REFERENCES

[1] A. Bensoussan and J. L. Lions, *Inequations Variationnnelles en Controle Stochastique*. Paris: Dunod, 1978.
[2] J. Chow and P. Kokotovic, "Near optimal feedback stabilization of a class of nonlinear singularly perturbed systems," *SIAM J. Contr.*, vol. 16, Sept. 1978.
[3] —, "A two stage Lyapunov–Bellman feedback design of a class of nonlinear systems," to be published.
[4] W. Fleming and R. Rishel, *Deterministic and Stochastic Control Theory*. New York: Springer, 1975.
[5] O. A. Ladyzhenskaya and N. N. Ural'tseva, *Linear and Quasi Linear Elliptic Equations*. New York: Academic, 1968.

Alain Bensoussan, for a photograph and biography, see this issue, p. 993.

The Certainty Equivalence Property in Stochastic Control Theory

HENK VAN DE WATER AND JAN C. WILLEMS, FELLOW, IEEE

Abstract—**In this paper we will give a general formulation of the certainty equivalence principle for stochastic optimal control problems. Special attention is paid to the question: "What do we mean by a certainty equivalence control law?" It is then shown that in this context the LQG-problem is indeed certainty equivalent.**

INTRODUCTION

THE PIONEERING work of Bellman with the emphasis on models using state variables and the method of dynamic programming has brought about a theory of dynamic decision making under uncertainty and in particular stochastic and adaptive control. One of the ideas in this area with a great deal of practical and conceptual appeal is that of "certainty equivalence." The basic idea here is to define a decision policy by putting the uncertainty equal to its expected value given the observations. While on first sight the basic intuitive content behind this idea seems quite clear, it is not an easy matter to give precise mathematical formulations of this principle.

Many authors [1], [2], [5] have sought to give a general formalization of this notion, first stated by Simon [6] and Theil [7], and in this process they have obtained further cases in which the certainty equivalence principle is valid. Unfortunately, some of these contributions show serious disadvantages in their formulation, particularly because the results are often not "basis free." (What is precisely meant by this is shown in paragraph 10), the "Nonexamples.")

In the present paper we will set up a mathematical framework for stating a certainty equivalence principle which does not suffer from these drawbacks.

Manuscript received April 13, 1981.
H. van de Water is with the School for Organization and Management, University of Groningen, Groningen, The Netherlands.
J. C. Willems is with the Mathematics Institute, University of Groningen, Groningen, The Netherlands.

We will use the following notations.

i) If A is a matrix, then its transpose is denoted by A^T. If $A=A^T$ and $x^TAx\geqslant 0\ \forall x$, then we write $A\geqslant 0$. If in addition $x^TAx=0\Rightarrow x=0$, then we write $A>0$.

ii) If $T\subset\mathbb{R}$, $t\in T$ and y: $T\to Y$, then y_t^- denotes the past of y defined by y_t^-: $T\cap(-\infty,t)\to Y$ with $y_t^-(\tau):=y(\tau)$ for $\tau<t$. As usual, U^T denotes all maps from T into U, i.e., $U^T:=\{f|f\colon T\to U\}$. (Since they appear in a different context, this will not be confused with transposition.) Let $\boldsymbol{U}\subset U^T$, and $\boldsymbol{F}$: $\boldsymbol{U}\to Y^T$; then $\boldsymbol{F}$ is said to be *nonanticipating* if $\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{U}$ and $\boldsymbol{u}(t')=\boldsymbol{v}(t')$ for $t'\leqslant t$ implies $(\boldsymbol{Fu})(t')=(\boldsymbol{Fv})(t')$ for $t'\leqslant t$ and *strictly nonanticipating*, if $\boldsymbol{u}(t')=\boldsymbol{v}(t')$ $t'<t$ implies $(\boldsymbol{Fu})(t')=(\boldsymbol{Fv})(t')$ for $t'\leqslant t$. In a nonanticipating map the past and the present values of the input determine the present output value, whereas in a strictly nonanticipating map the strict past determines the present output value.

iii) Let $\{\Omega,\boldsymbol{A},\boldsymbol{P}\}$ be a probability space; then $\boldsymbol{E}\{\cdot\}$ denotes integration with respect to the measure $\boldsymbol{P}$.

iv) Finally, we will denote as usual by $\boldsymbol{L}_p^m(t_0,t_1):=\{\boldsymbol{x}|\boldsymbol{x}\colon (t_0,t_1)\to\mathbb{R}^m,\ \int_{t_0}^{t_1}|x_i(t)|^p dt<\infty,\ i=1,\cdots,m)\}$ and, if $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are given Banach spaces, $\boldsymbol{L}(\boldsymbol{B}_1,\boldsymbol{B}_2):=\{\boldsymbol{L}|\boldsymbol{L}\colon \boldsymbol{B}_1\to\boldsymbol{B}_2,\ \boldsymbol{L}$ linear and continuous$\}$.

I. STOCHASTIC CONTROL: PROBLEM FORMULATION

In this section, we will explain what we mean by an optimal stochastic control law. In order to do this we first define an uncertain dynamical system and a control law.

1) Let $T\subset\mathbb{R}$ denote the time axis, U the control alphabet (this is the set, usually, but not necessarily infinite

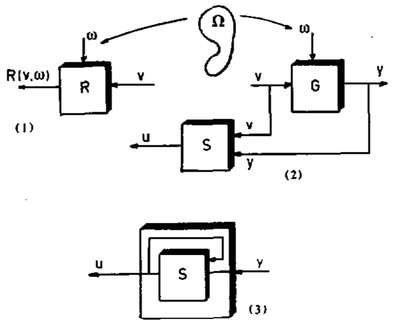


Fig. 1. An uncertain system G with feedback control law F.

where the inputs take their values), $U \subset U^T$ the control space, Y the measurement output alphabet, $Y \subset Y^T$ the output space, and Ω be a set called the uncertainty space.

Definition: An *uncertain dynamical system* is defined by a map G: $U \times \Omega \to Y$ called the *plant* such that $\forall \omega \in \Omega$, $G(\cdot, \omega)$ is strictly nonanticipating.

We will denote an uncertain dynamical system by its plant G.

2) We will view a control policy as a recipe for deciding the control on the basis of the observations, which respects the information flow constraints imposed by the following problem definition.

