# IFAC

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

WARSZAWA 1969

# Optimal Stochastic Control

Fourth Congress of the International Federation of Automatic Control
Warszawa 16–21 June 1969

TECHNICAL SESSION
56

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

# Optimal Stochastic Control

TECHNICAL SESSION No 56

FOURTH CONGRESS OF THE INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents



Wydawnictwa Czasopism Technicznych NOT - Polska

Zakład Poligraficzny WCT NOT. Zam. 65/69.

# LINEAR DIFFERENTIAL GAMES WITH COMPLETELY OPTIMAL STRATEGIES AND THE SEPARATION PRINCIPLE

**Pierre Faurre**

Adjoint Director
Automation Center of the School of Mining
35,rue St. Honoré, 77-Fontainebleau, France
Council Engineer at the SAGEM and at the IRIA

There is no need presently to underline the practical interest of the optimum control theory. It can be anticipated that the differential games theory also stimulated by military problems and being, as it seems, in the course of new development, will have more and more practical applications (agressive or cooperative games).

This paper treats linear differential games with quadratic performance indexes. A direct proof of the existence of optimum strategies for the case when an associated Riccati equation has a solution, is found. In spite of fact, that this result is known, the proof given doesn't make any use of dynamic programming or calculus of variations methods; moreover it demonstrates that the strategies obtained are "completely optimal" in the sense defined later. Finally, this direct method in extended to stochastic case for which the separation principle, classical in control theory, is proved.

## 1. INTRODUCTION TO GAMES

1.1. Suppose there are two gamblers (denoted by 1 and 2). The gambler 1 selects a variable $u_1$ and 2 selects $u_2$, both not knowing adversary's choice.

Performance index or outcome is defined by the function

$$V = V(u_1, u_2)$$

which gambler 1 strives to minimize, and 2 - to maximize.

1.2. A "policy of worth case" involves that:

a) The gambler 1 plays minimax , i.e. he selects

$$u_1^* = \arg\min_{u_1} \left\{ \max_{u_2} V \right\}$$

The corresponding minimax is denoted by $V_1^*$. 1 is therefore assured that whatever is the choice of 2

$$V \leq V_1^*$$

b) The gambler 2 plays maximin, i.e.

$$u_2^* = \arg\max_{u_2} \left\{ \min_{u_1} V \right\}$$

Denote the respective maximin by $V_2^*$. 2 is therefore assured that, whatever is the choice of 1,

$$V \geq V_2^*$$

It is obvious that in all cases:

$V_2^* \leq V_1^*$ (because of $V_2^* \leq V(u_1^*, u_2^*) \leq V_1^*$) .

We say that the game has the value $V^*$ if

$$V_1^* = V_2^* \quad (= V^*)$$

For natural reasons the variables $u_1^*$ and $u_2^*$ are called the optimal strategies.

1.3. EXAMPLE No. 1. For $u_1$ and $u_2$ both real, arbitrary, consider

$$V = u_1^2 - u_2^2$$

then

$$V_1^* = V_2^* = 0 \quad \text{and} \quad u_1^* = u_2^* = 0$$

1.4. EXAMPLE No. 2. Now suppose

$$V = u_1^2 + u_1 u_2 - u_2^2 = (u_1 + \frac{u_2}{2})^2 - \frac{5}{4} u_2^2 = \frac{5}{4} u_1^2 - (u_2 - \frac{u_1}{2})^2$$

we have still $V_1^* = V_2^* = 0$ and $u_1^* = u_2^* = 0$.

1.5. An essential difference between these two examples is to be aperceived however.

In example No. 1 there is no advantage for either gambler 1 or 2 to know the adversary's choice: optimal selection re-mains always $u_1 = u_1^*$ or $u_2 = u_2^*$.

In example No. 2, on the contrary, if 1 knows (by "intuition" or by spying ...) that 2 selected $u_2 = \alpha$, then, in order to minimize V, 1 has to play reasonably

$$u_1 = -\frac{\alpha}{2} \neq u_1^*$$

1.6. DEFINITION. A game is called the game with completely optimal strategies, if there exist $u_1^*$ and $u_2^*$ such that $\forall u_1, \ \forall u_2$

$$V(u_1^*, u_2) \leq V(u_1, u_2)$$
$$V(u_1, u_2^*) \geq V(u_1, u_2)$$

The game in the example No. 1 in therefore the game with completely optimal strategies.

## 2. LINEAR DIFFERENTIAL GAME WITH A QUADRATIC PERFORMANCE INDEX

2.1. Let us consider now a linear differential system (not necessarily stationary)

$$\dot{x} = Fx + G_1u_1 + G_2u_2 \qquad (1)$$

starting from initial condition

$$x(\tau) = \xi \qquad (2)$$

($x$ is a state vector of dimension $n$, called the state of game, $u_1$ and $u_2$ are vectors of dimensions $m_1$ and $m_2$ respectively).

At every time the gamblers 1 and 2 select their controls $u_1$ and $u_2$ taking into account the actual state $x$ of the game; in other words, 1 and 2 play using the strategies:

$$u_1 = u_1(x, t), \ u_2 = u_2(x, t)$$

The quadratic performance index is defined by:

$$V = \int_{\tau}^{T} \left\{ x'Qx + u_1'R_1u_1 - u_2'R_2u_2 \right\} ds + x'(T)\, A\, x(T) \tag{3}$$

with

$$R_1 > 0 \;, \quad R_2 > 0 \tag{4}$$

($Q$, $R_1$, $R_2$ and $A$ are symmetric matrices, functions of time eventually).

The gambler 1 strives to minimize $V$ and 2 to maximize $V$.

This problem is interesting in itself [5], or can be used, as in calcus of variations [6,7], to a local study of extremals for nonlinear differential games (theory of the second variation).

2.2. Let us define the Riccati equation associated to the above problem, as being the equation:

$$\dot{P} + F'P + PF - PG_1R_1^{-1}G_1'P + PG_2R_2^{-1}G_2'P + Q = 0 \tag{5}$$

with terminal condition

$$P(T) = A \tag{6}$$

We define the strategies

$$u_1^{*}(x, t) = -R_1^{-1}G_1'Px \tag{7}$$

$$u_2^{*}(x, t) = R_2^{-1}G_2'Px \tag{8}$$

and prove the following

2.3. LEMMA (fundamental formula)

<u>If the Riccati equation (5), (6) has a solution in the interval</u> $[\tau, T]$, <u>then</u>

$$V = \xi' P(\tau)\xi + \int_{\tau}^{T} \left\{ (u_1 - u_1^{*})' R_1(u_1 - u_1^{*}) + - (u_2 - u_2^{*})' R_2(u_2 - u_2^{*}) \right\} ds \tag{9}$$

2.4. Proof. We denote by

$$u_1 = u_1^{*} + \tilde{u}_1 \;, \quad u_2 = u_2^{*} + \tilde{u}_2 \tag{10}$$

then

$$\dot{x} = \tilde{F}x + G_1\tilde{u}_1 + G_2\tilde{u}_2 \tag{11}$$

with

$$\tilde{F} = F - G_1R_1^{-1}G_1'P + G_2R_2^{-1}G_2'P \tag{12}$$

The performance index V takes a form

$$V = x'(T)\,A\,x(T) + \int_{\tau}^{T}\Big\{x'\big[Q + PG_1R_1^{-1}G_1'P - PG_2R_2^{-1}G_2'P\big]x - 2x'PG_1\tilde{u}_1 - 2x'PG_2\tilde{u}_2 + \tilde{u}_1'R_1\tilde{u}_1 - \tilde{u}_2'R_2\tilde{u}_2\Big\}\,ds \tag{13}$$

Taking into account that the Riccati equation (5) can be written as

$$\dot{P} = \tilde{F}'P + P\tilde{F} + PG_1R_1^{-1}G_1'P - PG_2R_2^{-1}G_2'P + Q = 0 \tag{14}$$

we have

$$V = x'(T)\,A\,x(T) + \int_{\tau}^{T}\Big\{-x'\big[\dot{P} + PF + F'P\big]x - 2x'PG_1\tilde{u}_1 - 2x'PG_2\tilde{u}_2 + \tilde{u}_1'R_1\tilde{u}_1 - \tilde{u}_2'R_2\tilde{u}_2\Big\}\,ds \tag{15}$$

or alternately

$$V = x'(T)\,A\,x(T) + \int_{\tau}^{T}\Big\{-\frac{d}{ds}(x'Px) + \tilde{u}_1'R_1\tilde{u}_1 - \tilde{u}_2'R_2\tilde{u}_2\Big\}ds$$

$$= \xi'P(\tau)\xi + \int_{\tau}^{T}\Big\{\tilde{u}_1'R_1\tilde{u}_1 - \tilde{u}_2'R_2\tilde{u}_2\Big\}ds$$

Q.E.D.

The above lemma clearly proves the following

2.5. THEOREM

<u>If the Riccati equation (5), (6) has a solution in the interval $[\tau, T]$, then the game is the game with completely optimal strategies and, moreover, these strategies ( (7) and (8) ) are unique</u>.

## 3. STOCHASTIC GAMES AND THE SEPARATION PRINCIPLE

3.1. We consider now a stochastic differential equation (in the Ito [9] sense)

$$\dot{x} = Fx + G_1 u_1 + G_2 u_2 + v$$

where $v$ is a gaussian white noise

$$\begin{cases} E\, v(t) = 0 \\ E\, v(t)\, v'(s) = C_v(t)\, \delta(t-s) \end{cases} \tag{17}$$

The initial state $x(t) = \xi$ is a random vector with mean $E\{\xi\} = \bar{\xi}$ and covariance

$$E\{(\xi - \bar{\xi})(\xi - \bar{\xi})'\} = \Lambda \tag{18}$$

At every time instant the gamblers 1 and 2 select their controls $u_1$ and $u_2$ belonging to a class of admissible control such that (16) posseses a solution in Ito's sense.

We shall suppose:

(i) that they know (or measure) the adversary's control,

(ii) that at each time instant they dispose of the observations on game state

$$y_1 = H_1 x + w_1 \quad \text{for } \underline{1} \tag{19}$$

and

$$y_2 = H_2 x + w_2 \quad \text{for } \underline{2} \tag{20}$$

where $w_1$ and $w_2$ are white noises with covariances

$$E\, w_1(t)\, w_1'(s) = C_{w_1}(t)\, \delta(t-s) \tag{21}$$

$$E\, w_2(t)\, w_2'(s) = C_{w_2}(t)\, \delta(t-s) \tag{22}$$

The performance index that 1 strives to minimize and 2 to maximize is now:

$$V = E\left\{ \int_{\tau}^{T} \left\{ x'Qx + u_1'R_1u_1 - u_2'R_2u_2 \right\} ds + x'(T)\, A\, x(T) \right\} \tag{23}$$

We denote by $\hat{x}_1(t)$ ( $\hat{x}_2(t)$ respectively) the best estimate of the game's state $x(t)$ that can be constructed by 1 (2 respectively), and by $\Sigma_1(t)$ ($\Sigma_2(t)$ respectively) - corresponding error covariance. We recall that $\hat{x}_1$ and $\hat{x}_2$ are generated by Kalman-Bucy filters [8].

3.2. We consider the Riccati equation associated to a corresponding deterministic problem, i.e. the equation (9), and we define strategies

$$u_1^* = -R_1^{-1}G_1'P\hat{x}_1 \tag{24}$$

$$u_2^* = R_2^{-1}G_2'P\hat{x}_2 \tag{25}$$

We introduce also the variables p and q defined by

$$\begin{cases} q(T) = \text{trace}\left\{P(\tau)\Lambda\right\} & (26) \\ \dfrac{dq}{dt} = -\ \text{trace}\left\{P G_1 R_1^{-1} G_1' P \Sigma_1 - P G_2 R_2^{-1} G_2' P \Sigma_2\right\} & (27) \end{cases}$$

$$\begin{cases} p(T) = 0 & (28) \\ \dfrac{dp}{dt} = -\ \text{trace}\left\{PC_v\right\} & (29) \end{cases}$$

By analogy to the fundamental formula (23) we have

3.3. LEMMA

<u>If the Riccati equation (5), (6) has a solution in the interval</u> $[\tau, T]$, <u>then</u>

$$V = \xi' P(\tau)\xi + p(\tau) + q(\tau) + E\left\{\int_\tau^T \left\{(u_1 - u_1^*)'R_1(u_1 - u_1^*) - (u_2 - u_2^*)'R_2(u_2 - u_2^*)\right\}ds\right\} \tag{30}$$

3.4. PROOF. As previously let us denote

$$u_1 = u_1^* + \tilde{u}_1 , \quad u_2 = u_2^* + \tilde{u}_2 , \tag{31}$$

and

$$x = \hat{x}_1 + \tilde{x}_1 , \quad x = \hat{x}_2 + \tilde{x}_2 \tag{32}$$

According to our hypotheses $\tilde{u}_1(t)$ ($\tilde{u}_2(t)$ respectively) is independent of $\tilde{x}_1(t)$ ($\tilde{x}_2(t)$ respectively).

Then

$$\dot{x} = \tilde{F}x + G_1R_1^{-1}G_1'P\tilde{x}_1 - G_2R_2^{-1}G_2'P\tilde{x}_2 + G_1\tilde{u}_1 + G_2\tilde{u}_2 + v \qquad (33)$$

We have also (analogy to (15))

$$V = E\Big\{x'(T)\, A\, x(T) + \int_{\tau}^{T}\Big\{-x'[\dot{P} + P\tilde{F} + \tilde{F}'P]x + \tilde{x}_1'PG_1R_1^{-1}G_1'P\tilde{x}_1$$

$$- \tilde{x}_2'PG_2R_2^{-1}G_2'P\tilde{x}_2 - 2x'PG_1\tilde{u}_1 - 2x'PG_2\tilde{u}_2 - 2x'PG_1R_1^{-1}G_1'P\tilde{x}_1$$

$$+ 2x'PG_2R_2^{-1}G_2'P\tilde{x}_2 + 2\tilde{x}_1'PG_1\tilde{u}_1 + 2\tilde{x}_2'PG_2\tilde{u}_2 + \tilde{u}_1'R_1\tilde{u}_1 +$$

$$- \tilde{u}_2'R_2\tilde{u}_2\Big\}ds\Big\} \qquad (34)$$

But we know that the stochastic differential of $x'Px$ (Ito [9, 10] differentiation rule) is given by

$$d(x'Px) = x'\dot{P}x\, dt + 2x'P\, dx + \text{trace}\, \{PC_v\} \qquad (35)$$

Replacing $dx$ by its formula (33) and noting that

$$E\left\{\tilde{x}_1'PG_1\tilde{u}_1\right\} = 0$$

due, as already mentioned, to the mutual independence of $\tilde{u}_1$ and $\tilde{x}_1$.

$E\left\{\int_{\tau}^{T} x'Pv\, dt\right\} = 0$ by known property of stochastic integral, we obtain

$$V = E\,\Big\{\xi'P(\tau)\xi + \int_{\tau}^{T}\Big\{\text{trace}\,\{PC_v\} + \tilde{x}_1'PG_1R_1^{-1}G_1'P\tilde{x}_1$$

$$-\tilde{x}_2'PG_2R_2^{-1}G_2'P\tilde{x}_2 + \tilde{u}_1'R_1\tilde{u}_1 - \tilde{u}_2'R_2\tilde{u}_2\Big\}\, ds\Big\} \qquad (36)$$

Taking into account that

$$E\,\left\{\xi'P(\tau)\xi\right\} = \bar{\xi}'P(\tau)\bar{\xi} + \text{trace}\,\{P(\tau)\Lambda\}$$

and

$$E\left\{\tilde{x}_1' P G_1 R_1^{-1} G_1' P \tilde{x}_1\right\} = \text{trace}\left\{ P G_1 R_1^{-1} G_1' P \Sigma_1 \right\}$$

we obtain at once formula (30) which was to be proved.

This allows us to establish.

3.5. THEOREM

The strategies (24) and (25) are completely optimal, or in other words, the principle of control and estimation separation is valid for the differential game considered.

Game value

$$V = \bar{\xi}' P(\tau) \bar{\xi} + p(\tau) + q(\tau)$$

differs from value of the associated deterministic game by two terms:

- the first, $p(\tau)$, due to random perturbations ( white noise $v$ ) acting upon the system,
- the second, $q(\tau)$, due to the lack of gamblers information on the state of game.

## 4. CONCLUSION

It is clear that the restriction of the results and previous proofs to the case of a single gambler (optimal control), and their extention to the case of $N$ gamblers divided into two teams, are trivial. Concerning the extention to $N$ gamblers, the system equations can be written as

$$\dot{x} = Fx + \sum_{1}^{N} G_i u_i$$

the performance index is

$$V = x'(T)\, A\, x(T) + \int_{\tau}^{T} \left\{ x'Qx + \sum u_i' R_i u_i \right\} ds$$

with

$$R_i > 0 \quad \text{for } i = 1, \ldots, k$$

$$R_i < 0 \quad \text{for } i = k+1, \ldots, N$$

and the first k gamblers strive to minimize V , while the last N-k gamblers want to maximize it.

The associated Riccati equation is

$$\dot{P} + F'P + PF - P\left(\sum_i G_i R_i^{-1} G_i'\right)P + Q = 0$$

the completely optimal strategies are

$$u_i = \varepsilon_i R_i^{-1} G_i' P x$$

with

$$\varepsilon_i = \begin{cases} -1 & \text{for } i = 1, \ldots, k \\ +1 & \text{for } i = k+1, \ldots, N \end{cases}$$

Number of problems yet remain to be treated concerning these classes of differential games.

The most immediate would be to extend the proof given by the author for the optimal control problem [11] to prove rigorously and simply that the existence of the Riccati equation solution is necessary for existence of the game problem solution.

A more complex problem would be to formulate and to treat the stochastic game problem with hypothese (i) withdrawn.

## 5. BIBLIOGRAPHY


1. Rufus Isaacs, Differential Games, Wiley (1965).
2. L.S. Pontryagin, On the Theory of Differential Games, Translation of London Mathematical Society Vol. 21 No. 4 (1966) pp. 193-246.
3. Peter D. Joseph, Julius T. Tou, On Linear Control Theory, AIEE Trans. Vol. 80 Pt. II, pp. 193-196 (1961).
4. T.L. Gunckel II, Gene F. Franklin, A General Solution for Linear Sampled-Data Control, J. of Basic Engineering (1963) pp. 197-203.
5. Y.C. Ho, A.E. Bryson, S. Baron, Differential Games and Optimal Pursuit-Evasion Strategies, IEEE Trans. on Automatic

Control, AC-10 (1965) pp. 385-389.

6. C. Caratheodory, Calculus of Variations and Partial Differential Equations, Vol. 2, Holden-Day (1967) .
7. R.E. Kalman, T.S. Englar, A User's Manual for the ASP,NASA CR-475 (1966).
8. R.E. Kalman, R.S. Bucy, New Results in Linear Filtering and Prediction Theory, J. of Basic Engineering (1961) pp. 95-108.
9. K. Ito, On Stochastic Differential Equations, Mem. Amer. Math. Soc. 4 (1951).
10. W.M. Wonham, Lecture Notes on Stochastic Control, Center for Dynamical Systems, Brown University (1967).
11. P. Faurre, Sur les points conjugués en commande optimale, C.R. de l'Académie des Sciences (to appear).

# STOCHASTIC OPTIMAL CONTROL WITH PARTIALLY KNOWN DISTURBANCES

T. J. Tarn
Washington University
St. Louis, Missouri
U. S. A.

## 1. Introduction

Early research in the field of stochastic optimal control was concerned with the optimal control of systems in which the parameters of the system and the noise disturbances were precisely known. The main result of these investigations was that, in linear systems with quadratic performance criterion and Gaussian random effects, the optimal stochastic controller is synthesized by cascading an optimal filter with a deterministic optimal controller[1,2,3,4].

Freimer[5], Tou[6], and Lin and Yau[7] investigated linear systems with signal adaptation, in which the reference input to the system is a function of a random variable with unknown statistics, and linear systems with self-adaptation, in which the coefficients of the system equation are functions of random variables with unknown statistics, and the system is subject to additive random disturbances with known statistics. When these solutions are applied to practical engineering problems, such as in many chemical control processes, the assumption of a known distribution of additive system disturbance or measurement noise is sometimes open to question. Smith[8] investigated the estimation problem of measurement noise variance. Aoki[9] gave examples for a control system in which the measurement noises have either unknown mean or unknown variance.

It is the purpose of this paper to show how filtering theory based on a Bayesian approach may be used to solve the problem of optimally controlling a linear discrete stochastic system in which the additive white Gaussian input has fixed but unknown mean and variance. The basic idea is to consider the unknown parameters as random variables whose a priori distributions are given, and the problem solution consists of recursive equations for sequentially computing the a posteriori distributions of these random variables based on measurements. From the a posteriori distributions estimates can be formed. This has computational advantages when estimates are required in real time.

Using Bellman's dynamic programming[10], an exact analytical solution of the feedback control law may be found. This solution serves as standard for evaluating approximate solutions.

## 2. Practical Motivation

We will show how the problem, which we treat in this paper, was motivated by a practical control process[11].

Consider the stirred tank reactor shown in Figure 1. The reaction occurring is A ⟶ B. For simplicity, assume a liquid-phase reaction.

A stream of constant volumetric flow rate F, which contains A, flow into a tank of constant holdup volume V. The concentration of the entering stream $C_I$(= mole/Vol.) varies with time. The outlet concentration $C_o$ maintains a desired value.

Assuming the density of the solution to be constant, the flow rate in must equal the flow rate out, since the holdup volume is fixed. The reaction will be isothermal irreversible first order, it proceeds at a rate $r = K\,C_o$, where r = moles A reacting/(volume)(time), K = reaction velocity constant, $C_o$ = concentration of A in reactor, moles/volume.

From the mass balance for A we have

$$FC_{I,n-1} = FC_{o,n} + VKC_{o,n} + V(C_{o,n} - C_{o,n-1}) , \tag{1}$$

with n the present time. When the system is at steady state, that is, $C_{o,n} = C_{o,n-1} = C_{o,s}$, then we have

$$FC_{I,s} = FC_{o,s} + VKC_{o,s} . \tag{2}$$

From (1) and (2) we obtain

$$F(C_{I,n-1} - C_{I,s}) = F(C_{o,n} - C_{o,s}) + VK(C_{o,n} - C_{o,s})$$

$$+ V[(C_{o,n} - C_{o,s}) - (C_{o,n-1} - C_{o,s})] . \tag{3}$$

Define the control $u_n = C_{I,n} - C_{I,s}$, the state $x_n = C_{o,n} - C_{o,s}$; we get

$$x_n = ax_{n-1} + bu_{n-1} , \tag{4}$$

where $a = \frac{V}{F+VK+V}$ , $b = \frac{F}{F+VK+V}$ .

Frequently, there are many random disturbances which affect this reaction: for example, the mixing may not be perfect, or there may be fluctuations in the inflow concentration. Assuming these random disturbances are additive independent Gaussian processes, $v_n$, then we obtain

$$x_{n+1} = ax_n + bu_n + v_n \quad . \tag{5}$$

To control this system, we want to choose $u_n$ based on all available data such that $E[\sum_{i=n}^{N-1} (x_i^2+u_i^2) \mid X_n, U_{n-1}]$ is minimized for all $n=0,\ldots,N-1$. This means that we want to keep the concentration deviations from steady state in both inflow and outflow to a minimum over N stages of time.

When we begin the process, we may not know the statistics of $v_n$; hence, we have to estimate these statistics to achieve optimal control.

## 3. Problem Statement

A discrete time linear system with additive white Gaussian disturbance and exact observation of the state can be described by

$$\begin{aligned} \underline{x}_{n+1} &= \underline{H}\,\underline{x}_n + \underline{b}\,u_n + \underline{v}_n \\ \underline{x}_o &= \underline{c} \quad (\underline{c} \text{ is a constant vector}) \end{aligned} \tag{6}$$

where $\underline{x}$ is the r-dimensional state vector, u is the scalar control, $\underline{H}$ is the rxr constant matrix, $\underline{b}$ is the r-dimensional constant vector, $\underline{v}$ is the r-dimensional white Gaussian disturbance vector with unknown parameters.

Given the initial state $\underline{c}$ and the a priori probability densities for the unknown parameters, the control $u_n$ must be chosen based on all available measured data $X_n = [x_o,\ldots,x_n]$ and $U_{n-1} = [u_o,\ldots,u_{n-1}]$, with n the present time, such that

$$V = E[\sum_{i=n}^{N-1} \underline{x}_i^t\,\underline{Q}\,\underline{x}_i + ku_i^2 \mid X_n, U_{n-1}] \quad , \quad n=0,\ldots,N-1 \tag{7}$$

is minimized, where $\underline{Q}$ is a nonnegative definite symmetric matrix and k is a positive constant.

## 4. Unknown Mean and Variance

In the scalar case, the system equation is

$$\begin{aligned} x_{n+1} &= x_n + u_n + v_n \\ x_o &= c \quad (c \text{ is a constant}) \end{aligned} \tag{8}$$

where $x_n$ is the state, $u_n$ is the control, $v_n$ is a sequence of independent Gaussian random variables with unknown mean m and unknown variance $\sigma^2$.

Filtering:

From (8) we know that the exact observations on the state are equivalent to the observations of a sequence of samples of the disturbance v. When the mean and variance of the disturbance v are unknown, we can treat them as random variables. Since the sample mean and sample variance of an independent Gaussian sequence have a joint normal-gamma density[12], we assume a joint a priori density for $(m,\frac{1}{\sigma^2})$ as the normal-gamma density defined by

$$\begin{aligned} p_{N\gamma}(m,\tfrac{1}{\sigma^2} \mid a,b,f,g) &= p_N(m \mid a,\tfrac{b}{\sigma^2})p_{\gamma 2}(\tfrac{1}{\sigma^2} \mid f,g) \\ &\propto (\tfrac{1}{\sigma^2})^{1/2}\exp-\tfrac{b}{2\sigma^2}(m-a)^2\cdot(\tfrac{1}{\sigma^2})^{g/2-1}\exp-\tfrac{fg}{2\sigma^2} \quad , \end{aligned} \tag{9}$$

where $-\infty < m < \infty$, $\frac{1}{\sigma^2} \geq 0$, $-\infty < a < \infty, b,f,g > 0$, $p_N$ denotes normal density, $p_{\gamma 2}$ denotes gamma-2 density, and $\propto$ denotes proportionality with a known constant ratio. When such an a priori density is assigned to $(m,\frac{1}{\sigma^2})$ with parameters $(a_o,b_o,f_o,g_o)$, the object of the filtering is to produce the a posteriori density for $(m,\frac{1}{\sigma^2})$ at each time instant after measuring x. Since both mean and variance are unknown, to get the recursive filter it takes two measurements to form the new statistics at each time instant. Consequently we can change the control only when we have the new statistics.

The probability of measuring $x_1$ and $x_2$ given $(m,\frac{1}{\sigma^2})$ is

$$p(x_1,x_2 \mid m,\tfrac{1}{\sigma^2}) \propto (\tfrac{1}{\sigma^2})\exp-\tfrac{1}{2\sigma^2}\sum_{i=1}^{2}(x_i-x_{i-1}-u_{i-1}-m)^2 \quad . \tag{10}$$

After $x_1$, $x_2$ have been measured, by the Bayes's rule, the a posteriori density of $(m,\frac{1}{\sigma^2})$ will be

$$\begin{aligned} p(m,\tfrac{1}{\sigma^2} \mid x_1,x_2) &\propto p(x_1,x_2 \mid m,\tfrac{1}{\sigma^2})\; p(m,\tfrac{1}{\sigma^2}) \\ &\propto (\tfrac{1}{\sigma^2})^{1/2}\exp-\tfrac{b_2}{2\sigma^2}(m-a_2)^2\cdot(\tfrac{1}{\sigma^2})^{g_2/2-1}\exp-\tfrac{f_2 g_2}{2\sigma^2} \quad , \end{aligned} \tag{11}$$

where

$$\begin{aligned} a_2 &= \frac{a_o b_o + 2s_1(2)}{2+b_o} \quad , \qquad b_2 = 2+b_o \quad , \\ f_2 &= \frac{f_o g_o + a_o^2 b_o + s_2(2) + 2s_1(2)^2 - a_2^2 b_2}{2+g_o} \quad , \qquad g_2 = 2+g_o \quad ; \end{aligned} \tag{12}$$

and

$$s_1(2) = \frac{1}{2} \sum_{i=1}^{2} (x_i - x_{i-1} - u_{i-1}) \ ,$$
$$s_2(2) = \sum_{i=1}^{2} (x_i - x_{i-1} - u_{i-1} - s_1(2))^2 \ ; \qquad (13)$$

thus the a posteriori density of $(m, \frac{1}{\sigma^2})$ is normal-gamma.

For a normal-gamma a posteriori density the parameters $(a,b,f,g)$ are sufficient statistics: these sum up all the information of the measurements. Because information is conserved, direct computation of these statistics may be taken as an optimal filtering procedure. The form of (11) is the same at each time instant, so that the filtering equations are

$$a_{2n} = \frac{a_{2n-2}b_{2n-2} + 2s_1(2n)}{2+b_{2n-2}} \ , \quad b_{2n} = 2+b_{2n-2} \ , \quad g_{2n} = 2+g_{2n-2} \ ,$$
$$f_{2n} = \frac{f_{2n-2}g_{2n-2} + a_{2n-2}^2 b_{2n-2} + s_2(2n) + 2s_1(2n)^2 - a_{2n}^2 b_{2n}}{2+g_{2n-2}} \ ; \qquad (14)$$

where

$$s_1(2n) = \frac{1}{2} \sum_{i=2n-1}^{2n} (x_i - x_{i-1} - u_{i-1}) \ ,$$
$$s_2(2n) = \sum_{i=2n-1}^{2n} [x_i - x_{i-1} - u_{i-1} - s_1(2n)]^2 \ ; \qquad (15)$$

and the probability density of $(m, \frac{1}{\sigma^2})$ after measuring $x_{2n}$ is

$$P(m, \frac{1}{\sigma^2} \mid X_{2n}) \propto (\frac{1}{\sigma^2})^{1/2} \exp - \frac{b_{2n}}{2\sigma^2}(m - a_{2n})^2 . (\frac{1}{\sigma^2})^{g_{2n}/2-1} \exp \frac{f_{2n}g_{2n}}{2\sigma^2} \ . \qquad (16)$$

From equation (8) we have

$$x_{2n+2} = x_{2n+1} + u_{2n+1} + v_{2n+1}$$
$$= x_{2n} + 2u_{2n} + v_{2n} + v_{2n+1} \ ; \qquad (17)$$

after we have measured $x_{2n}$, the probability density of $x_{2n+2}$ given $(m, \frac{1}{\sigma^2})$ is Gaussian

$$p(x_{2n+2} \mid m, \frac{1}{\sigma^2}, X_{2n}) = \frac{1}{\sqrt{2\pi}} (\frac{1}{2\sigma^2})^{1/2} \exp - \frac{1}{4\sigma^2}(x_{2n+2} - x_{2n} - 2u_{2n} - 2m)^2 \ . \qquad (18)$$

Then, multiplying (16) by (18) and integrating over m and $\frac{1}{\sigma^2}$ we get

$$p(x_{2n+2}|X_{2n}) = p_s(x_{2n+2} \mid x_{2n}+2u_{2n}+2a_{2n}, \frac{b_{2n}}{2f_{2n}(4+b_{2n})}, g_{2n}) \; ; \tag{19}$$

similarly we get

$$p(v_{2n} \mid X_{2n}) = p_s(v_{2n} \mid a_{2n}, \frac{b_{2n}}{f_{2n}(1+b_{2n})}, g_{2n}) \; , \tag{20}$$

where $p_s$ denotes the student density defined by

$$p_s(y \mid p,q,r) = \frac{r^{r/2}}{b(1/2,r/2)} [r+q(y-p)^2]^{-(r+1)/2} \sqrt{q} \; ,$$

$$-\infty < y < \infty, \; q,r > 0, \text{ with } E[y] = p, \; r > 1; \; Var[y] = \frac{r}{q(r-2)}, \; r > 2 \; . \tag{21}$$

In addition, we will need the conditional distributions $p(a_{2n+2} \mid X_{2n})$, $p(b_{2n+2} \mid X_{2n})$, $p(f_{2n+2} \mid X_{2n})$ and $p(g_{2n+2} \mid X_{2n})$. However it is easy to see from equation (14) and (15) that $b_{2n+2}$ and $g_{2n+2}$ are non-random while $a_{2n+2}$ and $f_{2n+2}$ both are functions of $x_{2n+2}$, hence both are random variables. To evaluate the probability densities of $a_{2n+2}$ and $f_{2n+2}$ we digress for a moment to evaluate the probability density of $(s_1(2n+2), s_2(2n+2))$ first. From Raiffa and Schlaifer[12] the joint density of $(s_1(2n+2), s_2(2n+2))$ given $(m,\frac{1}{\sigma^2})$ is the product of the independent densities of $s_1(2n+2)$ and $s_2(2n+2)$

$$\begin{aligned} p(s_1(2n+2), s_2(2n+2) \mid m, \tfrac{1}{\sigma^2}, 2, 1) \\ = p_N(s_1(2n+2) \mid m, \tfrac{2}{\sigma^2}) p_{\gamma 2}(s_2(2n+2) \mid \tfrac{1}{\sigma^2}, 1) \; , \end{aligned} \tag{22}$$

the unconditional joint density of $(s_1(2n+2), s_2(2n+2))$ will then be

$$\begin{aligned} p(s_1(2n+2), s_2(2n+2)) \\ = \int_0^\infty \int_{-\infty}^\infty p_N(s_1(2n+2) \mid m, \tfrac{2}{\sigma^2}) p_{\gamma 2}(s_2(2n+2) \mid \tfrac{1}{\sigma^2}, 1) \\ \cdot p_{N\gamma}(m, \tfrac{1}{\sigma^2} \mid a_{2n}, b_{2n}, f_{2n}, g_{2n}) dm d(\tfrac{1}{\sigma^2}) \end{aligned} \tag{23}$$

$$\propto s_2(2n+2)^{1/2-1} [f_{2n}g_{2n} + s_2(2n+2) + \frac{2b_{2n}}{2+b_{2n}} (s_1(2n+2) - a_{2n})^2]^{-(g_{2n}+2)/2}$$

We can solve equation (14) to get equations for $s_1(2n+2)$ and $s_2(2n+2)$ in terms of $a_{2n+2}$ and $f_{2n+2}$, substituting $s_1(2n+2)$ and $s_2(2n+2)$ into equation (23), we obtain

$$p(a_{2n+2}, f_{2n+2} \mid X_{2n}) \propto \frac{[f_{2n+2}g_{2n+2}-f_{2n}g_{2n}-\frac{b_{2n}}{2}(2+b_{2n})(a_{2n+2}-a_{2n})^2]^{1/2-1}}{(f_{2n+2}g_{2n+2})^{g_{2n+2}/2}} \quad . \tag{24}$$

Finally from (24) we obtain

$$p(a_{2n+2} \mid X_{2n}) = p_s(a_{2n+2} \mid a_{2n}, \frac{b_{2n}(2+b_{2n})}{2f_{2n}}, \; g_{2n}) \quad , \tag{25}$$

$$p(f_{2n+2} \mid X_{2n}) = p_{i\beta 1}(f_{2n+2} \mid \tfrac{1}{2}g_{2n}, \tfrac{1}{2}g_{2n+2}, \frac{f_{2n}g_{2n}}{g_{2n+2}}) \quad , \tag{26}$$

where $p_{i\beta 1}$ denotes inverted-beta-1 density defined by

$$p_{i\beta 1}(z \mid p,q,r) \;=\; \frac{1}{B(p,q-p)} \, \frac{(z-r)^{q-p-1}r^p}{z^q} \quad , \tag{27}$$

$0 \le r \le z \le \infty$, $q > p > 0$ with $E[z] = \frac{r(q-1)}{p-1}$, $Var[z] = \frac{r^2(q-1)(q-p)}{(p-1)^2(p-2)}$ .

Optimal Control:

To find the optimal control, define the cost functional

$$V \;=\; \sum_{i=0}^{N-1} \; (qx_i^2+ku_i^2) \quad , \tag{28}$$

where q and k are positive constants. It should be noted that due to the continuously acting random disturbance $v_n$, the cost functional is now a random variable, we can only consider its statistical properties. Hence at any stage n, the current control $u_n$ and future controls $u_i$, $i > n$ must be chosen so as to minimize

$$E[V \mid X_n, U_{n-1}], \text{ for all } n = 0,\ldots,N-1. \tag{29}$$

The sequence of controls which minimize $E[V \mid X_n, U_{n-1}]$ is the same as that which minimize $E[\sum_{i=n}^{N-1} qx_i^2+ku_i^2 \mid X_n, U_{n-1}]$, for $n = 0,\ldots,N-1$. Let us set

$$V_n = \min_{\substack{u_i \\ n\le i\le N-1}} \; E[\sum_{i=n}^{N-1} qx_i^2+ku_i^2 \mid X_n, U_{n-1}] \quad . \tag{30}$$

From previous discussion we know that all the information in the measurements is summarized in the sufficient statistics $(a_{2n}, b_{2n}, f_{2n}, g_{2n})$ and the controls are changed only at each even numbered measurement, $u_{2n-1} = u_{2n-2}$. Hence we can write (30) in the following form

$$V_{2n} = \min_{\substack{u_i \\ 2n \le i \le N-1}} E\left[\sum_{i=2n}^{N-1} q(x_i^2+x_{i+1}^2)+2ku_i^2 \,\middle|\, X_{2n}, U_{2n-2}\right] , \tag{31}$$

where $0 \le n \le \frac{N-1}{2}$. By application of Bellman's dynamic programming optimality principle we have

$$\begin{aligned} V_{2n} &= \min_{u_{2n}}[E[q(x_{2n}^2+x_{2n+1}^2)+2ku_{2n}^2 \mid X_{2n}, U_{2n-2}] \\ &+ \min_{\substack{u_i \\ 2n+2 \le i \le N-1}} E\left[\sum_{i=2n+2}^{N-1} q(x_i^2+x_{i+1}^2)+2ku_i^2 \,\middle|\, X_{2n}, U_{2n-2}\right] , \end{aligned} \tag{32}$$

since

$$\begin{aligned} &E\left[\sum_{i=2n+2}^{N-1} q(x_i^2+x_{i+1}^2)+2ku_i^2 \,\middle|\, X_{2n}, U_{2n-2}\right] \\ &\quad = E\left[ E\left[\sum_{i=2n+2}^{N-1} q(x_i^2+x_{i+1}^2)+2ku_i^2 \,\middle|\, X_{2n+2}, U_{2n}\right] \,\middle|\, X_{2n}, U_{2n-2}\right] , \end{aligned} \tag{33}$$

we obtain

$$\begin{aligned} V_{2n} &= \min_{u_{2n}}[2qx_{2n}^2+2qx_{2n}(u_{2n}+a_{2n})+2qa_{2n}u_{2n}+(q+2k)u_{2n}^2 \\ &+ q\left(a_{2n}^2 + \frac{g_{2n}(1+b_{2n})}{b_{2n}(g_{2n}-2)} f_{2n}\right) + E[V_{2n+2} \mid X_{2n}, U_{2n-2}]]. \end{aligned} \tag{34}$$

For determining the optimal control, $u_{2n}^*$, this yields the equation

$$q(x_{2n}+a_{2n}) + (q+2k)u_{2n}^* + E\left[\frac{\partial V_{2n+2}}{\partial x_{2n+2}} \,\middle|\, X_{2n}, U_{2n-2}\right] = 0. \tag{35}$$

We will now show that a solution of the form

$$\begin{aligned} V_{2n}(x_{2n}, a_{2n}, f_{2n}) &= \tfrac{1}{2}A_{2n}x_{2n}^2+B_{2n}x_{2n}+C_{2n}x_{2n}a_{2n} \\ &+ D_{2n}a_{2n}+E_{2n}a_{2n}^2 + F_{2n}f_{2n}+G_{2n} \end{aligned} \tag{36}$$

may be chosen; in this case we have

$$\frac{\partial V_{2n+2}}{\partial x_{2n+2}} = A_{2n+2}x_{2n+2}+B_{2n+2}+C_{2n+2}(a_{2n+2}+\frac{1}{2+b_{2n}}x_{2n+2})$$

$$+ \frac{1}{2+b_{2n}} D_{2n+2}+E_{2n+2}[\frac{2a_{2n}b_{2n}}{(2+b_{2n})^2} + \frac{2(x_{2n+2}-x_{2n}-2u_{2n})}{(2+b_{2n})^2}]$$

$$+ F_{2n+2}[\frac{x_{2n+2}-x_{2n}-2u_{2n}-2v_{2n}}{2+g_{2n}} + \frac{x_{2n+2}-x_{2n}-2u_{2n}}{2+g_{2n}}$$

$$- \frac{b_{2n+2}}{2+g_{2n}}(\frac{2a_{2n}b_{2n}}{(2+b_{2n})^2} + \frac{2(x_{2n+2}-x_{2n}-2u_{2n})}{(2+b_{2n})^2})] \quad , \qquad (37)$$

from (19), (20) and (25) we obtain.

$$E[\frac{\partial V_{2n+2}}{\partial x_{2n+2}} | X_{2n}, U_{2n-2}] = (A_{2n+2}+\frac{C_{2n+2}}{2+b_{2n}}) X_{2n}+B_{2n+2}+\frac{D_{2n+2}}{2+b_{2n}}$$

$$+ [2A_{2n+2}+(1+\frac{2}{2+b_{2n}})C_{2n+2}+\frac{2E_{2n+2}}{2+b_{2n}}] a_{2n} \qquad (38)$$

$$+ 2(A_{2n+2} + \frac{C_{2n+2}}{2+b_{2n}}) u^*_{2n} \quad ,$$

substituting (38) into (35), solving for $u^*_{2n}$ we obtain for the optimal control

$$u^*_{2n} = - \frac{Q+(q+R)a_{2n}+(q+P)x_{2n}}{q+2k+2P} \quad , \qquad (39)$$

where $P = A_{2n+2} + \frac{C_{2n+2}}{2+b_{2n}}$, $Q = B_{2n+2} + \frac{D_{2n+2}}{2+b_{2n}}$, $R = 2A_{2n+2}+\frac{4+b_{2n}}{2+b_{2n}}C_{2n+2}+\frac{2E_{2n+2}}{2+b_{2n}}$.

From equations (19), (25), (26) and (36) we have for the expected cost

$$E[V_{2n+2} | X_{2n}, U_{2n-2}]$$

$$= \frac{1}{2}A_{2n+2}x_{2n}^2 + (B_{2n+2}+\frac{b_{2n}C_{2n+2}}{2+b_{2n}}) x_{2n}+2A_{2n+2}x_{2n}u^*_{2n}$$

$$+ (2A_{2n+2}+\frac{3C_{2n+2}}{2+b_{2n}}) x_{2n}a_{2n} + (4A_{2n+2}+\frac{6C_{2n+2}}{2+b_{2n}}) a_{2n}u^*_{2n}$$

$$+ 2A_{2n+2}u^{*2}_{2n} + (2A_{2n+2}+\frac{6C_{2n+2}}{2+b_{2n}}) a_{2n}^2 + 2(B_{2n+2}+\frac{b_{2n}C_{2n+2}}{2+b_{2n}}) u^*_{2n}$$

$$+ \left(2B_{2n+2} + \frac{2b_{2n}C_{2n+2}}{2+b_{2n}} + D_{2n+2}\right) a_{2n} + E_{2n+2}a_{2n}^2$$

$$+ \left[\left( A_{2n+2}+\frac{2C_{2n+2}}{2+b_{2n}}\right) \frac{(4+b_{2n})g_{2n}}{(g_{2n}-2)b_{2n}} + \frac{2g_{2n}E_{2n+2}}{(g_{2n}-2)\ (2+b_{2n})b_{2n}}\right.$$

$$\left. + \frac{g_{2n}(\frac{1}{2}g_{2n+2}-1)}{g_{2n+2}(\frac{1}{2}g_{2n}-1)} F_{2n+2}\right] f_{2n} + G_{2n+2} \quad . \tag{40}$$

Combining (34), (36), (39) and (40) we obtain the following recursion equations

$$A_{2n} = 2\left[ 2q+\frac{1}{2}A_{2n+2}-\frac{(q+P)\ (2q+2A_{2n+2})}{q+2k+2P} + \frac{(q+P)^2(q+2k+2A_{2n+2})}{(q+2k+2P)^2}\right],$$

$$B_{2n} = B_{2n+2} + \frac{b_{2n}C_{2n+2}}{2+b_{2n}} + \frac{2Q(q+P)(q+2k+2A_{2n+2})}{(q+2k+2P)^2}$$

$$- \frac{Q(2q+2A_{2n+2})+2(q+P)\ (B_{2n+2}+\frac{b_{2n}C_{2n+2}}{2+b_{2n}})}{q+2k+2P} \quad ,$$

$$C_{2n} = 2q+2A_{2n+2}+\frac{3C_{2n+2}}{2+b_{2n}} + \frac{2(q+R)(q+P)(q+2k+2A_{2n+2})}{(q+2k+2P)^2}$$

$$- \frac{(q+R)(2q+2A_{2n+2})+(q+P)\ (2q+4A_{2n+2}+\frac{6C_{2n+2}}{2+b_{2n}})}{q+2k+2P} \quad ,$$

$$D_{2n} = 2B_{2n+2} + \frac{2b_{2n}C_{2n+2}}{2+b_{2n}} + D_{2n+2} + \frac{2Q(q+2k+2A_{2n+2})(q+R)}{(q+2k+2P)^2}$$

$$- \frac{Q\ (2q+4A_{2n+2}+\frac{6C_{2n+2}}{2+b_{2n}}) +2(q+R)\ (B_{2n+2}+\frac{b_{2n}C_{2n+2}}{2+b_{2n}})}{q+2k+2P} \quad ,$$

$$E_{2n} = q+2A_{2n+2} + \frac{6C_{2n+2}}{2+b_{2n}} + E_{2n+2} + \frac{(q+R)^2(q+2k+2A_{2n+2})}{(q+2k+2P)^2}$$

$$- \frac{(q+R)\ (2q+4A_{2n+2}+\frac{6C_{2n+2}}{2+b_{2n}})}{q+2k+2P} \quad ,$$

$$F_{2n} = \frac{g_{2n}}{(g_{2n}-2)b_{2n}} \quad \left[(4+b_{2n})\ \left(A_{2n+2} + \frac{2C_{2n+2}}{2+b_{2n}}\right) + \frac{2E_{2n+2}}{2+b_{2n}}\right]$$

$$+\frac{qg_{2n}(1+b_{2n})}{b_{2n}(g_{2n}-2)}+\frac{g_{2n}(\frac{1}{2}g_{2n+2}-1)}{g_{2n+2}(\frac{1}{2}g_{2n}-1)}F_{2n+2} \ ,$$

$$G_{2n}=G_{2n+2}+\frac{Q^2(q+2k+2A_{2n+2})}{(q+2k+2P)^2}-\frac{2Q\left(B_{2n+2}+\frac{b_{2n}C_{2n+2}}{2+b_{2n}}\right)}{q+2k+2P} \ , \qquad (41)$$

with the boundary conditions

$$A_{N-1}=2q, \ B_{N-1}=0, \ C_{N-1}=0, \ D_{N-1}=0, \ E_{N-1}=0, \ F_{N-1}=0, \ G_{N-1}=0.$$

Thus the exact analytical solution for the feedback control has been obtained; from (39) we see that the optimal control depends on the filtering, the deterministic control law can not be applied in this case. Examining the marginal densities of m and $\frac{1}{\sigma^2}$: integrating equation (9) over m, we get the marginal density of $\frac{1}{\sigma^2}$, $p_{\gamma 2}(\frac{1}{\sigma^2}|f, g)$, this is the gamma-2 density with $E[\frac{1}{\sigma^2}]=\frac{1}{f}$, Var $[\frac{1}{\sigma^2}]=\frac{2}{f^2 g}$; integrating equation (9) over $\frac{1}{\sigma^2}$, we get the marginal density of m, $p_s(m \,|\, a, \frac{b}{f}, g)$, this is the student density with $E[m]=a$, Var $[m]=\frac{fg}{b(g-2)}$; thus we see that the variance of the mean, m, is proportional to the parameter f, the variance of $\frac{1}{\sigma^2}$ is inversely proportional to $f^2$, in addition, f is a function of a and u, hence there is a very complicated situation existing between control and filtering. Based on the exact analytical solution which we have obtained, we may study the compromise between control and filtering.

## 5. Conclusions

The stochastic optimal control problem - the problem of optimally controlling a linear discrete system which is subject to white Gaussian disturbances with partially known statistics - requires the solution of two equations: the filtering equation which updates the conditional probability densities of the unknown statistics; and the control equation which yields the input as a functional of these densities.

By Bayes's rule the filtering equations consist of a set of recursion equations. This has computational advantages when estimates are required in real time. Using Bellman's dynamic programming algorithm, an exact analytical solution of the feedback control law may be found. This solution serves as standard for evaluating approximate solutions.

Note that we update the estimates of statistics after every second measurement. Estimates may be updated after every measurement but this leads to a very complicated non-recursive filter. Consequently we have restricted our control to change only after every second measurement.

With these restrictions on the control and filtering we have used the dynamic programming algorithm to get the analytical solution of the feedback control law. We have not given a proof that the cascading of such an estimator with such a controller constitutes an over-all optimum control policy, but we feel that a proof could be given.

Although the derivations for the problem which we considered are quite involved, the resulting control and filtering algorithms which we have obtained are very simple. The results which we derived are for the scalar system. We would expect that general vector systems can be treated in much the same manner; there will be additional computations caused by the algebra. Also this work can be extended to the case when there are noisy observation of the state[13].

## Acknowledgment

The author is grateful to Dr. D. Gorman and Dr. J. Zaborszky for their valuable suggestions and discussions. This work is a part of the author's dissertation presented to the Sever Institute of Washington University, Saint Louis, Missouri, in partial fulfillment of the requirements for the degree of Doctor of Science and was supported in part by the Applied Mathematics Branch of the U.S. Air Force Office of Scientific Research under Grant No. AF-AFOSR-482-66.

References


1. Joseph, P. D., Tou, J. T., "On Linear Control Theory", AIEE Trans. Appl. Ind. 80, 193 (1961).
2. Florentin, J. J., "Partial Observability and Optimal Control", J. Elec. and Control, 13, 263 (1962).
3. Gunckel, T. L., and Franklin, G. F., "A General Solution for Linear Sampled-Data Control", Trans. ASME, Jour. of Basic Engr. 85D, 197 (1963).
4. Meier, L., "Combined Optimum Control and Estimation Theory", NASA Contractor Report, NASA CR-426, April, 1966.
5. Freimer, M., "A Dynamic Programming Approach to Adaptive Control Processes", IRE Nat. Conv. Record, Part 4, 10 (1959).
6. Tou, J. T., "System Optimization via Learning and Adaptation", International Jour. of Control, 2, 21 (1965).
7. Lin, T. T., and Yau, S. S., "Bayesian Approach to the Optimization of Adaptive Systems", IEEE Trans. on Sys. Sc. and Cyb. 3, 77 (1967).
8. Smith, G. L., "Sequential Estimation of Observation Error Variances in a Trajectory Estimation Problem", AIAA Jour., Vol. 5, 11, 1964 (1967).
9. Aoki, M., "Optimization of Stochastic Systems", Academic Press, New York, 1967.
10. Bellman, R., and Kalaba, R., "Dynamic Programming and Modern Control Theory", Academic Press, New York, 1964, p. 42.
11. Coughanowr, D., and Koppel, L., "Process Systems Analysis and Control", McGraw-Hill, New York, 1965, p. 71.
12. Raiffa, H., and Schlaifer, R., "Applied Statistical Decision Theory", Harvard, School of Business Administration, Boston, 1961.
13. Tarn, T. J., "Stochastic Optimal Control with Imperfectly Known Disturbances", D.Sc. Thesis, June 1968, Washington University, St. Louis, Mo.

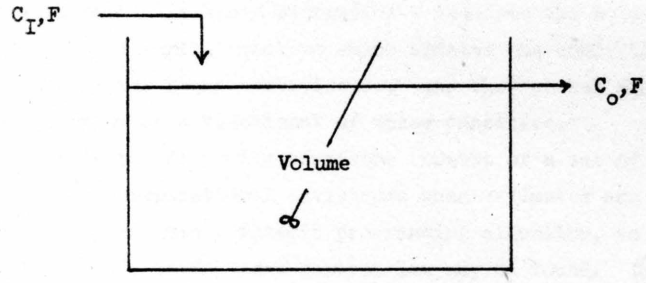




Figure 1. Continuous Stirred Tank Reactor.

(Theory-Stochastic Theory)

# AN APPROXIMATE METHOD OF STATE ESTIMATION AND CONTROL FOR NONLINEAR DYNAMICAL SYSTEMS UNDER NOISY OBSERVATIONS

by

Yoshifumi Sunahara
Department of Mechanical Engineering
Faculty of Polytechnic Sciences
Kyoto Institute of Technology
Kyoto, Japan

## 1. Introduction and symbolic conventions

There is no needs to say that dynamical systems to be controlled exhibit various kinds of nonlinear characteristics and may operate in a random environment whose stochastic characteristics undergo drastic changes. Thus, the general problem to be solved is to find the control of a noisy nonlinear dynamical system in some optimal fashion, given only partial and noisy observations of the system state and, possibly, only an incomplete knowledge of the system. It has already been shown under such conditions as linearity of the dynamical system, noisy observation and performance criterion given by a quadratic cost functional that the optimal control problem and the optimal estimation problem of the system state from the noise-corrupted observations may independently be solved.[1-3] However, this is, in general, not the case for the optimal control of nonlinear dynamical systems, and the over-all problem of optimal control and estimation must be carried out simultaneously. Since the establishment of the precise technique for the state estimation and the optimal control of nonlinear dynamical systems is almost impossible, in this paper, the author will introduce the reader to an approximate method which will be shown to play an important role in the realization of a broad class of stochastic optimal control.

Vector and matrix notations follow the usual manner, that is, lower case letters a, b and c,... will denote column vectors with i-th real components $a_i$, $b_i$ and $c_i$, etc. Capital letters A, B, C and G,... denote matrices with elements $a_{ij}$, $c_{ij}$ and $g_{ij}$, etc. If M is a matrix, then M' denotes its transpose. The symbol, $|M|$, denotes the determinant of the matrix M.

Certain algebraic quantities such as algebras, fields,... are expressed by the symbols, $\mathcal{Y}$, $\mathcal{F}$,..., etc. The symbol, $\mathcal{Y}_t$, denotes the smallest σ-algebra of ω sets with respect to which the random variables $y(\tau)$ with $\tau \leq t$ are measurable, where ω is the generic point of the probability space Ω. The mathematical expectation is denoted by $E$. The conditional expectation of a random variable conditioned by $\mathcal{Y}_t$ is simply expressed by "$\hat{\cdot}$" such that

$E\{x(t)|Y_\tau\} = \hat{x}(t|\tau)$, where $\tau \leq t$. For convenience of the present description, the principal symbols used here are listed below:

t: time variable, particularly the present time

$t_0$: the initial time at which observations start

x(t) and y(t): n-dimensional vector stochastic processes representing the system states and the observations respectively.

u(t): the control vector taking values in a convex compact subset $U \subset E^m$(m-dimensional Euclidean space)

w(t) and v(t): $d_1$- and $d_2$-dimensional Brownian motion processes respectively

C(t), G(t) and R(t): nxm, $nxd_1$ and $nxd_2$ matrices whose components depend on t

f[t,x(t)] and h[t,x(t)]: n-dimensional vector valued nonlinear functions respectively

$\hat{x}(t|t)$: optimal estimate of x(t) conditioned by $Y_t$, i.e., $E\{x(t)|Y_t\} \triangleq \hat{x}(t|t)$

P(t|t): an error covariance matrix in optimal estimate of x(t) conditioned by $Y_t$, i.e., $P(t|t) \triangleq \text{cov.}\ [x(t)|Y_t]$

## 2. Mathematical models and problem statement

Guided by a well-known state space representation concept, the dynamics of an important class of dynamical systems can be described by a nonlinear vector differential equation,

$$\frac{dx(t,\omega)}{dt} = f[t,\ x(t,\omega)] - C(t)u(t) + G(t)\gamma(t,\omega), \tag{2.1}$$

where $\gamma(t,\omega)$ is a $d_1$-dimensional Gaussian white noise disturbance. For the economy of descriptions, we shall omit to write the symbol $\omega$ here and below because of no confusion.

We shall start with a precise version of Eq.(2.1), namely the stochastic differential equation of Itô-type,[4]

$$dx(t) = f[t,\ x(t)]dt - C(t)u(t)dt + G(t)dw(t), \tag{2.2}$$

where the $d_1$-dimensional Brownian motion process w(t) has been introduced here along the relation between a Brownian motion process and a white noise or a sufficiently wide (but finite) band Gaussian random process $\gamma(t)$, (for more detail see the references[5-6])

$$w(t) \cong \int^t \gamma(s)ds. \tag{2.3}$$

We suppose that observations are made at the output of the nonlinear system with additive Gaussian disturbance. The observation process y(t) is the n-dimensional vector random process determined by

$$dy(t) = h[t,\ x(t)]dt + R(t)dv(t), \tag{2.4}$$

where we assume that the system noise w(t) and the observation noise v(t) are mutually independent.

In practical terms, the problem is to control x(t) in such a way as to minimize a real valued functional,

$$J(t, \mathcal{Y}_t) = \mathcal{E}\{\int_t^T L[s,x(s),u(s)]ds \,|\, \mathcal{Y}_t\}, \; [t_0 \le t \le T], \qquad (2.5)$$

based on the a priori probability distribution of $x(t_0)$, provided that the process $y(s)$ for $t_0 \le s \le t$ is acquired as the observation process, where $y(t_0) = 0$ and where L and $L_u$ are bounded, uniformly Hölder continuous in t and uniformly Lipschitz continuous in x and where $L_{uu}$ is bounded and continuous on $[t_0, T] \times E^n \times U$ (the y(s)-process contructs $\mathcal{Y}_t$). The subscript denotes differentiation here and below.

We shall consider the case where the state variables x(t) are completely observable. Usually, in this case, the optimal control must be assumed to depend on x(s), where $t_0 \le s \le t$. Bearing this fact in mind, we shall proceed to establish the solution of the stochastic differential equation (2.2).

Let $\psi(t,\cdot)$ be an m-dimensional vector stochastic process, such that, for each $t \in [t_0, T]$, $\psi(t,\cdot)$ is measurable and

$$\int_{t_0}^T \mathcal{E}\{\|\psi(t,\cdot)\|^2\}dt < \infty , \qquad (2.6)$$

where $\|\cdot\|$ expresses the norm in $E^m$. Let $\psi$ denotes the class of the $\psi(t)$-process. For some $\psi \in \Psi$, we call the u(t) admissible and write $u \in U$, if

$$u(t) = \psi(t,\cdot) \text{ for } t \in [t_0, T]. \qquad (2.7)$$

For the security of mathematical development in the sequel, the following hypotheses are additionally made:[7]

H-1: The component of the function $f[\cdot, \cdot]$ and $[h[\cdot, \cdot]$ are Baire functions with respect to the pair $(t,\xi)$ for $t_0 \le t \le T$ and $-\infty < \xi < \infty$, where $x(t)=\xi$

H-2: The functions $f[\cdot, \cdot]$ and $h[\cdot, \cdot]$ satisfy a uniform Lipschitz conditions in the variable $\xi$ and are bounded respectively by

$$\|f(t,\xi)\| \le K_1(1+\xi'\xi)^{1/2} \qquad (2.8a)$$

and

$$\|h(t,\xi)\| \le K_2(1+\xi'\xi)^{1/2}, \qquad (2.8b)$$

where both $K_1$ and $K_2$ are real positive constants and are independent of both t and $\xi$ respectively.

H-3: $x(t_0)$ is a random variable independent of the w(t)-process

H-4: All parameter matrices are measurable and bounded on the finite time interval $[t_0, T]$

H-5: $\{R(t)R(t)'\}^{-1}$ exists and this is bounded on $[t_0, T]$.

With the property (2.6) and the hypotheses H-1 to H-5, Eq.(2.2) has exactly a unique continuous solution x(t). A precise interpretation of Eq. (2.2) is given by the stochastic integral equation of Itô-type:[4]

$$x(t) = x(t_0) + \int_{t_0}^t f[s,x(s)]ds - \int_{t_0}^t C(s)u(s)ds + \int_{t_0}^t G(s)dw(s). \qquad (2.9)$$

## 3. Quasi-linear stochastic differentials and an approximation to non-linear filtering equations[8]

In this section, the development of the discussion requires that, until further notice, we set the control $u(t)$ equals to zero in Eq.(2.2). When $u(t)=0$, the symbol is temporarily changed from $x(t)$ to $z(t)$. With this symbolic change, Eq.(2.2) is

$$dz(t) = f[t, z(t)]dt + G(t)dw(t) \tag{3.1}$$

and also Eq.(2.4) is written by

$$dy(t) = h[t, z(t)]dt + R(t)dv(t), \tag{3.2}$$

where the same symbol $y(t)$ has been used as in Eq.(2.4) because of economy of notations.

The problem considered here is to find the minimal variance estimate of the state variable $z(t)$, provided that the process $y(s)$ for $t_0 \le s \le t$ is acquired as the observation process, where $y(t_0) = 0$.

We expand the function in Eq.(3.1) into

$$f[t, z(t)] = a(t) + B(t)\{z(t) - \hat{z}(t|t)\} + e(t), \tag{3.3}$$

where $a(t)$, $B(t)$ are an n-dimensional vector and an nxn matrix respectively, and where $e(t)$ denotes the collection of n-dimensional vector error terms, and where $\hat{z}(t|t) = E\{z(t)|\mathcal{Y}_t\}$. We shall determine $a(t)$ and $B(t)$ in such a way that the conditional expectation of the squared norm of $e(t)$ conditioned by $\mathcal{Y}_t$, $E\{\|e(t)\|^2|\mathcal{Y}_t\}$, becomes minimal with respect to $a(t)$ and $B(t)$. It is a simple exercise to show in the calculus of variation that the necessary and sufficient conditions for $\min. E\{\|e(t)\|^2|\mathcal{Y}_t\}$ are given by

$$a(t) = E\{f[t, z(t)]|\mathcal{Y}_t\} \tag{3.4a}$$

and

$$B(t) = E[\{f[t,z(t)] - \hat{f}[t,z(t)]\}\{z(t) - \hat{z}(t|t)\}'|\mathcal{Y}_t]\bar{P}(t|t)^{-1}, \tag{3.4b}$$

where

$$\bar{P}(t|t) = \text{cov.}\ [z(t)|\mathcal{Y}_t] \tag{3.5}$$

The scalar expressions of (3.4) are as follows:

$$a_i(t) = E\{f_i[t, z(t)]|\mathcal{Y}_t\} = \hat{f}_i[t, z(t)] \tag{3.6a}$$

$$\sum_{\nu=1}^{n} b_{i\nu}(t)E[\{z_\nu(t) - \hat{z}_\nu(t|t)\}\{z_j(t) - \hat{z}_j(t|t)\}|\mathcal{Y}_t]$$

$$= E[\{f_i[t, z(t)] - \hat{f}_i[t, z(t)]\}\{z_j(t) - \hat{z}_j(t|t)|\mathcal{Y}_t\} \tag{3.6b}$$

where $\hat{z}_j(t|t) = E\{z_j(t)|\mathcal{Y}_t\}$ and $i, j=1,2,\ldots,n$. Using $a(t)$ and $B(t)$ determined by (3.4) and (3.5), we approximate Eq.(3.1) by

$$z(t) = z(t_0) + \int_{t_0}^{t} [a(s) + B(s)\{z(s) - \hat{z}(s|s)\}]ds + \int_{t_0}^{t} G(s)dw(s). \tag{3.7}$$

The same procedure is applicable to the observation process given by

Eq.(3.2). Through the expansion of the function, $h[t, z(t)]$, in the form;

$$h[t, z(t)] = h_1(t) + H_2(t)\{z(t) - \hat{z}(t|t)\} + e_h(t), \tag{3.8}$$

the following conditions can easily be obtained so as to minimize $E\{\|e_h(t)\|^2 |\mathcal{Y}_t\}$ with respect to $h_1(t)$ and $H_2(t)$:

$$h_1(t) = E\{h[t, z(t)]|\mathcal{Y}_t\} \triangleq \hat{h}[t, z(t)] \tag{3.9a}$$

$$H_2(t) = E[\{h[t,z(t)] - \hat{h}[t,z(t)]\}\{z(t) - \hat{z}(t|t)\}'|\mathcal{Y}_t]\bar{P}(t|t)^{-1}. \tag{3.9b}$$

We shall assume here that, for $t\varepsilon[t_0, T]$, the conditional probability density function $p\{z(t)|\mathcal{Y}_t\}$, is Gaussian with the mean value $\hat{z}(t|t)$ and covariance matrix $\bar{P}(t|t)$, i.e.,

$$p\{z(t)|\mathcal{Y}_t\} = (2\pi)^{-\frac{n}{2}}|\bar{P}(t|t)|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}\{z - \hat{z}(t|t)\}'\bar{P}(t|t)^{-1} \times\{z - \hat{z}(t|t)\}]. \tag{3.10}$$

With the help of (3.10), both $a(t)$ and $B(t)$ can be obtained in the form $a(t) = a(t, \hat{z}(t|t), \bar{P}(t|t))$ and $B(t) = B(t, \hat{z}(t|t), \bar{P}(t|t))$ or $b_{ij}(t) = \partial a_i(t)/\partial \hat{z}_j(t|t)$. A striking fact is that the random variables $a(t)$ and $B(t)$ are not independent but depend mutually on the state estimate $\hat{z}(t|t)$ and the error covariance matrix $\bar{P}(t|t)$. From this point of view, in reality, more precise symbols, $a(t,\hat{z}(t|t), \bar{P}(t|t))$ and $B(t,\hat{z}(t|t), \bar{P}(t|t))$ should be introduced. However, for the economy of description, we merely denote these by $a(t)$ and $B(t)$ without indicating the dependence on both $\hat{z}(t|t)$ and $\bar{P}(t|t)$. Both $h_1(t)$ and $H_2(t)$ also follow this symbolic convention.

From Eq.(3.7), we may thus define here the following n-dimensional quasi-linear stochastic differentials of Itô-type for Eq.(3.1),

$$dz(t) = B(t)z(t)dt + \{a(t) - B(t)\hat{z}(t|t)\}dt + G(t)dw(t), \tag{3.11}$$

and for the observation process (3.2),

$$dy(t) = H_2(t)z(t)dt + \{h_1(t) - H_2(t)\hat{z}(t|t)\}dt + R(t)dv(t). \tag{3.12}$$

However, respective draft terms in Eqs.(3.3) and (3.8) still remain unknown. We shall thus proceed to solve the problem including the computation of the state estimate $\hat{z}(t|t)$ and the error covariance matrix $\bar{P}(t|t)$.

Let $\Phi(t, t_0)$ be the fundamental matrix associated with the homogeneous differential equation, $dz(t)/dt = B(t)z(t)$. The solution of Eq.(3.11) can formally be written as

$$z(t) = \Phi(t, t_0)z(t_0) + \int_{t_0}^{t} \Phi(t, s)\{a(s) - B(s)\hat{z}(s|s)\}ds + \int_{t_0}^{t} \Phi(t, s)G(s)dw(s). \tag{3.13}$$

We write for the second term of the right side of Eq.(3.13)

$$\zeta(t) = -\int_{t_0}^{t} \Phi(t, s)\{a(s) - B(s)\hat{z}(s|s)\}ds \qquad (3.14)$$

and introduce a new stochastic process

$$\xi(t) = z(t) + \zeta(t). \qquad (3.15)$$

Combining Eq.(3.13) with (3.14) and noting that $\xi(t_0) = \hat{z}(t_0)$, from Eqs. (3.14) and (3.15), the $\xi(t)$-process is of Ito-type and the stochastic differential is

$$d\xi(t) = B(t)\xi(t)dt + G(t)dw(t). \qquad (3.16)$$

On the other hand, it follows from Eq.(3.12) that

$$y(t) = \int_{t_0}^{t} H_2(s)z(s)ds + \int_{t_0}^{t} \{h_1(s) - H_2(s)\hat{z}(s|s)\}ds + \int_{t_0}^{t} R(s)dv(s). \qquad (3.17)$$

Let the second term of the right side of Eq.(3.17) be $\zeta_y(t)$ and define $\eta_y(t) \triangleq y(t) - \zeta_y(t)$. Then we obtain

$$d\eta_y(t) = H_2(t)z(t)dt + R(t)dv(t) \qquad (3.18)$$

with $\eta_y(t_0) = 0$. With $\eta_y(t)$ determined by Eq.(3.18), define a new stochastic process $\eta(t)$ by its stochastic differential,

$$d\eta(t) = d\eta_y(t) + H_2(t)\zeta(t)dt, \qquad (3.19)$$

and $\eta(t_0) = 0$. Using Eqs.(3.15) and (3.18), Eq.(3.19) becomes

$$d\eta(t) = H_2(t)\xi(t)dt + R(t)dv(t). \qquad (3.20)$$

Since $\zeta(t)$ is $Y_t$-measurable, it follows from Eq.(3.15) that

$$\hat{\xi}(t|t) = E\{\xi(t)|Y_t\} = \hat{z}(t|t) + \zeta(t). \qquad (3.21)$$

Let $H_t$ be the σ-algebra of ω sets generated by the random variables $\eta(s)$ for $t_0 \le s \le t$. Then the $y(t)$-process is $H_t$-measurable and thus

$$E\{\xi(t)|Y_t\} = E\{\xi(t)|H_t\} \triangleq \hat{\xi}(t|t). \qquad (3.22)$$

Now we consider that the $\xi(t)$-process is the fictitious state variables determined by Eq.(3.16) and that Eq.(3.20) denotes the observations which are made on the $\xi(t)$-process. This situation implies that the current estimate $\hat{\xi}(t|t)$ is given by [9,10]

$$d\hat{\xi} = B(t)\hat{\xi}dt + P_\xi(t|t)H_2(t)'\{R(t)R(t)'\}^{-1}\{d\eta - H_2(t)\hat{\xi}dt\}, \qquad (3.23)$$

where

$$P_\xi(t|t) = \text{cov.}\ [\xi(t)|H_t]. \qquad (3.24)$$

Substituting Eq.(3.20) into Eq.(3.23) and using Eqs.(3.12) and (3.21), it follows that

$$dz = \hat{f}[t, z(t)]dt + \bar{P}(t|t)H_2(t)'\{R(t)R(t)'\}^{-1}(dy - \hat{h}dt). \qquad (3.25)$$

where Eqs.(3.6a) and (3.9a) have been used. By combining (3.21) with (3.24), we have

$$\bar{P}(t|t) = \text{cov. } [z(t)|\mathcal{Y}_t] = P_\xi(t|t) \tag{3.26a}$$

and the version of $d\bar{P}(t|t)/dt$ is

$$\frac{d\bar{P}}{dt} = B\bar{P} + \bar{P}B' + GG' - \bar{P}H_2'\{RR'\}^{-1}H_2\bar{P}. \tag{3.26b}$$

Eqs.(3.25) and (3.26) describe the dynamic structure of a quasi-linear filter for generating a current estimate $\hat{z}(t|t)$ with the respectively given initial values, $\hat{z}(t_0|t_0)$ and $\bar{P}(t_0|t_0)$. In Appendix, the quantitative aspect of approximated fashion of filter dynamics is shown, including comparative discussions on various structures of filter dynamics.

## 4. Quasi-optimal control

In this section, the control term $u(t)$ in Eq.(3.1) is revived, noting that the symbol changes naturally from $z(t)$ to $x(t)$.

Let the function L in (2.5) be

$$L(t,x,u) = x'M(t)x + u'N(t)u, \tag{4.1}$$

where M and N are respectively measurable, locally bounded, positive semi-definite and positive definite symmetric matrices. In the case where both the dynamical system and the observation are respectively determined by linear stochastic differentials, it has already been verified that the optimal control exists and this is $u^o(t) = \hat{\psi}^o[t, \hat{x}(t|t)] = N(t)^{-1}C(t)'Q(t)\hat{x}(t|t)$, where Q is the unique solution of a certain matrix Riccati equation.[1,3] In the case of nonlinear regulator problems considered, the quasi-optimal control may be found out by an extensive use of the quasi linearization technique developed in the previous section to the version of stochastic control.

It is apparent that the x(t)-process has the quasi-linear stochastic differential,

$$dx(t) = B(t)x(t)dt + \{a(t)-B(t)\hat{x}(t|t)\}dt - C(t)\psi(t)dt + G(t)dw(t), \tag{4.2}$$

where the definition of the admissible control given by (2.7) has been taken into account with the simplified notation $\psi(t)$.

$$dy(t) = h_1(t)dt + H_2(t)\{x(t) - \hat{x}(t|t)\}dt + R(t)dv(t). \tag{4.3}$$

Furthermore, with the help of Eq.(3.25), it can easily be shown that the state estimation $\hat{x}(t|t)$ for the nonlinear system described by Eq.(4.2)

$$d\hat{x} = \hat{f}dt - C\psi dt + PH_2'(RR')^{-1}(dy - \hat{h}dt), \tag{4.4}$$

where the version of dP/dt has the same form as given by Eq.(3.26b).

In the present case, the basic process is $\hat{x}(t|t)$ $(t_0 \le t \le T)$ with the stochastic differential (4.4); the cost rate function is given by (4.1) and the performance index by (2.5).

Combining the stochastic linearization technique with the line of attack on the linear regulator problem, we shall suppose that $u(t) = \hat{\psi}[t, \hat{x}(t|t)]$. It has been proved by solving the following Bellman's equation[10] that the optimal control $\hat{\psi}^o$ and $V(t, \xi)$ exist

$$\min_{u\epsilon U}\{\hat{L}(t,\xi,u) + V_t(t,\xi) + \hat{\mathcal{L}}_{\hat{\psi}}V(t,\xi)\} = 0, \tag{4.5a}$$

with terminal condition

$$V(T, \xi) = 0, \tag{4.5b}$$

where

$$V(t,\xi) = E\{\int_t^T \hat{L}[s,\hat{x}(t|t), \hat{\psi}^o[s,\hat{x}(s|s)]]ds|\hat{x}(t|t) = \xi\}, \tag{4.6}$$

$$\hat{L} = E\{L[s, x(s), \hat{\psi}^o[s, \hat{x}(s|s)]]|\hat{x}(s|s) = \xi\} \tag{4.7}$$

and $\hat{\mathcal{L}}_{\hat{\psi}}$ denotes the differential generator of the $\hat{x}(t|t)$-process given by[11]

$$\hat{L}_{\hat{\psi}}(V) \equiv \frac{1}{2}tr\,\{\textstyle\sum(t)'V_{\xi\xi}\sum(t)\} + \{a(t) - C(t)\hat{\psi}(t,\xi)\}'V_\xi \tag{4.8}$$

with

$$\textstyle\sum(t) = PH_2'(RR')^{-1}R \tag{4.9}$$

because of (4.4) and the fact that the differential $dy-\hat{h}dt$ in Eq.(4.4) may be replaced by the suitably scaled differential of a Brownian motion process.

In the case where the function L is given by (4.1), it follows from (4.7) that

$$\hat{L}[s,\xi,\hat{\psi}(s,\xi)] = tr\,M(s)P(s|s) + \xi'M(s)\xi + \hat{\psi}'N(s)\hat{\psi}. \tag{4.10}$$

We shall suppose that Bellman's equation (4.5) has a solution

$$V(t,\xi) = \xi'\Pi(t)\xi + 2\xi'\alpha(t) + \beta(t), \tag{4.11}$$

where $\Pi(t)$, $\alpha(t)$ and $\beta(t)$ will be determined as the solutions of matrix differential equations which will be given later. Applying (4.8), (4.10) and (4.11) to (4.5) and performing the minimization of Eq.(4.5), the optimal control is

$$\hat{\psi}^o(t,\xi) = \{N(t)^{-1}C(t)'\Pi(t)\}\xi + N(t)^{-1}C(t)'\alpha(t), \tag{4.12}$$

and $\Pi(t)$, $\alpha(t)$ satisfy

$$\frac{d\Pi(t)}{dt} - \Pi(t)C(t)N(t)^{-1}C(t)'\Pi(t) + M(t) = 0 \tag{4.13}$$

$$\frac{d\alpha(t)}{dt} - \Pi(t)C(t)N(t)^{-1}C(t)'\alpha(t) + \Pi(t)a(t) = 0 \tag{4.14}$$

for $t_0 \le t \le T$ with

$$\Pi(T) = 0, \alpha(T) = 0. \tag{4.15}$$

Furthermore, $\beta(t)$ in (4.11) satisfies

$$\frac{d\beta(t)}{dt} + tr\,[\textstyle\sum(t)'\Pi(t)\sum(t)] + tr\,[M(t)P(t|t)] + 2a(t)'\alpha(t)$$
$$- \alpha(t)'C(t)N(t)^{-1}C(t)'\alpha(t) = 0 \tag{4.16}$$

for $t_0 \le t \le T$ with

$$\beta(T) = 0 \tag{4.17}$$

and this is necessary to compute (4.11), with $\Pi(t)$ and $\alpha(t)$. In Eqs.(4.13) and (4.14), both $\Pi(t)$ and $\alpha(t)$ are actually independent of the dynamic characteristics of an observation mechanism, $h(t,x)$ and $R(t)$. Hence the optimal control depends on the cost rate function matrices M and N and on the system dynamics $f(t,x)$. However, a serious diffculty arises in the version of numerical computation on Eqs.(4.12), (4.13), (4.14), (4.15) and (4.16). In fact, the computation of (4.12) with Eqs.(4.13) and (4.15) has to start with the pre-assigned initial values of the state estimation $\hat{x}(t_0|t_0)$ and error covariance $P(t_0|t_0)$ and, furthermore, with $\Pi(t_0)$ and $\alpha(t_0)$ which are determined by the so-called trial and error method.

5. An illustrative example

For the purpose of exploring the quantitative aspects, we shall consider here the one-dimensional case. The dynamical system considered here is schematically shown by block diagram in Fig.1. From Fig.1, the stochastic differential equation of the dynamical system is given by

$$dx = f(-x)dt + udt + gdw \tag{5.1}$$

with

$$f(x) = \sin x, \tag{5.2}$$

where $K_1 = K_2 = K_3 = 1$. The observation process is

$$dy = xdt + rdv. \tag{5.3}$$

Application of (3.4a) and (3.4b) to the present case gives

$$a(t) = -\sin\hat{x}\, \exp(-0.5p) \tag{5.4}$$

$$b(t) = -\cos\hat{x}\, \exp(-0.5p). \tag{5.5}$$

From Eqs.(4.4) and (3.26b), the approximated filter dynamics and related error covariance are determined by

$$d\hat{x} = -\sin\hat{x}\, \exp(-0.5p)dt - udt + pr^{-2}(dy - \hat{x}dt) \tag{5.6}$$

and

$$\frac{dp}{dt} = -2\cos\hat{x}\, \exp(-0.5p) + g^2 - p^2r^{-2}. \tag{5.7}$$

Letting n = 1 and m = 1, in (4.1), we have

$$\hat{\psi}^o(t, \xi) = \pi(t)\xi + \alpha(t) \tag{5.8}$$

and

$$V(t, \xi) = \pi(t)\xi^2 + 2\alpha(t)\xi + \beta(t), \tag{5.9}$$

where

$$\frac{d\pi(t)}{dt} = \pi^2(t) - 1 \tag{5.10a}$$

$$\frac{d\alpha(t)}{dt} = \pi(t)\alpha(t) - \pi(t)a(t) \tag{5.10b}$$

$$\frac{d\beta(t)}{dt} = -\sigma^2(t|t)\pi(t) - 2a(t)\alpha(t) + \alpha^2(t) - p(t|t). \qquad (5.10c)$$

Equations (5.6) to (5.10) are simulated on a digital computer with the subroutine for the generation of random disturbance, $\gamma(t)$ and $\theta(t)$. Fig.2(a) shows the running values of the state estimation $\hat{x}(t|t)$(in figures presented here and below, the symbols $\hat{x}(t|t)$ and $p(t|t)$ are simply denoted by $\hat{x}(t)$ and $p(t)$), for the pre-assigned control interval [0,1.0](sec). The sample path behavior of the true system is also shown as the run $x(t)$. However, the $x(t)$-process is, in practice, in accessible and this is only for comparative observation. The dotted run in Fig.2(a) shows the sample path behavior of the quasi-linearized system. Comparison of the sample paths of the quasi-linear system and filter dynamics with the that of true system, actully reveals that, as time goes on, the pursuit behavior of the $\hat{x}(t)$-process to the inaccessible $x(t)$-process becomes improved with the elevated accuracy of the stochastic linearization. The optimal control signal run is also plotted on Fig.2(a). Figure 2(b) shows the error covariance of filtering action $p(t|t)$, and also $\pi(t)$, $\alpha(t)$ which may be adopted as a successful set of trial and error method. Figure 3 shows the numerical results of digital simulation studies starting with differently initial values from Fig.2.

## 6. Conclusion

The technique started with the stochastic linearization of the dynamical system and with that of the observation dynamics. Based on the linearized system dynamics, a class of finite dimensional approximations to the optimal filter has been drived. The optimal control has been obtained for the linearized system by means of solving Bellman's equation. In general, the optimal control depends parametrically on both the conditional averaged behavior $\hat{f}[t,x] = a(t)$ of nonlinear action and the choice of performance index factors M, N. Through the analytical development and the numerical results, we may conclude that the approximation procedure has desirable properties in realizing the feedback configuration of stochastic optimal control.

## References

[1]. Joseph, P. D. and Tou, J.T.: On linear control theory, AIEE. Trans. Appl. Ind. Pt.II, 80, 193-196 (1961)

[2]. Schultz, P. R.: Optimal control in the presence of measurement errors and random disturbances, Ph. D. Thesis, Dept of Eng., Univ. of California, Los Angeles, California, (1963)

[3]. Wonham, W. M.: On the separation theorem of stochastic control; Tech. Reps. Control and Information System Lab., Elect. Res. Center, NASA, Rep. No. PM-38, Camb. Mass. Jan. (1968)

[4]. Itô, K.: On stochastic differential equations; Mem. Math. Soc., 4, (1951)

[5]. Yaglom, A.M.: An introduction to the theory of random functions, Prentice-Hall, N.J. (1962)

[6]. Wong, E. and Zakai, M.: On the relation between ordinary and stochastic differential equations; Int. J. Eng. Sci. 3, (1965)

[7]. Doob, J.L.: Stochastic processes: John Wiley, N.Y. (1953)

[8]. Sunahara, Y.: An approximate method of state estimation for nonlinear dynamical systems; Center for Dynamical Systems, Tech. Reps., Div. of Appl. Math., Brown Univ., No.67-8, (1967)

[9]. Kalman, R. E.: A new approach to linear filtering and prediction problems; Trans. ASME, J. Basic Eng., 82D, (1960), p.p. 33-45.

[10]. Kalman, R. E. and Bucy, R. S.: New results in linear filtering and prediction theory; Trans. ASME, J. Basic Eng., 83D, (1961), p.p. 95-108.

[11]. Dynkin, E. B.: Markov processes, I, II; Academic Press, N. Y., (1965)

## Appendix: Comparative discussions of approximations to nonlinear filters

As we can observe in this paper, the determination of a filter dynamics is extremely important in solving the optimal control for dynamical systems under noisy observations.

Up to the present time, several trials have been made on the physical realization of optimal nonlinear filters in an approximate form of finite dimensional filters.[A-1∿A-3] The major differences in the derivation lie in the estimation criteria and in the approximation procedure applied. Although the most familiar technique is the introduction of Taylor series expansion on a nonlinear function, the basic notion of the approximation described here is the expansion of the nonlinear function and the determination of the coefficients by means of the minimal square error criterion, including the Gaussian assumption to the conditional probability density function. This implies that the infinite dimensional filter is approximated by the two dimensional filter consisting of the first and second moments.

To make comparative discussions more clear, two examples are shown.

[Example-1]. We shall consider once again the example in section 5. Letting $u \equiv 0$, the approximated filter dynamics is determined by

$$d\hat{z} = -\sin\hat{z}\,\exp(-0.5\bar{p})dt + \bar{p}r^{-2}(dy - \hat{z}dt) \qquad (A.1)$$

and the related error covariance is given by Eq.(5.7) where the symbol should be changed from $\hat{x}$ to $\hat{z}$.

As another possible method of approximation, we shall consider the method of Taylor expansion.[A-1] We expand the nonlinear function into the following form,

$$f_i(z) = f_i(\hat{z}) + f_{ij}^{(1)}(\hat{z})(z_j-\hat{z}_j) + \frac{1}{2}f_{ijk}^{(2)}(z_j-\hat{z}_j)(\hat{z}_k-\hat{z}_k) \quad (A.2)$$

where $f_{ij}^{(1)} = \partial f_i/\partial z_j$ and $f_{ijk}^{(2)} = \partial^2 f_i/\partial z_j \partial z_k$, and where $i,j,k = 1,2,\ldots,n$. It follows from (A-2) that

$$\hat{f}_i = f_i(\hat{z}) + \frac{1}{2}f_{ijk}^{(2)}(\hat{z})\bar{P}_{jk} \quad (A.3)$$

where $\bar{P}_{jk}$ expresses the $(j,k)$ element of covariance matrix $\bar{P}$. Bearing (A.3) in mind, a somewhat tedious calculation shows the results,

$$d\hat{z} = (-\sin\hat{z} + \frac{1}{2}\sin\hat{z}\bar{p})dt + \bar{p}r^{-2}(dy - \hat{z}dt) \quad (A.4)$$

$$\frac{d\bar{p}}{dt} = -2\bar{p}\cos\hat{z} - \bar{p}^2 r^{-2} + g^2. \quad (A.5)$$

A numerical version of comparative results of filter dynamics determined by Eqs.(A.1) and (5.7) with those given by Eqs.(A.4) and (A.5) is shown by Figs. A-1 and A-2.

[Example 2]. Let us consider the two-dimensional $x(t)$-process with the vector nonlinear stochastic differential,

$$dz_1 = z_2 dt,\ dz_2 = [-z_2 + f(\dot{-z_1})]dt + gdw \quad (A.6)$$

with $z=z_1$, $dz/dt = z_2$ and with

$$f(z_1) = z_1 - \frac{z_1^3}{8} \quad (A.7)$$

The observation process is

$$dy_1 = z_1 dt + \gamma_1 dv_1 \quad dy_2 = z_2 dt + \gamma_2 d_2 . \quad (A.8)$$

In this case, it is a simple exercise to obtain

$$a_1 = \hat{z}_2,\ a_2 = -\hat{z}_2 - \hat{z}_1 + \frac{1}{8}\hat{z}_1^3 + \frac{3}{8}p_{11}\hat{z}_1 \quad (A.9)$$

and

$$b_{11} = 0,\ b_{12} = 1,\ b_{21} = -1 + \frac{3}{8}z_1^2 + \frac{3}{8}p_{11},\ b_{22} = 0. \quad (A.10)$$

Owing to limited space, the approximated filter dynamics and error covariance matrix are listed on Figs.A-3 and A-4.

## References

(A-1) Schwartz, L.: Approximate continuous nonlinear minimal-variance filtering; Huges Aerospace Tech. Reps., SSD60472R, Dec. 1966

(A-2) Cox, H.: On the estimation of state variables and parameters for noisy systems; IEEE Trans., AC-9, 1964

(A-3) Kushner, H.J.: Approximations to optimal nonlinear filters; 1967 JACC Preprints of papers, 1967

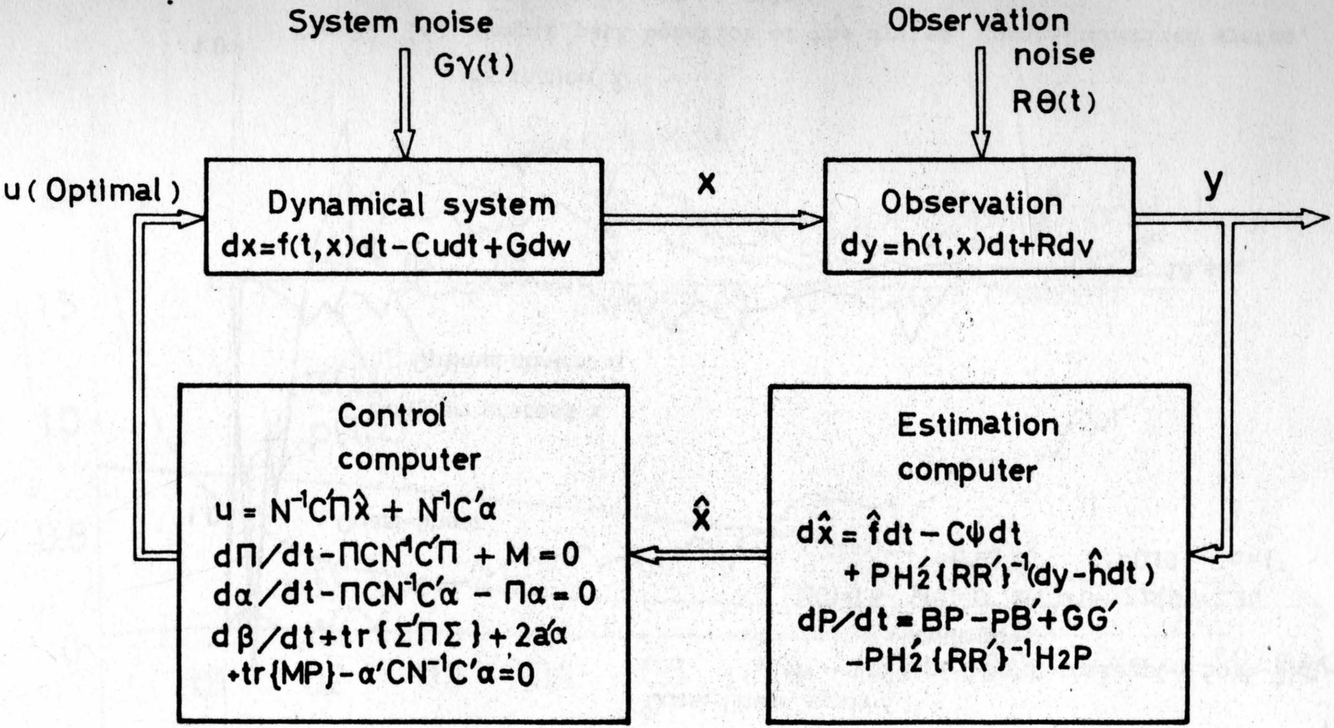



Fig.1. Over-all configuration of optimal control for the nonlinear dynamical system under noisy observations

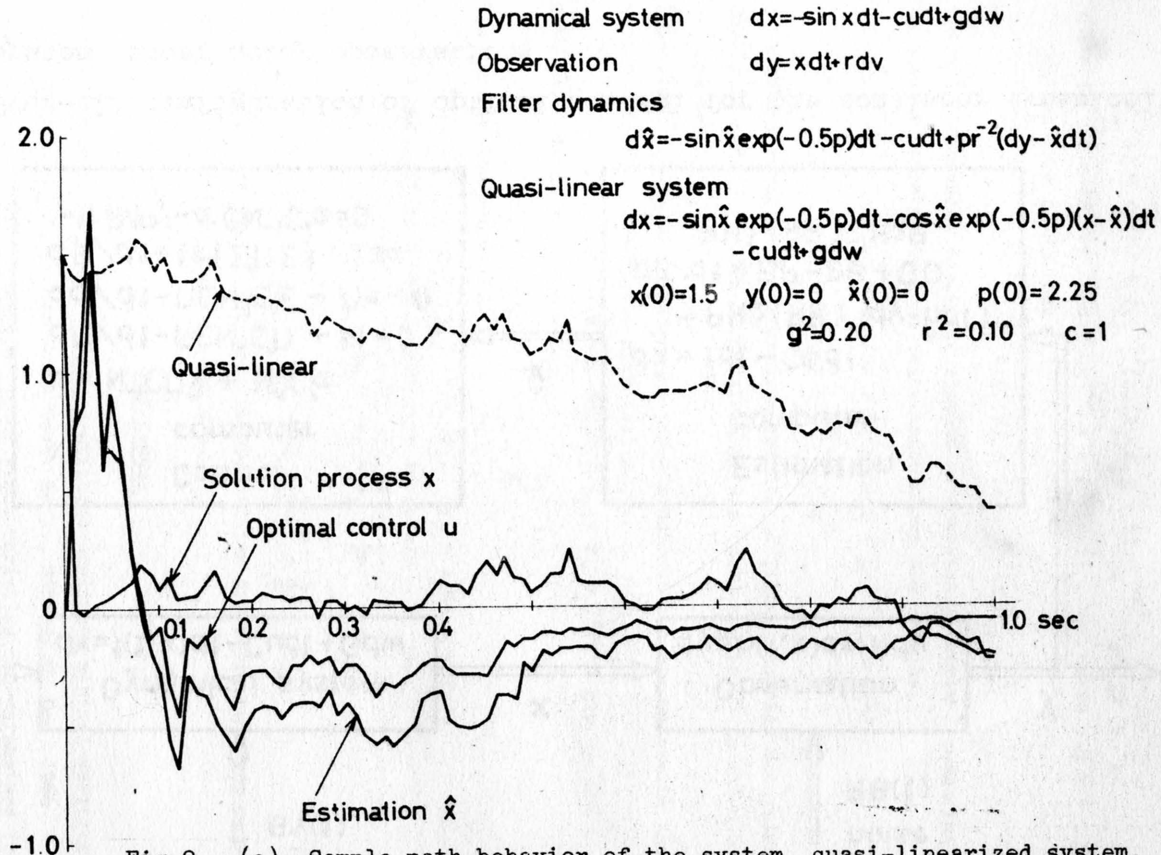



Fig.2. (a) Sample path behavior of the system, quasi-linearized system, filter and optimal filter

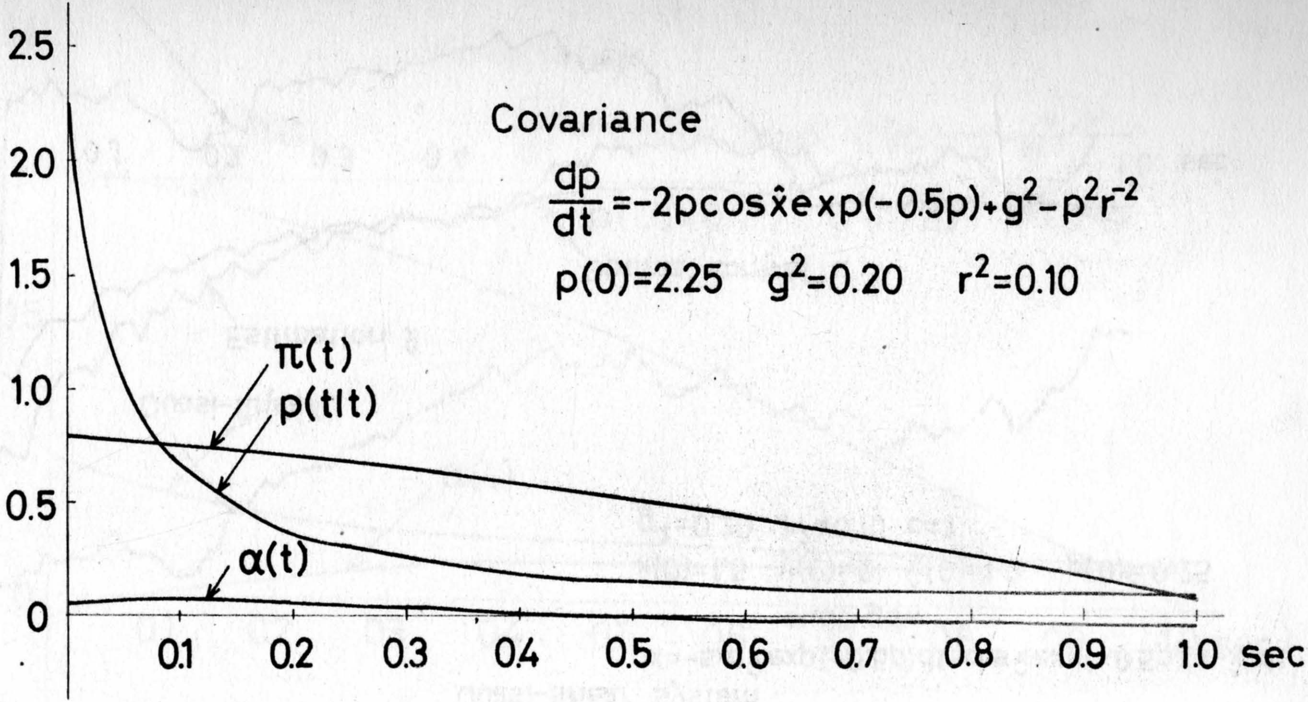



Fig.2. (b) Error covariance $P(t|t)$ and convergence of $\pi(t)$ and $\alpha(t)$

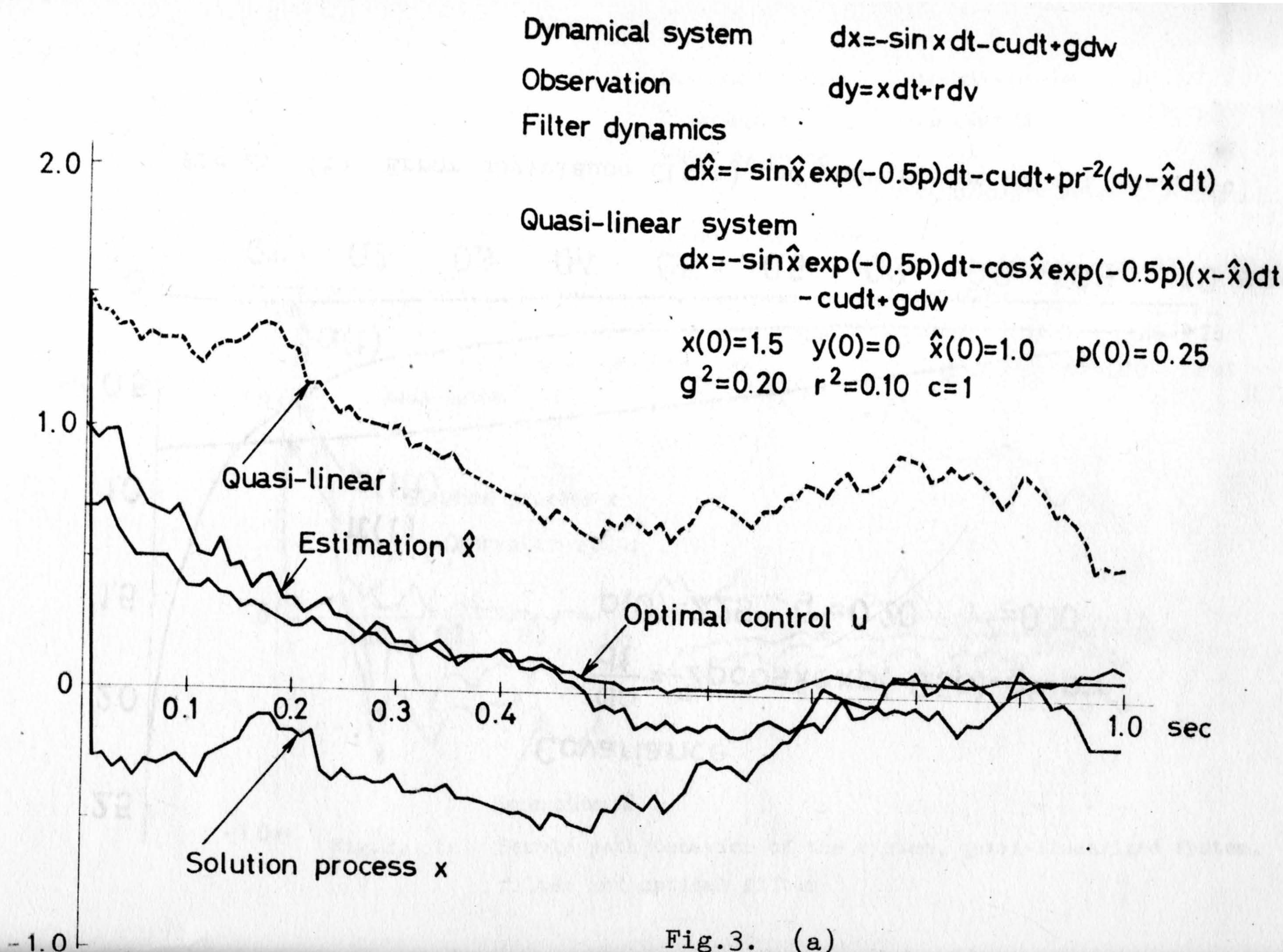

Dynamical system dx=-sin x dt-cudt+gdw
Observation dy=xdt+rdv
Filter dynamics
dx̂=-sinx̂ exp(-0.5p)dt-cudt+pr^-2(dy-x̂dt)
Quasi-linear system
dx=-sinx̂ exp(-0.5p)dt-cosx̂ exp(-0.5p)(x-x̂)dt
-cudt+gdw
x(0)=1.5 y(0)=0 x̂(0)=1.0 p(0)=0.25
g^2=0.20 r^2=0.10 c=1
2.0
1.0
0
-1.0
0.1
0.2
0.3
0.4
1.0 sec
Quasi-linear
Estimation x̂
Optimal control u
Solution process x

Fig.3. (a)

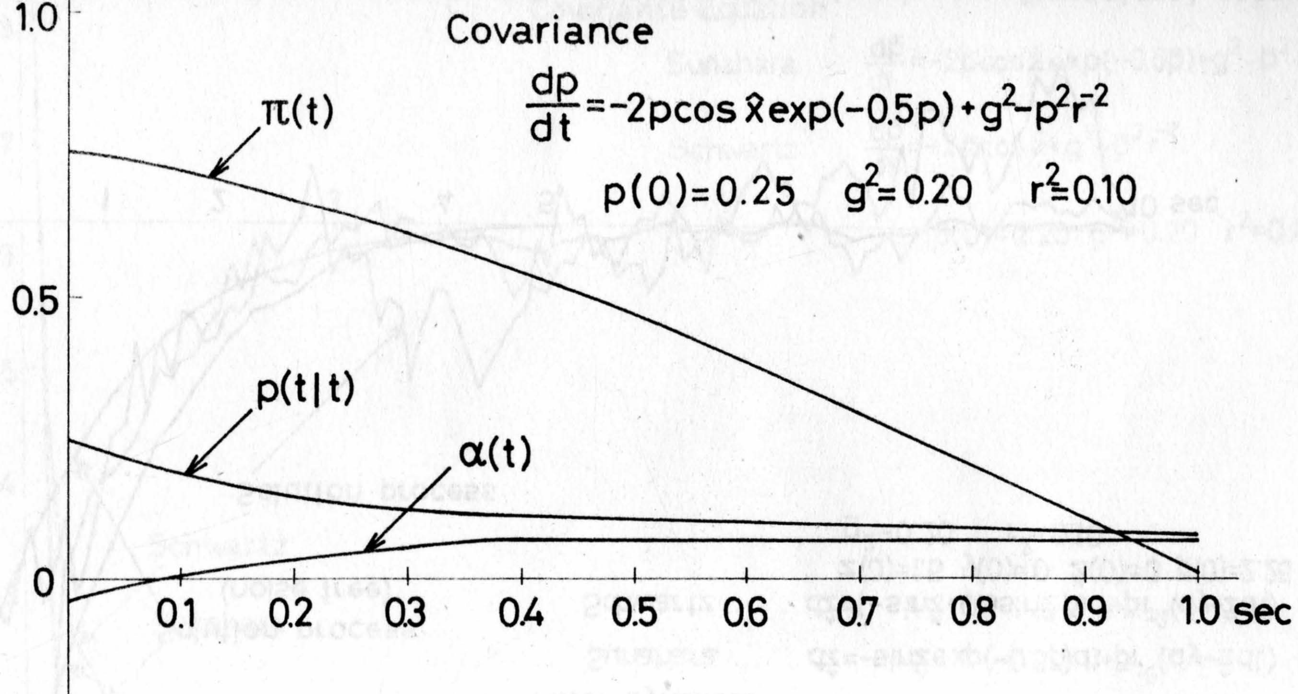

Covariance
dp/dt = -2p cos x̂ exp(-0.5p) + g²-p²r⁻²
p(0)=0.25   g²=0.20   r²=0.10
π(t)
p(t|t)
α(t)
1.0
0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0 sec

Fig.3. (b)

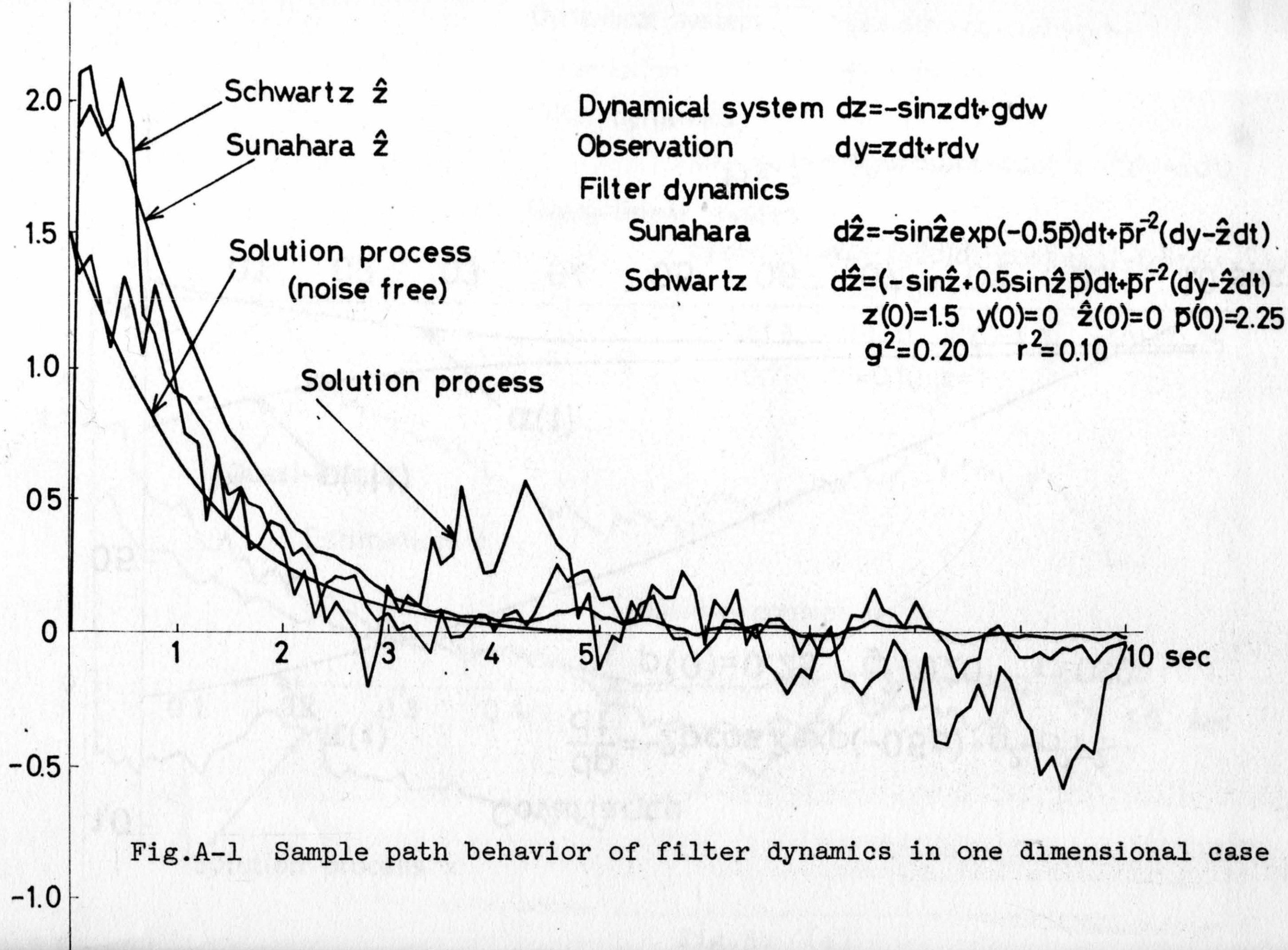



Fig.A-1 Sample path behavior of filter dynamics in one dimensional case

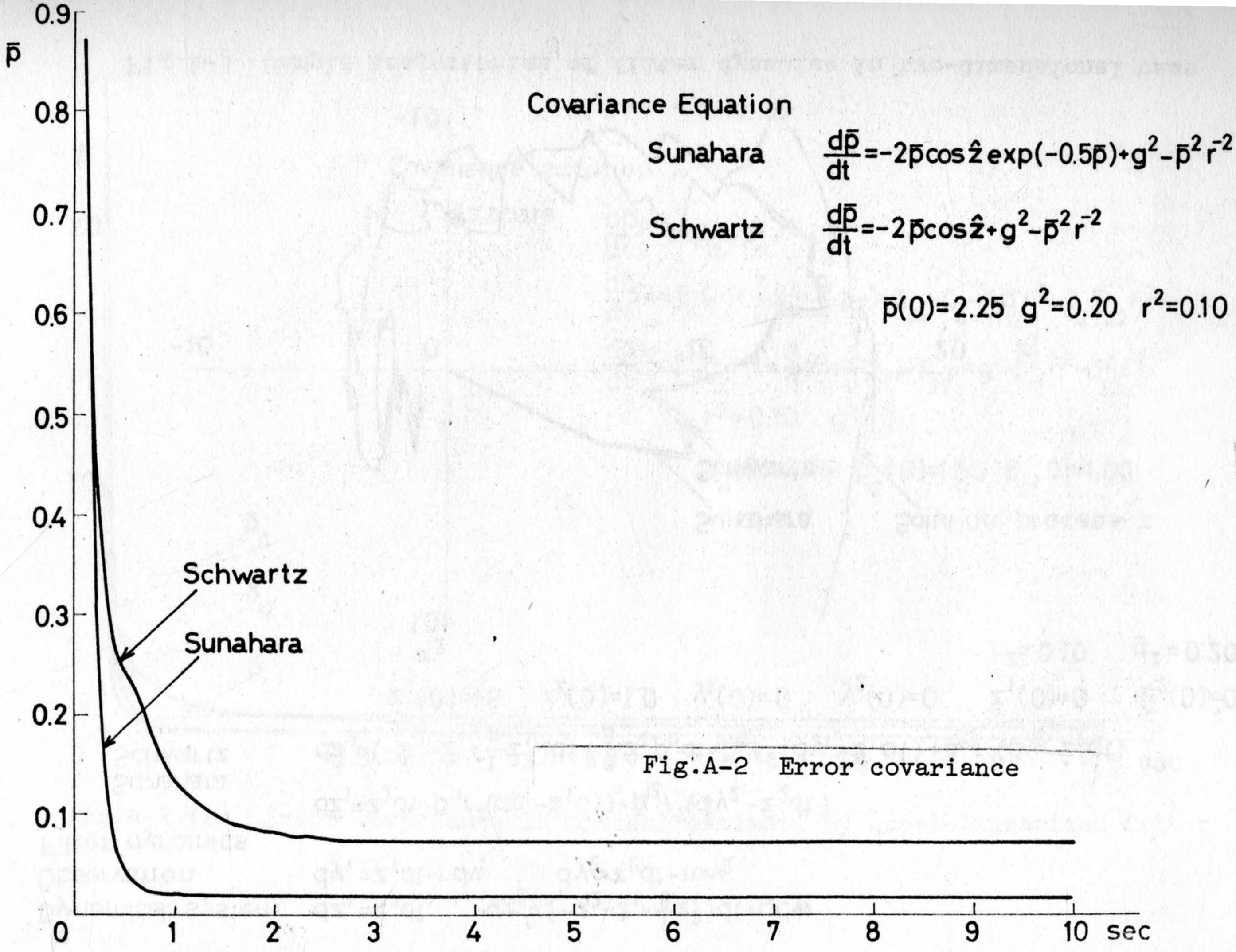



Fig.A-2 Error covariance

Dynamical system $dz_1 = z_2 dt \qquad dz_2 = (-z_2 - z_1 + \frac{1}{8} z_1^3)dt + g dw$

Observation $dy_1 = z_1 dt + r dv_1 \qquad dy_2 = z_2 dt + r dv_2$

Filter dynamics

Sunahara Schwartz

$$d\hat{z}_1 = \hat{z}_2 dt + \bar{p}_{11} r^{-2}(dy_1 - \hat{z}_1 dt) + \bar{p}_{12} r^{-2}(dy_2 - \hat{z}_2 dt)$$

$$d\hat{z}_2 = (-\hat{z}_2 - \hat{z}_1 + \frac{1}{8}\hat{z}_1^3)dt + \frac{3}{8}\hat{z}_1 \bar{p}_{11} dt + \bar{p}_{12} r^{-2}(dy_1 - \hat{z}_1 dt) + \bar{p}_{22} r^{-2}(dy_2 - \hat{z}_2 dt)$$

$z_1(0)=1.5 \quad z_2(0)=1.0 \quad y_1(0)=0 \quad y_2(0)=0 \quad \hat{z}_1(0)=0 \quad \hat{z}_2(0)=0$

$r^2 = 0.10 \quad g^2 = 0.20$

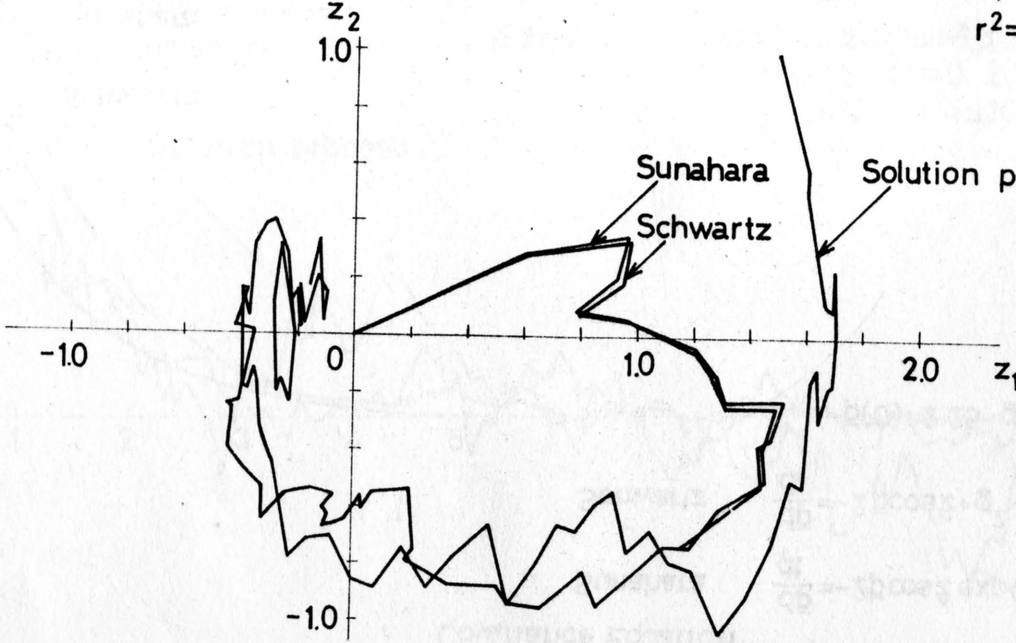



Fig.A-3 Sample trajectories of filter dynamics in two-dimensional case

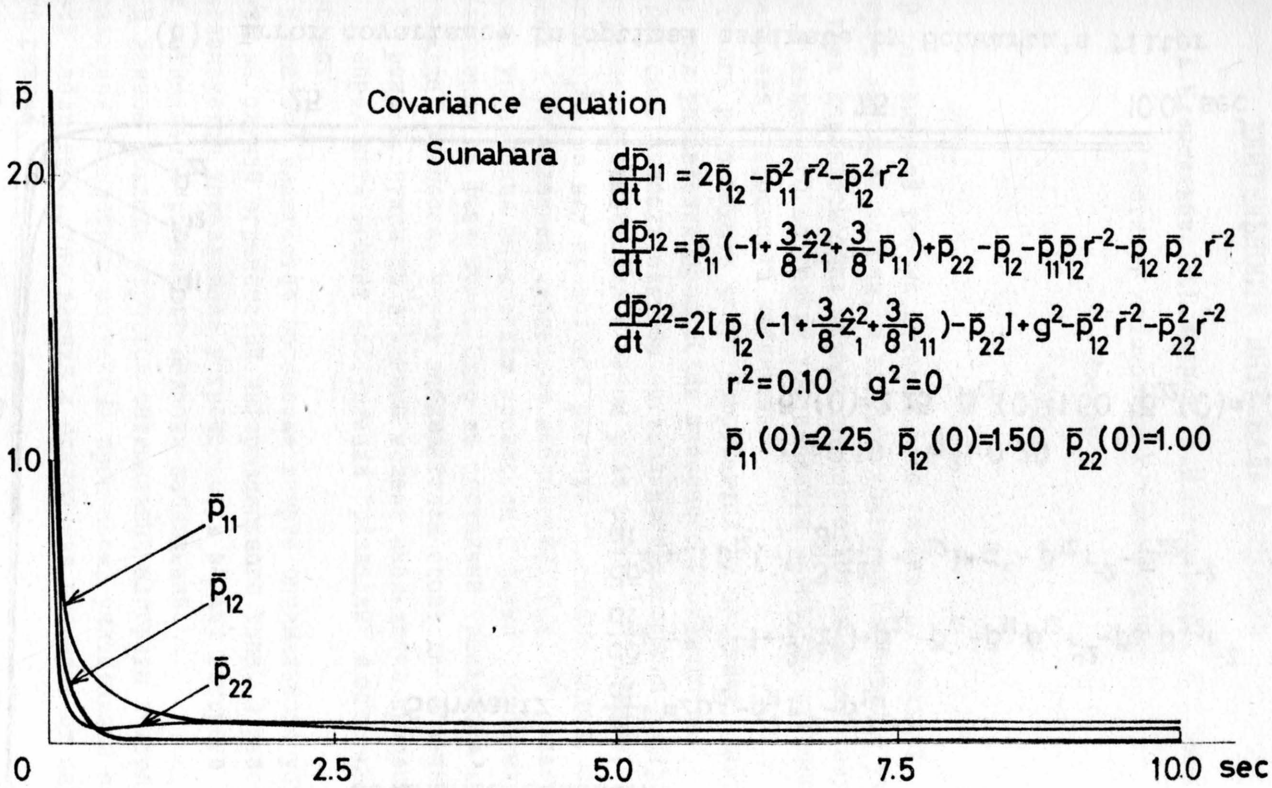
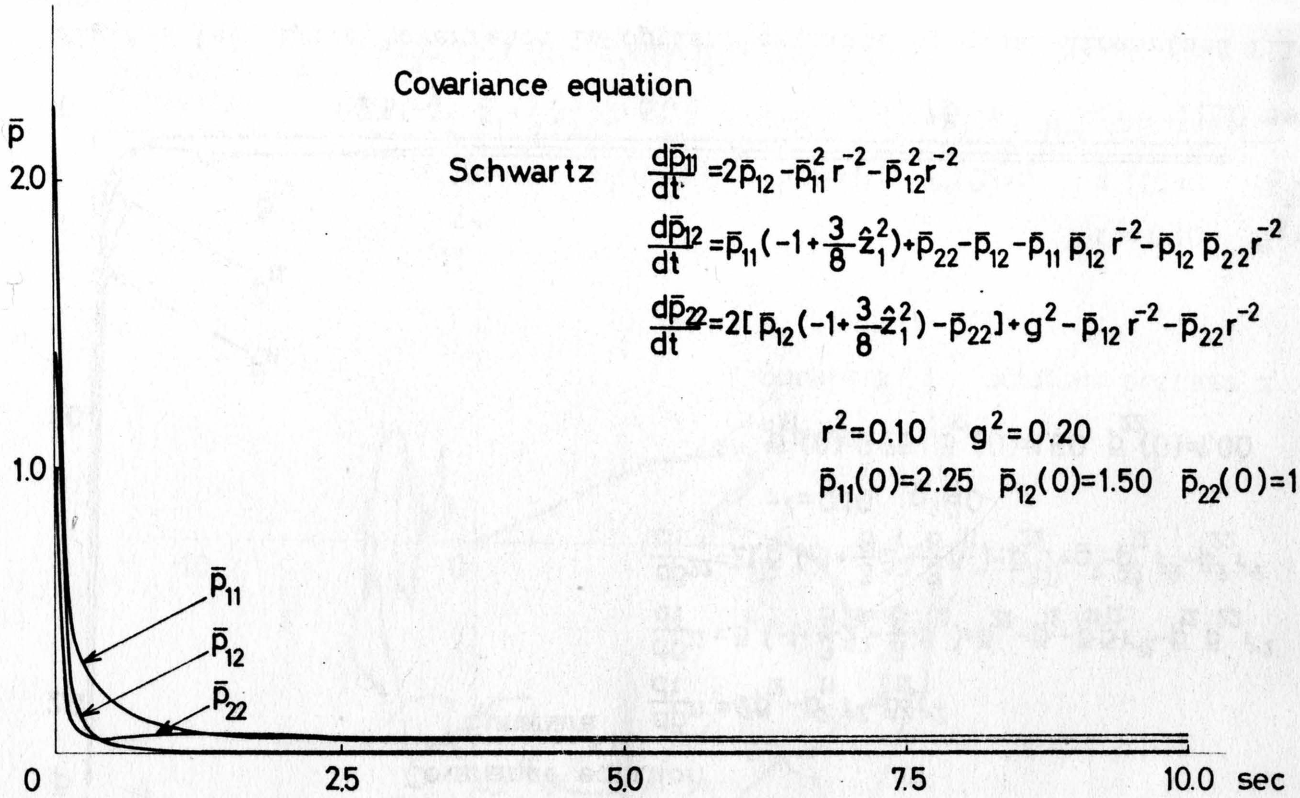



Fig.A-4 (a) Error covariance in optimal estimate by quasi-linearized filter




(b) Error covariance in optimal estimate by Schwartz's filter

# УПРАВЛЕНИЕ СТОХАСТИЧЕСКИМИ ПРОЦЕССАМИ ПРИ РЕГУЛИРУЕМОЙ ДЛИТЕЛЬНОСТИ ИНТЕРВАЛА КОНТРОЛЯ

А.А.Клементьев, Е.П.Маслов, А.М.Петровский, А.И.Яшин

Институт автоматики и телемеханики

Москва

СССР

В современной теории оптимального управления большой интерес вызывают задачи, в которых критерий строится с учетом трех характеристик: штрафа за несоответствие действительного и желаемого состояния управляемого процесса, стоимости управления и стоимости наблюдения.

Такого рода задачи часто встречаются на практике, например, при управлении процессами массового производства.

Целесообразность синтеза оптимального алгоритма контроля и управления отмечается в целом ряде работ [I-4]. Однако, лишь в некоторых из них приводится решение отдельных задач, связанных с управлением процессом наблюдений. Так, в[4] решается задача управления марковским процессом с двумя состояниями, наблюдаемым на фоне шума. Одно из состояний трактуется как "разладка" процесса. Перед наблюдателем стоят две задачи: I/ решить, нужны ли в данный момент наблюдения процесса, 2/ определить момент наступления "разладки" и остановить процесс.

В статье[5] излагается обобщение теории оптимального управления на случай ограничений информационного типа. В работе синтезируется управляющее устройство, в задачу которого входит оптимальное разбиение области наблюдений.

В данном докладе строится оптимальный алгоритм контроля и управления дискретным случайным процессом, учитывающий стоимости наблюдения, управления и отклонения процесса от заданного режима.

Работа опирается на методологию дуального управления [2]. Ниже приводится постановка задачи и дается ее решение для случая, когда об управляемом процессе имеется неполная информация, а помехи наблюдению отсутствуют.

## I. Постановка задачи

I. Изучается дискретно-непрерывная система. Все величины, фигурирующие в системе, определены лишь в дискретные моменты времени $n = 0, 1, 2, \ldots$ . Значение любой из величин в произвольный момент времени $t = n$ снабжается индексом $n$ . Число тактов управления процессом конечно и равно $N$ .

2. Решается байесовская задача, - априорные плотности всех случайных величин считаются известными. Погрешности контроля и управления полагаются равными нулю.

3. Статистические свойства управляемого одномерного процесса $\{\eta_n\}$ считаются известными с точностью до случайного вектора $\lambda$ параметров.

4. Контроль за управляемым процессом сводится к наблюдению его координаты $\eta_n$, $n = 0, 1, 2, \ldots$ .

Для простоты также предположим, что процесс $\{\eta_n\}$ должен отслеживать некоторый детерминированный заданный процесс $\{\theta_n\}$.

Физически процедура контроля и управления процессом $\{\eta_n\}$ выглядит следующим образом. К концу произвольного $(n-1)$ -го такта управляющее устройство, обладая некоторой информацией о ходе процесса принимает одно из двух решений:

I. Не контролировать процесс в момент времени $t = n-1$ и строить оптимальное управление по прошлой информации.

2. Осуществить процедуру контроля. В этом случае наблюдается координата $\eta_{n-1}$ , а оптимальное управление строится с учетом этого наблюдения и предыдущей информации.

Введем случайную величину:

$$x_n = \begin{cases} 1, & \text{если принимается решение о контроле процесса в момент времени } t = n-1, \\ 0, & \text{в противном случае,} \end{cases}$$

и примем следующую систему обозначений:

$\eta_n$ - координата управляемого процесса в момент времени $t=n$ ,

$u_n$ - управление на $n$ -ом такте;

$y_n$ - результат наблюдения в начале $n$ -го такта /в момент времени $t=n-1$ /.

Таким образом, каждый такт характеризуется четырьмя параметрами. Три из них, - решение $x_n$ о контроле, наблюдение $y_n$ и управление $u_n$ , - относятся к началу такта, а четвертый параметр, - координата $\eta_n$ процесса, - к концу такта. Будем записывать со стрелкой наверху последовательность величин, поступивших на вход системы, например, $\vec{\eta}_n=(\eta_1,\eta_2,...,\eta_n)$ .

Рассмотрим подробнее процесс наблюдений. Из дальнейшего станет ясно, что данная задача решается методом динамического программирования. В соответствии с этим методом [2,4,6], минимизация функционала осуществляется "попятным" движением от последнего такта к первому. Поэтому решение $x_n$ о контроле и управление $u_n$ на $n$ -ом такте, опираясь на информацию о предыдущем ходе процесса, по сути дела, зависят от еще не выбранной совокупности ( $\vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}$) . Лучшее, что можно в таком случае найти, - это построить некоторую зависимость $x_n$ и $u_n$ от этой совокупности в общем виде. Так как при выборе вектора ( $x_n, u_n$) заранее не известно, сколько раз в течение предыдущих тактов процесс $\{\eta_k\}$ контролировался, то для построения в общем виде алгоритма контроля и управления в данной работе предлагается некоторая формализованная схема синтеза последовательности наблюдений. Суть ее состоит в следующем.

Пусть принято решение произвести контроль координаты $\eta_{n-1}$, т.е. $x_n = 1$ . В этом случае результат наблюдения $y_n$ совпадает с координатой процесса $\eta_{n-1}$: $y_n = \eta_{n-1}$ . Решение $x_n = 0$ приводит к отсутствию контроля координаты $\eta_{n-1}$ . Но факт <u>непоступления</u> информации о процессе $\{\eta_n\}$ , - с точки зрения накопления информации о $\{\eta_n\}$ , - эквивалентен <u>поступлению</u> информации о некотором гипотетическом процессе, никак не связанном с $\{\eta_n\}$ . В частности, в данной работе будем полагать, что этим гипотетическим процессом служит некоторая случайная последовательность $\{\varepsilon_n\}$, не зависимая от $\{\eta_n\}$. Пусть также $\{\varepsilon_n\}$ состоит из независимых величин.

Поэтому формально можно записать

$$y_n = \begin{cases} \eta_{n-1}, & \text{при } x_n = 1 \\ \varepsilon_n, & \text{при } x_n = 0, \quad n = 1, 2, \dots, N, \end{cases} \qquad /1/$$

или в виде одного выражения

$$y_n = x_n \eta_{n-1} + (1 - x_n)\varepsilon_n, \quad n = 1, 2, \dots, N. \qquad /2/$$

К концу $(n-1)$ -го такта управляющее устройство обладает информацией в виде последовательности решений $\vec{x}_{n-1} = (x_1, x_2, \dots, x_{n-1})$, последовательности управлений $\vec{u}_{n-1} = (u_1, \dots, u_{n-1})$ и последовательности наблюдений $\vec{y}_{n-1} = (y_1, y_2, \dots, y_{n-1})$, где под наблюдением $y_k$ понимается формальное равенство $y_k = x_k \eta_{k-1} + (1 - x_k)\varepsilon_k$. Кроме того, управляющему устройству известна последовательность $\{\vec{\theta}_m\}$ значений эталонного детерминированного процесса для любого $m$.

На основе этой информации управляющее устройство принимает решение

$$x_n = x_n(\vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m) \qquad /3/$$

о контроле координаты $\eta_{n-1}$. Если устройство приняло решение $x_n = 0$, то координата $\eta_{n-1}$ не контролируется, $y_n = \eta_{n-1}$. Если же принимается решение $x_n = 1$, то осуществляется процедура контроля, $y_n = \eta_{n-1}$. В обоих случаях далее находится оптимальное управление

$$u_n = u_n(\vec{x}_n, \vec{y}_n, \vec{u}_{n-1}, \vec{\theta}_m) \qquad /4/$$

На следующем такте все повторяется вновь.

Критерием оптимальности в данной работе служит критерий минимума полного риска. Он образуется следующим образом.

На каждом такте существования процесса $\{\eta_n\}$ определяются три типа возможных потерь:

1. Потери, связанные с отклонением координаты $\eta_n$ от заданного режима $\theta_n$.
2. Потери, связанные с управлением $u_n$.
3. Потери, связанные с контролем координаты $\eta_{n-1}$.

Потери на отклонение координаты $\eta_n$ от $\theta_n$ определяются некоторой функцией потерь, зависящей от $\eta_n, \theta_n$ и, в общем случае, также от номера такта $n$ :

$$C_{1n} = C_1(n, \theta_n, \eta_n), \quad n = 1, 2, \dots, N. \qquad /5/$$

Потери на управление на $n$ -ом такте определяются функцией потерь

$$C_{2n} = C_2(n, u_n), \quad n = 1, 2, \dots, N; \qquad /6/$$

и с учетом случайной величины $x_n$ потери на контроль - функцией потерь

$$C_{3n} = x_n C_3(n), \quad n = 1, 2, \dots, N. \qquad /7/$$

Следуя терминологии[2], удельной функцией потерь на $n$ -ом такте назовем выражение вида

$$C_n = C_1(n, \theta_n, \eta_n) + C_2(n, u_n) + x_n C_3(n), \quad n = 1, 2, \dots, N. \qquad /8/$$

/В принципе, функция $C_n$ может зависеть от отдельных потерь и не аддитивно/.

Общей функцией потерь назовем выражение

$$C_\Sigma = \sum_{n=1}^{N} [C_1(n, \theta_n, \eta_n) + C_2(n, u_n) + x_n C_3(n)] \qquad /9/$$

Оптимальной считается такая процедура контроля и управления, для которой полный риск /математическое ожидание величины $C_\Sigma$ /

$$R_\Sigma = M\{C_\Sigma\} = M\{C_1(1, \theta_1, \eta_1) + C_2(1, u_1) + x_1 C_3(1)\} +$$

$$+ \dots + M\{C_1(N, \theta_N, \eta_N) + C_2(N, u_N) + x_N C_3(N)\} \qquad /10/$$

минимален.

Требуется построить последовательность правил решения

$$\Gamma_n = \Gamma_n^x(x_n | \vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m) \cdot \Gamma_n^u(u_n | \vec{x}_n, \vec{y}_n, \vec{u}_{n-1}, \vec{\theta}_m) \qquad /11/$$

и, соответственно, пар управлений $(x_n, u_n)$ , минимизирующих полный риск $R_\Sigma$. Если обозначить минимальный полный риск $R_\Sigma^*$, то данная задача сводится к поиску

$$R_\Sigma^* = \min_{\vec{x}_N, \vec{u}_N} R_\Sigma \qquad /12/$$

## 2. Вывод основных соотношений

Запишем выражение для удельного риска $R_n$, понимая под этим риск на $n$ -ом такте

$$R_n = \int\limits_{\Omega(x_n, u_n, \eta_n)} [C_1(n, \eta_n, \theta_n) + C_2(n, u_n) + x_n C_3(n)] \cdot P(x_n, \eta_n, u_n)\, d\Omega \qquad /13/$$

Здесь и далее $\Omega(\cdot)$ означает область совместного изменения величин, стоящих в скобках, а $d\Omega$ - ее бесконечно малый элемент. Условимся также, что функции $P(\cdot)$, имеющие разные аргументы, представляют собою, в общем случае, различные функции, несмотря на то, что они обозначены одной и той же буквой. Функции $P(\cdot)$ суть совместные /условные или безусловные/ плотности вероятности случайных величин. Плотность

$$P(x_n, \eta_n, u_n) = \int\limits_{\Omega(\bar{\lambda}, \vec{y}_n, \vec{x}_{n-1}, \vec{u}_{n-1})} P(\vec{x}_n, \vec{u}_n, \bar{\lambda}, \eta_n, \vec{y}_n)\, d\Omega \qquad /14/$$

Воспользуемся методикой дуального управления[2] и представим многомерную плотность $P(\vec{x}_n, \vec{u}_n, \bar{\lambda}, \eta_n, \vec{y}_n)$ в виде произведения ряда одномерных плотностей. В результате выражение для удельного риска $R_n$ примет следующий вид:

$$R_n = \int\limits_{\Omega(\bar{\lambda}, \vec{y}_n, \vec{x}_n, \vec{u}_n, \eta_n)} [C_1(n, \eta_n, \theta_n) + C_2(n, u_n) + x_n C_3(n)] \cdot P(\bar{\lambda}) \cdot P(\eta_n | \bar{\lambda}, \vec{x}_n, \vec{y}_n, \vec{u}_n) \cdot$$

$$\cdot \prod_{i=1}^{n} P(y_i | \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) \cdot \prod_{i=1}^{n} [\Gamma_i^x(x_i | \vec{x}_{i-1}, \vec{u}_{i-1}, \vec{y}_{i-1}, \vec{\theta}_m) \cdot$$

$$\cdot \Gamma_i^u(u_i | \vec{x}_i, \vec{y}_i, \vec{u}_{i-1}, \vec{\theta}_m)]\, d\Omega, \quad n = 1, 2, \ldots, N. \qquad /15/$$

Полный риск

$$R_\Sigma = \sum_{n=1}^{N} R_n \qquad /16/$$

Задача выбора $2N$ -мерного вектора $(\vec{x}_N, \vec{u}_N)$, минимизирующего полный риск $R_\Sigma$ решается методом динамического программирования.

Вначале определяется последняя пара $(x_N, u_N)$. В выражении /I6/ от $(x_N, u_N)$ зависит лишь последнее слагаемое $R_N$. Следовательно, оптимальная пара $(x_N^*, u_N^*)$ определяется из условия минимума $R_N$. Пусть $2(N-1)$ -мерный вектор $(\vec{x}_{N-1}, \vec{u}_{N-1})$ задан. При $n=N$

$$R_N = \int\limits_{\Omega(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^x \cdot \Gamma_i^u) \Big\{ \int\limits_{\Omega(\eta_N, \bar{\lambda}, x_N, u_N, y_N)} [C_1(N, \eta_N, \theta_N) + C_2(N, u_N) + x_N C_3(N)] \cdot$$
$$\cdot P(\bar{\lambda}) \cdot P(\eta_N | \bar{\lambda}, \vec{x}_N, \vec{y}_N, \vec{u}_N) \cdot \prod_{i=1}^{N} P(y_i | \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) \cdot$$
$$\cdot (\Gamma_N^x \cdot \Gamma_N^u) d\Omega \qquad /I7/$$

где

$$\Gamma_i^x \triangleq \Gamma_i^x(x_i | \vec{x}_{i-1}, \vec{y}_{i-1}, \vec{u}_{i-1}, \vec{\theta}_m), \quad \Gamma_i^u \triangleq \Gamma_i^u(u_i | \vec{x}_i, \vec{y}_i, \vec{u}_{i-1}, \vec{\theta}_m)$$

На каждом такте управляющее устройство принимает одно из двух решений: $x_n = 0$ или $x_n = 1$. Поэтому при $n = N$ решающее правило

$$\Gamma_N^x(x_N | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) = \gamma(x_N = 1 | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m)\, \delta(x_N - 1) +$$
$$+ \gamma(x_N = 0 | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m)\, \delta(x_N - 0), \qquad /I8/$$

где $\gamma(x_N = j | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m)$, $j = 0, 1,$ – вероятности принять соответствующие решения при наличии информации $(\vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m)$; $\delta(x_N - j)$ -дельта-функции.

Подставляя равенство /I8/ в /I7/ и интегрируя по $x_N$, находим

$$R_N = \int\limits_{\Omega(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^x \Gamma_i^u) \cdot \Big\{ \gamma(x_N = 1 | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) \int\limits_{\Omega(\eta_N, \bar{\lambda}, y_N, u_N)} [C_1(N, \theta_N, \eta_N) +$$
$$+ C_2(N, u_N) + C_3(N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N | \bar{\lambda}, \vec{y}_N, \vec{x}_{N-1}, x_N = 1, \vec{u}_N) \cdot \prod_{i=1}^{N} P(y_i | \bar{\lambda}, \vec{x}_i,$$
$$\vec{y}_{i-1}, \vec{u}_{i-1}) \cdot \Gamma_N^u(u_N | \vec{x}_{N-1}, x_N = 1, \vec{y}_N, \vec{u}_{N-1}, \vec{\theta}_m) d\Omega +$$
$$+ \gamma(x_N = 0 | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) \cdot \int\limits_{\Omega(\eta_N, \bar{\lambda}, y_N, u_N)} [C_1(N, \theta_N, \eta_N) + C_2(N, u_N)] \cdot P(\bar{\lambda}) \cdot$$
$$\cdot P(\eta_N | \bar{\lambda}, \vec{y}_N, \vec{x}_{N-1}, x_N = 0, \vec{u}_N) \cdot \prod_{i=1}^{N} P(y_i | \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) \cdot \qquad /I9/$$
$$\cdot \Gamma_N^u(u_N | \vec{x}_{N-1}, x_N = 0, \vec{y}_N, \vec{u}_{N-1}, \vec{\theta}_m) d\Omega \Big\} d\Omega$$

Рассмотрим подробнее формулу /I9/. Из /2/ следует, что при $x_N = 1$ : $y_N = \eta_{N-1}$ . Поэтому

$$P(\eta_N | \bar{\lambda}, \vec{y}_N, \vec{x}_{N-1}, x_N=1, \vec{u}_N) = P(\eta_N | \bar{\lambda}, \vec{y}_{N-1}, \eta_{N-1}, \vec{x}_{N-1}, x_N=1, \vec{u}_N) \qquad /20/$$

$$P(y_N | \bar{\lambda}, \vec{y}_{N-1}, \vec{x}_{N-1}, x_N=1, \vec{u}_{N-1}) = P(\eta_{N-1} | \bar{\lambda}, \vec{x}_{N-1}, x_N=1, \vec{y}_{N-1}, \vec{u}_{N-1}) \qquad /2I/$$

Далее, при $x_N = 0$ : $y_N = \varepsilon_N$ . В силу независимости процессов $\{\eta_n\}$ и $\{\varepsilon_n\}$:

$$P(\eta_N | \bar{\lambda}, \vec{y}_N, \vec{x}_{N-1}, x_N=0, \vec{u}_N) = P(\eta_N | \bar{\lambda}, \vec{y}_{N-1}, \varepsilon_N, \vec{x}_{N-1}, x_N=0, \vec{u}_N) =$$
$$= P(\eta_N | \bar{\lambda}, \vec{y}_{N-1}, \vec{x}_{N-1}, x_N=0, \vec{u}_N) \qquad /22/$$

$$P(y_N | \bar{\lambda}, \vec{y}_{N-1}, \vec{x}_{N-1}, x_N=0, \vec{u}_{N-1}) = P(\varepsilon_N) \qquad /23/$$

где $P(\varepsilon_N)$ - априорная плотность вероятности случайной величины $\varepsilon_N$ .

Выше отмечалось, что в том случае, когда принимается решение не контролировать координату $\eta_{n-1}$ , управление $u_n$ определяется прошлой информацией. Это означает, что решающее правило

$$\Gamma_N^u (u_N | \vec{x}_{N-1}, x_N=0, \vec{y}_N, \vec{u}_{N-1}, \vec{\theta}_m) = \Gamma_N^u (u_N | \vec{x}_{N-1}, x_N=0, \vec{y}_{N-1}, \varepsilon_N, \vec{u}_{N-1}, \vec{\theta}_m) =$$
$$= \Gamma_N^u (u_N | \vec{x}_{N-1}, x_N=0, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) \qquad /24/$$

Таким образом, во втором слагаемом формулы /I9/ от $\varepsilon_N = y_N$ зависит лишь плотность $P(\varepsilon_N)$ . Подставим выражения /20/÷/24/ в /I9/, проинтегрируем по $\varepsilon_N$ и введем функции

$$\alpha_N = \int\limits_{\Omega(\eta_N, \bar{\lambda})} [C_1(N, \theta_N, \eta_N) + C_2(N, u_N) + C_3(N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N | \bar{\lambda}, \vec{y}_{N-1},$$

$$\eta_{N-1}, \vec{x}_{N-1}, x_N=1, \vec{u}_N) \cdot P(\eta_{N-1} | \bar{\lambda}, \vec{x}_{N-1}, x_N=1, \vec{y}_{N-1}, \vec{u}_{N-1}) \cdot \qquad /25/$$
$$\cdot \prod_{i=1}^{N-1} P(y_i | \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) \, d\Omega$$

$$\beta_N = \int\limits_{\Omega(\eta_N, \bar{\lambda})} [C_1(N, \theta_N, \eta_N) + C_2(N, u_N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N | \bar{\lambda}, \vec{y}_{N-1}, \vec{x}_{N-1}, x_N = 0, \vec{u}_N) \cdot \prod_{i=1}^{N-1} P(y_i | \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) d\Omega \qquad /26/$$

С учетом этих обозначений выражение для риска $R_N$ приобретает следующий вид:

$$R_N = \int\limits_{\Omega(\vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^x \cdot \Gamma_i^u) \cdot \Phi_N (\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) d\Omega \qquad /27/$$

где

$$\Phi_N \triangleq \Phi_N (\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) = \gamma(x_N = 1 | \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot$$
$$\cdot \int\limits_{\Omega(\eta_{N-1}, u_N)} \alpha_N \cdot \Gamma_N^u (u_N | \vec{u}_{N-1}, \vec{x}_{N-1}, x_N = 1, \vec{y}_{N-1}, \eta_{N-1}, \vec{\theta}_m) d\Omega +$$
$$+ \gamma(x_N = 0 | \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \int\limits_{\Omega(u_N)} \beta_N \cdot \Gamma_N^u (u_N | \vec{u}_{N-1}, \vec{x}_{N-1}, x_N = 0, \vec{y}_{N-1}, \vec{\theta}_m) d\Omega \qquad /28/$$

Оптимизация $R_N$ относительно $(x_N, u_N)$ сводится к оптимизации $\Phi_N$. Начнем с выбора оптимальных управлений $u_N$. Наложим следующее ограничение. Будем искать управляющее устройство в классе систем, обладающих относительно $u_n$, $n = 1, 2, ..., N$, регулярной стратегией. Обозначим через $u_n^{0*}$ – оптимальное управление, отвечающее решению $x_n = 0$, а через $u_n^{1*}$ – оптимальное управление, отвечающее решению $x_n = 1$, $n = 1, 2, ..., N$. Тогда при $n = N$

$$\left.\begin{aligned} &\Gamma_N^u (u_N | \vec{u}_{N-1}, \vec{x}_{N-1}, x_N = 1, \vec{y}_{N-1}, \eta_{N-1}, \vec{\theta}_m) = \delta(u_N - u_N^{1*}), \\ &\Gamma_N^u (u_N | \vec{u}_{N-1}, \vec{x}_{N-1}, x_N = 0, \vec{y}_{N-1}, \vec{\theta}_m) = \delta(u_N - u_N^{0*}). \end{aligned}\right\} \qquad /29/$$

После подстановки /29/ в /28/ и интегрирования по $u_N$ в формуле для $\Phi_N$ от оптимального управления $u_N^{1*}$ оказывается зависящей функция $\alpha_N$, а от оптимального управления $u_N^{0*}$ – функция $\beta_N$; следовательно, оптимальным управлением $u_N^{1*}$ служит значение $u_N$, доставляющее минимум $\alpha_N$, а оптимальным управлением $u_N^{0*}$ – значение $u_N$, минимизирующее $\beta_N$.

Пусть

$$\alpha_N^* = \min_{u_N^* \in \Omega(u_N)} \alpha_N \; ; \qquad \beta_N^* = \min_{u_N^* \in \Omega(u_N)} \beta_N \qquad /30/$$

где $\Omega(u_N)$ - область допустимых управлений на $N$ -ом такте. Подставляя /30/ в /28/, находим

$$\Phi_N^* = \min_{u_N} \Phi_N = \gamma(x_N = 1 | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) \int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega + \gamma(x_N = 0 | \vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) \beta_N^* \qquad /31/$$

Пусть управляющее устройство обладает относительно решения $x_n$ также регулярной стратегией: из $\gamma(x_n = 1 | \vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m) = 1$ с вероятностью единица следует $\gamma(x_n = 0 | \vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m) = 0$, и наоборот.

Тогда оптимизация $\Phi_N^*$ относительно $x_N$ сводится к сравнению функций $\int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega$ и $\beta_N^*$ и выбору меньшей из них. При

$$\int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega > \beta_N^* \qquad /32/$$

принимается решение $x_N = 0$, а при

$$\int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega < \beta_N^* \; - \qquad /33/$$

решение $x_N = 1$ . При равенстве этих функций выбор решения $x_N$ произволен. В результате двойной минимизации образуется функция

$$\Phi_N^{**} = \min_{x_N, u_N} \Phi_N = \min \left[ \beta_N^*, \int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega \right] \qquad /34/$$

Переходим к определению предпоследней пары $(x_{N-1}, u_{N-1})$ . Прежде всего отметим, что по аналогии с формулами /25/,/26/, /28/ функции $\alpha_n$ , $\beta_n$ , $\Phi_n$ могут быть построены для любого такта, $n = 1, 2, \ldots, N$ . Для этого достаточно в соответствующих местах заменить индекс $N$ на $n$ . Пусть $2(N-2)$ -мерный вектор $(\vec{x}_{N-2}, \vec{u}_{N-2})$ выбран. В формуле /16/ от вектора $(x_{N-1}, u_{N-1})$ зависит сумма $S_{N-1} = R_{N-1} + R_N^*$, где $R_N^* = \min\limits_{x_N, u_N} R_N$. Эта сумма равна

$$S_{N-1} = \int\limits_{\Omega(\vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2})} \prod_{i=1}^{N-2} (\Gamma_i^x \cdot \Gamma_i^u) \cdot \Phi_{N-1}(\vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m)\, d\Omega + \int\limits_{\Omega(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^x \cdot$$

$$\cdot \Gamma_i^u)\, \Phi_N^{**}(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m)\, d\Omega =$$

$$= \int\limits_{\Omega(\vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2})} \prod_{i=1}^{N-2} (\Gamma_i^x \cdot \Gamma_i^u) \cdot F_{N-1}(\vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m)\, d\Omega \qquad /35/$$

где с учетом условий /20/-/24/ и обозначений /25/, /26/,/28/ при $n = N-1$:

$$F_{N-1} \triangleq F_{N-1}(\vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m) = \gamma(x_{N-1} = 1 \mid \vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m) \cdot$$

$$\cdot \Big\{ \int\limits_{\Omega(\eta_{N-2}, u_{N-1})} [\alpha_{N-1} + \Phi_N^{**}(\vec{x}_{N-2}, x_{N-1} = 1, \vec{y}_{N-2}, \eta_{N-2}, \vec{u}_{N-1}, \vec{\theta}_m)] \cdot$$

$$\cdot \Gamma_{N-1}^u (u_{N-1} \mid \vec{x}_{N-2}, x_{N-1} = 1, \vec{y}_{N-2}, \eta_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m)\, d\Omega \Big\} +$$

$$+ \gamma(x_{N-1} = 0 \mid \vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m) \cdot \Big\{ \int\limits_{\Omega(u_{N-1})} [\beta_{N-1} + \int\limits_{\Omega(\varepsilon_{N-1})} \Phi_N^{**}(\vec{x}_{N-2}, x_{N-1} = 0,$$

$$\vec{y}_{N-2}, \varepsilon_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m)\, d\Omega] \cdot \Gamma_{N-1}^u (u_{N-1} \mid \vec{x}_{N-2}, x_{N-1} = 1, \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m) \Big\} d\Omega \qquad /36/$$

В выражении /36/ от вектора $(x_{N-1}, u_{N-1})$ зависит лишь функция $F_{N-1}$. Поэтому достаточно рассмотреть лишь ее минимизацию относительно $(x_{N-1}, u_{N-1})$. Введем функции

$$\mathcal{H}_{N-1} = \alpha_{N-1} + \Phi_N^{**}(\vec{x}_{N-2}, x_{N-1} = 1, \vec{y}_{N-2}, \eta_{N-2}, \vec{u}_{N-1}, \vec{\theta}_m) \qquad /37/$$

$$\varphi_{N-1} = \beta_{N-1} + \int\limits_{\Omega(\varepsilon_{N-1})} \Phi_N^{**}(\vec{x}_{N-2}, x_{N-1} = 0, \vec{y}_{N-2}, \varepsilon_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m)\, d\Omega \qquad /38/$$

Из формул /36/-/37/ видно, что оптимальное управление $u_{N-1}^{1*}$ находится из условия минимума $\mathcal{H}_{N-1}$ относительно $u_{N-1}$, а оптимальное управление $u_{N-1}^{0*}$ - из условия минимума $\varphi_{N-1}$ относительно $u_{N-1}$. Пусть

$$\mathcal{H}_{N-1}^* = \min_{u_{N-1}^* \in \Omega(u_{N-1})} \mathcal{H}_{N-1} \quad ; \quad \varphi_{N-1}^* = \min_{u_{N-1}^* \in \Omega(u_{N-1})} \varphi_{N-1} \qquad /39/$$

Тогда

$$F_{N-1}^* = \min_{u_{N-1}} F_{N-1} = \gamma(x_{N-1}=1 \mid \vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m) \cdot \int\limits_{\Omega(\eta_{N-2})} \mathcal{H}_{N-1}^* d\Omega +$$

$$+ \gamma(x_{N-1}=0 \mid \vec{x}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m) \cdot \varphi_{N-1}^* \qquad /40/$$

Оптимизация $F_{N-1}^*$ относительно $x_{N-1}$ сводится к сравнению функций $\varphi_{N-1}^*$, $\int\limits_{\Omega(\eta_{N-2})} \mathcal{H}_{N-1}^* d\Omega$ и выбору меньшей из них. При $\int\limits_{\Omega(\eta_{N-2})} \mathcal{H}_{N-1}^* d\Omega > \varphi_{N-1}^*$ принимается решение $x_{N-1} = 0$, а при $\int\limits_{\Omega(\eta_{N-2})} \mathcal{H}_{N-1}^* d\Omega < \varphi_{N-1}^*$ – решение $x_{N-1} = 1$. При равенстве этих функций выбор решения $x_{N-1}$ произволен. В результате оптимизации получается функция $F_{N-1}^{**} = \min\limits_{x_{N-1}, u_{N-1}} F_{N-1}$.

Рассуждая совершенно аналогично, нетрудно показать, что для нахождения оптимальной пары $(x_{N-k}^*, u_{N-k}^*)$, $k = 1, 2, \ldots, N-1$, необходимо совершить следующие действия.

1. Построить пару функций

$$\mathcal{H}_{N-k} = \alpha_{N-k} + F_{N-k+1}^{**}(\vec{x}_{N-k-1}, x_{N-k}=1, \vec{y}_{N-k-1}, \eta_{N-k-1}, \vec{u}_{N-k}, \vec{\theta}_m);$$

$$\varphi_{N-k} = \beta_{N-k} + \int\limits_{\Omega(\varepsilon_{N-k})} F_{N-k+1}^{**}(\vec{x}_{N-k-1}, x_{N-k}=0, \vec{y}_{N-k-1}, \varepsilon_{N-k}, \vec{u}_{N-k}, \vec{\theta}_m) d\Omega .$$

2. Минимизируя $\mathcal{H}_{N-k}$ и $\varphi_{N-k}$ по $u_{N-k}$, найти $u_{N-k}^{1*}$ и $u_{N-k}^{0*}$, соответственно.

3. Вычислить минимальные значения

$$\mathcal{H}_{N-k}^* = \min_{u_{N-k}^* \in \Omega(u_{N-k})} \mathcal{H}_{N-k} \; ; \quad \varphi_{N-k}^* = \min_{u_{N-k}^* \in \Omega(u_{N-k})} \varphi_{N-k}$$

4. Сравнить $\varphi_{N-k}^*$ и $\int\limits_{\Omega(\eta_{N-k-1})} \mathcal{H}_{N-k}^* d\Omega$ и выбрать меньшую функцию. При

$$\int\limits_{\Omega(\eta_{N-k-1})} \mathcal{H}_{N-k}^* d\Omega > \varphi_{N-k}^*$$

принять решение $x_{N-k} = 0$, а при

$$\int\limits_{\Omega(\eta_{N-k-1})} \mathcal{H}_{N-k}^* d\Omega < \varphi_{N-k}^*$$

решение $x_{N-k} = 1$. При равенстве функций выбор решения $x_{N-k}$ произволен.

Пример. Пусть $\{\eta_n\}$ - случайный процесс, в отсутствие управления отвечающий формуле $\eta_n = \lambda n + h_n$, где $\{h_n\}$ - последовательность независимых нормальных величин со статистиками $(0, \sigma^2)$, а $\lambda$ - неизвестный случайный параметр, распределенный по нормальному закону $P(\lambda) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\{-(\lambda-\lambda_0)^2 / 2\sigma_1^2\}$

Пусть процесс $\{\eta_n\}$ начинается из случайного начального условия $\eta_0$, распределенного по нормальному закону со статистиками $(0, \sigma_2^2)$. Пусть, для простоты, рассматривается двухшаговый процесс, $N = 2$, а на первом шаге принимается решение $x_1 = 1$. Тогда $y_1 = \eta_0$. При наличии управлений образуется последовательность

$$\begin{aligned} \eta_1 &= u_1 + \lambda + \eta_0 + h_1, \\ \eta_2 &= u_2 + u_1 + 2\lambda + \eta_0 + h_2. \end{aligned} \qquad /\text{п.I}/$$

Каждому члену последовательности /п.I/ ставится в соответствие наблюдение

$$y_n = x_n \eta_{n-1} + (1 - x_n)\varepsilon_n$$

Пусть гипотетическая последовательность $\{\varepsilon_n\}$ распределена по нормальному закону со статистиками $(0, \sigma_\varepsilon^2)$. Выберем функции потерь

$$c_{1n} = \eta_n^2, \quad c_{2n} = u_n^2, \quad c_{3n} = 1, \quad n = 1, 2 \qquad /\text{п.2}/$$

и примем, что при $x_n = 0$: $u_n = 0$.

Пропуская промежуточные выкладки, приведем сразу окончательный результат. Оказывается, что при этой постановке возможны две оптимальные стратегии:

$$\text{I.} \left\{ x_1^* = 1,\ u_1^{1*} = -0{,}6\eta_0 - 0{,}8\lambda_0;\ x_2^* = 1,\ u_2^{1*} = -0{,}5(u_1^{1*} + \eta_0) - \right. \qquad /\text{п.3}/$$

$$\left. - \frac{\sigma^2\lambda_0 + \sigma_1^2(\eta_1 - \eta_0 - u_1^{1*})}{\sigma^2 + \sigma_1^2} \right\}$$

Минимальный полный риск при этой стратегии равен

$$R_{\Sigma 1}^* = 2 + 2\sigma^2 + 3\sigma_1^2 + \frac{2\sigma_1^2\sigma^2}{\sigma_1^2 + \sigma^2} + \frac{7}{5}\lambda_0^2 + \frac{3}{5}\sigma_2^2 \qquad /\text{п.4}/$$

2. $\{ x_1^* = 1,\ u_1^{1*} = -\frac{2}{3}\eta_0 - \lambda_0;\ x_2^* = 0,\ u_2^{0*} = 0 \}$ /п.5/

Минимальный полный риск в этом случае равен

$$R^*_{22} = 1 + 2\sigma^2 + 5\sigma_1^2 + 2\lambda_0^2 + \frac{2}{3}\sigma_2^2$$ /п.6/

Эффект адаптации в данной системе сводится к изучению неизвестного параметра $\lambda$ и приводит к уменьшению дисперсии его распределения. При $0 \le \sigma_1^2 \le 1/2$ оптимальна вторая стратегия, а при $\sigma_1^2 \ge 1/2$ - оптимальна первая стратегия; $\sigma^2 = \sigma_2^2 = \lambda_0^2 = 1$ .

## ЛИТЕРАТУРА


I. Пугачев В.С. Теория случайных функций и ее применение к задачам автоматического управления. Изд. 3-е, Физматгиз, I962.

2. Фельдбаум А.А. Основы теории оптимальных автоматических систем. Изд. 2-е, Изд-во "Наука", I966.

3. Цыпкин Я.З. Адаптация, обучение и самообучение в автоматических системах. Автоматика и телемеханика, I966, № I .

4. Стратонович Р.Л. Условные марковские процессы и их применение к теории оптимального управления. Изд-во МГУ, I966.

5. Стратонович Р.Л. Экстремальные задачи теории информации и динамическое программирование. Техническая кибернетика, I967, № 5 .

6. Ховард. Д. Динамическое программирование и марковские процессы. ИЛ, I960.

# SYNTHESIS OF CONTINUOUS-TIME STOCHASTIC CONTROL SYSTEMS

I.G. Cumming
Department of Electrical Engineering
University of Toronto
Toronto, Canada

## 1. Introduction

This paper discusses the design and synthesis of a controller for a physical random system, and in doing so, discusses one of the problems of applying modern continuous-time stochastic control theory to practice.

The steps in the design of a controller for a physical system are outlined in Fig. 1:

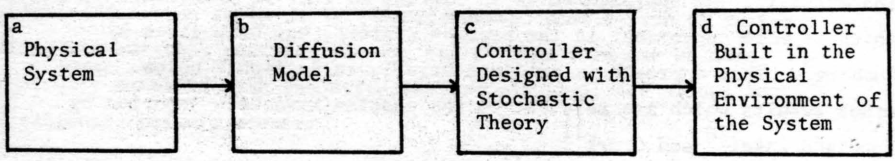



Fig. 1 Steps in the Design of a Controller

A physical system or process is defined to be a continuous-time random process described by ordinary differential equations whose right hand sides contain noise components with a high but finite upper frequency. As such, it is physically realizable, and is typical of many random processes found in practice.

A given physical system which is to be controlled is shown symbolically in Fig. 1a, and Fig. 1d represents the final objective, the system with its controller attached. To design the controller which optimizes some system performance criterion, continuous-time stochastic control theory[1-4] is used in Fig. 1c. However, as the development of the theory requires that the processes involved be Markov processes, the theory cannot be directly applied to the physical system of Fig. 1a, but must be applied to a Markov process which is closely related to the physical system.

As the physical system is a continuous-time process with high bandwidth noise, a closely related Markov process is a continuous-time process with white noise. Such a process is called a diffusion process[5], and is described by a stochastic differential equation of the Ito[5] or Stratonovich[6] type. If the diffusion process models the physical system in the sense that their relevant statistical properties are approximately the same, then the stochastic control theory can be applied to the physical system by way of the diffusion model, as shown in Fig. 1b. A similar problem occurs when the stochastic controller, designed in Fig. 1c for a white noise system, must be built in Fig. 1d for a coloured noise system.

With this motivation, the purpose of this paper is to discuss the relation between physical processes and diffusion processes, and to illustrate how this relation affects the synthesis of a stochastic control system. The relation between physical and diffusion processes was first discussed by Stratonovich[7] on a physically rigorous level, and later by Khasminskii[8]. An approach which is similar to that of Stratonovich, but which is more convenient in the present context, has been given by Cumming[9]. This approach is outlined briefly in Section 2 below. Some other results which are relevant to the problem have been reported by Wong and Zakai[10] and Clark[11].

Having discussed the relation between physical and diffusion processes, it is used in Section 3 to form a diffusion model of a given physical system, allowing the use of stochastic control theory to design an optimal controller for the physical system. In Section 4, the relation is also used to translate the controller designed by stochastic theory (which is still a diffusion process) into a physically realizable controller which can be built onto the given system. Finally, the performance of a controller so designed is shown to be superior to one designed by not taking the theory of Section 2 into account.

## 2. The Relation Between Diffusion Processes and Physical Processes.

This section will state the conditions under which a diffusion process and a physical process are approximately statistically equivalent, and the resulting relation between them. The criterion of statistical equivalence will be that the first order probability density functions of the two processes will be approximately equal over almost all of the time range of interest. This criterion is sufficient for stochastic control problems where performance criteria such as mean or integral squared error are of interest

Form of Physical Process: Many of the continuous-time stochastic systems found in practice can be described by ordinary differential equations (o.d.e.) in the state vector form

$$\dot{X}(t) = g(X,t) + G(X,t)y(t), \tag{2.1}$$

where X(t) is the n-dimensional "state" of the system, g(X,t) and G(X,t) are the known system dynamics, and y(t) is a high-bandwidth, zero mean, m-dimensional noise vector specified by the matrix correlation function

$$E[y(t)y^T(t-\tau)] = R(t,\tau), \tag{2.2}$$

where $E[\cdot]$ is the ensemble expectation operator and $(\cdot)^T$ denotes transpose. The initial conditions of (2.1) are not given as they are not important to the discussion below.

The form (2.1) is restrictive in the sense that the noise y(t) must be factorable from the system dynamics in the manner shown. For example, a r.h.s. term such as X(t) sin (y(t)) is not allowed in the present form unless a new noise $y'(t) = \sin(y(t))$ can be defined and specified by a correlation function as in (2.2). Stratonovich[7] uses a form of system equation without this restriction, but the present form can be more conveniently compared with the diffusion process (2.3) and allows for non-stationary noise processes.

Form of Diffusion Process: A continuous Markov or diffusion process is described by the Ito stochastic differential equation (s.d.e.)[5].

$$dx(t) = f(x,t)\, dt + F(x,t)\, dw(t) \tag{2.3}$$

where x(t) is the n-dimensional state of the process, f(x,t) and F(x,t) are the dynamics of the process, d. is a stochastic increment in the Ito sense, and w(t) is an m-dimensional Wiener process with the incremental property

$$E[dw(t)\, dw(t)^T] = 2D(t)dt \tag{2.4}$$

The Stratonovich form[6] of the s.d.e. (2.3) is given as equation (1") in Cumming[12].

Incremental Properties: The statistical properties of the diffusion process (2.3) are specified by two incremental properties

$$E_x[\delta x(t)] = f(x,t)\delta t + o(\delta t), \tag{2.5}$$

and

$$E_x[\delta x(t)\delta x(t)^T] = 2FDF^T(x,t)\delta t + o(\delta t), \tag{2.6}$$

where $\delta\cdot$ is a forward difference operator over a time increment $\delta t$, $o(\cdot)$ denotes of order higher than $(\cdot)$, and $E_x[\cdot]$ is the ensemble conditional expectation operator, given $x(t)$.

The existence of the parameters (2.5,6) is a basic property of continuous Markov processes as these parameters specify the form of the Fokker-Planck equation of the diffusion process $x(t)$. In contrast, analogous incremental properties for the non-Markov physical process (2.1) do not exactly specify the statistical properties of $X(t)$, but if the bandwidth of the physical noise $y(t)$ is high enough, the incremental properties of $X(t)$ do specify the statistics of $X(t)$ to a sufficiently high accuracy. In this case, the physical process $X(t)$ is said to be near-Markov, and the statistical equivalence of $X(t)$ and $x(t)$ is ensured by equating the incremental properties (2.5,6) of the processes.

The justification of this statistical equivalence is discussed by Stratonovich[7] and Cumming[9]. It is necessary to make the following assumptions about the physical process $X(t)$.

A1: A positive parameter $\tau_{cor}$, the correlation time of the physical noise $y(t)$, exists and is the smallest number for which $R_{ij}(t,\tau)$ is effectively zero for $|\tau|>\tau_{cor}$, all i, j and time t within the range of interest. More precisely, $\tau_{cor}$ must be big enough so that

$$\int_{-\tau_{cor}}^{\tau_{cor}} R(t,\tau)d\tau \doteq \int_{-\infty}^{\infty} R(t,\tau)d\tau. \tag{2.7}$$

The parameter $\tau_{cor}$ is related to the upper frequency $\omega_y$ of the noise $y(t)$ by

$$\tau_{cor} = \gamma\omega_y^{-1}, \tag{2.8}$$

A2: A parameter $\tau_{rel}$, the relaxation time of the physical process $X(t)$, exists and

$$\tau_{rel} > 10\ \tau_{cor}. \tag{2.9}$$

Stratonovich[7] defines $\tau_{rel}$ in terms of g, G, and D, but his definition is difficult to apply for all but the simplest systems. It is simpler to define

$$\tau_{rel} = \gamma\omega_x^{-1}, \tag{2.10}$$

where $\omega_x$ is the upper frequency of the system $X(t)$, defined analogously to $\omega_y$. The assumption then becomes

$$\omega_y > 10\omega_x \tag{2.11}$$

which ensures that the system X(t) is near-Markov. The noise y(t) then affects the system in a similar way to white noise.

A3: The functions g, G, $G_t$ and $G_x$ are of finite variation within the time range of interest. The subscripts t and X denote partial derivatives.

A4: The integral

$$A(t) = \int_0^\infty R(t,\tau)\, d\tau \tag{2.12}$$

exists and the non-stationarity of y(t) is sufficiently slow that A(t) changes a negligible amount over time intervals as small as $\tau_{cor}$. Then

$$A(t) + A^T(t) \doteq \int_{\infty}^{\infty} R(t,\tau)\, d\tau. \tag{2.13}$$

A(t) is called the characteristic matrix of the physical noise y(t), and contains all the statistical information of y(t) needed to estimate the statistical properties of X(t) under the present assumptions.

A5: A time increment $\delta t$ is chosen over which the incremental properties of X(t) are evaluated, and

$$\tau_{cor} < \delta t < \tau_{rel}. \tag{2.14}$$

With these assumptions, approximate expressions can be derived[9] for the incremental statistics of X(t),

$$E_x[\delta X_i(t)] \doteq (g_i + \sum_j^n \sum_{k,\ell}^m \frac{\partial G_{ik}}{\partial X_j} G_{j\ell} A_{k\ell})\, \delta t,\ i = 1, n, \tag{2.15}$$

and $E_x[\delta X\, \delta X^T(t)] \doteq G(A + A^T)G^T \delta t,$ (2.16)

where the summations have an implied lower limit of one, and the functions are evaluated at X and t.

The expressions are necessarily only approximate, as the future statistics of a non-Markovian process are not completely known when X(t) is given. However, the assumptions A1-5 ensure that these expressions have a useful accuracy.

Statistical Equivalence: If the expressions (2.15, 16) match those of the diffusion process (2.5, 6) the physical process X(t) and the diffusion process x(t) are statistically equivalent in the following sense. The first order probability density functions of X(t) and x(t) are approximately equal over the time range of interest, with the possible exception of times within $\tau_{cor}$ of arbitrary initial conditions of the systems. This latter exception has a negligible affect on the applications to follow.

Matching the incremental properties of x(t) and X(t), the systems are statistically equivalent if

$$1. \quad 2D(t) = A(t) + A^T(t). \tag{2.17}$$

This ensures that the physical noise y(t) and the white noise $\dot{w}(t)$ have equal low frequency spectral density.

$$2. \quad F(x,t) = G(X,t), \tag{2.18}$$

and

$$3. \quad f_i(x,t) = g_i(X,t) + \sum_j^n \sum_{k,\ell}^m \frac{\partial G_{ik}}{\partial X_j} G_{j\ell}(X,t) A_{k\ell}(t), \; i=1,n. \tag{2.19}$$

The term involving the summation on the r.h.s. of (2.19) is the interesting term in the relation between physical processes and diffusion processes, and will be called the bias term. This bias term accounts for the effect of the short-term correlation between the noise y(t) and state X(t) in the term Gy(t) of the physical system, which the term Fdw(t) of the Ito equation does not reflect on its own accord.

Comments on the Assumptions: The factor of 10 used in A2 is chosen as a safe figure, for in several examples tested, no noticeable improvement in statistical equivalence occurred when the noise bandwidth exceeded ten times that of the system. It is not clear how fast the accuracy deteriorates as the noise bandwidth is reduced towards the bandwidth of the system, and unless explicit error estimates can be obtained, each example will have to be treated as a special case. If the noise y(t) and the system output X(t) are available, the relation (2.9) can be tested experimentally by measuring the maximum correlation times of the two signals. Also, the characteristic matrix A of (2.12) can be estimated experimentally by averaging the product of the noise and its integral, as

$$A(t) = E[y(t)\, Y^T(t)], \tag{2.20}$$

where

$$Y(t) = \int_{-\infty}^{t} y(s)\, ds. \tag{2.21}$$

## 3. Design of a Stochastic Control System

In this section, a typical design exercise as outlined in the Introduction will be detailed. As an example, the optimal controller for a linear system with random coefficients will be designed, for this is one of the more interesting problems for which an optimal control solution is currently known. For simplicity, a first order system is considered.

<u>Control Problem</u>: Given the physical system of Fig. 2 described by the o.d.e.

$$\dot{X}(t) = [a + y_1(t)]X(t) + y_2(t), \tag{3.1.}$$

and the ability to control X(t) through the addition of the control term b u(t) on the r.h.s. of (3.1), find the stationary control u(t) which minimizes the expected rate of increase of the cost function

$$J(t) = \int_0^t [u^2(s) + qX^2(s)]\, ds, \tag{3.2}$$

when y(t) is a 2-dimensional stationary physical noise possessing a sufficiently high upper frequency that the assumptions of Section 2 are satisfied, and the characteristic matrix A is known.

In order to use stochastic control theory to design the controller, we must first approximate the given system (3.1) by a diffusion process which is amenable to theoretical analysis. As the performance criterion depends on the probability density of X(t), the method of Section 2 can be used to obtain a relevant diffusion model for (3.1).

<u>Diffusion Model</u>: As the assumptions of Section 2 are satisfied, and the bias term of (2.19) is $A_{11}X(t) + A_{12}$, a diffusion model for the physical process (3.1) is given by the Ito s.d.e.

$$dx(t) = [(a + A_{11})\, x(t) + A_{12}]dt + x(t)dw_1(t) + dw_2(t), \tag{3.3}$$

where w(t) is a two-dimensional Wiener process with the incremental property

$$E[dw(t)dw^T(t)] = 2\, D\, dt = (A + A^T)dt. \tag{3.4}$$

<u>Optimal Control Design</u>: The optimal control of the Markov process (3.3) is found by the method of dynamic programming[3,4], where the incremental properties (2.5,6) of (3.3) describe the behaviour of the process. Note that the physical noise parameter $A_{12}$ enters into the control calculation, which is interesting as this information is not contained in the matrix D which characterizes the white noise $\dot{w}(t)$.

Under mild restrictions[1], an optimal stationary controller exists, and is given by

$$u(t) = -\frac{1}{2} b\, k_1 - b\, k_2\, x(t), \tag{3.5}$$

where $$k_2 = b^{-2}[a + A_{11} + D_{11} + ((a + A_{11} + D_{11})^2 + b^2 q)^{\frac{1}{2}}], \tag{3.6}$$

and $$k_1 = \frac{(2A_{12} + 4D_{12})\, k_2}{b^2 k_2 - a - A_{11}} \quad . \tag{3.7}$$

Non-optimal Control Designs: If the physical process (3.1) is not modelled properly, non-optimal control designs can result. Authors who consider theoretical control problems generally do not discuss the relation between physical processes and diffusion processes, but only discuss the control of a diffusion process described by Ito or Stratonovich s.d.e.'s. On reading their papers, one could easily fall into the trap of interpreting their equations as being directly applicable to physical processes. This is equivalent to interpreting the o.d.e. (3.1) of the physical system directly as an s.d.e., on replacing $y(t)$ by white noise $\mathring{w}(t)$ of equal low frequency spectral density, 2D. Two types of error can result.

Non-Optimal Design 1: This design can occur if (3.1) is interpreted as a Stratonovich s.d.e. This is equivalent to specifying the physical noise $y(t)$ only by its low frequency spectral density, 2D, instead of by the more detailed information of the characteristic matrix A, and the non-optimal model and control design are obtained by replacing A by D in equations (3.3-7).

Note that as $A + A^T = 2D$, no error is made in this design if A is a symmetric matrix. However, many noise vectors found in practice do not have a symmetric characteristic matrix, as, for example, when one component of $y(t)$ lags another. Also, no error is made if the bias term of (2.19) is zero.

Non-Optimal Design 2: This design can occur if (3.1) is interpreted as an Ito s.d.e. This is equivalent to ignoring the bias term of (2.19) and the non-optimal model and control design are obtained by setting A to zero in equations (3.3-7).

Note that the error in this design is zero only when the bias term of (2.19) is zero, but this term is non-zero whenever the noise magnitude depends on the state $X(t)$. The error that can be made in this design is usually more serious than that of Design 1, but the use of Design 2 was largely eliminated when the work of Stratonovich became known.

