Minimax —or Feared Value— L^1/L^∞ Control

Pierre Bernhard[1]
ESSI,
930 route des colles, BP 145,
06903 Sophia Antipolis Cedex,
France

January 1, 1999

[1]I3S, University of Nice-Sophia Antipolis and CNRS, France

Abstract

This is the third paper, after [7, 8], where we attempt to develop and exploit the parallel between stochastic control and min-max control induced by the use of the max-plus algebra, using the concept of feared value as the parallel to expected value. The present paper builds on the main formula of [8], the results of which are a subset of those given here. Its new contribution is twofold. On the one hand we clarify the role of the integral part of the cost in that parallel. This leads to a more extensive theory of the so called (imperfect information) minimax L^∞ control problem than apparently available in the literature including a certainty equivalence theorem. On the other hand, we extend the parallel to the continuous time case as much as we can. In that direction, the present paper slightly extends the classical framework of the variational inequality of stopping time games, and gives a formal treatment of the partial information case. Altogether, this might be the first new results obtained with the tool of mathematical fear.

1 Introduction

This paper follows our previous works [7, 8] in our attempt to perfect the parallel between stochastic control and minimax control, using the concept of cost measure and feared value as the parallel to probability measures and expected value. In the process, new results seem to flow naturally.

The concept of cost measure has been introduced by various authors working on the concept of (max-plus) algebra, or idempotent algebra. A bibliography can be found in [8]. Two of the most important references for the type of application we make here of these ideas are [1, 3]. The concept we call "feared value" can obviously be found in these papers, but it seems that we were the first to emphasize it as the main tool to investigate minimax control, giving it the name we use here, in [7].

In [8], we have succeeded in giving a completely parallel treatment of stochastic and minimax control of discrete time systems with imperfect information, up to the point where essentially the same separation principle, with the same proof, applies to both. However, this good parallel was obtained at the expense of restricting the performance index to a purely terminal one. Although we know that there is no lack of generality in doing so, yet it would be nicer to extend the parallel to the case with a running cost, or integral cost, added to the terminal cost. This is what we do here. The natural parallel is a generalization of the so-called L^∞ control problem, where the criterion is a max over time of a function of the state and controls. But because of the integral penalization term inherent to the fear operator, we end up with a theory of minimax partial observation control of a mixed L^1/L^∞ performance index, that can also be viewed as a stopping time problem. This yields results that generalize somewhat those available in that domain, and in particular a separation principle for those problems.

In a second part, we try to see what can be extended to the continuous time case. The parallel is there imperfect, and there are good reasons for that. Yet something can be done. In the perfect information case, this leads us to an extension of the classical variational inequality of stopping time games (see [4, 11]). In the incomplete information case, we give a formal treatment that at least points to the equations to which harder mathematical work should give a precise meaning, including a separation principle.

1.1 Notations

Consider a cost-space, i.e. a topological space Ω endowed with a cost measure K from its borelian $\mathcal{B}$ to $\mathbb{R}^-$, having a density c. Remember that then for any $A \in \mathcal{B}$, one has $K(A) = \sup_{\omega \in A} c(\omega)$. Let also $x = X(\omega)$, where $X(\cdot)$ is a continuous function from Ω to $\mathbb{R}^n$, be a decision variable, with induced cost density Q. That is,

$$Q(x) = K(\{\omega \in \Omega \mid X(\Omega) = x\}) = \sup_{\omega \in X^{-1}(x)} c(\omega)\,.$$

For any function g of x and possibly other variables, we call *mathematical fear* and write

$$\mathbb{F}_x^Q g := \sup_x [g(x) + Q(x)].$$

When no ambiguity is possible, we may write it $\mathbb{F}_x g$ or simply $\mathbb{F} g$.

The upper index denoting the cost law will mainly be used when a conditioning modifies the cost law. For instance, the conditional law *given another event*, say $\omega \in \Omega'$ is

$$R(x) = \sup_{\omega \in \Omega' \cap X^{-1}(x)} c(\omega) - \sup_{\omega \in \Omega'} c(\omega)$$

and the mathematical fear with respect to that cost density will be denoted, of course, $\mathbb{F}_x^R$, or if the function has only x as a variable just $\mathbb{F}^R$, or, by a slight abuse of notations, $\mathbb{F}^{\Omega'}$.

We shall make ample use of the fact that the fear operator is (max,+) linear.

2 Discrete time

2.1 The problem

2.1.1 The system

Let a discrete time partially observed disturbed control system be given by

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

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

where $x_t \in \mathbb{R}^n$ is the state vector at time t, u_t is the control vector at time t, to be chosen within a set $\mathsf{U} \subset \mathbb{R}^m$, $w_t \in \mathbb{R}^l$ is a disturbance vector at time t, may be constrained to belong to a set W, and $y_t \in \mathsf{Y} \subset \mathbb{R}^p$ is the observed output at time t.

We shall write $\mathbf{u} \in \mathcal{U}$ for the time sequence $\{u_t\}_{t \in [0, T-1]} \in \mathsf{U}^T$ (The upper index T is indeed a cartesian power, as it should, and *contrary* to the notations we introduce next and use in the rest of the paper) and similarly for $\mathbf{w} \in \mathcal{W}$ and $\mathbf{y} \in \mathcal{Y}$.

We shall need partial sequences defined as follows:

$$u^t = (u_0, u_1, \ldots, u_t)\,,$$

and similarly for all time sequences. (As a consequence, $\mathbf{u} = u^{[T-1]}$.) We shall let $u^t \in \mathsf{U}^t$ ([1]), $w^t \in \mathsf{W}^t$, $y^t \in \mathsf{Y}^t$.

Let also $\omega = (x_0, \mathbf{w})$ denote the disturbances a priori unknown to the controller, and $\omega \in \Omega = \mathbb{R}^n \times \mathcal{W}$. We also use $\omega^t = (x_0, w^t) \in \Omega^t = \mathbb{R}^n \times \mathsf{W}^t$.

We shall need the input-state and input-output maps of system (1)(2), that we call ϕ and η respectively, meaning that

$$x_t = \phi_t(x_0, u^{t-1}, w^{t-1}) = \phi_t(u^{t-1}, \omega^{t-1})\,, \tag{3}$$
$$y_t = \eta_t(x_0, u^{t-1}, w^{t-1}) = \eta_t(u^{t-1}, \omega^{t-1})\,. \tag{4}$$

Finally, we shall use ϕ^t and η^t to mean the sequences $\{\phi_\tau\}_{\tau=1,\ldots,t}$ and $\{\eta_\tau\}_{\tau=1,\ldots,t}$.

The problem shall always be to choose a control sequence to achieve a certain goal, based upon the knowledge of the noise corrupted output. And of course, the controller shall have to be causal, but with perfect recall: no past information is forgotten at any time. We shall even restrict it to be strictly causal. Thus an admissible strategy will be a sequence of maps $\{\mu_t : \mathsf{U}^{t-1} \times \mathsf{Y}^{t-1} \to \mathsf{U}\}_{t\in[0,T-1]}$ defining the control sequence through

$$u_t = \mu_t(u^{t-1}, y^{t-1})\,.$$

We shall let $\mathcal{M}$ denote the class of such admissible strategies.

To any admissible strategy and any $\omega \in \Omega$ corresponds a unique trajectory $\mathbf{x}$ and a unique control sequence $\mathbf{u}$. So that, although this is an abuse of notations, we shall write such things as $\phi_T(\mu, \omega)$ where what we mean is the final state on the trajectory generated by μ and ω.

