On the Performance Index of Feared Value Control

Pierre Bernhard

August 5, 1997

Abstract

This is yet another step in the continuing attempt to perfect the parallel between stochastic control and min-max control using the concept of feared value as the parallel to expected value. The (modest) contribution of this short paper is twofold: on the one hand clarify the role of the integral part of the cost in that parallel, on the other hand extend the parallel to the continuous time case as much as possible.

1 Introduction

This to be considered as another interim report in the continuing 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.

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 [3]. The concept of 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 [2]. In [3], 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 runing cost, or integral cost, added to the terminal cost. This is what we do here.

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, and this is what we investigate in the second part.

2 Discrete time

2.1 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:

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

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

As in our previous two papers on that topic, we let $\omega = (x_0, \mathbf{w})$ denote the disturbations a priori unknown to the controller, and let $\omega \in \Omega = \mathbb{R}^n \times \mathsf{W}^{T+1}$. We also use $\omega^t = (x_0, w^t) \in \Omega^t = \mathbb{R}^n \times \mathsf{W}^t$.

The problem shall always be to choose a control sequence to achieve a certain goal, based on 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 $\varphi_T(\mu, \omega)$ where what we mean is the final state on the trajectory generated by that μ and ω.

2.2 The stochastic problem

In stochastic control, it is assumed that Ω is endowed with a (known) probability distribution. Usually, we assume that x_0 and $\mathbf{w}$ are independant, and moreover that $\mathbf{w}$ is a white process, so that the probability on Ω is entirely specified by a probability density P_0 over $\mathbb{R}^n$ governing x_0, and a set of probability densities Π_t, $t = 0, \dots, T$ over W governing the w_t's.

The mathematical expectation of any function $\psi(\omega)$ is thus

$$\mathbb{E}\psi = \int_\Omega \psi(x_0, w_0, w_1, \dots, w_{T-1}) P_0(x_0)\Pi_0(w_0)\Pi_1(w_1)\cdots\Pi_{T-1}(w_{T-1}) dx_0\, dw_0\, dw_1 \dots dw_{T-1}\,.$$

In [3], we considered a performance index

$$J(\mathbf{u}, \omega) = M(x_T) = M \circ \varphi_T(\mathbf{u}, \omega)\,.$$

Here, we wish to find the minimax parallel of the case where an integral cost is added. Let therefore

$$J(\mathbf{u}, \omega) = M(x_T) + \sum_{t=0}^{T-1} L_t(x_t, u_t, w_t)\,. \tag{3}$$

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

The problem we shall consider is to minimize

$$G(\mu) = \mathbb{E}J(\mu, \omega)\,.$$

However, we shall not repete the classical theory, as can be found in [3], with this augmented performance index. It is well known that it just results in a term $+L_t$ being added to the right hand side of Bellman's equation, be it in the state feedback theory, or in the partial information theory.

We only write here these two Bellman's equations, using the same notations as in [3]:

We first introduce the *full information* Bellman return function V_t defined by the classical dynamic programming recursion :

$$\begin{aligned} \forall x \in \mathbb{R}^n, \qquad V_T(x) &= M(x)\,, && (4)\\ \forall t \in [0, T-1]\,, \forall x \in \mathbb{R}^n\,, \quad V_t(x) &= \inf_u \mathbb{E}_w^{\Gamma_t}[V_{t+1}(f_t(x,u,w)) + L_t(x,u,w)]\,. && (5) \end{aligned}$$

The infimum of the performance index $G(\varphi)$ is $\mathbb{E}_x^{P_0} V_0(x)$ (where we recall that the probability density P_0 of x_0 is a data). Furthermore, if the minimum is reached for all (t, x) in (5), then the argument $\varphi_t^*(x)$ of the minimum is an optimal *state feedback strategy.*

In the partial information case, the Bellman return function U is a function of the conditional cost probability density $P_t \in \mathcal{P}_t$, itself obtained through a non linear recursive filter of the form

$$P_{t+1} = \mathcal{F}_t(P_t, u_t, y_t) \tag{6}$$

initialized at P_0. Together with Π_t, it generates through (2) an "a priori" probability density Δ_t on the output to come y_t. Then the sequence $\{U_t\}$ is obtained by the recurrence relation

$$\begin{aligned} \forall P \in \mathcal{P}_T, \qquad U_T(P) &= \mathbb{E}_x^P M(x)\,, && (7)\\ \forall t \in [0, T-1]\,, \forall P \in \mathcal{P}_t\,, \quad U_t(P) &= \inf_u \mathbb{E}_y^{\Delta_t} U_{t+1}(\mathcal{F}_t(P,u,y)) + \mathbb{E}_{x,w}^{P,\Pi_t} L_t(x,u,w)\,. && (8) \end{aligned}$$

We can state the following theorem:

Theorem 1 *If there exists a sequence of functions $\{U_t\}$ from $\mathcal{P}_t$ into $\mathbb{R}$ satisfying equations (7)(8), then the optimal cost is $U_0(P_0)$.*

Moreover, assume that the minimum in u is attained in (8) above at $u = \hat{\mu}_t(P)$. Then (6) and

$$u_t = \hat{\mu}_t(P_t) \tag{9}$$

define an optimal controller for the stochastic control problem.

