



A Separation Theorem for Expected Value and Feared Value Discrete Time Control

Pierre Bernhard

► To cite this version:

Pierre Bernhard. A Separation Theorem for Expected Value and Feared Value Discrete Time Control. RR-2764, INRIA. 1995. inria-00073927

HAL Id: inria-00073927

<https://inria.hal.science/inria-00073927>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Separation Theorem for Expected Value and Feared Value Discrete Time Control

Pierre Bernhard

N° 2764

Décembre 1995

PROGRAMME 5



***apport
de recherche***

A Separation Theorem for Expected Value and Feared Value Discrete Time Control

Pierre Bernhard

Programme 5 — Traitement du signal, automatique et productique
Projet Miaou

Rapport de recherche n° 2764 — Décembre 1995 — 21 pages

Abstract: We show how the use of a parallel between the ordinary $(+, \times)$ and the $(\max, +)$ algebras, Maslov measures that exploit this parallel, and more specifically their specialization to probabilities and the corresponding *cost measures* of Quadrat, offer a completely parallel treatment of stochastic and minimax control of disturbed nonlinear discrete time systems with partial information. This paper is based upon, and improves, the discrete time part of the earlier paper [8].

Key-words: Dynamical System, Minimax Control, Max-Plus Algebra, Dynamical Games

(Résumé : *tsvp*)

Un théorème de séparation pour la commande en espérance mathématique et en frayer mathématique d'un système dynamique à temps discret.

Résumé : Nous montrons comment l'utilisation d'un parallèle entre l'algèbre "classique" $(+, \times)$ et l'algèbre $(\max, +)$, des mesures de Maslov qui exploitent ce parallèle, et plus particulièrement leur spécialisation aux *mesures de coût* de Quadrat, permet un traitement entièrement parallèle de la commande stochastique ou minimax en feedback de sortie d'un système à temps discret perturbé. Cet article améliore une communication au symposium de l'International Society of Dynamic Games, Saint Jovite, 1994, publié sous la référence [8] ci-dessous.

Mots-clé : Système Dynamique, Commande Minimax, Algèbre Max-Plus, Jeux Dynamiques

1 Introduction

Minimax control, or worst case design, as a means of dealing with uncertainty is an old idea. It has gained a new popularity with the recognition, in 1988, of the fact that H_∞ -optimal control could be cast into that concept. Although some work in that direction existed long before (see [9]), this viewpoint has vastly renewed the topic. See [3] and related work.

Many have tried to extend this work to a nonlinear setup. Most prominent among them perhaps is the work of Isidori, [15, 16] but many others have followed suit : [25, 4, 5] and more recently [18, 19, 20]. This has contributed to a renewed interest in nonlinear minimax control.

If we decide that minimax is an alternative to stochastic treatment of disturbances, it makes sense to try to establish a parallel. To that end, we build upon the work of Maslov and coworkers [21, 22] on Maslov measures, revisiting in fact Choquet's capacities [11], and more specifically on the work of Quadrat and coworkers [23, 2, 1] who developed the dual in that parallel to probability theory, introducing a special class of Maslov measures, *cost measures*. One can also refer to the work of Del Moral [13, 14]

2 Quadrat's morphism

In a series of papers [23, 2, 1], Quadrat and coauthors have fully taken advantage of the morphism introduced between the ordinary algebra $(+, \times)$ and the $(\min, +)$, or alternatively the $(\max, +)$, algebra to develop a *decision calculus* parallel to probability calculus. It has been pointed out by Quadrat and coauthors that a possible way of understanding that morphism was through Cramer's transform. We shall not, however, develop that way of thinking here, but merely rely on the algebraic similarity between the two calculus, as stressed by the theory of Maslov measures.

Let us briefly review some concepts, based on [1].

2.1 Cost measure

The parallel to a probability measure is a *cost measure*. Let Ω be a topological space, $\mathcal{A}$ a σ -field of subsets, $K : \mathcal{A} \rightarrow \mathbb{R} \cup \{-\infty\}$ is called a cost measure if it satisfies the following axioms :

- $K(\emptyset) = -\infty$
- $K(\Omega) = 0$

- for any family of (disjoint) elements A_n of $\mathcal{A}$,

$$K(\cup A_n) = \sup_n K(A_n) .$$

(It is straightforward to see that the word “disjoint” can be omitted from this axiom).

The first and third axioms are that of *Maslov measures*. The second one specializes it as a *cost measure*. One may notice the parallel with a probability measure. In the first two axioms, the 0 of probability measures, the neutral element of the addition, is replaced by the neutral element of the max operator : $-\infty$, and the 1, the neutral element of the product, is replaced by the neutral element of the sum, 0. In the third axiom, the sum of the measures of the disjoint sets is replaced by the max.

In the theory of Maslov measures, the habit is to denote the max as $\oplus$, the sum as a product, $-\infty$ as $\mathbb{O}$, 0 as $\mathbb{I}$ so that $a \oplus \mathbb{O} = a$, $a \cdot \mathbb{I} = a$. However, one has $a \oplus a = a$ so that this school refers to *idempotent algebra* and *idempotent measures*. This leads to a theory of integration which is equivalent to the penalized supremums that we shall use in the minimax theory below, but stressing further the parallel with classical measures and integration. This parallel continues with the following concept:

The function $\Gamma : \Omega \rightarrow \mathbb{R} \cup \{-\infty\}$ is called a *cost density* of K if we have

$$\forall A \in \mathcal{A}, \quad K(A) = \sup_{\omega \in \Omega} \Gamma(\omega) .$$

One has the following theorem (Maslov, Akian)

Proposition 1 *Every cost measure defined on the open sets of a Polish space Ω admits a unique maximal extension to 2^Ω , this extension has a density, which is a concave u.s.c. function.*

2.2 Feared value

The wording *Feared value*¹ is used here to stress the parallel with *expected value*.