2.1.2 The performance index

The set Ω is assumed to be endowed with a cost measure governing the decision variable ω. We assume that x_0 and $\mathbf{w}$ are independent, and that $\mathbf{w}$ is a white sequence, so that the cost measure is entirely specified by a cost density Q_0 over $\mathbb{R}^n$ governing x_0, and a sequence of cost densities $\{\Gamma_t\}$ over

[1] It is here that our notations are inconsistent, since U^t therefore stands for the cartesian power $t+1$ of U.

W governing the w_t's. And the mathematical fear of any function $\psi(\omega)$ is defined as

$$\mathbb{F}\psi = \sup_\omega \left[\psi(\omega) + Q_0(x_0) + \sum_{t=0}^{T-1} \Gamma_t(w_t) \right]$$

Remember also that cost densities are always normalized with their maximum at zero. We shall assume that all functions we use are upper semi continuous, and that the maxima are well defined. (For instance, the cost densities might have a compact support.)

We know that in the parallel we exploit, the algebra $(+, \times)$ is to be replaced by the algebra $(\max, +)$. Therefore, the natural equivalent to the classical performance index is

$$J(\mathbf{u}, \omega) = \max\{M(x_T), \max_{0 \le t < T} L_t(x_t, u_t, w_t)\} = \max_{0 \le t \le T} L_t(x_t, u_t, w_t), \qquad (5)$$

where we have, for convenience, let $L_T = M$, as we shall from now on.

Therefore, the criterion we shall strive to minimize will be

$$H(\mu) = \mathbb{F}J(\mu, \omega) \qquad (6)$$

It is worthwhile, to point out the following fact. We are interested in

$$\mathbb{F}_{x_0}\mathbb{F}_{\mathbf{w}}J(\mathbf{u}, \omega) = \max_{x_0} \max_{w_0 \ldots w_{T-1}} \left[J(\mathbf{u}, \omega) + \sum_{k=0}^{T-1} \Gamma_k(w_k) + Q_0(x_0) \right].$$

The above expression involves the quantity $\mathbb{F}_{\mathbf{w}}J$ which can be expanded into

$$\mathbb{F}_{\mathbf{w}}J = \max_{w_0 \ldots w_{T-1}} \max_t \left[L_t(x_t, u_t, w_t) + \sum_{k=0}^{T-1} \Gamma_k(w_k) \right].$$

Now, this is equal to the same expression where we limit the summation sign to t instead of $T-1$:

Proposition 1

$$\mathbb{F}_{\mathbf{w}}J = \max_{w_0 \ldots w_{T-1}} \max_t \left[L_t(x_t, u_t, w_t) + \sum_{k=0}^{t} \Gamma_k(w_k) \right].$$

As a matter of fact, the Γ_k's are always non positive. Therefore,

$$\max_t \left[L_t(x_t, u_t, w_t) + \sum_{k=0}^{t} \Gamma_k(w_k) \right] \ge \max_t \left[L_t(x_t, u_t, w_t) + \sum_{k=0}^{T-1} \Gamma_k(w_k) \right] \qquad (7)$$

But assume that for a sequence $\mathbf{w}$ and a time $\hat{t}$,

$$L_{\hat{t}}(x_{\hat{t}}, u_{\hat{t}}, w_{\hat{t}}) + \sum_{k=0}^{\hat{t}} \Gamma_k(w_k) > \max_t \Big[L_t(x_t, u_t, w_t) + \sum_{k=0}^{T-1} \Gamma_k(w_k) \Big] \tag{8}$$

Pick the same sub-sequence $\{w_k\}$ up to $k = \hat{t}$, and for $k > \hat{t}$ pick w_k such that $\Gamma_k(w_k) = 0$. The state trajectory up to $\hat{t}$ is unchanged. Moreover, for that sequence,

$$L_{\hat{t}}(x_{\hat{t}}, u_{\hat{t}}, w_{\hat{t}}) + \sum_{k=0}^{T-1} \Gamma_k(w_k) = L_{\hat{t}}(x_{\hat{t}}, u_{\hat{t}}, w_{\hat{t}}) + \sum_{k=0}^{\hat{t}} \Gamma_k(w_k)$$

so that, necessarily

$$\max_t \Big[L_t(x_t, u_t, w_t) + \sum_{k=0}^{T-1} \Gamma_k(w_k) \Big] \geq L_{\hat{t}}(x_{\hat{t}}, u_{\hat{t}}, w_{\hat{t}}) + \sum_{k=0}^{\hat{t}} \Gamma_k(w_k)$$

contradicting the assumption (8). Therefore, we have

$$\forall t,\ L_t(x_t, u_t, w_t) + \sum_{k=0}^{t} \Gamma_k(w_k) \leq \max_t \Big[L_t(x_t, u_t, w_t) + \sum_{k=0}^{T-1} \Gamma_k(w_k) \Big] ,$$

which together with (7) yields the proposition.

To improve the parallel with the forthcoming continuous time case, it is also useful to notice that this may be written in terms of

$$\bar{L}_t(x, u) := \mathbb{F} L_t(x, u, w) = \max_w [L_t(x, u, w) + \Gamma_t(w)] .$$

We also have

Proposition 2

$$\mathbb{F}_{\mathbf{w}} J = \max_{\mathbf{w}} \max_t \Big[\bar{L}_t(x_t, u_t) + \sum_{k=0}^{t-1} \Gamma_k(w_k) \Big] . \tag{9}$$

Again, let $\hat{t}$ be the time when the $\max_t$ is reached. Exactly as the later w_t's can be chosen to make Γ_t null, because they have no effect on the rest of the performance index, yielding in effect Proposition 1, also in the maximizing sequence $\mathbf{w}$ we shall have for $w_{\hat{t}}$ the w that maximizes $L_{\hat{t}} + \Gamma_{\hat{t}}$ since this is the only term in the sum that depends on w_t.

This last form is usefull in that it shows that there is indeed no gain in generality in taking L_t to depend on w_t. We might as well consider only the problem in $\bar{L}_t$.

Finally, it only takes a carefull reading to check that in all the sequel, the Γ_t's may depend as well on x_t and u_t, without invalidating our calculations. So that although we shall write $\Gamma_t(w)$, the problem we consider is really to minimize over $\mathcal{M}$

$$H(\mu)=\mathbb{F}J(\mu,\omega)=\max_{\omega}\left\{\max_{\tau\in[0,T]} L_\tau(x_\tau,u_\tau,w_\tau)+\sum_{\tau=0}^{T-1}\Gamma_\tau(x_\tau,u_\tau,w_\tau)+Q(x_0)\right\} \tag{10}$$

or any of the equivalent form given by the propositions above. However, at this time, the Γ_t's are restricted to be normalized, i.e. $\max_w \Gamma_t(x,u,w)=0$ for all (x,u).

If one wants to investigate a general L^1/L^∞ performance index such as (10), without the normalization assumption on the Γ_t's, it suffices to let

$$\sup_w \Gamma_t(x,u,w)=\gamma_t(x,u)\,, \qquad \Gamma_t=\Gamma_t^0+\gamma_t$$

and introduce an extra state variable $\xi_t=\sum_{k=0}^{t-1}\gamma_t$, ruled by the dynamical equation

$$\xi_{t+1}=\xi_t+\gamma_t(x_t,u_t)\,,\quad \xi_0=0\,.$$

Let also $L_t^0(x,\xi,u,w)=L_t(x,u,w)+\gamma_t(x,u)+\xi$. Now we have one extra state variable, but the performance index written in terms of L^0 and Γ^0 has exactly the previous form, and Γ^0 is by construction normalized.

It is a simple matter to show that the extended state value function in the perfect information case can be written

$$V_t^0(x,\xi)=V_t(x)+\xi$$

where V_t satisfies exactly the recursion (11)(12) below. The imperfect information case is slightly more complicated to work out in that case. But we can still show rather simple relations such as $\sup_\xi Q_t^0(x,\xi)=Q_t(x)$, and Λ_t is preserved.

2.2 Perfect information

Let us first consider the simpler problem where the controller (choosing u) has access to the exact state, and therefore may control in state feedback. Hence x_0 is known to him. We therefore treat it as a fixed, known variable, and let it vary as a decision variable only in the end.

We have an (extended) Isaacs equation:

$$\forall x \in \mathbb{R}^n\,, \quad V_T(x) \;=\; M(x)\,, \tag{11}$$

$$\forall t \in [0, T-1], \forall x \in \mathbb{R}^n,$$

$$V_t(x) \;=\; \inf_u \mathbb{F}_w^{\Gamma_t} \max\{V_{t+1}(f_t(x,u,w)), L_t(x,u,w)\} \tag{12}$$

We may state the following theorem

Theorem 1 *If the backwards recursion (11),(12) generates a bounded Value function V, then, the infimum of the problem (6) is given by $\mathbb{F}^{Q_0} V_0$ (recall that the initial state cost density Q_0 is given). Moreover, if the minimum in u is reached at $\varphi^*(t,x)$ in (12), then this is an optimal state feedback strategy.*

Proof Let us sketch the proof of the theorem. Let $\mathbf{u}$ be a fixed control sequence, and assume that at each instant of time, w_t coincides with the maximizing w in the $\mathbb{F}_w$ operation of (12). According to (12), we have along the trajectory $\mathbf{x}$ thus generated

$$V_0(x_0) \le \max\{V_1(x_1)\,, L_0(x_0,u_0,w_0)\} + \Gamma_0(w_0)$$
$$= \max\{V_1(x_1) + \Gamma_0(w_0)\,, L_0(x_0,u_0,w_0) + \Gamma_0(w_0)\}\,.$$

Use the same relation written between $t=1$ and $t+1=2$ to substitute for V_1 in the rhs above. It comes

$$V_0(x_0) \le \max\{V_2(x_2) + \Gamma_1(w_1) + \Gamma_0(w_0)\,,$$
$$L_1(x_1,u_1,w_1) + \Gamma_1(w_1) + \Gamma_0(w_0)\,, L_0(x_0,u_0,w_0) + \Gamma_0(w_0)\}\,,$$

and so on recursively. (We have freely moved an added term to a max inside the max operator, and collapsed $\max\{\max\{\ldots\},\ldots\}$ into a single max operation, thus using the properties of linearity and associativity of the $(\max,+)$ algebra.) In the end, we end up with

$$V_0(x_0) \le \max_t \Big[L_t(x_t,u_t,w_t) + \sum_{k=0}^{t} \Gamma_k(w_k)\Big]\,,$$

with $L_T(x,u,w) = M(x)$ using (11). Use the proposition to conclude that a fortiori

$$V_0(x_0) \le \mathbb{F}_{\mathbf{w}} J(x_0,\mathbf{u},\mathbf{w})\,. \tag{13}$$

But if u_t is chosen minimizing the r.h.s of (12), the $\leq$ signs above are all replaced by $=$ signs, showing that that strategy yields $V_0(x_0) = J(x_0, \mathbf{u}, \mathbf{w})$ for the sequence $\mathbf{w}$ generated by the above procedure.

There remains to assume that u keeps using that state feedback strategy and choosing an arbitrary sequence $\mathbf{w}$ to have the opposite inequality signs in the above calculations, that reduce to equal signs if w is chosen as the maximizing one, to conclude that indeed

$$V_0(x_0) = \mathbb{F}_{\mathbf{w}} J(x_0, \varphi^*, \mathbf{w}) \,,$$

which, together with (13), concludes the proof upon taking the mathematical fear with respect to x_0 of both sides.

2.3 Imperfect information

We now turn to the case where the minimizer only knows the output (2). The solution follows that proposed in [8] with the same modification as above. That is, we introduce the conditional state cost density $Q_t \in \mathcal{Q}$ in identically the same fashion as in [8]. It can be computed recursively from Q_0 in real time via

$$Q_{t+1}(x) = \sup_{\xi, w \left| \begin{smallmatrix} f_t(\xi, u_t, w) = x \\ h_t(\xi, w) = y_t \end{smallmatrix} \right.} [Q_t(\xi) + \Gamma_t(w)] - \Lambda_t(y_t)$$

where Λ_t is the cost measure induced on y_t by Q_t through (2):

$$\Lambda_t(y) = \sup_{x,w | h_t(x,w) = y} [Q_t(x) + \Gamma_t(w)] \,.$$

We write simply this relation as [2]

$$Q_{t+1} = \mathcal{G}_t(Q_t, u_t, y_t) \,. \tag{14}$$

Then we introduce a dynamic programming recursion for a cost function $U_t(Q_t)$:

$$\forall Q \in \mathcal{Q} \,, \quad U_T(Q) \;=\; \mathbb{F}_x^Q M(x) \,, \tag{15}$$

$$\forall t \in [0, T-1], \forall Q \in \mathcal{Q}, \tag{16}$$

$$U_t(Q) \;=\; \inf_u \max \Big\{ \mathbb{F}_y^{\Lambda_t} U_{t+1}(\mathcal{G}_t(Q, u, y)) \,,\, \mathbb{F}_{x,w}^{Q;\Gamma_t} L_t(x, u, w) \Big\} .$$

[2] Notice that $Q_t(x)$ is finite if and only if x is in the set of states at time t compatible with the prior information.

(Remember that $\mathbb{F}^{Q,\Gamma_t}_{x,w} L_t = \mathbb{F}^{Q}_{x} \bar{L}_t$ so that all calculations below could be slightly simplified using $\bar{L}_t$.)

The theorem is as expected:

Theorem 2 *If the recursion (15),(16) defines a sequence of functions $\{U_t\}$ from $\mathcal{Q}$ to $\mathbb{R}$, then the optimal partial information cost is $U_0(Q_0)$. Moreover, if the min is attained in (16) for every $(t,P) \in [0,T-1] \times \mathcal{P}$, this together with (14) initialized at Q_0, defines an optimal control strategy for problem (5),(6).*

Proof The proof relies on the formula of imbeded conditional fears of [8] : write $Q_{t+1}[y]$ for $\mathcal{G}_t(Q_t,u,y)$, one has for any function ψ

$$\mathbb{F}^{\Lambda_t}_{y} \mathbb{F}^{Q_{t+1}[y]}_{x} \psi(x) = \mathbb{F}^{Q_t,\Gamma_t}_{x,w} \psi(f_t(x,u,w))\,. \tag{17}$$