2.3 The minimax problem

We now turn our attention to the minimax case. Now, Ω is assumed to be endowed with a cost measure governing the decision variable ω. We still assume that x_0 and $\mathbf{w}$ are independant, 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 W governing the w_t's. And the mathematical fear of any function $\psi(\omega)$ is defined as

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

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 performance index (3) is now

$$J(\mathbf{u},\omega)=\max\{M(x_T)\,,\,\max_t L_t(x_t,u_t,w_t)\}\,. \tag{10}$$

Of course, exactly as in the previous case, there is no real need to distinguish between the notations M and L_T. It is however convenient to keep this parallel with the continuous time case.

As a consequence, the problem we consider is to minimize over $\mathcal{M}$

$$H(\mu)=\mathbb{F}J(\mu,\omega)\,. \tag{11}$$

2.3.1 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. We have an (extended) Isaacs equation:

$$\forall x\in\mathbb{R}^n\,,\qquad V_T(x)\;=\;M(x)\,, \tag{12}$$

$$\forall t\in[0,T-1]\,,\forall x\in\mathbb{R}^n\,,\quad V_t(x)\;=\;\inf_u \mathbb{F}_w^{\Gamma_t}\max\{V_{t+1}(f_t(x,u,w)),L_t(x,u,w)\}\,. \tag{13}$$

We may state the following theorem

Theorem 2 *If the backwards recursion (12),(13) generates a bounded Value function V, then, the infimum of the problem (11) is given by $\mathbb{F}V_0(x_0)$ (recall that the initial state cost density Q_0 is given). Moreover, if the minimum in u is reached at $\varphi^*(t,x)$ in (13), then this is an optimal state feedback strategy.*

We shall sketch the proof which is straightforward. It is worthwhile, however, to point out the following fact. We are interested in

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

For the sake of simplicity, let us write $L_T(x,u,w)$ for $M(x)$. The above expression involves the quantity $\mathbb{F}_{\mathbf{w}}J$ which can be expanded into

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

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

Proposition 1

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

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

$$\max_t\Big[L_t(x_t,u_t,w_t)+\sum_{k=0}^{t}\Gamma_k(w_k)\Big]\geq\max_t\Big[L_t(x_t,u_t,w_t)+\sum_{k=0}^{T-1}\Gamma_k(w_k)\Big] \tag{14}$$

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

$$L_{\hat{t}}(x_{\hat{t}},u_{\hat{t}},w_{\hat{t}})+\sum_{k=0}^{\hat{t}}\Gamma_k(w_k)>\max_t\Big[L_t(x_t,u_t,w_t)+\sum_{k=0}^{T-1}\Gamma_k(w_k)\Big] \tag{15}$$

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 (15). 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 (14) yields the proposition.

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 one in the $\mathbb{F}_w$ operation of (13). According to (13), we have along the trajectory $\mathbf{x}$ thus generated

$$\begin{aligned}V_0(x_0)&\leq\max\{V_1(x_1)\,,L_0(x_0,u_0,w_0)\}+\Gamma_0(w_0)\\&=\max\{V_1(x_1)+\Gamma_0(w_0)\,,L_0(x_0,u_0,w_0)+\Gamma_0(w_0)\}\,.\end{aligned}$$

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

$$\begin{aligned}V_0(x_0)\leq\max\{&V_2(x_2)+\Gamma_1(w_1)+\Gamma_0(w_0)\,,\\&L_1(x_1,u_1,w_1)+\Gamma_1(w_1)+\Gamma_0(w_0)\,,L_0(x_0,u_0,w_0)+\Gamma_0(w_0)\}\,,\end{aligned}$$

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

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

with $L_T(x,u,w) = M(x)$ using (12). Use the proposition to conclude that a fortiori

$$V_0(x_0) \le \mathbb{F}_{\mathbf{w}} J(x_0, \mathbf{u}, \mathbf{w})\,. \tag{16}$$

But if u_t is chosen minimizing the r.h.s of (13), the $\le$ 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 chosing an arbitrary sequence $\mathbf{w}$ to have the opposite inequality signs in the above calculations, that reduce to equal signs if w choses the maximizing one, to conclude that indeed

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

which, together with (16), concludes the proof upon taking the mathematical fear with respect to x_0 of both sides.

