B^{-1}E'\big[\Pi x + (P-\Pi)\hat{x}\big] = B^{-1}E'(P\hat{x} + \Pi\tilde{x}), \tag{37}$$

(where $\tilde{x} = x - \hat{x}$*), with* P, Π *and* Σ *symmetric matrices solutions of*

i) *the classical game theoretic Riccati equation for* P*:*

$$\dot{P} + PF + F'P - PGR^{-1}G'P + PEB^{-1}E'P + Q = 0, \qquad P(t_1) = A, \tag{38}$$

ii) *a pair of coupled Riccati equations, with two-point boundary values for* Π *and* Σ*:*

$$\begin{aligned}\dot{\Pi} + \Pi(F - \Sigma H'W_2^{-1}H) + (F' - H'W_2^{-1}H\Sigma)\Pi + \Pi EB^{-1}E'\Pi \\ + P\Sigma H'W_2^{-1}H + H'W_2^{-1}H\Sigma P + Q = 0, \qquad \Pi(t_1) = A\end{aligned} \tag{39}$$

(40) $$\dot{\Sigma}-(F-EB^{-1}E'\Pi)\Sigma-\Sigma(F'-\Pi EB^{-1}E')+\Sigma H'W_2^{-1}H\Sigma-W_1=0, \qquad \Sigma(t_0)=0.$$

Proof. Try to use the lemma with

$$V(x,\hat{x},t)=x'P(t)x+(x-\hat{x})'\big(P(t)-\Pi(t)\big)(x-\hat{x}).$$

(The natural form for V would be $x'Px+x'\tilde{P}\hat{x}+\hat{x}'\tilde{P}'x+\hat{x}'\hat{P}\hat{x}$. However, the fact that $\mathrm{E}(\tilde{x}'\Pi x)=\mathrm{E}(\tilde{x}'\tilde{P}\hat{x})=0$ strongly suggests to write $\hat{x}=x-\tilde{x}$, and rewriting V in terms of x and $\tilde{x}$ only. The particular form $P-\Pi$ for the wheighting matrix of $\tilde{x}^2$ has been used to make some formulas simpler.)

Notice that indeed, (35), (40) is the Kalman filter of (32), (33) with $v(t)$ replaced by $\hat{\psi}(x(t),\hat{x}(t),t)$ as in (37). It is known, then, that the conditional law of x is a gaussian law of covariance Σ, depending on u only through its mean $\hat{x}$.

Finally the only fact which is not trivial to check is the inequation (31). Use the fact that

$$\int_{\mathbf{R}^n}(u'G'Px+u'Ru)d\pi_{\hat{x}}(x)=u'G'P\hat{x}+u'Ru$$

so that the minimization is easy to carry out, as if there were no integration.

We end up with the following remark:

Remark. If $H=0$, i.e. the maximizer has no measurement, then he plays open loop. Notice that then equation (39) simplifies into just the Riccati equation of the maximization problem. It is a known fact ,[10], that in the noise free case, playing one's open loop control generated by the pair of optimal strategies is indeed still optimal if the one-sided Riccati equation has a solution over the game time interval .This result is extended here to the noisy dynamics case.

Bibliography

[1] W.M.Willman, "Formal solutions for a class of stochastic pursuit evasion games", *IEEE Transactions on Automatic Control* **A.C.14** (1969), pp 504–509.

[2] A.Bagchi and G.J.Olsder, "Linear stochastic pursuit-evasion games", *Journal of Applied Mathematics and Optimization* **7** (1981), pp 95–123.

[3] R.D.Behn & Y-C. Ho, "On a class of linear stochastic differential games", *IEEE Transactions on Automatic Control* **A.C. 13** (1968) pp227–239.

[4] P.R.Kumar & J.H. van Schuppen, "On Nash equilibrium solutions in stochastic dynamic games", *IEEE Transactions on Automatic Control* **A.C. 25** (1980), pp 1146–1149.

[5] I.B.Rhodes & D.Luenberger, "Differential games with inperfect state information", *IEEE Transactions on Automatic Control* **A.C. 14**, (1969) pp 29–38.

[6] P.Faurre, "Jeux différentiels à stratégies complètement optimales", *4th IFAC World Congress, Warsaw* (1969).

[7] G.Papavassilopoulos, "On linear quadratic continuous-time Nash games", *Journal of Optimization Theory and Applications* **42** (1984) pp 525–549.

[8] J.Lévine, "incomplete information in differential games and team problems", *8th IFAC World Congress, Kyoto* (1981).

[9] J.Lévine, "Sur quelques structures d'information intervenant en jeux, dans le problème d'équipes ou de controle et en filtrage", *thèse de doctorat d'État, Université Paris 9-Dauphine* (1984).

[10] P.Bernhard, *Commande optimale, décentralisation et jeux dynamiques*, Dunod, Paris (1972).

[11] P.Bernhard, "Principe d'optimalité et information nécéssaire pour la commande dans le cas le plus défavorable", *cahier du CEREMADE 7923, Université Paris Dauphine*, (1979)

[12] A.L.Colomb, "Étude de jeux à deux joueurs en information incomplète", *thèse, Université de Provence, Marseille, France* (1986)

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique