A min-max certainty equivalence principle for nonlinear discrete-time control problems

Pierre Bernhard
INRIA-Sophia Antipolis
2004 route des lucioles, B.P. 93,
06902 Sophia-Antipolis cedex, France

September, 1993

Abstract We extend to a setup comparable to that provided in [4] for the continuous time case, the main result of [3] in the discrete time case, thus obtaining the most comprehensive certainty equivalence principle for the discrete time problem to date.

Keywords discrete time systems, robust control, minimax, partial information, certainty equivalence

1 Introduction

The first min-max certainty equivalence principle ever seems to be due to Whittle [6], and concerns the standard discrete time linear quadratic problem. (Although it appeared in the investigation of a stochastic exponential quadratic or "risk sensitive" problem.) Since then, the author introduced systematically this concept, for both discrete time and continuous time problems in [2], and it was used in [1]. The continuous time case was further studied in [5] and [4]. In [3], we proposed a more general theory, with a certainty equivalence principle as one of its consequences, but in a less general setup. Here, we extend the result of [3] to a set up comparable to that of [4], thus providing the most general such result to date for discrete time systems, with a very simple proof.

2 The framework

2.1 The system

We are given a disturbed discrete time control system in $I\!R^n$

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

The time t is a non negative integer, the control input u, with value u_t at time t, ranges over U, and the disturbance input w over W. Moreover, initial state x_0 is also considered a disturbance ranging over a set $\mathsf{X}_0 \subset I\!R^n$.

The control sequence $\{u_t\}_{t\in\mathbb{N}}$, or equivalently the time function $t \mapsto u_t \in \mathsf{U}$ is in $\mathcal{U}$ and similarly the disturbance sequence $\{w_t\}_{t\in\mathbb{N}}$ belongs to $\mathcal{W}$. They will often be written $u_{(\cdot)}$ and $w_{(\cdot)}$.

The complete disturbance is

$$\omega = (x_0, w_{(\cdot)}) \in \Omega = \mathsf{X}_0 \times \mathcal{W}\,.$$

Let

$$w^\tau = \{w_t, 0 \le t \le \tau\} \in \mathcal{W}^\tau$$

be the restriction to $[0,\tau]$ of the sequence $w_{(\cdot)}$, and similarly for other time sequences. Let also

$$\omega^\tau = (x_0, w^\tau) \in \Omega^\tau\,.$$

2.2 The observation and admissible strategies

The controller does not have a complete knowledge of the past disturbance nor of the state. An *observation process* has been defined that to each pair $(u_{(\cdot)}, \omega)$ makes correspond a sequence $\{\Upsilon_t\}_{t\in\mathbb{N}}$ of subsets of Ω. The observation process is assumed to enjoy the following three properties :

Hypothesis A

A1 Consistency :

$$\forall (u_{(\cdot)}, \omega), \forall t, \quad \omega \in \Upsilon_t\,,$$

A2 Strict non anticipativeness :

$$\omega \in \Upsilon_t \Leftrightarrow \omega^{t-1} \in \Upsilon_t^{t-1}\,,$$

A3 Perfect recall :

$$\Upsilon_{t+1} \subset \Upsilon_t\,.$$

The set $\mathcal{M}$ of admissible strategies for the controller will be that of functions μ of the form

$$u_{(\cdot)} = \mu(\Upsilon_{(\cdot)}) : u_t = \mu_t(\Upsilon_t)\,. \tag{2}$$

A typical instance is when an output

$$y_t = h_t(x_t, w_t)$$

is measured by the controller, who is allowed to use strictly causal controls of the form $u_t = \mu_t(y^{t-1})$. From the output sequence y_{t-1} observed up to time $t-1$ the controller can infer the equivalence class Υ_t of disturbances which, together with the past controls u_{t-1} used, generate that same output.

2.3 The performance index

A terminal set $\mathsf{T} \subset \mathbb{N} \times \mathbb{R}^n$ is given. The problem terminates at the first time instant such that the state reaches T,

$$t_f = \min\{t \mid (t, x_t) \in \mathsf{T}\}\,. \tag{3}$$

In the sequel, x_f always stands for x_{t_f}.

It turns out to be convenient to introduce the section $\mathsf{T}_t \in \mathbb{R}^n$ of T at time t as

$$\mathsf{T}_t = \{x \mid (t, x) \in \mathsf{T}\}$$

so that the termination condition $(t, x_t) \in \mathsf{T}$ can equivalently be written $x_t \in \mathsf{T}_t$.

A performance index is given as [1]

$$J(u_{(\cdot)}, \omega) = M_{t_f}(x_f) + \sum_{t=0}^{t_f-1} L_t(x_t, u_t, w_t) + N(x_0)\,, \tag{4}$$

[1]the term $N(x_0)$ is unnecessary. It can be absorbed in L_0. It turns out to be convenient to keep it there in applications. See the linear quadratic case in [1] for instance.

The controller's objective is to minimize the worst possible case, thus to chose a control strategy μ^* such that

$$\max_{\omega\in\Omega} J(\mu^*,\omega) = \min_{\mu\in\mathcal{M}} \max_{\omega\in\Omega} J(\mu,\omega)\,. \tag{5}$$

Instead of restricting x_0 to X_0, we may equivalently let $N(x) = -\infty$ for all x not in X_0. This causes the maximizing ω to always have $x_0 \in \mathsf{X}_0$.

2.4 An alternate performance index

All the sequel extends to a problem with no target set and performance index

$$J(u_{(\cdot)},\omega) = \min_{t_f}[M_{t_f}(x_f) + \sum_{t=0}^{t_f-1} L_t(x_t,u_t,w_t) + N(x_0)]\,,$$

Isaacs'equation (6) is then replaced by its obvious generalization

$$V_t(x) = \min\left\{M_t(x), \min_u \max_w [V_{t+1}(f_t(x,u,w)) + L_t(x,u,w)]\right\}$$

and T by the set $\bar{\mathsf{T}} = \{(t,x) \mid V_t(x) = M_t(x)\}$.

Such a formulation lets one solve the qualitative problem of whether one can insure that the state reach a given target set defined by $T(x) \leq 0$. Let $M = T$, $L = 0$, $\forall x \in \mathsf{X}_0, N(x) = 0$. The sign of the minimax value of the game yields the answer.

We do not develop this in the sequel, the adaptation is straightforward.

2.5 Minima and maxima

We assume that a proper set of hypotheses hold to insure the existence of the minima and maxima we use hereafter. One possibility is to assume that f_t and L_t are of class C^1 for all t, as well as M and N, and that U, W, and X_0 are compact. But this does not account for the classical linear quadratic case, where U, W, and X_0 are whole vector spaces, existence of the extrema being insured by the behavior at infinity of M, L, and N.

3 The auxiliary problem

3.1 Basic formulation

We consider the full information dynamic game defined by (1), (3), and (4), but without the $N(x_0)$ term which has no meaning in a full information game, and state feedbacks as admissible strategies. Let $V_t(x)$ be its upper value. It satisfies Isaacs' equation

$$V_t(x) = \min_{u\in\mathsf{U}} \max_{w\in\mathsf{W}} \left[V_{t+1}\left(f_t(x,u,w)\right) + L_t(x,u,w)\right] , \tag{6}$$

with the boundary condition

$$\forall (t,x) \in \mathsf{T}, \quad V_t(x) = M_t(x) \, . \tag{7}$$

We notice that the game with the original cost J as in (4) and ω as maximizing control has an upper value

$$A_0 = \max_{x\in\mathsf{X}_0} [V_0(x) + N(x)]. \tag{8}$$

We now introduce an assumption concerning the full information game (we shall say what to do if it is not satisfied)

Hypothesis B We assume that the minimum in u in (6) is, for each (t,x), reached at a unique point $u = \phi_t^*(x)$.