Fix a control sequence $\mathbf{u}$, and assume any control sequence $\mathbf{w}$. It generates a sequence $\{Q_t\}$. Equation (16) written at time $T-1$ yields

$$U_{T-1}(Q_{T-1}) \le \max\left\{ \mathbb{F}^{\Lambda_{T-1}}_{y} U_T(Q_T)\,,\, \mathbb{F}^{Q_{T-1},\Gamma_{T-1}}_{x,w} L_{T-1}(x,u_{T-1},w) \right\}.$$

Use (15) to substitute in the first term of the r.h.s. above, and make use of (17). It reads

$$\mathbb{F}^{\Lambda_{T-1}}_{y} U_T(Q_T) = \mathbb{F}^{\Lambda_{T-1}}_{y} \mathbb{F}^{Q_T}_{x} M(x) = \mathbb{F}^{Q_{T-1},\Gamma_{T-1}}_{x,w} M(f_{T-1}(x,u_{T-1},w))\,.$$

By $(\max,+)$ linearity, the two symbols $\mathbb{F}^{Q_{T-1},\Gamma_{T-1}}_{x,w}$ collapse into a single one with the max inside, and it comes

$$U_{T-1}(Q_{T-1}) \le \mathbb{F}^{Q_{T-1},\Gamma_{T-1}}_{x,w} \max\{M(x_T)\,,\, L_{T-1}(x,u_{T-1},w)\}\,.$$

Notice that

$$\mathbb{F}^{Q_{T-1},\Gamma_{T-1}}_{x,w} = \mathbb{F}^{Q_{T-1}}_{x} \mathbb{F}^{\Gamma_{T-1}}_{w}$$

so that the right hand side above is again a mathematical fear with respect to x for the cost density $Q_{T-1} = \mathcal{G}_{T-2}(Q_{T-2},u_{T-2},y_{T-2})$. So that upon using (16) at time $T-2$, (17) will apply again:

$$U_{T-2}(Q_{T-2}) \le \mathbb{F}^{Q_{T-2},\Gamma_{T-2}}_{x,v} \max\Big\{ \mathbb{F}^{Q_{T-1}}_{w} \max\{M(x_T), L_{T-1}(x_{T-1},u_{T-1},w)\}\,, \\ L_{T-2}(x,u_{T-2},v) \Big\}\,.$$

One should be carefull that in the formula above, the mathematical fear operations involve variables x, v, and w, while x_{T-1} stands for $f_{T-2}(x,u_{T-2},v)$

and x_T for $f_{T-1}(x_{T-1}, u_{T-1}, w)$. Using also the fact that $\mathbb{F}_v \mathbb{F}_w \psi(v) = \mathbb{F}_v \psi(v)$, the last inequality can be written as

$$U_{T-2}(Q_{T-2}) \leq \mathbb{F}_x^{Q_{T-2}} \mathbb{F}_{v,w}^{\Gamma_{T-2},\Gamma_{T-1}} \max\Big\{ M(x_T) \, , \, L_{T-1}(x_{T-1}, u_{T-1}, w) \, , \\ L_{T-2}(x_{T-2}, u_{T-1}, v) \Big\} \, .$$

Proceeding in that fashion down to time 0, it finally comes;

$$U_0(Q_0) \leq \mathbb{F}_{x_0}^{Q_0} \mathbb{F}_{\mathbf{w}} \max_t \{ L_t(x_t, u_t, w_t) \} = \mathbb{F}_\omega J(\mathbf{u}, \omega) \, .$$

(We have again used the convention $L_T(x, u, w) = M(x)$.)

The end of the proof proceeds as previously : check that using the s-trategy advocated by the theorem, the inequality signs are all replaced by equality signs, so that indeed, $U_0(Q_0)$ is the minimum value. If the infimum is finite but not attained in (16), choose an ε-efficient strategy, i.e. a strategy that guarantees that we are at most at ε/T of the infimum at each instant of time. This yields a cost no more than $U_0(Q_0) - \varepsilon$.

2.4 Certainty equivalence

A separation theorem can be derived from this result in the same vein as in [8], although we shall choose a slightly modified statement, more parallel to the stochastic case. The difference with [8], in addition to the inclusion of L_t, is in the treatment of the sequence $\{R_t\}$ of the theorem. Notice however that the only reasonable case where this theorem may apply seems to be when this sequence is constant and equal to $-\infty$, meaning that the condition of the theorem is that a saddle point holds for S_t. This condition, however, is not as unlikely as its stochastic counterpart, since several conditions are known, the most famous one being the Von-Neumann-Sion condition, insuring that fact. See, e.g., [12].

Define

$$S_t(x, u) = \mathbb{F}_w^{\Gamma_t} \max\{ V_{t+1} \circ f_t(x, u, w) \, , \, L_t(x, u, w) \} + Q_t(x) \, .$$

Theorem 3 *If there exists a sequence of numbers* $\{R_t\}$, $R_t \in \mathbb{R} \cup \{-\infty\}$ *with* $R_T \leq \inf_x M(x)$ *such that,*

$\forall \mathbf{u} \in \mathcal{U}$, $\forall \omega \in \Omega$, $\forall t \in [0, T-1]$,

$$\max \left\{ \inf_u \max_x S_t(x, u) \, , \, R_{t+1} \right\} = \max \left\{ \max_x \inf_u S_t(x, u) \, , \, R_t \right\}$$

then an optimal control strategy is generated by minimizing at each time step the conditional feared value of $\max\{V_{t+1} \circ f_t \, , \, L_t\}$.

Proof The proof goes by checking that

$$U_t(Q) = \max\{\mathbb{F}_x^Q V_t(x)\,,\, R_t\}\,,$$

which in turn will prove the result in view of the previous theorem.

Indeed, the terminal condition (15) is satisfied, in view of (11) and the fact that one always has $\mathbb{F}M \geq \inf M$.

As a recurence hypothesis, assume that the equality above holds at time $t+1$, for all (reachable) Q. Substitute in the r.h.s. of (16). One quickly get

$$U_t(Q) = \inf_u \max\{\mathbb{F}_y^{\Lambda_t}\mathbb{F}_x^{Q_{t+1}[y]}V_{t+1}\,,\, \mathbb{F}_{x,w}^{Q,\Gamma_t}L_t\,,\, R_{t+1}\}\,,$$

where $Q_{t+1}[y]$ stands for $\mathcal{G}_t(Q,u,y)$ for short. Then replace the first mathematical fear in the r.h.s. above using (17) to get

$$U_t(Q) = \inf_u \mathbb{F}_x^Q\mathbb{F}_w^{\Gamma_t}\max\{V_{t+1}\circ f_t\,,\, L_t\,,\, R_{t+1}\}\,.$$

Because, for any function $\psi(u)$ and constant r one has

$$\max\{\inf_u \psi(u)\,,\, r\} = \inf_u \max\{\psi(u)\,,\, r\}\,,$$

then the l.h.s. of the equality in the hypothesis of the theorem may be re-ordered to make this hypothesis read

$$\inf_u \mathbb{F}_x^Q\mathbb{F}_w^{\Gamma_t}\max\{V_{t+1}\circ f_t, L_t, R_{t+1}\} = \max\Big\{\mathbb{F}_x^Q \inf_u \mathbb{F}_w^{\Gamma_t}\max\{V_{t+1}\circ f_t, L_t\}\,,\, R_t\Big\}.$$

Upon using (12) the result follows.

3 Continuous time

While [7] has a section on continuous time, we chose to forego that problem in [8] because we were not able to get a nice parallel with the stochastic case. We show here how close we can get.

The treatment will be in a large extent formal, as questions pertaining to the regularity of the functions involved are much more delicate here than in the discrete time case, but will nevertheless be as carelessly ignored as in the discrete time case. We shall implicitly make any regularity assumption needed to make our calculations, as our aim is to exhibit the equations one might hope to prove. Finding milder regularity assumptions on the one hand, and a reasonable set of conditions under which they may be shown to hold on the second hand, is a major undertaking yet to be begun. In particular, it will require advances in infinite dimensional viscosity solutions of PDE's (see [10]). An alternative is in the use of set-valued non smooth analysis. See e.g. [2].

The set of admissible state feedbacks may be chosen in the implicit way we explained in [6] and admissible closed loop strategies in a similar way.

3.1 The problem

The dynamical system considered is now continuous-time, so that (1) is replaced by[3]

$$\dot{x} = f_t(x, u, w)\,, \tag{18}$$

for the partial information problem, the observation scheme remains as in the discrete time case (2), the notations $\mathbf{u}$, $\mathbf{w}$ stand for the whole time functions over $[0, T]$. We shall keep the notation x_t for the more classical $x(t)$.

We shall consider (almost) the same performance index as in the discrete time case:

$$J(\mathbf{u}, \omega) = \max\{M(x_T)\,,\ \sup_{t\in[0,T]} L_t(x_t, u_t, w_t)\}\,. \tag{19}$$

The time variable t runs over the continuous time interval $[0, T]$. This creates a difficulty because the control and disturbance variables might be discontinuous at the time when the $\sup_t$ is reached. One way around that difficulty would be to consider the essential supremum. We choose a different approach. We may consider that the time at which the performance index L_t is evaluated to define J is part of the choice of the "opponent", i.e. the disturbance. This is consistent with the fact that we seek the $\min_u \max_{\omega,t} L_t$. In that case, the maximizer may choose to make a discontinuity in w at *its* chosen final time t^* in order to get a larger income. Thus it will insure itself a payoff

$$J = \sup_w L_{t^*}(x_{t^*}, u_{t^*}, w)\,.$$

We shall later on somewhat alter that in the precise definition of the "feared" payoff.

To avoid a difficulty with a discontinuity of u, and as the minimizer is not aware of the t^* the disturbance will choose, we may assume that the control function $\mathbf{u}$ is constrained to be continuous from the left (while the disturbance $\mathbf{w}$ will be continuous from the right).

[3]The index t at f_t, L_t,... should not be mistaken for a time partial derivative. Its use is the same as in the discrete time case, to denote dependence upon t.

We wish now to consider the problem of minimizing

$$H(\mu) = \mathbb{F}_\omega J(\mu,\omega)\,.$$

We must be carefull in the precise definition of the mathematical fear here. Notice that in the stochastic case, that we wish to parallel, it is difficult to give a meaning to an integral performance index where the integrand L_t would depend on the "white noise" $w(t)$, unless we take that dependence to be linear : $L_t = \bar{L}_t(x) + \lambda_t(x)w$. Then, the integral payoff is a diffusion process. It is easy to check that in that case, the expectation of the integral is the integral of the "expected instantaneous cost" $\bar{L}_t(x)$.

It is natural to follow that idea, but formally we do not need any linearity. Let

$$\bar{L}_t(x_t,u_t) := \mathbb{F}_w^{\Gamma_t} L_t(x_t,u_t,w) \tag{20}$$

and choose as the payoff

$$H(\mu) = \mathbb{F}_\omega \max\{M(x_T)\,,\ \sup_{t\in[0,T]} \bar{L}_t(x_t,u_t)\}\,, \tag{21}$$

with $\bar{L}$ defined by (20) above. Thus, as previously said, we let the maximizer choose t^* as well as $w(t^*)$, but moreover, as usual in that theory there is a penalty $\Gamma_{t^*}(w)$ associated to that choice.

Furthermore, at the level of our formal treatment, and as previously, Γ_t may everywhere be taken to depend as well on x_t and u_t, although this makes the regularity issues only worse.

Finally, as in the discrete time case, we may notice that we also have

$$\mathbb{F}_\omega J(\mathbf{u},\omega) = \sup_\omega \left\{ \sup_{t\in[0,T]} \left[\bar{L}_t(x_t,u_t) + \int_0^t \Gamma_s(x_s,u_s,w_s)\,ds \right] + Q(x_0) \right\}, \tag{22}$$

where as previously we have set $\bar{L}_T(x,u) = M(x)$.

3.2 Perfect information

Let us first investigate the complete information problem, where we seek a state feedback strategy $u_t = \varphi_t(x_t)$. We introduce the related Isaacs equation:

$$\forall x \in \mathbb{R}^n\,, \quad V_T(x) = M(x)\,, \tag{23}$$

$$\forall t \in [0,T]\,,\ \forall x \in \mathbb{R}^n\,,$$
$$\inf_u \mathbb{F}_w^{\Gamma_t} \max\left\{ \frac{\partial V_t(x)}{\partial t} + \frac{\partial V_t(x)}{\partial x} f_t(x,u,w)\,,\ L_t(x,u,w) - V_t(x) \right\} = 0\,. \tag{24}$$

We may state the following result:

Theorem 4 *If there exists a C^1 function $(t,x) \mapsto V_t(x)$ satisfying the partial differential equation (23),(24), then the optimal cost in the full information problem is $\mathbb{F}V_0(x_0)$, and if the infimum in u in (24) is reached by an admissible state feedback, say $\varphi^*_t(x)$, it is optimal.*

Let us prove that result. We shall write

$$\frac{dV_t(x)}{dt} := \frac{\partial V_t(x)}{\partial t} + \frac{\partial V_t(x)}{\partial x} f_t(x,u,w)\,.$$

Notice first that since $\mathbb{F}$ and max commute, the second term in the max of (24) is just $\bar{L}_t(x,u) - V_t(x)$.

Pick an arbitrary control function $\mathbf{u}$, and a fixed x_0. Assume moreover that $u(t)$ does not belong to the minimizing u's over a time interval $[0,\tau]$. There are such disturbances that insure that either $dV_t/dt + \Gamma_t(w)$ or $L_t - V_t$ is positive. Hence, either $\bar{L}_0(x_0,u_0) > V_0(x_0)$, and then a fortiori $J > V_0(x_0)$, or $\bar{L}_0(x_0,u_0) \le V_0(x_0)$, but then $dV/dt + \Gamma_t$ is positive. And it will remain nonnegative at least untill $\bar{L}_t = V_t$, or $t = T$ whichever happens first. Let

$$\hat{t} = \inf\{t \mid \bar{L}_t = V_t\}\,,$$

assumed first to be less than T. Then, because $dV/dt + \Gamma_t$ was positive in a right neighborhood of 0 and nonnegative untill $\hat{t}$, we have that

$$V_{\hat{t}}(x_{\hat{t}}) + \int_0^{\hat{t}} \Gamma_s \, ds > V_0(x_0)\,,$$

and since

$$V_{\hat{t}}(x_{\hat{t}}) = \bar{L}_{\hat{t}}(x_{\hat{t}}, u_{\hat{t}})\,,$$

a fortiori, $J + \int \Gamma_s \, ds > V_0(x_0)$. And if there is no such $\hat{t} < T$, then $V_T(x_T) + \int_0^T \Gamma_t \, dt > V_0(x_0)$, which, in view of (23) again proves that $J + \int \Gamma_s \, ds > V_0(x_0)$. Hence, if $\mathbf{u}$ is not chosen as minimizing in (24), the augmented payoff obtained for some disturbances is larger than $V_0(x_0)$. Taking the mathematical fear also w.r.t. x_0 yields a fortiori $\mathbb{F}J > \mathbb{F}V_0(x_0)$.

Assume now that there exists an admissible state feedback strategy $\varphi^*_t(x)$ that provides the $\min_u$ in (24). Then for any disturbance $\mathbf{w}$, both terms in the max of (24) are nonpositive. Thus, on the one hand

$$\frac{dV_t(x)}{dt} + \Gamma_t(w_t) \le 0\,,$$

so that

$$\forall t \in [0,T]\,,\; V_t(x_t) + \int_0^t \Gamma_s(w_s)\, ds \le V_0(x_0)$$

and in particular in view of (23)

$$M(x_T) + \int_0^T \Gamma_s(w_s)\,ds \le V_0(x_0)$$

and on the other hand,

$$\forall t \in [0,T]\,,\ \bar{L}_t(x_t,u_t) \le V_t(x_t)\,,$$

so that using the previous result

$$\forall t \in [0,T]\,,\ \bar{L}_t(x_t,u_t) + \int_0^t \Gamma_s(w_s)\,ds \le V_0(x_0)\,.$$

Therefore, it follows that, even taking the worst disturbance,

$$\mathbb{F}_{\mathbf{w}} J(x_0,\varphi^*,\mathbf{w}) \le V_0(x_0)\,.$$

Now, for the worst disturbance at each instant of time either $dV/dt = 0$ or $L_t = V_t$, both remaining non positive. If these two functions are measurable in t, this defines time intervals over which one of these two situations prevails: either $L_t = V_t$ and $V_t + \int \Gamma_s\,ds$ is nonincreasing, therefore so is $L_t + \int \Gamma_s\,ds$, or $V_t + \int \Gamma_s\,ds$ is constant, while L_t is no more than V_t. Integrating and using (23) in case L_t remains allways less than V_t yields the fact that then $\mathbb{F}_{\mathbf{w}} J(x_0,\varphi^*,\mathbf{w}) = V_0(x_0)$, hence $\mathbb{F}J(\varphi^*,\omega) = \mathbb{F}V_0(x_0)$.

Before we close this section, we make a final remark. In section 2.2, the equation (12) can also be written as

$$\inf_u \mathbb{F}_w \max\{V_{t+1}(x_{t+1}) - V_t(x_t)\,,\ L_t(x_t,u,w) - V_t(x_t)\} = 0\,,$$

so that equation (24), which can be written as

$$\inf_u \mathbb{F}_w \max\Big\{\frac{dV_t(x_t)}{dt}\,,\ L_t(x_t,u,w) - V_t(x_t)\Big\} = 0\,,$$

should come as no surprise.

The parallel is less perfect with the stochastic case, however, where the performance index (19) should be replaced by the classical

$$M(x_T) + \int_0^T L_t(x_t,u_t,w_t)\,dt\,,$$

yielding the classical Bellman equation

$$\inf_u \mathbb{E}\Big[\frac{dV_t(x_t)}{dt} + L_t(x_t,u,w)\Big] = 0\,.$$

3.3 Imperfect information

As in the discrete time case, we introduce a *conditional state cost density* $Q_t(\xi)$ and its dynamics. But this time we need go in some detail concerning the later.

Equations (18) and (2) define maps

$$x_t = \phi_t(u^t, \omega^t)\,, \qquad \text{and} \qquad y_t = \eta_t(u^t, \omega^t)\,.$$

We shall also use the time functions restricted to [0,t):

$$x^t = \phi^t(u^t, \omega^t)\,, \qquad \text{and} \qquad y^t = \eta^t(u^t, \omega^t)\,.$$

For any ξ in $\mathbb{R}^n$, we define the sets of *conditional compatible disturbances*

$$\Omega_t[\xi \mid u^t, y^t] = \{\omega \in \Omega \mid \eta^t(u^t, \omega^t) = y^t \,\text{and}\, \phi_t(u^t, \omega^t) = \xi\}\,.$$

The *conditional worst past cost function* is

$$W_t(\xi) = \sup_{\omega \in \Omega_t[\xi | u^t, y^t]} \left[\int_0^t \Gamma_s(w_s)\, ds + Q_0(x_0)\right]\,.$$

We assume that $W_t(\cdot)$ remains a C^1 (quasi-)concave function with a finite maximum, and let

$$R_t := \max_{\xi \in \mathbb{R}^n} W_t(\xi) \quad \text{and} \quad \widehat{X}_t = \{x \in \mathbb{R}^n \mid W_t(x) = R_t\} \tag{25}$$

to define finally the *conditional state cost density* as

$$Q_t(x) = W_t(x) - R_t\,. \tag{26}$$

Notice that $W_0 = Q_0$, and $R_0 = 0$, so that our notations are consistent. Clearly, $W_t(\cdot)$, R_t, and thus $Q_t(\cdot)$ are functions of (u^t, y^t).

Define also the sets

$$\mathsf{V}_t(x \mid y) = \{w \in \mathsf{W} \mid h_t(x, w) = y\} \tag{27}$$

With our assumption that W_t remains a C^1 function, it obeys a forward Bellman equation:

$$\frac{\partial W_t(x)}{\partial t} = \max_{w \in \mathsf{V}_t(x|y)} \left[-\frac{\partial W_t}{\partial x} f_t(x, u_t, w) + \Gamma_t(w)\right]\,.$$

We may also notice that according to Danskin's theorem (see[9]), we have

$$\dot{R}_t = \max_{\hat{x}\in\widehat{X}_t} \frac{\partial W_t}{\partial t}(\hat{x})$$

By the definition of $\widehat{X}$, $(\partial W_t/\partial x)(\hat{x}) = 0$, so that

$$\dot{R}_t = \max_{\hat{x}\in\widehat{X}_t} \max_{w\in\mathsf{V}_t(\hat{x}|y)} \Gamma_t(w)$$

The r.h.s. above is a function of y. It is nonpositive, and obviously has a zero maximum in y (just pick $y = h_t(\hat{x}, \bar{w})$ with $\Gamma_t(\bar{w}) = 0$). We interpret it as a cost density on y induced in a particular way by the cost density Q_t on x. In that respect, notice that if Q is a cost density, so is pQ for any positive number p. We would normally write

$$\Lambda^{pQ}(y) = \max_x \max_{w\in\mathsf{V}(x|y)} [pQ(x) + \Gamma(w)]$$

the cost density on y induced by pQ. According to classical penalization theory, it is easy to see that the cost density (28) is the limit of the above as $p \to \infty$. As a consequence, we shall write it

$$\Lambda_t^{\infty}(y) = \max_{\hat{x}\in\widehat{X}_t} \max_{w\in\mathsf{V}_t(\hat{x}|y)} \Gamma_t(w)\,. \tag{28}$$

leaving the Q implicit in the notation. We shall denote $\mathbb{F}_y^{\infty}$ or $\mathbb{F}_y^{\infty Q}$ the corresponding mathematical fear operator.

It is concievably feasible to follow in real time the evolution of Q_t as a function of the available information according to the nonlinear PDE

$$\frac{\partial Q_t(x)}{\partial t} = \max_{w\in\mathsf{V}_t(x|y_t)} \left[-\frac{\partial W_t}{\partial x} f_t(x, u_t, w) + \Gamma_t(w)\right] - \Lambda_t^{\infty}(y_t)\,.$$

Denote

$$\frac{dQ_t}{dt} = \left\{x \mapsto \frac{\partial Q_t(x)}{\partial t}\right\}$$

we shall write the above PDE as

$$\frac{dQ}{dt} = \mathcal{G}_t(Q, u_t, y_t)\,. \tag{29}$$

(It is a not-so-simple matter at this time to convince oneself that the arguments in $\mathcal{G}$ above are indeed those on which this derivative depends.)

We are now in a position to state the dynamic programming equation, bearing on a Value function $U_t(Q)$ from the set $\mathcal{Q}$ of cost densities over $\mathbb{R}^n$ into $\mathbb{R}$. We assume that $U_t(Q)$ has both a partial derivative in t and a continuous Frechet derivative in Q in the topology of pointwise convergence over $\mathcal{Q}$, denoted $D_Q U$. [4]

$$\forall Q \in \mathcal{Q}, U_T(Q) = \mathbb{F}^Q M(x)\,, \tag{30}$$

$$\forall t \in [0,T]\,,\, \forall Q \in \mathcal{Q}\,,$$
$$\inf_u \max\left\{\mathbb{F}_y^{\infty Q}\left[\frac{\partial U_t(Q)}{\partial t} + D_Q U_t(Q)\mathcal{G}_t(Q,u,y)\right], \mathbb{F}_{x,w}^{Q,\Gamma_t} L_t(x,u,w) - U_t(Q)\right\} = 0 \tag{31}$$

Theorem 5 *If for all admissible controls the functions $W_t(\cdot)$ remain C^1, and if there exists a regular enough function $(t,Q) \mapsto U_t(Q)$ satisfying (30),(31) above, then the optimal value of the imperfect information game is $U_0(Q_0)$. If moreover, the minimum in u is attained in (31) at $\mu_t^*(Q)$ and if this, together with (29) initialized at Q_0, constitutes an admissible strategy, then it is optimal.*

Assume that μ^* exists and is admissible. (It is indeed causal, admissibility pertains to the existence of solutions to (18) and (29)). Assume we pick $u_t = \mu_t^*(Q_t)$ for all t, where of course Q_t is given by (29). Pick a disturbance $\{w_t\}$, and consider the trajectories $\{u_t\}$, $\{x_t\}$, $\{y_t\}$, and $\{Q_t\}$ generated. We have, on the one hand,

$$\frac{dU_t(Q_t)}{dt} + \Lambda_t^\infty(y_t) \leq 0\,,$$

or, recalling that $\Lambda_t^\infty(y_t) = \dot{R}_t$, and integrating

$$\forall t \in [0,T]\,,\, U_t(Q_t) + R_t \leq U_0(Q_0)\,. \tag{32}$$

In particular, for $t = T$, and taking (30) into account,

$$\max_x [M(x) + Q_T(x)] + R_T \leq U_0(Q_0)\,.$$

Now, recall that, by definition,

$$Q_t(x) + R_t = W_t(x) =$$
$$\max_\omega \left[\int_0^t \Gamma_s(w_s)\,ds + Q_0(x_0) \mid \phi_t(\mathbf{u},\omega) = x\,,\, \eta^t(\mathbf{u},\omega) = y^t\right].$$

[4]Precisely, what is requested is that, if there exists a function $\dot{Q}(x)$ such that $\forall x$, $Q_{t+h}(x) = Q_t(x) + \dot{Q}(x)h + o(h)$, then, $U_t(Q_{t+h}) = U_t(Q_t) + hD_Q U_t(Q_t)\dot{Q} + o(h)$.

Therefore, whatever the actual x_T, we conclude that

$$M(x_T) + \int_0^T \Gamma_s(w_s)\,ds + Q_0(x_0) \le U_0(Q_0)\,. \tag{33}$$

On the second hand, we have

$$\forall t \in [0,T]\,,\ \mathbb{F}_{x,w}^{Q_t,\Gamma_t} L_t(x,u_t,w) \le U_t(Q_t)\,.$$

Together with (32), this yields

$$\forall t \in [0,T]\,,\ \bar L_t(x_t,u_t) + Q_t(x_t) + R_t \le U_0(Q_0)\,.$$

Hence, and for every $\omega \in \Omega$,

$$\sup_{t\in[0,T]} \left\{ \bar L_t(x_t,u_t) + \int_0^t \Gamma_s(w_s)\,ds + Q_0(x_0) \right\} \le U_0(Q_0)\,.$$

As previously, this is easily seen to be equivalent to

$$\forall \omega \in \Omega\,,\quad \sup_t \bar L_t(x_t,u_t) + \int_0^T \Gamma_s(w_s)\,ds + Q_0(x_0) \le U_0(Q_0)\,. \tag{34}$$

Now, (33) and (34) together show that, upon playing according to μ^*, the controller insures that

$$\mathbb{F}J(\mu^*,\omega) \le U_0(Q_0)\,.$$

Fix now an $\mathbf{u}$ and an ω such that for an open interval of time $(0,\tau)$, u_t does not belong to the argmax in (31) with Q_t for Q. Then either

$$\bar L_0(x_0,u_0) + Q_0(x_0) > U_0(Q_0)\,,$$

and this is enough to ascertain that

$$\mathbb{F}J(\mathbf{u},\omega) > U_0(Q_0)\,,$$

or $\bar L_0(x_0,u_0) + Q_0(x_0) \le U_0(Q_0)$ but then, for a positive time interval,

$$\frac{dU_t(Q_t)}{dt} + \dot R_t > 0\,.$$

In that case, either $d(U_t+R_t)/dt \ge 0$ until $t=T$, and therefore $U_T(Q_T) + R_T > U_0(Q_0)$, or it lasts only until a time $\hat t$ when $\mathbb{F}_x^{Q_{\hat t}} \bar L_{\hat t}(x,u_{\hat t}) = U_{\hat t}(Q_{\hat t})$. Let $\bar x$ provide the $\max_x$ in $\mathbb{F}_x^{Q_{\hat t}} \bar L_{\hat t}$. Notice that $Q_{\hat t}(x)$ is finite only for those x that are compatible with the past information. Therefore, there exists an

$\bar{\omega}$ that yields the same $y^{\hat{t}}$ and hence the same $Q_{\hat{t}}$ as the one considered here, and such that $\phi_{\hat{t}}(u^{\hat{t}}, \bar{\omega}) = \bar{x}$. For that $\bar{\omega}$ we have

$$\bar{L}_{\hat{t}}(x_{\hat{t}}, u_{\hat{t}}) + Q(x_{\hat{t}}) + R_{\hat{t}} > U_0(Q_0)\,.$$

Given the definition of $W_{\hat{t}} = Q_{\hat{t}} + R_{\hat{t}}$, may be for yet another $\tilde{\omega}$ compatible with the same past information and $x_{\hat{t}} = \bar{x}$,

$$\bar{L}_{\hat{t}}(x_{\hat{t}}, u_{\hat{t}}) + \int_0^{\hat{t}} \Gamma_s(\tilde{w}_s)\, ds + Q_0(\tilde{x}_0) > U_0(Q_0)\,.$$

In every cases,

$$\mathbb{F}J(\mathbf{u}, \omega) > U_0(Q_0)\,,$$

Hence the result is proved.