Definition: A *control law* is a map F: $Y \to U$ which is

i) nonanticipating;

ii) such that the implicit equation in u;

$$u = FG(u, \omega)$$

has for all $\omega \in \Omega$, unique solution on U (this solution depends of course on F and ω and so we will denote it by $F_u(\omega, F)$);

iii) such that the information flow of G is not anticipated in the closed-loop system—by this we mean if $\exists t \in T$, such that $\forall u \in U$, $G(u, \omega_1)(t') = G(u, \omega_2)(t')$ for $t' \leq t$, then $F_u(\omega_1, F)(t') = F_u(\omega_2, F)(t')$ for $t' \leq t$ (in the discrete time case $T = [t_0, t_0 + 1, \cdots, t_1]$, $-\infty < t_0 \leq t_1 \leq \infty$ and with $U = U^T$ and $Y = Y^T$ assumptions ii) and iii) follow from assumption i)—see Lemma A of the Appendix).

Properties i) and ii) are of course natural assumptions. Property iii) guarantees for instance that if $G(u, \omega) = G'(u, w(\omega))$, w: $\Omega \to W^T$ with W the disturbance input alphabet and w the disturbance, then $F_u(\cdot, F) = F'_u(w(\cdot), F)$ with $F'_u(\cdot, F)$ nonanticipating in w.

Assumption ii) is nothing more than a natural generalization of this constraint.

We will call the maps F_u and F_y defined by $F_y(\omega, F) := G(\omega, F_u(\omega, F))$ the *closed-loop system functions*. The above definitions are illustrated in Fig. 1.

3) In order to choose a control law for an uncertain plant one needs to define a family $\mathbb{F}$ of *admissible control laws* and a *decision rule*. This rule is usually obtained by modeling the system performance. Thus, one defines the *cost function*

$$J: U \times \Omega \to \mathbb{R}^e.$$

($\mathbb{R}^e = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$. It is reasonable not to assume *a priori* that the performance of the system is finite, e.g., in infinite time problems.) This induces a map

$$\hat{J}: \mathbb{F} \times \Omega \to \mathbb{R}^e$$

defined by $\hat{J}(F, \omega) := J(F_u(\omega, F), \omega)$. The object is to minimize $\hat{J}$ by choosing $F \in \mathbb{F}$. However, since there may not be any $F^* \in \mathbb{F}$, such that $\hat{J}(F^*, \omega) \leq \hat{J}(F, \omega)$ for all $\omega \in \Omega$ and $F \in \mathbb{F}$, one may turn to the average performance which brings us in a very natural way to stochastic control by modeling the uncertainty space Ω as a probability space.

4) Assume that the uncertainty space Ω is a probability space $\{\Omega, A, P\}$, and that for all admissible control laws $F \in \mathbb{F}$, $\hat{J}(F, \cdot)$ is an (extended) real random variable such that $J_{AV}(F) := E\{\hat{J}(F, \cdot)\}$ is well defined. This obviously induces a map

$$J_{AV}: \mathbb{F} \to \mathbb{R}^e.$$

The problem in optimal stochastic control is then to find a control law $F^* \in \mathbb{F}$, such that

$$J^*_{AV} := J_{AV}(F^*) \leq J_{AV}(F)$$

for all $F \in \mathbb{F}$; J^*_{AV} is called the *optimal performance* and F^* an *optimal control law*.

5) In terms of the above notation the continuous finite time interval LQG-problem reads as follows:

$$T = [t_0, t_1], \qquad -\infty < t_0 \leq t_1 < \infty,$$
$$U = \mathbb{R}^m, \quad U = U_2^m(t_0, t_1),$$
$$Y = \mathbb{R}^p, \quad Y = L_2^p(t_0, t_1)$$

and the plant G is defined through the state $x \in \mathbb{R}^n$ by

$$dx(t, \cdot) = A(t)x(t, \cdot)\,dt + B(t)u(t)\,dt + G(t)\,dw_1(t, \cdot) \tag{5.1}$$

$$dy(t, \cdot) = C(t)x(t, \cdot)\,dt + D(t)\,dw_2(t, \cdot) \tag{5.2}$$

$$x(t_0, \cdot) = x_0(\cdot), \qquad y(t_0) = 0.$$

Here w_1 and w_2 are, respectively, n-dimensional and p-dimensional disturbance inputs defined on the uncertainty space Ω.

In order to give a mathematical meaning to (5.1) and (5.2) we assume that $\{w_1(t, \cdot)\ t_0 \leq t \leq t_1\}$ and $\{w_2(t, \cdot)\ t_0 \leq t \leq t_1\}$ are, respectively, n-dimensional and p-dimensional independent standard Wiener processes defined on a probability space $\{\Omega, A, P\}$ and that $x_0(\cdot)$ is an n-dimensional Gaussian random vector independent of both $w_1(t, \cdot)$ and $w_2(t, \cdot)$. Furthermore, A, B, C are matrix valued functions on $[t_0, t_1]$ with the entries of $B, C \in L_2(t_0, t_1)$ and of $A \in L_1(t_0, t_1)$. This then defines via the definition of the Itô-stochastic integral [11] a plant G: $U \times \Omega \to Y$ of the form

$$(u, \omega) \overset{G}{\mapsto} Lu + a(\omega) \tag{5.3}$$

with L a (deterministic) bounded linear operator and a a p-dimensional Gaussian $L_2(t_0, t_1)$ stochastic process on $[t_0, t_1]$ [9], [10]. The cost functional is given by

$$\int_t^{t_1} \left[x^T(t, \cdot) R_1(t) x(t, \cdot) + u(t) R_2(t) u(t) \right] dt$$

with $R_1 = R_1^T$, $R_2 = R_2^T$, and R_1, R_2 appropriate dimensional matrix valued functions on $[t_0, t_1]$ with the entries of $R_1 \in L_2(t_0, t_1)$ and $R_2 \in L_\infty(t_0, t_1)$, respectively. (As it stands the problem is well defined. However, in order for an optimal stochastic control to exist, more conditions on J need to be satisfied.)

Note that in the class of mathematical models discussed here one uses the measure in $\{\Omega, A, P\}$ in order to define G. In any case one ends up with a situation of type (5.3) and it is not unnatural to consider (5.3) as the primary "experimental" description of G and to consider (5.1) and (5.2) as a mathematical reformulation for (5.3) after Ω is made into a probability space.

The class of admissible control laws consists of all (possibly nonlinear) operators F: $L_2^p(t_0, t_1) \to L_2^m(t_0, t_1)$ which are nonanticipating and such that $I\text{-}LF$ has a nonanticipating inverse on $L_2^p(t_0, t_1)$. (This is what i), ii), and iii) reduce to in this case. In particular, this includes all nonanticipating operators with local Lipschitz constant less then unity; see [12].) Since Ω was already assumed to be a probability space in order to give (5.1) and (5.2) a mathematical meaning-the averaging of 4) in the average performance can now be carried out.

This example shows that the framework sketched in 4) does not imply that if the model is given by an Itô equation, then it should allow a sample pathwise interpretation of the solution. Rather a measure on Ω should then enter in order to interpret the model and a measure (which will be but—to make the point—would not forceably need to be the same) will enter in order to interpret the "average" performance.

6) Although the above setting of optimal stochastic control is quite natural and the underlying ideas well known, there has been a tendency in control theory, particularly when it is approached from a purely probabilistic point of view, to use alternate approaches. Sometimes one finds formulations in which a family of σ-algebras, A_t for $t \in T$ is given and the control is a random process $u \in U$ such that $u(t, \cdot)$ is A_t-measurable for all t. In this context A_t describes the information available to the controller at time t. This formulation ignores the important fact that the control itself may influence the information available in the future and this formulation is hence of very limited use and generality. This possibility of manipulating the information received is, in fact, the basis of the so-called *dual effect* in stochastic control [2]. A second formulation is to call a stochastic process $u \in U$ and admissible control if $u(t, \cdot)$ is A_t-measurable where A_t is the σ-field generated by the corresponding observation process y of course obtained by taking $y = G(u, \cdot)$. The inadequacy of this definition is illustrated by considering the optimal control problem with plant

$$\dot{y} = u \quad y(t_0) = 0$$

and cost function

$$\int_{t_0}^{t_1} |u(t) - w(t)|^2 \, dt$$

where w is Brownian motion. As can be seen after some rewriting, this problem is actually of the LQG-type. In the formulation suggested in the beginning of this paragraph $u = w$ will be an admissible control (which gives an excellent performance) which is of course quite unacceptable. Our framework would give $u = 0$ as the optimal control law, which is more like it.

We feel that the formulation given in (1)–(4) is logical both from mathematical and more importantly also from the "practical" point of view where it corresponds to the idea that a control policy is a processor (often a device-a-computer) which generates decisions on the basis of the available observations and not some type of random generator (or a measure selector) suggested by more probabilistic formulations. A related formulation is found in [4]. Finally, it ought to be mentioned that by and large all available formulations are equivalent for discrete time systems.

II. The Certainty Equivalence Control Law

In this section we will define, essentially on the level of generality used before, what we mean by the "certainty equivalent control law."

7) The basic idea is the following. Assume that we face the decision of choosing a control $u \in U$ so as to minimize the cost functional $J(u, \omega)$. In the case of perfect information, i.e., if $\omega \in \Omega$ is known, this leads to the decision $u^*(\omega)$ choosen, such that $J(u^*(\omega), \omega) \leq J(u, \omega)$ for all $u \in U$. However, if we only know ω only through the observations $y(\omega)$ and we cannot solve the problem of finding the corresponding optimal feedback control law $u^*(y)$ it is not totally without merit to use the control law

$$\hat{u}(y) := E\{u^*(\omega) | y(\omega) = y\}$$

where we have assumed that this conditional expectation is well defined. This idea is called certainty equivalence and will now be formalized in a dynamic framework.

8) We will consider the discrete and continuous time case separately. Let $T = \{t_0, t_0 + 1, \cdots, t_1\}$ with $-\infty < t_0 \leq t_1 \leq \infty$, and assume that the plant G [see 1)] and the cost functional J [see 3)] are given and that the control input alphabet U is a subset of a finite-dimensional vector space (or any space, such that expectations can be defined). For simplicity we will also assume that $U = U^T$ and $Y = Y^T$.

Let $v \in U$, $t \in T$ and consider now the following "deterministic" optimal control problem:

$$\underset{u \in U_{v,t}}{\text{minimize}} \, J(u, \omega)$$

where

$$U_{v,t} := \{u \in U | u(t') = v(t') \text{ for } t' < t\}.$$

Let $u^*(v, t, \omega)$ be a minimizing control (hence, assumed to exist). By considering its value at time t this defines a map R: $U \times \Omega \to U$ by

$$R(v, \omega)(t) := u^*(v, t, \omega)(t).$$

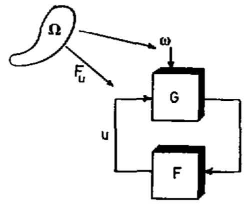


Fig. 2. The idea of certainty equivalence at time t [(1) and (2)] and for all t (3).

This $\boldsymbol{R}(\boldsymbol{v}, \omega)(t)$ is the control which we would use at time t if we faced the situation in which there is perfect information (ω known) but the previous decisions (the control $\boldsymbol{v}$ before t) have already be decided upon. The observations up to time t which would have occured under the above assumptions are given by $\boldsymbol{G}(\boldsymbol{v}, \omega)$ and $\boldsymbol{G}(\boldsymbol{v}, \omega)(t')$ for $t' \leqslant t$ would be available at time t. The certainty equivalence idea explained in 7) would suggest to use at time t the control value

$$\boldsymbol{E}\{\boldsymbol{R}(\boldsymbol{v}, \cdot)(t) | \boldsymbol{G}(\boldsymbol{v}, \cdot)(\tau) = \boldsymbol{y}(\tau) \ \tau \leqslant t\}.$$

(We assume this conditional expectation to be well defined.) This then defines a map $\boldsymbol{S}$: $\boldsymbol{U} \times \boldsymbol{Y} \to \boldsymbol{U}$ defined by $\boldsymbol{S}(\boldsymbol{v}, \boldsymbol{y})(t) := \boldsymbol{E}\{\boldsymbol{R}(\boldsymbol{v}, \cdot)(t) | \boldsymbol{G}(\boldsymbol{v}, \cdot)(\tau) = \boldsymbol{y}(\tau) \quad \tau \leqslant t\}$. This map is nonanticipating in $\boldsymbol{y}$ and strictly nonanticipating in $\boldsymbol{v}$. Consider now the simultaneous equations

$$\boldsymbol{S}(\boldsymbol{v}, \boldsymbol{y}) = \boldsymbol{u}; \qquad \boldsymbol{u} = \boldsymbol{v}$$

(which says nothing more than that the certainty equivalence idea is used for all t). By Lemma A of the Appendix these equations define a unique map

$$\boldsymbol{F}_{ce}\colon \boldsymbol{Y} \to \boldsymbol{U}$$

which also defines a control law in the sense explained in 2). We will call $\boldsymbol{F}_{ce}$ the *certainty equivalence control law.* Using this control law is called the certainty equivalence *principle*, and we will say that the certainty equivalence *property* holds for a stochastic control problem if the optimal control law $\boldsymbol{F}^* = \boldsymbol{F}_{ce}$.

The above definitions are illustrated in Fig. 2.

9) In the continuous time case the ideas are similar, but the details are somewhat more technical. Assume that $T = [t_0, t_1]$ (where $-\infty < t_0 \leqslant t_1 \leqslant \infty$) and that U is a subset of a finite dimensional vector space. Let $\boldsymbol{u}^*(\boldsymbol{v}, t, \omega)$ be defined analogously as in 8) and assume that it is continuous from the right. This defines, as in 8), a map $\boldsymbol{R}$: $U \times \Omega \to U^T$ and $\boldsymbol{S}$: $\boldsymbol{U} \times \boldsymbol{Y} \to U^T$. By assuming now that there exists a unique nonanticipating control law $\boldsymbol{F}_{ce}$: $\boldsymbol{Y} \to \boldsymbol{U}$, such that $\boldsymbol{u} = \boldsymbol{F}_{ce}(\boldsymbol{y})$ solves the equations: $\boldsymbol{S}(\boldsymbol{v}, \boldsymbol{y}) = \boldsymbol{u}$; $\boldsymbol{u} = \boldsymbol{v}$, for all $\boldsymbol{y} \in \boldsymbol{Y}$ one then obtains the certainty equivalence control law in the continuous time case. Thus, the only difference with the discrete time case is that the solvability of the equation $\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}$ has to be postulated while in the discrete time case it follows from Lemma A of the Appendix.

10) "Nonexamples": Previous attempts to formulate a general certainty equivalence principle were based on the idea of replacing all the random quantities in the system equations by their mean in order to obtain an "equivalent deterministic problem" which would in general be easier to solve. Thus, the obtained deterministic control law obtained this way is then used in the stochastic control problem. (Already here we are confronted with the question: What is precisely meant by "use"?). If this action results in optimal performance the problem is called certainty equivalent. However, this idea of equivalent deterministic system is not well defined as is easily shown by the following example:

$$x(k, \cdot) := a(\cdot) \qquad k = t_0, t_0 + 1, \cdots, t_1;$$
$$\cdot a(\cdot)\colon \Omega \to \mathbb{R}^+ := [0, \infty).$$

Define $b := \sqrt{a}$ then the system may be written in the form

$$x(k, \cdot) = b^2(\cdot).$$

Replacing in both the equations the random quantities by their mean results in $x_d(k) = \boldsymbol{E}\{a(\cdot)\}$ and $x_d(k) = [\boldsymbol{E}\{b(\cdot)\}]^2$. In general, $[\boldsymbol{E}\{b(\cdot)\}]^2 \neq \boldsymbol{E}\{b^2(\cdot)\}$; hence, we have obtained two different deterministic systems.

This ambiguity of what to consider noise input does not occur when we consider an equivalent deterministic system defined in terms of sample paths, as suggested in 8) and 9).

Consider, therefore, the following discrete time stochastic control problem:

$$\begin{aligned} x(k+1, \cdot) &= a(k, \cdot) x(k, \cdot) + b\boldsymbol{u}(k) \qquad x(0) := x_0 \\ y(k, \cdot) &= x(k, \cdot) \\ &\{a(k, \cdot) \ k = 0, \cdots, t_1\} \end{aligned} \tag{10.1}$$

a sequence of independent identically distributed random variables, $a(k, \cdot)\colon \Omega \to \mathbb{R}^+$.

The cost criterion $\boldsymbol{J}$ is given by $x^2(t_1, \cdot)$. Assume $\omega \in \Omega$ known, this results in a deterministic optimal control $\boldsymbol{u}^{od}(t_1 - 1, \omega) = -a(t_1 - 1, \omega)/bx(t_1 - 1, \omega)$ and $\boldsymbol{u}^{od}(t, \omega)$ arbitrary for $t = 0, \cdots, t_1 - 2$.

Solving the stochastic problem one obtains using dynamic programming:

$$(\boldsymbol{u}^{os}(y))(t) = -\frac{\boldsymbol{E}\{a(t, \cdot) | \, y(t, \cdot)\}}{b} x(t, \cdot).$$

Using the notation $\boldsymbol{u}^{od}(t_1 - 1) = \boldsymbol{h}(a(t_1 - 1, \omega), x(t_1 - 1, \omega))$, we see that $\boldsymbol{u}^{os}(t_1 - 1) = \boldsymbol{h}(\hat{a}(t_1 - 1), \hat{x}(t_1 - 1))$ where $\hat{a}$ and $\hat{x}$ denote, respectively, the conditional mean of a and x given the observation $y(t_1 - 1, \cdot)$. We conclude that this problem is certainty equivalent according to the classical definition.

Now, let us assume that for one reason or an other it may be necessary to define $c := \sqrt{a}$; the above system then reads

$$x(k+1, \cdot) = c^2(k, \cdot) x(k, \cdot) + b\boldsymbol{u}(k). \tag{10.2}$$

The deterministic problem (i.e., $\omega \in \Omega$ known) is solved by

$$\boldsymbol{u}^{od}(t_1 - 1) = -\frac{c^2(t_1 - 1, \omega)}{b} x(t_1 - 1, \omega)$$

and solving the stochastic problem gives

$$u^{os}(t_1-1)=-\frac{E\{c^2(t_1-1,\cdot)|y(t_1-1,\cdot)\}}{b}x(t_1-1,\cdot).$$

Using the certainty equivalence control law in the classical sense yields

$$u^{ce}(t_1-1)=-\frac{\left(E\{c(t_1-1,\cdot)|y(t_1-1,\cdot)\}\right)^2}{b}x(t_1-1,\cdot).$$

In general, $E\{c^2(t_1-1,\cdot)|y(t_1-1,\cdot)\}\neq(E\{c(t_1-1,\cdot)|y(t_1-1,\cdot)\})^2$. Hence, $u^{ce}(t_1-1)\neq u^{os}(t_1-1)$ from which it follows that our system is not certainty equivalent. Yet (10.1) and (10.2) stand for the same stochastic systems. [In our sense the system is certainty equivalent regardless whether we use (10.1) or (10.2)!]

The above examples clearly show that the crucial point is the question: What is meant by a certainty equivalence control law? It is precisely this question we have tried to give an answer in 8) and 9).

III. Certainty Equivalence for the LQG-Problem

In this section it will be shown that the LQG-problem is certainty equivalent.

11) Consider first the discrete time LQG-problem. Let $T=[t_0,t_0+1,\cdots,t_1]$, $-\infty<t_0\leq t_1<\infty$, $U=\mathbb{R}^m$, $\boldsymbol{U}=U^T$, $Y=\mathbb{R}^p$, $\boldsymbol{Y}=Y^T$ and the plant $\boldsymbol{G}$ be defined analogously as in 5) through the state $x\in\mathbb{R}^n$ by

$$\begin{aligned} x(t+1)&=A(t)x(t)+B(t)u(t)+w_1(t)\\ y(t)&=C(t)x(t)+w_2(t)\\ x(t_0)&=x_0\quad y(t_0)=0 \end{aligned}\tag{11.1}$$

with $\{x_0,w_1(t_0),\cdots,w_1(t_1),w_2(t_0),\cdots,w_2(t_1)\}$ mutually independent Gaussian random vectors, $E\{x_0\}:=\bar{x}_0 E\{(x_0-\bar{x}_0)(x_0-\bar{x}_0)^T\}:=\Sigma(t_0)$; w_1,w_2 white noise with zero mean and covariance $V_1(t)$ and $V_2(t)$, respectively, and $A(t)$, $B(t)$, and $C(t)$ matrices of appropriate dimension for $t\in T$. The cost functional is given by

$$\sum_{t_0}^{t_1-1}\left[x^T(t+1)R_1(t+1)x(t+1)+u^T(t)R_2(t)u^T(t)\right]+x^T(t_1)P_1x(t_1)\tag{11.2}$$

with $P_1=P_1^T$, $R_1(t)=R_1^T(t)$, and $R_2(t)=R_2^T(t)$ matrices of appropriate dimension, $R_1(t)\geq0$, $R_2(t)>0$ for $t\in T$ and $P_1\geq0$. The admissible control laws consist of all nonanticipating maps from $\boldsymbol{Y}$ into $\boldsymbol{U}$. Assume, as in all of LQG optimal control theory, that the discrete time Riccati equation

$$P(t-1)=[A(t-1)-B(t-1)F(t-1)]^T\cdot[R_1(t)+P(t)]A(t-1)\qquad P(t_1)=P_1\tag{11.3}$$

$$F(t-1)=\left[B^T(t-1)(R_1(t)+P(t))B(t-1)+R_2(t-1)\right]^{-1}\cdot B^T(t-1)[R_1(t)+P(t)]A(t-1)\tag{11.4}$$

has a solution for $t_0\leq t\leq t_1$.

We will now prove the following.

Theorem 1: The discrete time LQG-problem is certainty equivalent.

Proof: The proof is based on two lemmas. The first lemma is no more than an exercise in (deterministic) LQ-theory and is proven explicitly in [9].

Lemma 1: For the problem under consideration the control $u^*(v,t,\omega)$ introduced in 8) is given for $t\in\{t,t+1,\cdots,t_1\}$ by

$$u^*(v,t,\omega)(t')=-F(t')x(t',\omega)+L(t',\omega)\tag{11.5}$$

where F is defined by (11.4) and L is given by

$$L=-\tilde{F}w_1-\hat{\hat{F}}\tag{11.6}$$

with

$$\tilde{F}(t):=\left[B^T(t)(R_1(t+1)+P(t+1))B(t)+R_2(t)\right]^{-1}\cdot B^T(t)(R_1(t+1)+P(t+1))$$

$$\hat{\hat{F}}(t):=\left[B^T(t)(R_1(t+1)+P(t+1))B(t)+R_2(t)\right]^{-1}\cdot B^T(t)N(t,\omega)$$

where $N(t,\omega)$ is the solution of the backward difference equation

$$\begin{aligned} N(t,\omega)&=[A(t)-B(t)F(t)]^TN(t+1,\omega)\\ &\quad+[A(t)-B(t)F(t)]^T\cdot[R_1(t+1)+P(t+1)]w_1(t,\omega)\\ N(t_1,\omega)&=0. \end{aligned}\tag{11.7}$$

Finally, x in (11.5) is defined recursively by the plant equations (11.1) with

$$u(t')=\begin{cases} v(t') & \text{for } t'<t\\ -F(t')x(t',\omega)+L(t',\omega) & \text{for } t'\geq t. \end{cases}$$ ■

With Lemma 1 at hand it is now easy to identify what the map $\boldsymbol{R}$ of 8) is. Thus,

$$R:(u,\omega)\mapsto\tilde{u}$$

with $\tilde{u}$ given by

$$\tilde{u}=-Fx+L\tag{11.8}$$

with x recursively defined by

$$\begin{aligned} x(t+1,\omega)&=A(t)x(t,\omega)+B(t)u(t)+w_1(t,\omega)\\ x(t_0,\omega)&=x_0(\omega) \end{aligned}$$

and L given by (11.6) for $t_0\leq t\leq t_1$.

We now turn to the computation in the LQG case of the map $\boldsymbol{S}$ as defined in 8) in general.

Lemma 2: Let $\tilde{u}$ be defined as above and let y be given by the plant equations (11.1) (In both equations we consider u as a deterministic function.) Then $E\{\tilde{u}(t)|y_t^-\}$ is given by $E\{\tilde{u}(t)|y_t^-\}=-F(t)E\{x(t,\omega)|y_t^-\}$.

Proof: Solving (11.7) and substitution of the result into (11.6) shows that $L(t,\omega)$ is linearly dependent on

$w_1(t,\omega), w_1(t+1,\omega), \cdot, w_1(t_1,\omega)$. Moreover, the measurements up to time t are functions of the past $w_1(t_0,\omega),\cdots,w_1(t_1-1,\omega)$ and of $w_2(t,\omega)$. Hence, $L(t,\omega)$ is independent of y_t^-. □

From Lemma 2 it follows that $\boldsymbol{S}$: $\boldsymbol{U}\times\boldsymbol{Y}\to\boldsymbol{U}$ is defined by $\boldsymbol{S}(\boldsymbol{u},\omega)(t)=-F(t)\hat{x}(t)$ with $\hat{x}(t):=\boldsymbol{E}\{x(t)|\,y_t^-\}$. Evaluation of $\hat{x}(t)$ is simply the discrete time Kalman filter and is recursively defined by

$$\hat{x}(t+1)=A(t)\hat{x}(t)+B(t)\boldsymbol{u}(t)+\tilde{L}(t)\cdot[y(t)-C(t)\hat{x}(t)]$$
$$\hat{x}(t_0)=x_0$$

with $\tilde{L}(t)$ the Kalman filter gain obtained, as is well known, via the solution of a Riccati type equation (see, for example [3]). It is important to observe that $\tilde{L}(t)$ is deterministic.

In order to compute the certainty equivalent control law it now suffices to solve the equations $\boldsymbol{S}(\boldsymbol{u},y)=\boldsymbol{u}$ which after substitution in the above expression for $\boldsymbol{S}$ yields that $\boldsymbol{F}_{ce}$: $\boldsymbol{Y}\to\boldsymbol{U}$ is defined by

$$(\boldsymbol{F}_{ce}y)(t)=-F(t)\hat{x}(t)$$

with $\hat{x}(t)$ recursively defined by

$$\hat{x}(t+1)=[A(t)-B(t)F(t)]\hat{x}(t)+\tilde{L}(t)\cdot[y(t)-C(t)\hat{x}(t)]$$
$$\hat{x}(t_0)=\boldsymbol{E}\{x_0\}$$

which defines as is well known [3] the optimal feedback control law for the LQG problem. We conclude that the LQG problem is certainty equivalent. This ends the proof of Theorem 1. ■

12) We will consider now the continuous time LQG-problem as introduced in 5). We will also assume that

i) $\exists\epsilon>0$ such that $R_2(t)=R_2^T(t)\geqslant\epsilon I$ for $t\in T$ and

ii) the Riccati differential equation

$$\dot{K}(t)=-A^T(t)K(t)-K(t)A(t)+K(t)B(t)\cdot R_2^{-1}(t)B^T(t)K(t)-R_1(t)\qquad K(t_1)=0\quad(12.1)$$

has a unique solution for $t_0\leqslant t\leqslant t_1$.

We then have the following.

Theorem 2: The continuous time LQG-problem is certainty equivalent.

Proof: The proof is again based on several lemmas. The first lemma is the analog of Lemma 1 of the discrete time problem. However, to prove it we need some stochastic calculus. We will omit the proof here and refer the reader to [9] for the details.

Lemma 3: For the continuous time problem the control $\boldsymbol{u}^*(v,t,\omega)$ introduced in 9) is given for $t'\in[t,t_1]$ by

$$\boldsymbol{u}^*(v,t,\omega)(t')=-R_2^{-1}(t')B(t')K(t')x(t',\omega)+R_2^{-1}(t')B(t')r(t_0+t_1-t',\omega)\quad(12.2)$$

where $K(t)$ is given by (12.1), and $r(t)$ is the solution of the stochastic differential equation:

$$dr(t,\cdot)=\left[A(t_0+t_1-t)-B(t_0+t_1-t)R_2^{-1}(t_0+t_1-t)\cdot B^T(t_0+t_1-t)K(t_0+t_1-t)\right]^T r(t,\cdot)\,dt-K(t_0+t_1-t)G(t_0+t_1-t)dw'(t,\cdot)$$
$$r(t_0)=0\quad(12.3)$$

and $\{w'(t,\cdot)\ t_0\leqslant t\leqslant t_1\}$ the standard Wiener process defined by

$$w'(t,\cdot):=w_1(t_1,\cdot)-w_1(t_0+t_1-t,\cdot).\quad(12.4)$$

Finally x in (12.2) is defined by the plant equations (5.1) with

$$\boldsymbol{u}(t')=\begin{cases}v(t') & \text{for } t'<t\\ \text{the right-hand side of (12.2)} & \text{for } t'\geqslant t\end{cases}.$$

The map $\boldsymbol{R}$ is thus given by $\boldsymbol{R}$: $(\boldsymbol{u},\omega)\mapsto\tilde{\boldsymbol{u}}$ with $\tilde{\boldsymbol{u}}$ defined by

$$\tilde{\boldsymbol{u}}(t)=-R_2^{-1}(t)B(t)K(t)x(t)+R_2^{-1}(t)B(t)r(t_0+t_1-t)\quad(12.5)$$

with x defined by

$$dx(t,\cdot)=A(t)x(t,\cdot)\,dt+B(t)\boldsymbol{u}(t)\,dt+G(t)\,dw_1(t,\cdot)$$
$$x(t_0,\cdot)=x_0(\cdot)$$

and $r(t,\cdot)$ given by (12.3), $t_0\leqslant t\leqslant t_1$. ■

To compute the map $\boldsymbol{S}$ as defined in q for the continuous time problem we need the following lemma.