When a stochastic disturbance is introduced into a problem model, in order to derive a controller design for instance, it comes with a given probability distribution. We shall always assume here that these distributions have densities. Let therefore $w \in \mathbb{W}$ be a stochastic variable. Let its probability distribution be $\Pi(\cdot)$. Let ψ be a function of w with values in $\mathbb{R}$. We define its expected value as

$$\mathbb{E}_w \psi := \int \psi(w) \Pi(w) dw$$

¹In French *Frayeur mathématique* parallel to *Espérance mathématique*

and we omit the subscript w to $\mathbb{E}$ when no ambiguity results.

Similarly, let a disturbance w be given together with a cost distribution $\Gamma(\cdot)$. The *feared value* of a function ψ from $\mathbf{W}$ into $\mathbf{R}$ is defined as

$$\mathbb{F}_w \psi := \max_w [\psi(w) + \Gamma(w)]$$

which is the formula dual to that of the expected value in Quadrat's morphism.

The “Fear” operator enjoys the linearity properties one would expect in the $(\max, +)$ algebra :

$$\mathbb{F}(\max\{\phi, \psi\}) = \max\{\mathbb{F}\phi, \mathbb{F}\psi\},$$

and if λ is a constant,

$$\mathbb{F}(\lambda + \psi) = \lambda + \mathbb{F}\psi.$$

Induced measure Let a decision variable x have a cost measure $C(x)$. Let $y = h(x)$. Then, y is itself a decision variable, with an induced cost measure that can be deduced from the inspection of the feared value of an arbitrary function $\phi(y)$:

$$\mathbb{F}\phi = \max_x [\phi(h(x)) + C(x)] = \max_y \left[\phi(y) + \max_{x \in h^{-1}(y)} C(x) \right].$$

Hence we conclude that the cost measure of y is

$$D(y) = \max_{x \in h^{-1}(y)} C(x) \tag{1}$$

Observe that this is indeed a cost measure, since it is normalized:

$$\max_y D(y) = \max_x C(x) = 0.$$

Conditioning ² We may introduce the *conditioning* by an event. Let $v \in \mathbf{V}$ be a decision variable, $B \subset \mathbf{V}$ be a subset of $\mathbf{V}$, and as usual $K(B) = \max_{v \in B} \Gamma(v)$. Let

$$\Gamma(v|B) = \begin{cases} \Gamma(v) - K(B) & \text{if } v \in B, \\ -\infty & \text{if } v \notin B. \end{cases}$$

This is the “natural” cost density on v if it is constrained to belong to B .

²This is a “naive” theory of conditioning, based upon Bayes rule. For an equivalent of the full stochastic theory of conditional expectations based upon a sigma-field representation, see Del Moral [13, 14]

This yields

$$K(A|B) := \max_{v \in A} \Gamma(v|B) = \max_{v \in A \cap B} \Gamma(v) - K(B),$$

leading to Bayes formula:

$$K(A \cap B) = K(A|B) + K(B).$$

Let now a pair of decision variables (v, w) ranging over sets $\mathbf{V} \times \mathbf{W}$ have a joint cost density $\Gamma(v, w)$. We may define the marginal laws Γ_1 and Γ_2 for v and w respectively as

$$\Gamma_1(v) = \max_{w \in \mathbf{W}} \Gamma(v, w), \quad \Gamma_2(w) = \max_{v \in \mathbf{V}} \Gamma(v, w).$$

Let K denote the cost measure generated by Γ on $\mathbf{V} \times \mathbf{W}$. One has for instance, for a subset $A \subset \mathbf{V}$, $K(A \times \mathbf{W}) = \mathbb{F}_v^{\Gamma_1}(\mathbb{1}_A(v))$. Here, the symbol $\mathbb{F}_v^{\Gamma_1}$ means Feared value with respect to v endowed with the cost density Γ_1 , and $\mathbb{1}_A$ is the concave characteristic function of A , equal to 0 if $v \in A$ and to $-\infty$ otherwise.

We now apply the previous concept of conditioning to the cylindrical condition set $B = \mathbf{V} \times \{w\}$. This leads to the *conditional cost density*

$$\Gamma(v|w) = \Gamma(v, w) - \Gamma_2(w)$$

and notice that this is indeed a (normalized) cost density. The decision variables v and w are said to be *independant* if $\Gamma(v|w) = \Gamma_1(v)$, i.e. if $\Gamma(v, w) = \Gamma_1(v) + \Gamma_2(w)$.

A sequence of stochastic variables $\{w_t\}$, $t = 0 \dots T-1$ also denoted $w_{[0, T-1]}$, are said to be *independant* if their joint probability density is the product of their individual probability densities Π_t :

$$\Pi(w_{[0, T-1]}) = \prod_{t=0}^{T-1} \Pi_t(w_t)$$

leading to the following formula, where J is a function of the whole sequence

$$\mathbb{E}J(w_{[0, T-1]}) = \int J(w_{[0, T-1]}) \prod_{t=0}^{T-1} \Pi_t(w_t) dw_{[0, T-1]}.$$

In a similar fashion, a sequence of independant decision variables $w_{[0, T-1]}$ with cost densities Γ_t will have a joint cost density Γ equal to the sum of their individual cost densities:

$$\Gamma(w_{[0, T-1]}) = \sum_{t=0}^{T-1} \Gamma_t(w_t)$$

leading to the dual formula

$$\mathbb{E} J(w_{[0,T-1]}) = \max_{w_{[0,T-1]}} [J(w_{[0,T-1]}) + \sum_{t=0}^T \Gamma_t(w_t)] .$$

3 The discrete time control problem

We consider a partially observed two input control system

$$x_{t+1} = f_t(x_t, u_t, w_t) , \quad (2)$$

$$y_t = h_t(x_t, w_t) , \quad (3)$$

where $x_t \in \mathbb{R}^n$ is the state at time t , $u_t \in \mathbf{U}$ the (minimizer's) control, $w_t \in \mathbf{W}$ the disturbance input, and $y_t \in \mathbf{Y}$ the measured output. We shall call $\mathbf{U}$ the set of input sequences over the time horizon $[0, T-1]$: $\{u_t\}_{t \in [0, T-1]}$ usually written as $u_{[0, T-1]} \in \mathbf{U}$, and likewise for $w_{[0, T-1]} \in \mathbf{W}$. The initial state $x_0 \in \mathbf{X}_0$ is also considered part of the disturbance. We shall call $\omega = (x_0, w_{[0, T-1]})$ the combined disturbance, and $\Omega = \mathbf{X}_0 \times \mathbf{W}$ the set of disturbances.

The solution of (2) (3) above shall be written as

$$x_t = \phi_t(u_{[0, T-1]}, \omega) ,$$

$$y_t = \eta_t(u_{[0, T-1]}, \omega) .$$

Finally, we shall call u^t a partial sequence $(u_0, u_1, \dots, u_t)$ and $\mathbf{U}^t$ the set of such sequences³, likewise for $w^t \in \mathbf{W}^t$ and $y^t \in \mathbf{Y}^t$. Also, we write $\omega^t = (x_0, w^t) \in \mathbb{R}^n \times \mathbf{W}^t =: \Omega^t$.

The solution of (2) and (3) may alternatively be written as

$$x_t = \phi_t(u^{t-1}, \omega^{t-1}) , \quad (4)$$

$$y_t = \eta_t(u^{t-1}, \omega^t) . \quad (5)$$

We shall also write

$$x^t = \phi^t(u^{t-1}, \omega^{t-1}) , \quad (6)$$

$$y^t = \eta^t(u^{t-1}, \omega^t) , \quad (7)$$

to refer to the partial sequences solution of (2) and (3)

³notice the slight inconsistency in notations, in that our $\mathbf{U}^t$ is the cartesian $(t+1)$ power of $\mathbf{U}$. Other choices of notations have their drawbacks too.

Admissible controllers will be *strictly causal output feedbacks* of the form $u_t = \mu_t(u^{t-1}, y^{t-1})$. We denote by $\mathcal{M}$ the class of such controllers.

A performance index is given. In general, it may be of the form

$$J(x_0, u_{[0,T-1]}, w_{[0,T-1]}) = M(x_T) + \sum_{t=0}^{T-1} L_t(x_t, u_t, w_t) .$$

However, we know that, to the expense of increasing the state dimension by one if necessary, we can always bring it back to a purely terminal payoff of the form

$$J(x_0, u_{[0,T-1]}, w_{[0,T-1]}) = M(x_T) = M \circ \phi_T(u_{[0,T-1]}, \omega) . \quad (8)$$

The data of a strategy $\mu \in \mathcal{M}$ and of a disturbance $\omega \in \Omega$ generates through (2)(3) a unique pair of sequences $(u_{[0,T-1]}, w_{[0,T-1]}) \in \mathbf{U} \times \mathbf{W}$. Thus, with no ambiguity, we may also use the abusive notation $J(\mu, \omega)$. The aim of the control is to minimize J , in some sense, “in spite of the unpredictable disturbances”.

We want to compare here two ways of turning this unprecise statement into a meaningful mathematical problem.

Stochastic control In the first approach, *stochastic control*, we modelize the unknown disturbance as a random variable, more specifically here a random variable x_0 with a probability density $P_0(x)$ and an independant white stochastic process $w_{[0,T-1]}$ of known instantaneous probability distribution Π_t . (We notice that nothing in the sequel prevents Π_t from depending on x_t and u_t .) The criterion to be minimized is then

$$G(\mu) := \mathbb{E}_\omega J(\mu, \omega) . \quad (9)$$

This can be expanded into

$$G(\mu) = \int M(x_T) \left(\prod_{t=0}^{T-1} \Pi_t(w_t) \right) P_0(\xi) dw_{[0,T-1]} d\xi$$

This can be characterized as a *weighted average* since the perturbation values are weighted through their probability density.

Minimax control In the second approach, we are given the cost density Q_0 of x_0 , and the cost densities Γ_t of the w_t 's. (Again, Γ_t might depend on x_t and u_t .) The criterion to be minimized is then

$$H(\mu) := \mathbb{E}_\omega J(\mu, \omega) , \quad (10)$$

which can be expanded into

$$H(\mu) := \max_{w_{[0, T-1]}} \max_{\xi} \left[M(x_T) + \sum_{t=0}^{T-1} \Gamma_t(w_t) + Q_0(\xi) \right]$$

This can be characterized as a *penalized max*, since the perturbation values are penalized through their cost density.

Remark If all cost measures of the disturbances are taken constant, (hence 0 by normalization), then $H(\mu)$ is, if it exists, the *guaranteed value* given only the sets over which the perturbations range. Therefore, minimizing it is insuring the best possible guaranteed value.

4 Dynamic programming

4.1 Stochastic dynamic programming

4.1.1 Full information

We quickly recall here for reference purposes the classical solution of the full information problem via dynamic programming.

In the case where the controller has perfect and instantaneous knowledge of the current state, and is therefore allowed to use state feedback control strategies of the form $u_t = \varphi_t(x_t)$, it is well known that a solution of the optimal control problem is obtained as follows.

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

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

$$\forall x \in \mathbb{R}^n, \quad V_t(x) = \inf_u \mathbb{E}_{w_t} V_{t+1}(f_t(x, u, w_t)). \quad (12)$$

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 (12), then the argument $\varphi_t^*(x)$ of the minimum is an optimal strategy.

4.1.2 Partial information

Let us turn now to the (almost as classical) solution of the partial information, stochastic problem. One has to introduce the *conditional state probability measure*,

and, assuming it is absolutely continuous with respect to the Lebesgue measure, its density P . Let, thus, $P_t(x) dx$ be the conditional probability measure of x_t given y^{t-1} , or *a priori* state probability distribution at time t , and $P_t^\eta(x) dx$ be the conditional state distribution given y^{t-1} and given that $y_t = \eta$, or *a posteriori* state probability distribution at time t . Notice also that our notations are consistent, in that at time $t = 0$, with no information available yet beyond the data of the problem, the probability density of x is P_0 .

Clearly, P_t is a function only of past measurements and controls, assuming these are *causal*, i.e. independant from future disturbances. As a matter of fact, we can give the filter that lets one compute it. Starting from P_0 , at each step, P_t^η can be obtained by Bayes rule. A standard condition for this step to be well posed is that, for all (t, x, w) , the map $w \mapsto h_t(x, w)$ be locally onto, and more specifically that the partial derivative $\partial h_t(x, w)/\partial w$ be invertible. It suffices here to notice that, because the information is increasing, (the information algebras are nested), we have, for any test function $\psi(\cdot) \in L^1(\mathbb{R}^n)$,

$$\mathbb{E}_y \int \psi(x) P_t^\eta(x) dx = \int \psi(x) P_t(x) dx =: \mathbb{E}_x^{P_t} \psi(x). \quad (13)$$

Then P_{t+1} is obtained by propagating $P_t^{y_t}$ through the dynamics. It suffices for our purpose to define this propagation by the dual operator: for any test function ψ ,

$$\mathbb{E}^{P_{t+1}} \psi := \int \psi(x) P_{t+1}(x) dx = \int \mathbb{E}_w \psi(f_t(x, u_t, w)) P_t^{y_t}(x) dx. \quad (14)$$

Hence, combining (13) and (14),

$$\mathbb{E}_y \mathbb{E}_x^{P_{t+1}} \psi(x) = \mathbb{E}_x^{P_t} \mathbb{E}_w^{\Pi_t} \psi(f_t(x, u_t, w)) \quad (15)$$

The above calculations also emphasize the dependance of the sequence $\{P_t\}$ on the control sequence $u_{[0, T-1]}$ and the observation sequence $y_{[0, T-1]}$. Let this define the function $\mathcal{F}_t$ as

$$P_{t+1} = \mathcal{F}_t(P_t, u_t, y_t). \quad (16)$$

Let $\mathcal{P}_t$ be the set of all possible such probability distributions P_t , for all sequences $u_{[0, T-1]}$ and all ω .

Via a standard dynamic programming argument, we can check that the Bellman return function U is obtained by the recurrence relation

$$\forall P \in \mathcal{P}_T, \quad U_T(P) = \mathbb{E}_x^P M(x), \quad (17)$$

$$\forall t \in [0, T-1], \forall P \in \mathcal{P}_t, \quad U_t(P) = \inf_u \mathbb{E}_y U_{t+1}(\mathcal{F}_t(P, u, y)). \quad (18)$$

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 (17)(18), then the optimal cost is $U_0(N)$.*

Moreover, assume that the minimum in u is attained in (18) above at $u = \hat{\mu}_t(P)$. Then (16) and

$$u_t = \hat{\mu}_t(P_t) \quad (19)$$

define an optimal controller for the stochastic control problem.

Proof Consider a causal sequence of controls $\{u_t\}$ and the associated sequence $\{P_t\}$. (The very definition and calculation procedure of P_t implies that u_t has to be independant from the future disturbances). Rewrite (18) as

$$U_t(P_t) \leq \mathbb{E}_{y_t} U_{t+1}(P_{t+1}) .$$

Apply this at time $t = T - 1$, using (17) and (15):

$$U_{T-1}(P_{T-1}) \leq \mathbb{E}_x^{P_{T-1}} \mathbb{E}_w^{\Gamma_{T-1}} M(f_{T-1}(x, u_{T-1}, w)) .$$

We may now iterate:

$$U_{T-2}(P_{T-2}) \leq \mathbb{E}_x^{P_{T-2}} \mathbb{E}_{w_{T-2}} [\mathbb{E}_{w_{T-1}} M(f_{T-1}(f_{T-2}(x, u_{T-2}, w_{T-2}), u_{T-1}, w_{T-1}))]$$

and so on until we get

$$U_0(P_0) \leq \mathbb{E}_x^{P_0} \mathbb{E}_{w_{[0, T-1]}} M \circ \phi_T(u_{[0, T-1]}, x_0, w_{[0, T-1]})$$

And according to formula (8), the rihgt-hand side above is just $\mathbb{E}J$. Furthermore, an appropriate choice of u_t at each step allows the controller to be arbitrarily close to equality in each of the inequalities, thus establishing $U_0(P_0)$ as the infimum of the criterion. Moreover, if the minimizing control (19) exists, all inequalities are equalities, and the infimum of the criterion is reached by that strategy, which is clearly causal, and thus optimal.

4.2 Minimax dynamic programming

4.2.1 Full information

Let us quickly review Isaacs' dynamic programming theory for the *full information* problem, i.e. for state feedback strategies.

We introduce the *full information* Isaacs Value function $V_t(x)$ which satisfies the classical Isaacs equation:

$$\begin{aligned} \forall x \in \mathbb{R}^n, \quad V_T(x) &= M(x), \\ \forall x \in \mathbb{R}^n, \quad V_t(x) &= \inf_u \mathbb{E}_{w_t} V_{t+1}(f_t(x, u, w_t)). \end{aligned}$$

Notice that we have an identical formula with that of subsection 4.1.1, simply replacing $\mathbb{E}$ by $\mathbb{F}$.

Notice also that we do not need that the Isaacs condition, i.e. the existence of a saddle point in the right hand side above, hold. If it does not, V is an upper value, which is what is needed in the context of minimax control.

4.2.2 Partial information

The theory reviewed here is a modification of results first published in [20], also derived in [6] in essentially the same way as here.

Let us consider now the problem of minimizing $H(\mu)$. We have to introduce the *conditional state cost measure* and its cost density Q (according to the concepts introduced in section 2.1 following [1]). It is defined as the *normalized* maximum possible past cost knowing the past information, as a function of current state. In the previous papers such as [20], [6, 7, 8], [12], the authors used the so-called *informational state* (James et al.) or *cost to go* (Başar, Bernhard) which is an *unnormalized* version of the present conditional state cost density. The result was a lack of symmetry with the stochastic case, and as a result less appealing formulas.

To be more precise, let us introduce the following subsets of Ω . Given a pair $(u^t, y^t) \in \mathcal{U}^t \times \mathcal{Y}^t$, and a subset A of $\mathbb{R}^n$, let

$$\Omega_t(A \mid u^t, y^t) = \{\omega \in \Omega \mid y^t = \eta^t(u^{t-1}, \omega^t), \text{ and } \phi_{t+1}(u^t, \omega^t) \in A\}. \quad (20)$$

For any $x \in \mathbb{R}^n$, we shall write $\Omega_t(x \mid u^t, y^t)$, or simply $\Omega_t(x)$ when no ambiguity results, for $\Omega_t(\{x\} \mid u^t, y^t)$. And likewise for $\Omega_{t-1}(x)$.

The conditional “worst cost to go” of A (or the “informational state”) is the Maslov measure $\sup_{\omega \in \Omega_{t-1}(A)} [Q_0(x_0) + \Gamma(w_{[0, T-1]})]$, and hence the associated Maslov density function is, for $t > 0$,

$$W_t(x) = \sup_{\omega \in \Omega_{t-1}(x)} \left[\sum_{k=0}^{T-1} \Gamma_k(w_k) + P_0(x_0) \right] = \sup_{\omega \in \Omega_{t-1}(x)} \left[\sum_{k=0}^{t-1} \Gamma_k(w_k) + P_0(x_0) \right].$$

Initialize this sequence with $W_0 = Q_0$ which constiutes the cost density of x_0 before any further information is available.

It is a simple matter to write recursive equations of the form

$$W_{t+1} = \tilde{\mathcal{G}}_t(W_t, u_t, y_t) .$$

In fact, $\tilde{\mathcal{G}}_t$ is defined by the following. Let for ease of notations

$$\mathbf{Z}_t(x \mid u, y) = \{(\xi, v) \in \mathbb{R}^n \times \mathbf{W} \mid f_t(\xi, u, v) = x, \quad h_t(\xi, v) = y\} ,$$

then we have

$$W_{t+1}(x) = \sup_{(\xi, v) \in \mathbf{Z}_t(x \mid u_t, y_t)} [W_t(\xi) + \Gamma_t(v)] . \quad (21)$$

As was probably first shown in [20], (also presented in a talk in Santa Barbara in july 1993), one can do simple dynamic programming in terms of this function W_t , ranging over the space $\mathcal{W}_t$. The value function U will now be obtained through the following relation

$$\forall W \in \mathcal{W}_T, \quad U_T(W) = \sup_x (M(x) + W(x)) , \quad (22)$$

$$\forall t \in [0, T-1], \forall W \in \mathcal{W}_t, \quad U_t(W) = \inf_u \sup_y U_{t+1}(\tilde{\mathcal{G}}_t(W, u, y)) . \quad (23)$$

Moreover, assume that the minimum in u is attained in (30) above at $u = \hat{\mu}_t(W)$. Then it defines an optimal feedback (19), with W_t now defined by (21), for the minimax control problem. The optimal cost is $U_0(N)$.

However, the formula (23) is *not* exactly dual to (18) in that it involves a $\sup_y$ instead of the expected $\mathbb{E}_y$. The problem stems from the fact that the cost to go W_t is *un-normalized*. Introduce thus

$$Q_t(x) = W_t(x) - \max_{\xi} W_t(\xi) .$$

Now Q_t is indeed a cost density. Let $\mathcal{Q}_t$ be the space of all possible such cost densities for every sequence $u_{[0, T-1]}$ and all ω .

Introduce the set

$$\mathbf{Y}_t(y) = \{(\xi, v) \in \mathbb{R}^n \times \mathbf{W} \mid h_t(\xi, v) = y\} = \bigcup_x \mathbf{Z}_t(x \mid u, y) . \quad (24)$$

Notice that although the rightmost set above seems to depend on u , it actually does not, since u only enters the definition of $\mathbf{Z}$ through the constraint $f_t(\xi, u, v) = x$. Let

$$\hat{Q}_{t+1}(x) = \max_{(\xi, v) \in \mathbf{Z}_t(x \mid u_t, y_t)} [Q_t(\xi) + \Gamma_t(v)] \quad (25)$$

and

$$\Lambda_t(y) = \max_{(\xi, w) \in Y_t(y)} [Q_t(\xi) + \Gamma_t(w)]. \quad (26)$$

Because of the last representation formula for $Y_t(y)$ in (24), we have

$$\max_x \hat{Q}_{t+1}(x) = \Lambda_t(y_t),$$

It follows that

$$Q_{t+1}(x) = \hat{Q}_{t+1}(x) - \Lambda_t(y_t) =: \mathcal{G}_t(Q_t, y_t, u_t)(x), \quad (27)$$

where $\hat{Q}_{t+1}$ and Λ_t are defined by (25) and (26) respectively.

Moreover, according to the formula 1 for induced cost measures, Λ_t is the cost measure induced on y_t by Q_t and Γ_t , i.e. the conditional cost measure given the past information (u^{t-1}, y^{t-1}) .

Then, we can now derive the analogous formula to (15). For ease of notations, let in that derivation $Q_{t+1}[y] = \mathcal{G}(Q_t, u_t, y)$. For a given u , we need to compute for a test function $\psi(x)$

$$\begin{aligned} \mathbb{F}_y \mathbb{F}_x^{Q_{t+1}[y]} \psi(x) &= \\ \max_y \left\{ \max_x \left[\psi(x) + \max_{(\xi, w) \in Z_t(x|u, y)} [Q_t(\xi) + \Gamma_t(w)] - \Lambda_t(y) \right] + \Lambda_t(y) \right\}. \end{aligned}$$

The Λ_t 's cancel out, and the last two max operations collapse into

$$\mathbb{F}_y \mathbb{F}_x^{Q_{t+1}[y]} \psi(x) = \max_y \max_{(\xi, w) \in Y_t(y)} [\psi(f_t(\xi, u, w)) + Q_t(\xi) + \Gamma_t(w)],$$

or finally

$$\mathbb{F}_y \mathbb{F}_x^{Q_{t+1}[y]} \psi(x) = \max_{\xi, w} [\psi(f_t(\xi, u, w)) + Q_t(\xi) + \Gamma_t(w)],$$

which is the required formula, that we may write

$$\mathbb{F}_y \mathbb{F}_x^{Q_{t+1}} \psi(x) = \mathbb{F}_x^{Q_t} \mathbb{F}_w^{\Gamma_t} \psi(f_t(x, u_t, w)). \quad (28)$$

(Compare with formula (15).)

We then easily derive the following theorem:

Theorem 2 *If there exists a sequence of functions $\{U_t\}$ from $\mathcal{Q}_t$ into $\mathbb{R}$ satisfying the following recursion*

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

$$\forall t \in [0, T-1], \forall Q \in \mathcal{Q}_t, \quad U_t(Q) = \min_u \mathbb{F}_y U_{t+1}(\mathcal{G}_t(Q, u, y)) \quad (30)$$

then the optimum value of the payoff $H(\mu)$ is $U_0(Q_0)$. Furthermore, if the minimum is always reached in (30) above, the argument of the minimum, $\hat{\mu}_t(Q)$ constitutes an optimal control strategy.

Proof The proof is identical to that of the stochastic case, only replacing $\mathbb{E}$ by $\mathbb{F}$, and invoking equation (28) instead of (15).

5 Separation theorem

5.1 Stochastic separation theorem

We are here in the stochastic setup. The performance criterion is G , V_t stands for the full information stochastic Bellman function as introduced in subsection 4.1.1, and P_t stands for the conditional state probability density.

We can state the following result.

Theorem 3 *Let*

$$S_t(x, u) := \mathbb{E}_{w_t} V_{t+1}(f_t(x, u, w_t)) P_t(x) .$$

If there exists a (decreasing) sequence of (positive) numbers R_t with $R_T = 0$ such that,

$$\forall t \in [0, T-1], \forall u_{[0, T-1]} \in \mathbf{U}, \forall \omega \in \Omega, \\ \int \min_u S_t(x, u) dx + R_t = \min_u \int S_t(x, u) dx + R_{t+1} ,$$

then the optimal control is obtained by minimizing the conditional expectation of the full information Bellman return function, i.e. choosing a minimizing u in the right hand side above.

Proof The proof relies on the following fact :

Lemma 1 *Under the hypothesis of the proposition, the function*

$$U_t(P) = \mathbb{E}_x^P V_t(x) + R_t \tag{31}$$

satisfies the dynamic programming equations (17)(18).

Let us check the lemma. Assume that $\forall P_{t+1} \in \mathcal{P}_{t+1}$,

$$U_{t+1}(P_{t+1}) = \mathbb{E}_x^{P_{t+1}} V_{t+1}(x) + R_{t+1} = \int V_{t+1}(x) P_{t+1}(x) dx + R_{t+1}$$

and apply (18), using (14)

$$U_t(P_t) = \min_u \mathbb{E}_y \int \mathbb{E}_{w_t} V_{t+1}(f_t(x, u, w_t)) P_t^y(x) dx + R_{t+1}$$

and, according to (13) this yields

$$U_t(P_t) = \min_u \int \mathbb{E}_{w_t} V_{t+1}(f_t(x, u, w_t)) P_t(x) dx + R_{t+1}.$$

Using the hypothesis of the proposition and Bellman's equation for V_t , it comes

$$U_t(P_t) = \int V_t(x) P_t(x) dx + R_t,$$

and the recursion relation holds.

The hypothesis of the theorem sounds in a large extent like wishfull thinking. It holds, as easily checked, in the linear quadratic case. (In that case, symmetry properties result in the certainty equivalence theorem.) There is little hope to find other instances. We state it here to stress the parallel with the minimax case.

5.2 Minimax separation theorem

This section is based upon [6, 7]. Essentially the same result appeared independantly in [17].

We are now in the minimax setup. The performance criterion is H , V_t stands for the full information Isaacs Value function as introduced in subsection 4.2.1, and Q_t stands for the conditional state cost density.

Theorem 4 *Let*

$$S_t(x, u) = \mathbb{E}_{w_t} V_{t+1}(f_t(x, u, w)) + Q_t(x).$$

If there exists a (decreasing) sequence of numbers R_t , such that,

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

$$\max_x \min_u S_t(x, u) + R_t = \min_u \max_x S_t(x, u) + R_{t+1},$$

then the optimal control is obtained by minimizing the conditional worst cost, future cost being measured according to the full information Isaacs Value function, i.e. taking a minimizing u in the right hand side above.

Proof The proof relies on the following fact :

Lemma 2 *Under the hypothesis of the proposition, the function*

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

satisfies the dynamic programming equations (29)(30).

Let us check the lemma. Assume that, $\forall Q_{t+1} \in \mathcal{Q}_{t+1}$,

$$U_{t+1}(Q_{t+1}) = \mathbb{F}_x^{Q_{t+1}} V_{t+1}(x) + R_{t+1} = \max_x [V_{t+1}(x) + Q_{t+1}(x)] + R_{t+1}$$

and apply (30), using (27)

$$\begin{aligned} U_t(Q_t) = & \min_u \max_y \left\{ \max_x \left[V_{t+1}(x) + \max_{(\xi, v) \in \mathcal{Z}_t(x|u, y)} [Q_t(\xi) + \Gamma_t(v)] - \Lambda_t(y) \right] + R_{t+1} \right. \\ & \left. + \Lambda_t(y) \right\}. \end{aligned}$$

As previously, the Λ_t cancel out, and the max operations may be merged into

$$\begin{aligned} U_t(Q_t) = \min_u \left(\max_{\xi, v} [V_{t+1}(f_t(\xi, u, v)) + \Gamma_t(v) + Q_t(\xi)] + R_{t+1} \right) \\ = \min_u \max_x S_t(x, u) + R_{t+1}. \end{aligned}$$

Then, using the hypothesis of the proposition and Isaacs equation for V , it comes

$$U_t(Q_t) = \max_x [V_t(x) + Q_t(x)] + R_t,$$

thus establishing the recursion relation.

The hypothesis of the proposition is not as unrealistic as in the stochastic case. It is satisfied in the linear quadratic case, but more generally, it can be satisfied if S is convex-concave, for instance, with $R_t = 0$. Moreover, in that case, the same u provides the minimum in both sides, yielding a certainty equivalence theorem.

5.3 An abstract formulation

It is known that in the stochastic control problem, some results, including derivation of the separation theorem, are more easily obtained using a more abstract formulation of the observation process, in terms of a family of σ -fields $\mathcal{Y}_t$ generated in the disturbance space. The axioms are that

- the brownian motion w_t is *adapted* to the family $\mathcal{Y}_t$,
- the family $\mathcal{Y}_t$ is *increasing*.

The same approach can be pursued in the minimax case. Instead of an explicit observation through an output (3), one may define the observation process in the following way. To each pair $(u_{[0,T-1]}, \omega)$ the observation process associates a sequence $\{\Omega_t\}_{t \in [0,T]}$ of subsets of Ω . The axioms are that, for any $(u_{[0,T-1]}, \omega)$, the corresponding family Ω_t satisfies the following properties.

- The process is *consistent*, i.e. $\forall t, \quad \omega \in \Omega_t$.
- The process is *strictly non anticipative*, i.e. $\omega \in \Omega_t \Leftrightarrow \omega^{t-1} \in \Omega_t^{t-1}$ where Ω_t^{t-1} stands for the set of restrictions to $[0, t-1]$ of the elements of Ω_t .
- The process is *with complete recall*: $\forall (u_{[0,T-1]}, \omega), \quad t < t' \Rightarrow \Omega_t \supset \Omega_{t'}$.

In the case considered above, we have

$$\Omega_t = \Omega(\mathbb{R}^n \mid u^t, y^t)$$

but the abstract formulation suffices, and allows one, for instance, to extend the minimax certainty equivalence principle to a variable end time problem. See [7] for a detailed derivation, and [24] for an application of (a continuous version of) theorem 2 to a free end time problem.