3.4 Certainty equivalence

We assume in this section that L_t is independant of u, a rather classical case in such problems. (This is the case, for instance, for "surveillance problems" where $L_t = d(x, \mathcal{C}_t)$ with d the distance, and $\mathcal{C}_t$ a (moving) target in $\mathbb{R}^n$.)

Then, essentially the same certainty equivalence theorem as in [7] holds.

Assume that for every $(\mathbf{u}, \omega) \in \mathcal{U} \times \Omega$ and for every $t \in [0, T]$, the maximum in

$$\max_x [V_t(x) + Q_t(x)]$$

is attained at a *unique* point $\hat{x}_t$ in $\mathbb{R}^n$. Then the control

$$u_t = \varphi_t^*(\hat{x}_t)\,,$$

with φ_t^* as in theorem 4, is optimal, and insures a payoff $\mathbb{F}^{Q_0} V_0$.

As in [7], the proof goes by checking that

$$U_t(Q) := \mathbb{F}^Q V_t$$

solves the equations (30),(31). It is shown in [7] that with that form,

$$\frac{\partial U_t(Q)}{\partial t} = \frac{\partial V_t(\hat{x}_t)}{\partial t}$$

and for a function $G(\cdot)$ from $\mathbb{R}^n$ into $\mathbb{R}$,

$$D_Q U_t(Q_t) G = G(\hat{x}_t)\,.$$

Notice also that

$$\frac{\partial W_t(x)}{\partial t} = \frac{\partial Q_t(x)}{\partial t}\,,$$

and that thus, recalling the definition of $\hat{x}_t$,

$$-\frac{\partial W_t(\hat{x}_t)}{\partial x} = \frac{\partial V_t(\hat{x}_t)}{\partial x}\,.$$

Checking (31) amounts to looking at

$$\max\Big\{\max_y\Big[\frac{\partial V_t(\hat{x}_t)}{\partial t} + \max_{w|y}\Big(\frac{\partial V_t(\hat{x}_t)}{\partial x} f_t(\hat{x}_t, u, w) + \Gamma_t(w) - \dot{R}_t(y)\Big) + \Lambda_t^\infty(y)\Big]\,,$$
$$\max_x[\bar{L}_t(x) + Q_t(x)] - \max_x[V_t(x) + Q_t(x)]\Big\}\,.$$

which simplifies into

$$\max\Big\{\max_w\Big[\frac{\partial V_t(\hat{x}_t)}{\partial t} + \frac{\partial V_t(\hat{x}_t)}{\partial x} f_t(\hat{x}_t, u, w) + \Gamma_t(w)\Big]\,,$$
$$\max_x[\bar{L}_t(x) + Q_t(x)] - [V_t(\hat{x}_t) + Q_t(\hat{x}_t)]\Big\}\,.$$

By definition, $u_t = \varphi_t^*(\hat{x}_t)$ provides the minimum in the first term of the max operator. The only new point in the proof has to do with the second element in the max operation of (31). Just notice that for every $x \in \mathbb{R}^n$, $\bar{L}_t(x) \le V_t(x)$, so that also

$$\bar{L}_t(x) + Q_t(x) \le V_t(x) + Q_t(x) \le V_t(\hat{x}_t) + Q_t(\hat{x}_t)\,.$$

If $\bar{L}_t$ and V_t coincide at $\hat{x}_t$, then

$$\max_x\{\bar{L}_t(x) + Q_t(x)\} = V_t(\hat{x}_t) + Q_t(\hat{x}_t)\,,$$

or alternatively

$$\mathbb{F}^{Q_t}\bar{L}_t = \mathbb{F}^{Q_t}V_t$$

while otherwise, the l.h.s. above is always less than or equal to the r.h.s.

This shows that indeed, as in (24), $\varphi_t^*(\hat{x}_t)$ insures that one of the two terms in the max is zero, while both are always nonpositive.

4 Conclusion

The dynamic programming equations (11,12) have been well known for some time, at least in the case where on the one hand L does not depend on u and w, and, on the other hand, the additive terms in the criterion are absent. They are then just a variational inequality (see [5]). They were usually written in a slightly different way. See for instance [4, 11] for the same application as here. The above theory mainly reformulates them to show how they are natural within the context of feared value control, providing the natural extension to the full performance index (22). The continuous time partial information case is still too sketchy. At least we have here the formal PDE's to investigate.

In discrete time, the full information case was most probably well known to many, although probably not with the full performance index (10). As it stands, it is fairly general, and embodies the "pure L^∞" problem by taking all the Γ_t's identically zero, and the "pure L^1" problem by taking all the L_t's —but not M if necessary— equal to $-\infty$.

The imperfect information case appears to be original here. The calculations we performed are greatly simplified by the use of the feared value and its algebraic properties, in particular its (max, +) linearity, and thanks to the lemma (17) it suggests.

The last question is : "how about practical applications ?" The answer is that there is a long way to go. We may expect particular cases to arise where things simplify, and in particular where the conditional state cost density is finite dimensional. Some instances have been found, beyond the classical linear quadratic case of $\mathcal{H}_\infty$ optimal control, for the problem without the L^∞ cost. In that respect, this theory may be considered as potentially more usable than that of nonlinear partial information stochastic control.

References

[1] M. Akian, J-P. Quadrat and M. Viot: Bellman Processes, *11th International Conference on Systems Analysis and Optimization*, Sophia Antipolis, France, 1994.

[2] "The method of characteristics revisited, a viability approach, a mini-course", *Cahiers du Centre de Recherche Viabilité Jeux Contrôle*, University Paris Dauphine, 1999.

[3] F. Baccelli, G. Cohen, G-J. Olsder, J-P. Quadrat: *Synchronization and Linearity*, Wiley, Chichester, U.K., 1992.

[4] E.N. Barron, “Differential Games with Maximum Cost”, Nonlinear Analysis, (**14**), pp 971–989, 1990.

[5] A. Bensoussan and J.L. Lions *Application des inéquations variationnelles en contrôle stochastique*, Dunod, Paris, 1978.

[6] P. Bernhard: “Differential games, Isaacs'equation”, in M.Singh ed.: *Encyclopaedia of Systems and Control*, pp 1010–1017, Pergamon, 1987.

[7] P. Bernhard “Expected Value, Feared Value and partial Information optimal Control”, in G.J.Olsder ed.: *New trends in Dynamic Games and Applications*, Annals of the International Society of Dynamic Games **3**, pp 3–24, Birkhauser, Boston, USA, 1995.

[8] P. Bernhard “A Separation Theorem for Expected Value and Feared Value Dicrete Time Control” COCV, Vol. 1, pp. 191-206, SMAI, http://www.emath.fr/cocv, 1996.

[9] P. Bernhard and A. Rapaport: “On a Theorem of Danskin with an Application to a Theorem of Von Neumann-Sion”, *Non Linear Analysis, Theory, Methods and Applications*, **24**, pp 1163–1181, Pergamon, 1995.

[10] M.R. James and J.S. Baras, “Partially Observed Differential Games, Infinite Dimensional HJI Equations, and Nonlinear H-Infinity Control”, SIAM J.on Control and Optimization, **34**, pp 1342-1364, 1996.

[11] A. Rapaport: “A Characterization of Barriers of Differential Games”, JOTA (**97**), 1998.

[12] S. Simons, “Minimax theorems and their proofs”, *Minimax and applications*, Ding Zhu Du & Panos M. Pardalos eds., pp1–23, Kluwer Academic Publishers, Dordrecht-Boston, 1995.