Lemma 4: Let $\tilde{\boldsymbol{u}}$ be defined by (12.5) and let y be given by the plant equations (5.1) and (5.2). (We consider $\boldsymbol{u}$ to be a deterministic function.) Then $\boldsymbol{E}\{\tilde{\boldsymbol{u}}(t)|\,y_t^-\}$ is given by

$$\boldsymbol{E}\{\tilde{\boldsymbol{u}}(t)|\,y_t^-\}=-R_2^{-1}(t)B(t)K(t)\boldsymbol{E}\{x(t)|\,y_t^-\}.\quad(12.6)$$

Proof: This easily follows from the fact that $r(t_0+t_1-t)$ is independent of $y(s,\cdot)$ $s\leqslant t$, with $y(t,\cdot)$ given by (5.2); which is a direct consequence of the definition of $r(t,\cdot)$ and $w'(t,\cdot)$ $t_0\leqslant t\leqslant t_1$ and the independence of $w_1(t,\cdot)$ and $w_2(t,\cdot)$ (for details see [9]). ■

The map $\boldsymbol{S}$: $\boldsymbol{U}\times\boldsymbol{Y}\to\boldsymbol{U}$ is given by $\boldsymbol{S}(\boldsymbol{u},y)(t)=-R_2^{-1}(t)B^T(t)K(t)\hat{x}(t,\cdot)$ with $\hat{x}(t,\cdot):=\boldsymbol{E}\{x(t,\cdot)|\,y_t^-\}$ solution of the continuous time Kalman filter

$$d\hat{x}(t,\cdot)=A(t)\hat{x}(t,\cdot)\,dt+B(t)\boldsymbol{u}(t)\,dt+\hat{L}(t)\cdot[dy(t,\cdot)-C(t)\hat{x}(t,\cdot)\,dt]$$
$$\hat{x}(t_0)=\boldsymbol{E}\{x(t_0,\cdot)\}\quad(12.7)$$

with $\hat{L}(t)$ the Kalman gain defined by $\hat{L}(t):=\Sigma(t)C^T(t)(D(t)D^T(t))^{-1}$ of

$$\dot{\Sigma}(t)=A(t)\Sigma(t)+\Sigma(t)A^T(t)+G(t)G^T(t)-\Sigma(t)C^T(t)\left(D(t)D^T(t)\right)^{-1}C(t)\Sigma(t)$$
$$\Sigma(t_0)=\boldsymbol{E}\{(x_0-\bar{x}_0)(x_0-\bar{x}_0)^T\}$$

with

$$\Sigma(t):=\boldsymbol{E}\{[x(t,\cdot)-\hat{x}(t,\cdot)][x(t,\cdot)-\hat{x}(t,\cdot)]^T\}.$$

We still have to compute the certainty equivalent control law F_{ce} by solving the equations $S(u, y)=u$. In general, these equations do not have a unique nonanticipating map as their solution (we explicitly assumed existence in the continuous time case!). However, for the case at hand existence may be shown. Substitution of $S(u, y)=u$ into (12.7) yields

$$d\hat{x}(t,\cdot)=\left[A(t)-B(t)R_2^{-1}(t)B^T(t)K(t)-\hat{L}(t)C(t)\right]\cdot\hat{x}(t,\cdot)\,dt+\hat{L}(t)dy(t,\cdot)$$

and this defines (notice that $\hat{L}(t)$ is deterministic) together with $S(u, y)=u$ a nonanticipating map F_{ce}: $Y\to U$ given by

$$(F_{ce}y)(t)=-R_2^{-1}(t)B^T(t)K(t)\hat{x}(t)$$

which yields as is well known the optimal feedback control law for the LQG problem. We conclude the certainty equivalence. This ends the proof of Theorem 2. □

IV. Conclusions

13) In this paper we have given a general framework for defining a certainty equivalent control law. The basic idea is that in certainty equivalent control policy one chooses, at every instant of time, as the control the conditional expectation (given the measurements until that moment) of what would be the optimal control if there were no uncertainty. Within this framework it is then not difficult to prove that the discrete as well as the continuous time LQG-problems are certainty equivalent. We have given here an explicit proof of the case of finite time intervals, but these results are easily extended to the infinite time case.

The framework used in the paragraphs 1)–4) of our paper may actually give an appealing framework for defining and studying other stochastic control principles as the separation principle and neutrality, caution and probing, and information state in stochastic control. Providing an adequate mathematical framework for these tantalizing intuitive ideas seems very much in the spirit of Bellman's contribution to decision making under uncertainty.

Appendix

Let $T:=\{t_0, t_0+1,\cdots,t_1\}$ $-\infty<t_0\leq t_1<\infty$ and U, V be given spaces and $U=U^T$, $V=V^T$. Let F: $U\times V\to U$ with F nonanticipating and strictly nonanticipating on U. We will now prove the following lemma.

Lemma A: Consider the following equation:

$$u=F(u,v). \qquad (*)$$

Then, there exists a map F_u: $V\to U$, such that $F_u(v)$ is the unique solution of $(*)$. Moreover, F_u is strictly nonanticipating.

Proof: The proof goes by induction. Take $t=t_0$; then $u(t_0)=F(u,v)(t_0)$. Hence, there exists a map ϕ: $V\to U$, such that $u(t_0)=\phi_0(v(t_0))$; where the existence of ϕ_0: $V\to U$ follows from the fact that F is strictly nonanticipating. Take $t=t_0+i$, $i=0,1,\cdots,k$ and let

$$u(t_0+i)=\phi_i(v(t_0),\cdots,v(t_0+i)) \qquad (**)$$

with ϕ_i: $V^{i+1}\to U$, $i=0,1,\cdots,k$.

We then have

$$u(t_0+k+1)=F(u,v)(t_0+k+1)$$
$$=\psi(u(t_0),\cdots,u(t_0+k),v(t_0),\cdots,v(t_0+k+1))$$

(where we have repeatedly made use of the fact that F is nonanticipating in V and strictly nonanticipating). Now using $(**)$ we get

$$u(t_0+k+1)=\phi_{k+1}(v(t_0),\cdots,v(t_0+k+1))$$

with ϕ_{k+1} defined by

$$\phi_{k+1}\colon (v(t_0),\cdots,v(t_0+k+1))\mapsto\psi(\phi_0(v(t_0)),\phi_1(v(t_0),v(t_0+1)),\cdots,\phi_k(v(t_0),\cdots,v(t_0+k)),v(t_0),\cdots,v(t_0+k+1)).$$