2.3.2 Imperfect information

We now turn to the case where the minimizer only knows the output (2). The solution follows that proposed in [3] 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 [3]. It can be computed recursively through an equation of the form

$$Q_{t+1} = \mathcal{G}_t(Q_t, u_t, y_t)\,. \tag{17}$$

This state cost density, together with the cost density Γ_t of w_t, induces through (2) a cost density Λ_t on y_t.

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

$$\forall Q \in \mathcal{Q}\,, \qquad U_T(Q) = \mathbb{F}_x^Q M(x)\,, \tag{18}$$
$$\forall t \in [0, T-1], \forall Q \in \mathcal{Q}\,, U_t(Q) = \inf_u \max\Big\{\mathbb{F}_y^{\Lambda_t} U_{t+1}(\mathcal{G}_t(Q,u,y))\,, \mathbb{F}_{x,w}^{Q,\Gamma_t} L_t(x,u,w)\Big\}. \tag{19}$$

The theorem is as expected:

Theorem 3 *If the recursion (18),(19) 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 (19) for every $(t,P) \in [0,T-1] \times \mathcal{P}$, this together with (17) initialized at Q_0, defines an optimal control strategy for problem (10),(11).*

The proof relies as in [3] on the fact that, writing $Q_{t+1}[y]$ for $\mathcal{G}_t(Q_t,u,y)$, one has for any function ψ

$$\mathbb{F}_y^{\Lambda_t}\mathbb{F}_x^{Q_{t+1}[y]}\psi(x) = \mathbb{F}_{x,w}^{Q_t,\Gamma_t}\psi(f_t(x,u,w))\,. \tag{20}$$

Fix a control sequence $\mathbf{u}$, and assume any control sequence $\mathbf{w}$. It generates a sequence $\{Q_t\}$. Equation (19) written at time $T-1$ yields

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

Use (18) to substitute in the first term of the r.h.s. above, and make use of (20). It reads

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

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

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

Notice that

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

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 (19) at time $T-2$, (20) will apply again:

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

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

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

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

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

(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 strategy 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 (19), 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 $U_0(Q_0) - \varepsilon$.

2.3.3 Separation theorem

In this section, we assume for simplicity that L_t does not depend on u. Then the separation theorem of [3], still holds unchanged, there is no point in stating it again. The proof extends almost trivially, again thanks to the $(\max,+)$ linearity of the mathematical fear operation. (See the continuous time section for more details).

3 Continuous time

While [2] has a section on continuous time, we chose to forego that problem in [3] 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 make any regularity assumption we need 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.

The set of admissible state feedbacks may be chosen in the implicit way we explained in [1] 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

$$\dot{x} = f_t(x, u, w)\,, \tag{21}$$

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

$$J(\mathbf{u}, \omega) = \max\{M(x_T)\,,\ \sup_{t\in[0,T]} L_t(x_t, u_t, w_t)\}\,. \tag{22}$$

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 that time. 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 disontinuity 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 contiunuous from the left (while the disturbance $\mathbf{w}$ will be continous from the right).

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. It turns out to be natural to decide that there is an "impulsive cost" $\Gamma_t(w_t)$ to the disturbance, associated with the choice of the time instant t^* when the payoff is juged, and to the choice of the discontinuity allwed to it at that time. This will be done through the following device: let $\mathbb{F}J$ be defined as

$$\mathbb{F}_\omega J(\mathbf{u}, \omega) = \max_\omega\Big[\max\{M(x_T), \sup_t[L_t(x_t, u_t, w_t) + \Gamma_t(w_t)]\} + \int_0^T \Gamma_s(w_s)\,ds\Big]. \tag{23}$$

Alternatively, if we prefer to stick with the definition that

$$\mathbb{F}_\mathbf{w}\psi(\mathbf{w}) = \max_\mathbf{w}\Big[\psi(\mathbf{w}) + \int_0^T \Gamma_s(w_s)\,ds\Big]\,,$$

we may decide that the impulsive cost is in the "runing cost", so that the payoff to be maximinimized is

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

Thus the payoff (23) may also be written as

$$H(\mu) = \mathbb{F}_\omega \max\{M(x_T)\,,\ \sup_{t\in[0,T]} \bar{L}_t(x_t, u_t)\}\,, \tag{24}$$

with the above definition of $\bar{L}$.