One may think of the subsets Ω_t as playing the role of the measurable sets of the σ -field $\mathcal{Y}_t$.

6 Conclusion

The present setup shows a remarkable parallel between the stochastic and the minimax case, fully exploiting “Quadrat’s morphism” between probabilities and maximisation. This goes all the way to the more powerful abstract formulation for the observation process, and to the separation theorem (which we therefore write as a singular).

We may make a further remark. In the linear quadratic theory, it has been thought that the minimax certainty equivalence theorem of [3] was less a “separation” theorem than in the stochastic case because the “filter” to compute $\hat{x}$ depends on the weighting matrices in the criterion. We see here that another viewpoint is possible, in which these weighting matrices are only the counterpart to the gaussian

law of the disturbances in the stochastic case, that also enter the Kalman filter. Then the exceptional fact with the LQG case is that it be possible to add a (quadratic) integral term to the (quadratic) terminal payoff.

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] F. Baccelli, G. Cohen, J-P. Quadrat, G-J. Olsder: *Synchronization and Linearity*, Wiley, Chichester, U.K., 1992.
- [3] T. Başar and P. Bernhard: *H_∞ Optimal Control and Related Minimax Design Problems, a Game Theory Approach*, second edition, Birkhauser, Boston, 1995.
- [4] J.A. Ball and J.W. Helton : “ H_∞ -Control for Nonlinear Plants: Connections with Differential Games”, 28th CDC, Tampa, Fla, 1989
- [5] J.A. Ball, J.W. Helton and M.L. Walker, “ H_∞ -Control For Nonlinear Systems with Output Feedback”, IEEE Trans. Automatic Control, **AC-38**, 1993, pp 546–559.
- [6] P. Bernhard “Sketch of a Theory of Nonlinear Partial Information Min-Max Control”, Research Report 2020, INRIA, France, 1993.
- [7] P. Bernhard: “A Discrete Time Min-max Certainty Equivalence Principle”, Systems and Control Letters, **24**, 1995, pp 229–234.
- [8] P. Bernhard: “Expected Values, Feared Values and Optimal Control”, in G-J. Olsder ed. : *New Trends in Dynamic Games and Applications*, Birkhauser, 1995
- [9] P. Bernhard and G. Bellec : “On the Evaluation of Worst Case Design, with an Application to the Quadratic Synthesis Technique”, 3rd IFAC symposium on Sensitivity, Adaptivity and Optimality, Ischia, Italy, 1973
- [10] P. Bernhard and A. Rapaport: “Min-max Certainty Equivalence Principle and Differential Games”, revised version of a paper presented at the Workshop on Robust Controller Design and Differential Games, UCSB, Santa Barbara, California, 1993. Submitted for publication.

- [11] G. Choquet: “Theory of capacities”, *Annales de l’Institut Fourier*, Grenoble, **t.5** pp 131–295, 1955
- [12] G. Didinsky, T. Başar, and P. Bernhard “Structural Properties of Minimax Policies for a Class of Differential Games Arising in Nonlinear H_∞ -control”, *Systems and Control Letters*, **21**, 1993, pp 433–441.
- [13] P. Del Moral: “Résolution particulière des problèmes d’estimation et d’optimisation non linéaires. PhD dissertation Université Paul Sabatier, Toulouse, 1994.
- [14] P. Del Moral: “Maslov Optimization Theory”, *Russian Journal of Mathematical Physics*, to appear, 1995
- [15] A. Isidori : “Feedback Control of Nonlinear Systems” 1st ECC, Grenoble, France, 1991
- [16] A. Isidori and A. Astolfi : Disturbance Attenuation and $\mathcal{H}_\infty$ Control via Measurement Feedback in Nonlinear Systems”, *IEEE trans. Automatic Control*, **AC 37**, pp 1283-1293, 1992
- [17] M. James “On the Certainty Equivalence Principle and the Optimal Control of Partially Observed Dynamic Games”, *IEEE Trans. on Automatic Control*, **AC 39**, pp 2321–2324, 1994.
- [18] M.R. James and J.S. Baras “Partially Observed Differential Games, Infinite Dimensional HJI Equations, and Nonlinear H_∞ Control”. To appear.
- [19] M.R. James, J.S. Barras and R.J. Elliott: “Output Feedback Risk Sensitive Control and Differential Games for Continuous Time Nonlinear Systems”, 32nd CDC, San Antonio, Texas, 1993
- [20] M.R. James, J.S. Barras and R.J. Elliott: “Risk Sensitive Control and Dynamic Games for Partially Observed Discrete Time Nonlinear Systems”, *IEEE Trans. on Automatic Control* **AC-39**, pp 780–792, 1994
- [21] V. Maslov: *Méthodes opératoriellles*, Éditions MIR, Moscow, 1987
- [22] V. Maslov and S.N. Samborskii: “Idempotent analysis”, *Advances in Soviet Mathematics*, **13**, AMS, Providence, 1992.

-
- [23] J-P. Quadrat: “Théorèmes asymptotiques en programmation dynamique”, *Compte Rendus de l’Académie des Sciences*, 311: pp 745–748, 1990.
 - [24] A. Rapaport and P. Bernhard: “Un jeu de poursuite-evasion avec connaissance imparfaite des coordonnées”, International Symposium on Differential Games and Applications, St Jovite, Canada, 1994.
 - [25] A.J. van der Schaft : “Nonlinear State Space $\mathcal{H}_\infty$ Control Theory", ECC 1993, Groningen, Netherlands



Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399