We introduce a fictitious observation process : let

$$\Gamma_t = \{\omega \in \Omega \mid \forall s \leq t, \ x_s \notin \mathsf{T}_s\}$$

It is straightforward to check that for $t < t_f$, this process satisfies hypothesis **A**. And also, because both $\Upsilon_{(\cdot)}$ and $\Gamma_{(\cdot)}$ satisfy hypothesis **A**, so does their intersection. Introduce therefore the modified observation process

$$\Omega_t = \Upsilon_t \cap \Gamma_t \, ,$$

it satisfies hypothesis **A** for $t \in [0, t_f - 1]$.

The *auxiliary problem* is defined for each t as the following maximization problem: let

$$G_t(u^{t-1},\omega) = G_t(u^{t-1},\omega^{t-1}) = \left[V_t(x_t) + \sum_{s=0}^{t-1} L_s(x_s,u_s,w_s) + N(x_0)\right] \tag{9}$$

$$A_t = \max_{\omega \in \Omega_t} G_t(u_{t-1}, \omega) \tag{10}$$

For each t, the minimizer knows the past controls u^{t-1} it has used, and the above problem can therefore be solved, yielding a solution only depending, beyond u^{t-1}, on the available information Ω_t. We call $\hat{\omega}_t$ a maximizing ω above, and $\hat{x}_t$ the current "worst state" x_t it leads to.

Notice that the notation A_t is consistent with the notation A_0 introduced in (8).

3.2 Alternate formulation

We need a further notation : for $\xi \notin \mathsf{T}_t$, let

$$\Omega_t(\xi) = \{\omega \in \Omega_t \mid x_t = \xi\}$$

As all the other subsets of Ω we have introduced, it is a function of the past controls u^{t-1}. Contrary to Ω_t, it may be empty even before termination.

We define the *conditional cost to come* function W from $\mathbb{N} \times \mathbb{R}^n$ into $\mathbb{R} \cup \{-\infty\}$ as

$$W_t(x) = \max_{\omega \in \Omega_t(x)} \left[\sum_{s=0}^{t-1} L_s(x_s, u_s, w_s) + N(x_0) \right]$$

where it is understood that as usual, the max over an empty set is $-\infty$. Notice that we have

$$\forall x \in \mathbb{R}^n, \quad W_0(x) = N(x)\,. \tag{11}$$

The problem (12) may be written as

$$A_t = \max_x \max_{\omega \in \Omega_t(x)} \left[V_t(x_t) + \sum_{s=0}^{t-1} L_s(x_s, u_s, w_s) + N(x_0) \right] ,$$

and using the definition of the conditional cost to come, as

$$A_t = \max_x [V_t(x) + W_t(x)]\,. \tag{12}$$

And $\hat{x}_t$, as defined above, is any maximizing x in (12).

4 The certainty equivalence principle

4.1 Main result

Introduce the notation

$$S_t(x,u) = \max_{w\in\mathsf{W}} \left[V_{t+1}\left(f_t(x,u,w)\right) + L_t(x,u,w) + W_t(x)\right] . \tag{13}$$

We introduce the main hypothesis :

Hypothesis C For $u_{(\cdot)}$ generated by the strategy $\hat{\mu}$ of the theorem, the function S_t has, for all $\omega \in \Omega$ and all t a saddle-point $\hat{x}_t, \hat{u}_t$, so that

$$\max_x S_t(x,\hat{u}_t) = \max_x \min_u S_t(x,u) . \tag{14}$$

(We shall see that $\hat{x}_t$ as defined here necessarily coincides with the definition in the auxiliary problem, so that this notation is consistent.)

We may now state the theorem :

Theorem 1 (Certainty Equivalence Principle) *Under hypotheses* **A**, **B**, *and* **C**, *let $\hat{x}_t$ be a worst case state as defined by the auxiliary control problem, then $u_t = \hat{\mu}_t(\Omega_t) := \phi_t^*(\hat{x}_t)$ is uniquely defined for all t and is an optimal controller in the sense of (5), leading to a value A_0 as in (8).*

Proof Unicity of $\phi^*(\hat{x}_t)$ easily follows from assumption **B**.

Let us investigate the behavior of A_t under the effect of the strategy $\hat{\mu}$ of the theorem. Take the expression of A_t in (12) and use Isaacs'equation (6) to replace V_t in terms of V_{t+1}. It comes

$$A_t = \max_x \min_{u\in\mathsf{U}} \max_{w\in\mathsf{W}} \left[V_{t+1}\left(f_t(x,u,w)\right) + L_t(x,u,w) + W(x)\right] ,$$

The maximum in x is by definition reached at $x = \hat{x}_t$ and the minimum in u at $u = \phi_t^*(\hat{x}_t) = \hat{\mu}_t(\Omega_t)$. Using (13),

$$A_t = \max_x \min_{u\in\mathsf{U}} S_t(x,u) = S_t(\hat{x}_t, \hat{\mu}_t(\Omega_t)) . \tag{15}$$

Now, we also have, using (10)

$$A_{t+1} = \max_{\omega\in\Omega_{t+1}} \left[V_{t+1}(x_{t+1}) + L_t(x_t,u_t,w_t) + \sum_{s=0}^{t-1} L_s(x_s,u_s,w_s) + N(x_0)\right] .$$

According to hypothesis A.3, we get

$$A_{t+1} \leq \max_{\omega\in\Omega_t} \left[V_{t+1}(x_{t+1}) + L_t(x_t,u_t,w_t) + \sum_{s=0}^{t-1} L_s(x_s,u_s,w_s) + N(x_0) \right] .$$

But according to hypothesis A.2, $\omega \in \Omega_t$ is equivalent to $\omega^{t-1} \in \Omega_t^{t-1}$, and thus w_t free. We therefore get

$$A_{t+1} \leq \max_{\omega^{t-1}\in\Omega_t^{t-1}} \max_{w\in\mathsf{W}} \left[V_{t+1}(x_{t+1}) + L_t(x_t,u_t,w_t) + \sum_{s=0}^{t-1} L_s(x_s,u_s,w_s) + N(x_0) \right]$$

Use (1) to substitute for x_{t+1}, and again write

$$\max_{\omega^{t-1}\in\Omega_t^{t-1}} [\cdots] = \max_x \max_{\omega^{t-1}\in\Omega_t^{t-1}(x)} [\cdots]$$

to obtain

$$A_{t+1} \leq \max_x \max_{w\in\mathsf{W}} \left[V_{t+1}\left(f_t(x,u_t,w)\right) + L_t(x,u_t,w) + W(x) \right] .$$

We are assessing the effect of the strategy $u_t = \hat{\mu}_t(\Omega_t)$. We therefore obtain

$$A_{t+1} \leq \max_x S_t(x, \hat{\mu}_t(\Omega_t)) . \tag{16}$$

Compare (15) with (16) using hypothesis **C**, and specifically (14). It results that $A_{t+1} \leq A_t$, and therefore, that

$$A_{t_f-1} \leq A_0. \tag{17}$$

Let ω be fixed in Ω, but arbitrary. Together with the strategy $\hat{\mu}$ they generate an observation process $\Omega_{(\cdot)}$. By definition, hypothesis **A.1** holds up to time $t_f - 1$, and according to the definition (10),

$$\forall \omega \in \Omega, \quad G_{t_f-1}(u^{t_f-1}, \omega) \leq A_{t_f-1} .$$

and also

$$\forall \omega \in \Omega, \quad G_{t_f}(\hat{\mu}, \omega) \leq \max_{\omega | \omega^{t_f-1} \Omega_{t_f-1}^{t_f-1}} G_{t_f}(\hat{\mu}, \omega) .$$

But observe that the proof above concerning A_{t+1} begins by dropping any constraint on w_t. Thus it is also true that

$$\max_{\omega|\omega^{t_f-1}\Omega^{t_f-1}_{t_f-1}} G_{t_f}(\hat{\mu},\omega) \leq A_{t_f-1}\,.$$

Now, comparing (9) and (4), and using (7), it comes $G_{t_f} = J$. This together with the above inequality and (17) yield

$$\forall\omega\in\Omega,\quad J(\hat{\mu},\omega)\leq A_0\,.$$

We have pointed out that A_0 is the value of the *full information* game. Hence no partial information strategy can do better, and $\hat{\mu}$ is optimal.

4.2 Sufficient conditions

We investigate some sufficient conditions that insure hypothesis **C**. The first corollary is obvious:

Corollary 1 *If for all $\omega\in\Omega$, the function S_t is , for all t convex s.c.i. in u, concave s.c.s. in x, and diverges to ∞ for $\|u\|\to\infty$ in U, and to $-\infty$ for $\|x\|\to\infty$, then the certainty equivalence principle holds.*

Proof Simply apply the Von Neumann-Sion Theorem.

Corollary 2 *If the set of first order necessary conditions concerning the extrema in Isaacs'equation and the maximization problem in (12) has a unique solution for all ω and all t, then the certainty equivalence principle holds.*

Proof Under the hypothesis of the corollary, the maximum in w in Isaacs' equation (6) is unique. Applying Danskin's theorem, the partial derivative of S_t in u is the same as that of the right hand side of (6), and using the unicity of the minimum in u, that in x is the same as that of the right hand side of (12). Thus the hypothesis of the theorem says that the first order necessary conditions for the minimum in u and the maximum in x of S_t have a unique solution. Using Danskin's theorem again, it classically follows that it has a saddle point. (These conditions are those needed to extend to non linear problems the proof of [1])

Corollary 3 *If the minimization problem in Isaacs equation is always convex, and the auxiliary problem concave for all t, (both, of course, having solutions) then the certainty equivalence principle holds.*

Proof Apply the theorem of Von Neumann-Sion to the function

$$(u,\omega) \mapsto G_{t+1}(u^{t-1}\cdot u,\omega)$$

and notice that

$$\max_{\omega} G_{t+1}(u^{t-1}\cdot u,\omega) = \max_{x} S_t(x,u)\,.$$

(This condition is that given in [2])

5 Conclusions

We have extended to a more general case, specifically variable end time, whether defined by a capture condition or taken as a minimization parameter, the discrete time certainty equivalence principle of [3]. Taking advantage of the fact that we focus on the case where the principle holds, we were able to give a simpler proof, very much in the spirit of that of [4]. We conjecture that condition **C** above is the most general that can be given. Notice that if the unicity condition **B** is not met, one should chose among the certainty equivalent controls the one that provides the saddle point in **C**. The rest of the proof holds unchanged.

References

[1] T. Başar and P. Bernhard: H_∞ *Optimal Control and Related Minimax Design Problems*, Birkhauser, Boston, Mas. 1991.

[2] P. Bernhard: A min-max certainty equivalence principle and its applications to continuous time, sampled data, and discrete time H_∞ optimal control, INRIA Research Report 1347, 1990.

[3] P. Bernhard: Sketch of a theory of nonlinear partial information min-max control, INRIA Research Report 2020, 1993. Submitted for publication.

[4] 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. INRIA Research Report 2019. Submitted for publication.

[5] G. Didinsky, T. Başar, and P. Bernhard: Structural Properties of Minimax Policies for a Class of Differential Games Arising in Nonlinear H^∞-Control and Filtering, *Proceedings of the 32nd CDC*, San Antonio, Texas, December 1993.

[6] P. Whittle, Risk sensitive Linear/Quadratic/Gaussian Control, *Advances in Applied Probabilities* **13**, pp 764–777, 1981.