Hence, there exists a unique map F_u: $V\to U$ defined by

$$(F_uv)(t)=\phi_{t-t_0}(v(t_0),\cdots,v(t)), \qquad t\in[t_0,t_0+1,\cdots,t_1]$$

which is nonanticipating. □

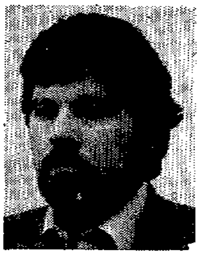
References

[1] M. Aoki, *Optimization of Stochastic Systems.* New York: Academic, 1967.
[2] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence and separation in stochastic control," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 494–550, 1974.
[3] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems.* New York: Wiley, 1972.
[4] A. Lindquist, "On feedback control of linear stochastic systems," *SIAM J. Contr.*, vol. 11, pp. 323–353, 1973.
[5] J. W. Patchel and O. L. R. Jacobs, "Separability, neutrality and certainty equivalence," *Int. J. Contr.*, vol. 13, pp. 337–342, 1971.
[6] H. A. Simon, "Dynamic programming under uncertainty with a quadratic criterion function," *Econometrica*, vol. 24, pp. 74–81, 1956.
[7] H. Theil, "A note on certainty equivalence in dynamic planning," *Econometrica*, vol. 25, pp. 346–349, 1957.
[8] E. Tse and Y. Bar-Shalom, "Generalized certainty equivalence and dual effect in stochastic control," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 817–819, 1975.
[9] H. van de Water, "The value of information in stochastic control," Ph.D. dissertation, Mathematics Inst., Univ. Groningen, Groningen, The Netherlands, 1980.
[10] J. C. Willems, "Recursive filtering," *Statistica Neerlandica*, vol. 32, pp. 1–39, 1978.
[11] E. Wong, *Stochastic Processes in Information and Dynamical Systems.* New York: McGraw-Hill, 1971.
[12] W. M. Wonham, "Random differential equations in control theory," In *Probabilistic Methods in Applied Mathematics*, vol. 2, A. T. Bharucha-Reid, Ed. New York: Academic, 1970, pp. 132–212.

Henk van de Water was born in Schiedam, The Netherlands, in 1947. He received the degree in mathematical engineering from the Delft University of Technology and the Ph.D. degree in systems and control from the University of Groningen, Groningen, The Netherlands, in 1974 and 1980, respectively.

From 1979 to 1980 he was a Lecturer of Mathematics at Erasmus University, Rotterdam, The Netherlands. Presently, he is Senior Lecturer at the School of Organization and Management, University of Groningen. His research interests lie in stochastic systems theory and applications of systems theory to organization and management problems.

Dr. van de Water is a member of the Dutch General Systems Society.



Jan C. Willems (S'66–M'68–SM'79–F'80) was born in Brugge, Belgium in 1939. He studied at the University of Gent in Belgium, the University of Rhode Island, and the Massachusetts Institute of Technology where he received his Ph.D. degree in electrical engineering in 1968.

From 1968 he was an Assistant Professor of electrical engineering at MIT, Cambridge, MA. In 1973 he was appointed to his present position of Professor of Systems and Control at the Mathematics Institute of the University of Groningen, The Netherlands. He has held brief visiting appointments at the University of Cambridge, the University of Florida, Harvard University, the University of Rome, the ETH, and Ben Gurion University.

Dr. Willems serves as a Consulting Advisor to the Mathematical Centre in Amsterdam and is member of the Editorial Board of several journals in the field of systems and control. He is a member of SIAM, the Dutch General Systems Society, and the Dutch "Wiskundig Genootschap." His research interests lie in systems theory and automatic control. His recent work is concerned with the geometric theory of linear systems, the synthesis of multivariable controllers, and problems of system representation.

A Singular Perturbation Approach to Modeling and Control of Markov Chains

RANDOLPH G. PHILLIPS AND PETAR V. KOKOTOVIC, FELLOW, IEEE

Abstract—**Finite state continuous time Markov processes with weak interactions are modeled as singularly perturbed systems. Aggregate states are obtained using a grouping algorithm. Two-time scale expansions simplify cost equations and lead to decentralized optimization algorithms.**

Introduction

MARKOV decision processes have played an important part in Bellman's development of dynamic programming [1]–[3]. Recent applications, such as management of hydrodams [4], [5] and queueing network models of computer systems [6]–[8], have accentuated the need for reduced order approximations of large scale Markov chains. In this regard particularly promising is a perturbational decomposition-aggregation method of Pervozvanski, Smirnov, and Gaitsgori [9]–[12], and Delebecque and Quadrat [5], [13]. The method assumes that the groups of strongly interacting states are known and treats the weak interactions between these groups as perturbations. The result is a short-term decomposition. Over a longer period the weak interactions become significant, while each group of the coupled states can be replaced by an aggregate state. A long-term aggregate model is thus obtained. In controlled Markov processes this time scale separation leads to hierarchical algorithms in which fast subsystem optimizations are coordinated at a slower aggregate level [5], [12].

This paper contributes to the further development of the perturbational decomposition-aggregation method. First, an explicit singular perturbation form of the model of a process with weak interactions is proposed. This form interprets earlier aggregation results and improves the accuracy of the aggregated model. Second, it is shown that a grouping algorithm developed for power systems [14] can be used to identify the groups of strongly interacting states. Third, the singularly perturbed form has simplified the treatment of the cost equations and decentralized algorithms in optimization problems.

Singular Perturbation Model

Consider an n-state Markov process in which the N groups of strongly interacting states have been identified, group j consists of n_j states and $\sum_{j=1}^{N} n_j = n$. We express the weak interactions between the states in different groups as multiples of a small positive scalar ϵ and form the continuous time model

$$\frac{dp}{d\tau} = p(A + \epsilon B) \tag{1}$$

where p is the n-dimensional row of probabilities p_i to be in state i at time τ. Thus,

$$\sum_{i=1}^{n} p_i = 1. \tag{2}$$

We assume that for $0 < \epsilon \leq \epsilon^*$ matrices A, ϵB, and $A + \epsilon B$

Manuscript received April 13, 1981. This work was supported by the Joint Services Electronics Program under Contract N00014-79-C-0424, in part by the National Science Foundation under Grant ECS-79-19396, and in part by the U.S. Department of Energy under Contract DE-AC01-79-ET29243.

R. G. Phillips was with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801. He is now with Bell Laboratories, Holmdel, NJ 07733.

P. V. Kokotovic is with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801.