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

$$\mathbb{F}_\omega J(\mathbf{u}, \omega) = \max_\omega \max\Big\{M(x_T) + \int_0^T \Gamma_s(w_s)\,ds\,,\ \sup_t \Big[\bar{L}_t(x_t, u_t) + \int_0^t \Gamma_s(w_s)\,ds\Big]\Big\}\,.$$

3.2 The complete information problem

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:

$$\forall x \in \mathbb{R}^n\,, \quad V_T(x) = M(x)\,, \tag{25}$$

$$\forall t \in [0,T]\,,\ \forall x \in \mathbb{R}^n\,,\quad \inf_u \mathbb{F}_w^{\Gamma_t} \max\left\{\frac{\partial V_t(x)}{\partial t} + \frac{\partial V_t(x)}{\partial x} f_t(x,u,w)\,,\ L_t(x,u,w) - V_t(x)\right\} = 0\,. \tag{26}$$

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 (25),(26), then the optimal cost in the full information problem is $\mathbb{F}V_0(x_0)$, and if the infimum in u in (26) is atained by an admissible state feedback, say $\varphi_t^*(x)$, it is optimal.*

Let us proove 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 (26) 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 (25) again proves that $J + \int \Gamma_s\, ds > V_0(x_0)$. Hence, if $\mathbf{u}$ is not chosen as minimizing in (26), 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 (26). Then for any disturbance $\mathbf{w}$, both terms in the max of (26) are nonpositive. Thus, on the one hand

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

so that

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

and in particular in view of (25)

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

and on the other hand,

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

so that using the previous result

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

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

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

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 (25) 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 2.3.1, the equation (13) can also be written as

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

or, for that matter, for any "step size" $h > 0$

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

so that equation (26), 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 (22) 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 (21) and (2) define maps

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

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

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

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

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

The *conditional worst pas cost* is

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

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

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

to define finally the *conditional state cost density* as

$$Q_t(x) = W_t(x) - R_t\,.$$

Notice that $W_0 = Q_0$, and $R_0 = 0$, so that our notations are consistent.

Define also the sets

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

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[4]), we have

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

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

$$\dot{R}_t = \max_{\hat{x}\in\widehat{X}_t} \max_{w\in\mathsf{V}_t(\hat{x}|y)} \Gamma_t(w) \tag{27}$$

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 (27) is the limit of the above as $p \to \infty$. As a consequence, we shall write it

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

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

$$\frac{dQ}{dt} = \mathcal{G}_t(Q, u_t, y_t)\,. \tag{28}$$

(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 $D_Q U$.

$$\forall Q \in \mathcal{Q}, U_T(Q) = \mathbb{F}^Q M(x)\,, \tag{29}$$

$$\begin{aligned}&\forall t \in [0, T]\,,\ \forall Q \in \mathcal{Q}\,,\\ &\inf_u \max\left\{\mathbb{F}_y^{\infty Q}\left[\frac{\partial U_t(Q)}{\partial t} + D_Q U_t(Q)\mathcal{G}_t(Q, u, y)\right]\,,\ \mathbb{F}_{x,w}^{Q,\Gamma_t} L_t(x, u, w) - U_t(Q)\right\} = 0\,.\end{aligned} \tag{30}$$

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 (29),(30) above, then the optimal value of the imperfect information game is $U_0(Q_0)$. If moreover, the minimum in u is attained in (19) at $\mu_t^*(Q)$ and if this, together with (28) 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 (21) and (28)). Assume we pick $u_t = \mu_t^*(Q_t)$ for all t, where of course Q_t is given by (28). 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,

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

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

$$\forall t \in [0,T]\,,\; U_t(Q_t) + R_t \le U_0(Q_0)\,. \tag{31}$$

In particular, for $t = T$, and taking (29) into account,

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

Now, recall that, by definition,

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

Therefore, whatever the actual x_T, we conclude that

$$M(x_T) + \int_0^T \Gamma_s(w_s)\,ds + Q_0(x_0) \le U_0(Q_0)\,. \tag{32}$$

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 (31), 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 $\omega \in \Omega$,

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

As previously, this is easily seen to be equivalent to

$$\forall \omega \in \Omega\,,\quad \sup_t \bar{L}_t(x_t,u_t) + \int_0^T \Gamma_s(w_s)\,ds + Q_0(x_0) \le U_0(Q_0)\,. \tag{33}$$

Now, (32) and (33) 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 (30) 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 $\varphi_{\hat{t}}(u^{\hat{t}},\bar{\omega})=\bar{x}$. 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 [2] holds.

Assume that for every $(\mathbf{u},\omega)\in\mathcal{U}\times\Omega$ and for evry $t\in[0,T]$, 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 [2], the proof goes by checking that

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

solves the equations (29),(30). It is shown in [2] that

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

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 (30) 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 (30). 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 (26), $\varphi_t^*(\hat{x}_t)$ insures that one of the two terms in the max is zero, while both are always nonpositive.

References

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

[2] 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

[3] 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

[4] 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