## 4. Construction of the Stochastic Controller

The system, with the stationary controller attached, is now given as a diffusion process described by the Ito s.d.e. (adding b times (3.5) to the r.h.s. of (3.3))

$$dx(t) = [(a + A_{11} - b^2k_2)x(t) + A_{12} - \tfrac{1}{2} b^2k_1] dt + x(t)dw_1(t) + dw_2(t) \quad (4.1)$$

using the appropriate design values of $k_1$ and $k_2$. The construction problem is to build a physical system which is the statistical equivalent of (4.1) within the environment containing the physical noise y(t).

The statistically equivalent physical system is obtained from the theory of Section 2. In fact, as the noise and its coefficients are the same as when the diffusion model (3.3) was chosen for the system (3.1), the the bias term of (2.19) is the same, $A_{11}x(t) + A_{12}$. This term, however, is now subtracted from (4.1) to obtain the statistically equivalent physical system given by the o.d.e.

$$\dot{X}(t) = (a + y_1(t) - b^2k_2)X(t) - \tfrac{1}{2} b^2k_1 + y_2(t) \quad (4.2)$$

Comparing (4.2) with the original system equation (3.1) indicates that the controller is to be built by supplying the feedback terms $-b^2k_2X(t) - \tfrac{1}{2} b^2k_1$.

The construction problem is particularly simple in this case, as the addition of the controller introduces no new noise terms and the s.d.e. to o.d.e. bias term is the same as in the modelling step. This is not always the case, as, for example, in the synthesis of the non-linear filter proposed by Wonham[13]. The construction of this filter is discussed by Cumming[9].

Another interesting problem is the synthesis of a controller when the control loop gain b has a noisy component. In this case, the bias term of (2.19) used to obtain the model (3.3) now depends on the feedback coefficient $k_2$, as yet undetermined. However, to determine $k_2$ the bias term must be known, and so the two must be obtained simultaneously. Present indications are that an iterative solution is possible for the stationary control case[14].

<u>Controller Performance</u>: The effectiveness of the controller is measured by the expected value of the rate of increase C of the cost (3.2),

$$C = E[u^2(t) + qX^2(t)]. \quad (4.3)$$

As the control u(t) is a function of X(t) and the design coefficients $k_1$ and $k_2$, the cost rate C can be conveniently expressed in terms of the first two moments of X(t), $m_1$ and $m_2$,

$$C = \tfrac{1}{4} b^2k_1^{\,2} + b^2k_1k_2m_1 + (b^2k_2^{\,2} + q)m_2. \quad (4.4)$$

The moments of X(t) are approximately the same as those of x(t) and the latter can be found from their differential equations[12]. They have the stationary values

$$m_1 = \frac{A_{12} - \frac{1}{2}b^2k_1}{b^2k_2 - a - A_{11}}, \tag{4.5}$$

and
$$m_2 = \frac{(A_{12} + 2D_{12} - \frac{1}{2}b^2k_1)m_1 + D_{22}}{b^2k_2 - a - A_{11} - D_{11}} \tag{4.6}$$

provided the denominator in each case is positive. If the denominator of (4.6) is non-positive, the system is unstable in the mean square sense and the controller is generally unacceptable—$k_2$ must be sufficiently large to prevent this. Note that the parameter noise $y_1(t)$ has a destabilizing influence on the system.

In order to compare the non-optimal control designs with the optimal one, it is convenient to evaluate the added cost rate $C^+$ which is the difference of C, (4.4), between the non-optimal and the optimal designs,

$$C^+ = C_{\text{non-opt}} - C_{\text{opt}}. \tag{4.7}$$

With the equations (3.6,7), and (4.4-7), it is easy to test the sensitivity of the added cost rate $C^+$ to any of the parameters involved. This sensitivity is discussed below for an arbitrary choice of parameters.

Non-optimal Design 1: The feedback coefficients for this design are obtained by setting A to D in equations(3.6,7), and differ from the optimal coefficients in $k_1$ only, as $A_{11} = D_{11}$. The added cost rate is

$$C^+ = \left(\frac{bk_2}{b^2k_2 - a - A_{11}}\right)^2 (D_{12} - A_{12})^2. \tag{4.8}$$

Note that $C^+$ is non-negative for all values of the parameters of the system, noise, controller and cost function, showing that this control design is indeed non-optimal. The added cost rate is zero only when $A_{12} = D_{12}$, the assumed design value, and varies with $A_{12}$ in a parabolic fashion.

Figures 3 and 4 show typical sensitivity curves which can be obtained from the equations above. The nominal system parameter values are $a = -1$, $b = .5$, $A_{11}$ and $D_{11} = .5$, $D_{12} = .4$ and $q = 1$. The common abcissae is the true value of $A_{12}$, and is allowed to vary between zero and $2D_{12}$ (although $A_{12}$ can extend beyond these limits, they are realistic limits for a given $D_{12}$). Each graph gives the optimal and the non-optimal values of the cost rate, C, the system mean $m_1$, and the system mean square, $m_2$.

The graphs show that when the true value of $A_{12}$ differs from $D_{12}$, the system with the non-optimal controller can have markedly differing statistics from the optimal one. A convenient performance measure is the percentage with which the non-optimal cost rate exceeds the optimal $100\ C^{+}/C$. This percentage has maximum values at each end of the abcissae range, being 11.8% when $A_{12} = 0$ and 3.5% when $A_{12} = 2D_{12}$.

The nominal system given had moderate values of this percentage. The percentage was sensitive to most of the system parameter values, and in particular, it increased when the system parameter a, or the noise parameters D, increased.

Non-optimal Design 2: This design will not be discussed in detail, except to state that the percentage increase in cost rate is much higher for this non-optimal design than Design 1 (for the nominal system parameter values used above, the percentage is 110.6%). The main difference of this design from Design 1, is that the feedback gain $k_2$ is lower, and is often not high enough to stabilize a system whose coefficient noise $y_1(t)$ has a significant destabilizing influence.

The performance figures were obtained from the theoretical equations above. The design method of this paper and the performance figures were confirmed through a digital simulation of the system.

## 5. Conclusions.

This paper has been concerned with the application of some of the modern stochastic control theory to physical problems. A difficulty arises because systems considered in theory and those met in practice are of different types, and the type of differential equation used to describe each type of system is different.

In particular, all continuous random processes in practice have noise components with a finite spectrum, and as such are described by ordinary differential equations, while for mathematical tractability, continuous random processes studied in theory usually have white noise components, and as such are described by stochastic differential equations. However, most of the authors who develop stochastic control theory use stochastic differential equations to describe their systems without discussing how their equations relate to practical systems. In failing to do so, these authors allow their readers to misinterpret the application of the theory.

To this end, the main purpose of this paper is to outline the relation between ordinary and stochastic differential equations. This is done in Section 2 by taking a general state vector form of ordinary differential equation describing a physical random process, and showing how, and under which conditions, a white noise process can be chosen which is essentially statistically equivalent to the physical process. The differences between the two processes arise from the manner in which the correlation between the noise and the state affect the system's statistics in each case, and a bias term is used to correct for this difference.

Having developed this relation, stochastic control theory can be applied to physical problems by a) making a white noise model of the original physical process, and b) constructing a physical system from the one designed theoretically. To illustrate this, the synthesis of an optimal controller for a linear system with a noisy coefficient is discussed. In particular, the paper shows how inferior control designs can result if the controls are designed without regard to the relation between the theoretical and practical processes.

References


1. Wonham, W.M., "Optimal Stationary Control of a Linear System with State-dependent noise," SIAM J. Control, 5, 3, p. 486, 1967.
2. Sancho, N.G.F., "Optimal Control of Markov Stochastic Systems which have Random Variation of Gain of Plant," Int. J. Control, 3, 5, p. 487, 1966.
3. McLane, P.J. and Graefe, P.W.U., "Optimal Regulation of a Class of Linear Stochastic Systems Relative to Quadratic Criteria," Int. J. Control, 5, 2, p. 135, 1967.
4. Ariaratnam, S.T., and Loh, N.K., "Optimal Control of Linear Stochastic Systems," Int. J. Control, 6, 1, p. 51, 1967.
5. Doob, J.L., "Stochastic Processes," Wiley, New York, p. 273, 1953.
6. Stratonovich, R.L., "A New Form of Representation of Stochastic Integrals and Equations," SIAM J. Control, 4, p. 362, 1966.
7. Stratonovich, R.L., "Topics in the Theory of Random Noise, Volume 1," Gordon and Breach, New York, 1963.
8. Khasminskii, R.Z., "A Limit Theorem for the Solutions of Differential Equations with Random Right-hand Sides," Th. Prob. Appl., 11, 3, p. 390, 1966.
9. Cumming, I.G., "Computing Aspects of Problems in non-linear Prediction and Filtering," Ph.D. thesis, Centre for Computing and Automation, Imperial College, London, 1967.
10. Wong, E. and Zakai, M., "On the Relation Between Ordinary and Stochastic Differential Equations and Applications to Stochastic Problems in Control Theory," Preprints, Third IFAC Congress, London, 1966.

11. Clark, J.M.C., "The Representation of Non-linear Stochastic Systems with Applications to Filtering," Ph.D. thesis, Department of Electrical Engineering, Imperial College, London, 1966.

12. Cumming, I.G., "Derivation of the Moments of a Continuous Stochastic System," Int. J. Control, 5, 1, p. 85, 1967.

13. Wonham, W.M., "Some Applications of Stochastic Differential Equations to Optimal Non-linear Filtering," SIAM J. Control, 2, 3, p. 347, 1965.

14. McLane, P.J., "Optimal Regulation of Stationary Linear Systems with Control-dependent Stochastic Disturbances", Informal Report, Dept. of Electrical Engineering, Univ. of Toronto, Canada, June, 1968.

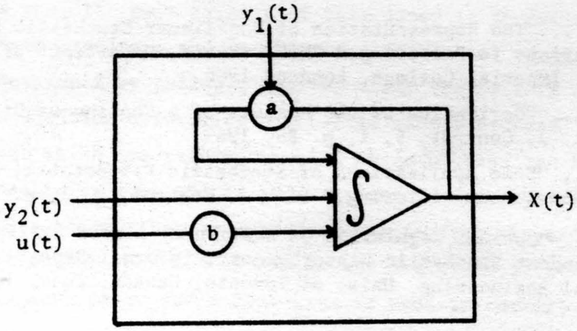



Fig. 2 Physical System of Equation (3.1)

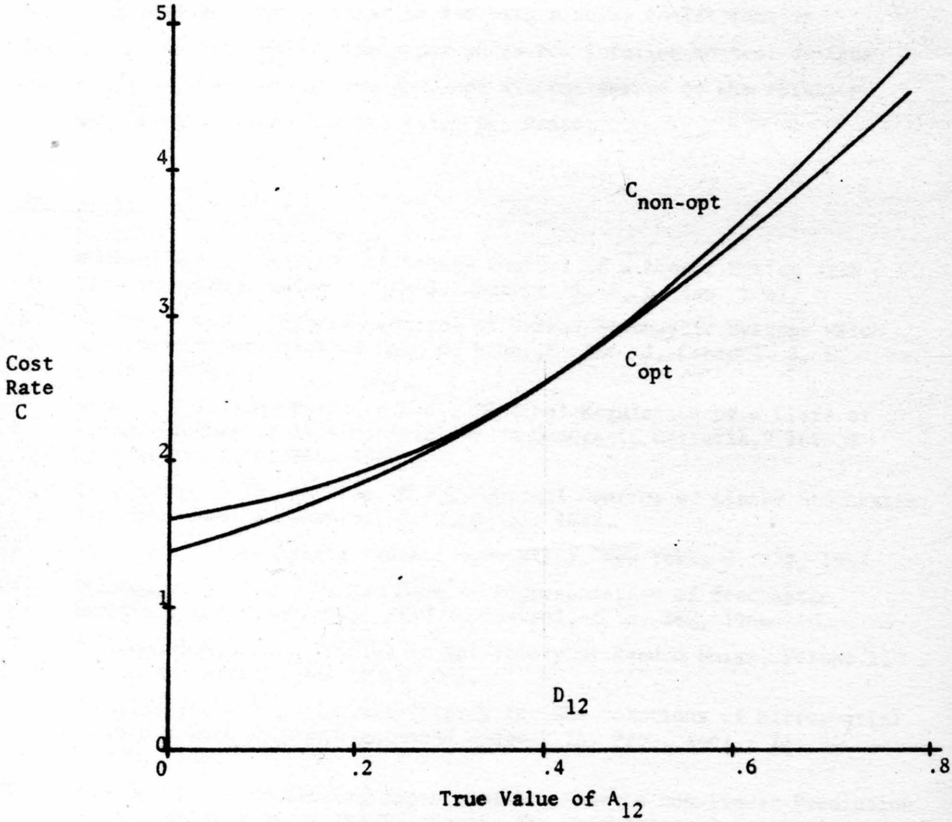



Fig. 3 Sensitivity of System Cost Rate to $A_{12}$

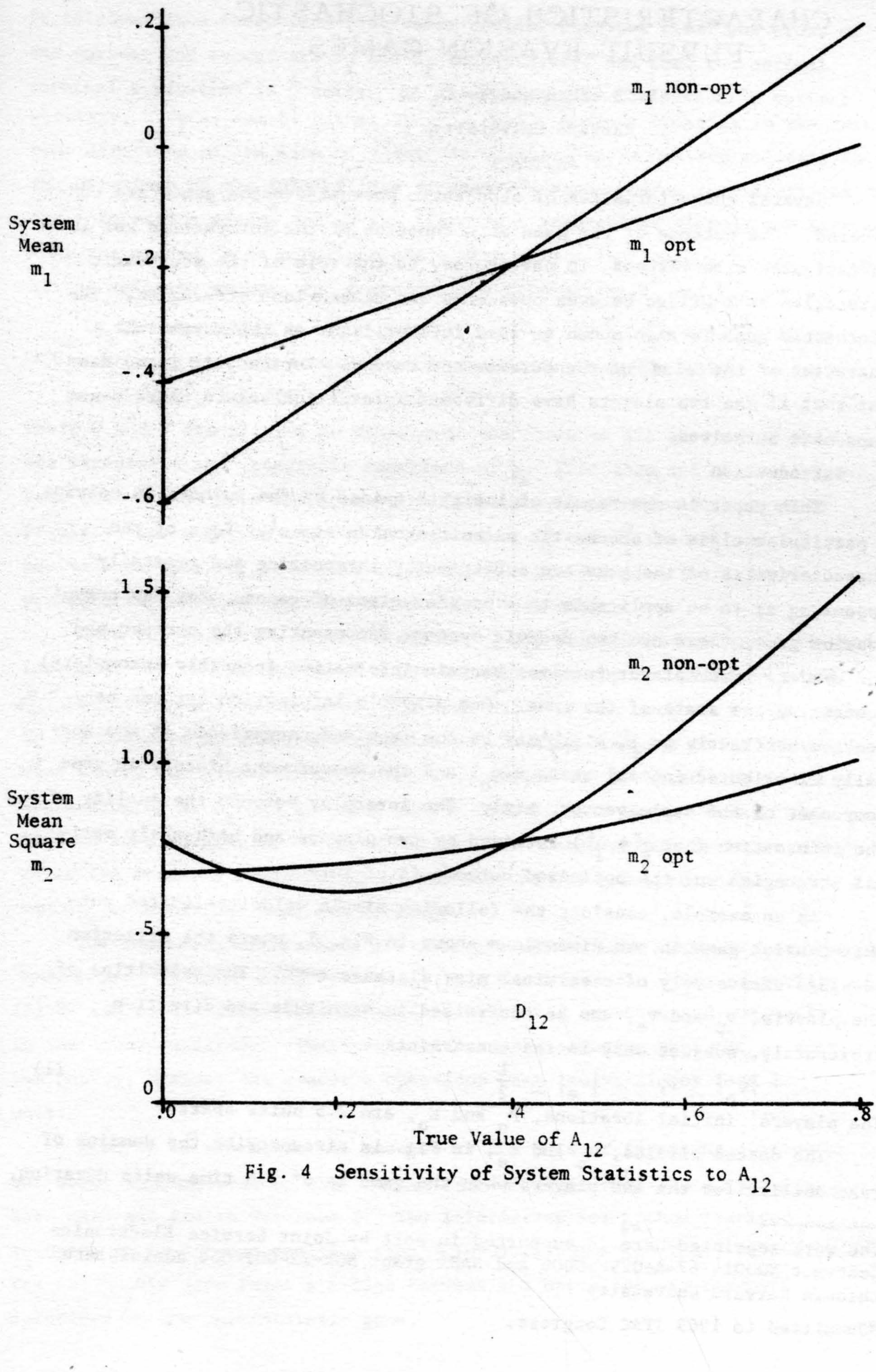



Fig. 4 Sensitivity of System Statistics to $A_{12}$

# CHARACTERISTICS OF STOCHASTIC PURSUIT-EVASION GAMES

R. D. Behn and Y. C. Ho

Harvard University

ABSTRACT

Several characteristics of stochastic pursuit-evasion games are considered. The outcome of the game as a function of the information set is investigated with respect, in particular, to the role of the stochastic strategies as a bridge between open-loop and closed-loop strategies. The stochastic game is also shown to shed further light on the asymmetric character of the roles of the pursuer and evader. Further, it is pointed out that if the two players have different information sets a nonzero-sum game must be solved.

## I. Introduction

This paper is the result of insights gained by the authors in solving a particular class of stochastic pursuit-evasion games[1]. Some of the characteristics of the game are sufficiently interesting and generally appealing as to be applicable to a broader class of games. For any pursuit-evasion game, there are two dynamic systems representing the pursuer and the evader. Each player receives certain information (possibly incomplete) concerning the state of the game. One player's information set can be denoted abstractly as $\mathcal{Z}'$, e.g. $\mathcal{Z}'$ may be the mean and convariance of the normally distributed initial state $x(t_o)$ and the measurement history on some component of the state vector, $x(t)$. The interplay between the quality of the information (set $\mathcal{Z}^p$ & $\mathcal{Z}^e$) received by two players and both their optimal strategies and the optimized outcome is of interest.

As an example, consider the following simple velocity-limited pursuit-evasion game in two dimensions shown in Fig. 1, where the criterion consists exclusively of a terminal miss distance term. The velocities of the players, $v_p$ and $v_e$, can be controlled in magnitude and direction arbitrarily, subject only to the constraints

$$|v_p| \leq 1, \qquad |v_e| \leq \frac{1}{2} \tag{1}$$

The players' initial locations, $P_o$ and $E_o$, are 2.5 units apart.

The dotted circles, $C_p$ and $C_e$, in Fig. 1a circumscribe the domains of reachability for the two players when the game is of 2.0 time units duration.

-----------

The work reprinted here is supported in part by Joint Service Electronics Contract N00014-67-A-0298-0006 and NASA grant NGR-22-007-068 administered through Harvard University

*Submitted to 1969 IFAC Congress.

It is clear from inspection that under optimal play the final positions of the pursuer and evader are $P_f$ and $E_f$ respectively, and that the optimal terminal separation is $\frac{3}{2}$ units. To accomplish this the pursuer's optimal strategy, $U^o(\mathfrak{z}^P)$, can be either $U_1^o(\mathfrak{z}^P)$: "apply maximum velocity in the current direction of the line of sight" or $U_2^o(\mathfrak{z}_2^P)$: "apply maximum velocity in the direction of the initial line of sight." With obvious specifications for the information sets $\mathfrak{z}_1^P$ and $\mathfrak{z}_2^P$ the first is a feedback or closed-loop strategy while the second is an open-loop one. With only a 180° change in the direction of the velocity vector, two corresponding strategies, $V_1^o$ and $V_2^o$, can be defined.

It is interesting to note that $J(U_1^o, V_1^o) = J(U_2^o, V_1^o) = J(U_1^o, V_2^o) = J(U_2^o, V_2^o)$ and that any strategy pair constitutes a saddle point

$$J(U^o,V) \leq J(U^o, V^o) \leq J(U,V^o) \quad (2)$$

where U and V are allowed to range over the class of all strategies which are measurable and integrable functions of $\mathfrak{z}_2$. From this and the well known fact that closed-loop and open-loop controls are equivalent in one sided problems, one might be tempted to infer that $U_1^o$ and $U_2^o$ (and $V_1^o$ and $V_2^o$) are equivalent. However, the game of 3. times units duration in Figure 1b presents an immediate counter example for this conjecture.

Here the pursuer's feedback strategy, $U_1^o$, still satisfies the left hand inequality of the saddle-point condition (2) when opposed by both $V_1^o$ and $V_2^o$, but his open-loop strategy, $U_2^o$, does not. If the pursuer merely heads in the positive $x_1$ direction--without continually checking to determine what course the evader is following--the evader can sneak around behind him ending up at $E_f'$ and 2. units rather than 1. unit away. Thus if the game is long enough, it is important that the pursuer receive information as to the evader's location during the play of the game; otherwise he cannot implement the feedback strategy.

The evader, however, does not open himself to such problems by operating open-loop. In other words his $V_2^o$ still satisfies the r. h. inequality of (2). He can merely determine the pursuer's initial location and run madly in the other direction. There exists no sneaky strategy that the pursuer can employ, against the evader's open-loop one, to get closer than 1. unit.

Now, if we visualize the case where the pursuer is making continuous but noisy measurements on the current line of sight with the noise being zero mean and finite variance Q. The information set $\mathfrak{z}_Q$ thus received occupies an in-between position from $\mathfrak{z}_1(Q=0)$ to $\mathfrak{z}_2(Q=\infty)$. Thus, we see that the stochastic game forms a bridge between the open-loop and feedback solutions of the deterministic game.

In the sections below, these phenomena and others are discussed quantitatively for a class of stochastic differential games.

II. The Game

The game considered here is the stochastic extension of the deterministic game first solved by Ho-Bryson-Baron in 1965[2]. The dynamic system is

$$\dot{y} = G_p(t)u - G_e(t)v, \qquad y(t_o) = y_o \tag{3}$$

and the criterion

$$\mathcal{J} = \frac{a^2}{2} ||y(t_f)||^2 + \frac{1}{2} \int_{t_o}^{t_f} (||u||^2_{R_p} - ||v||^2_{R_e})dt \tag{4}$$

It is well known that more general linear-quadratic games, either do not add anything conceptually new or can be reduced to the above form without loss of generality[2]. Hence, for notational simplicity the game of Eqs. (1) and (2) shall be treated here.

For the stochastic game, we shall assume in addition that:

(i) Initially, both players have a common assessment of the state $y(t_o)$ as a normal random variable with mean $\hat{y}(t_o)$ and covariance $P_o$.

(ii) The pursuer and evader are making respectively the measurements on the state

$$z_p = H_p y + w_p \tag{5}$$

$$z_e = H_e y + w_e \tag{6}$$

where $w_p$ and $w_e$ are independent white gaussian random processes with zero means and spectrums $Q_p$ and $Q_e$. Thus $\mathcal{Z}_p(t) = \{z_p(\tau),\ t_o \leq \tau \leq t;\ \hat{y}(t_o),\ P_o\}$ and $\mathcal{Z}_e(t) = \{z_e(\tau)\ t_o \leq \tau \leq t;\ \hat{y}(t_o),\ P_o\}$.

For the deterministic game mentioned above, the solution (i.e. the pair of optimal strategies $U^o$ and $V^o$) is obtained from the saddle-point condition (2) for the criterion $\mathcal{J}$ in (4). For the stochastic game, the payoff or outcome is still determined by (4). However, because of the stochastic nature of the problem neither player knows the outcome before or during the game, even though he may know the strategies employed. Consequently, both players must use an expected value operator to evaluate the criterion, and these two evaluations will be different, for the two players have different information sets over which to take the expectation. Thus at any time t, the pursuer's criterion (i.e. his evaluation of the outcome of the game) is

$$J_p(t, U, V) = E[\mathcal{J}(U, V) | \mathcal{Z}^p(t)] \tag{7}$$

while the evader's criterion or outcome evaluation is

$$J_e(t, U, V) = E[\mathcal{J}(U, V) | \mathcal{Z}^e(t)] \tag{8}$$

where U, V are measureable and integrable functions of the information set $\mathcal{Z}^p(t)$ and $\mathcal{Z}^e(t)$ respectively.

The pursuer would like to select a strategy which would minimize his evaluation of the expected payoff which is $J_p$. Conversely, the evader must maximize his evaluation of the expected payoff, $J_e$. Consequently, the solution must be obtained by solving a nonzero-sum game; strategies are sought which satisfy the equilibrium condition consisting of the pair of inequalities,

$$J_e(t, U^o, V) \leq J_e(t, U^o, V^o) \tag{9a}$$

$$J_p(t, U^o, V^o) \leq J_p(t, U, V^o), \tag{9b}$$

where the class of strategy pairs is restricted as above.

This does not mean that the game itself is not zero-sum; the payoff (4) certainly is. It is the fact that the two players have different information sets, and thus evaluate the outcome differently, that results in this characteristic. The same would be true of a zero-sum matrix game in which the two players had different thoughts concerning the payoff values in the matrix.

However, before the game begins, both players have the same evaluation of the outcome, since $\mathfrak{z}^p(t_o) = \mathfrak{z}^e(t_o)$

$$J = E_{y(t_o)}[\mathcal{G}] \tag{10}$$

where $E_{y(t_o)}$ means the expectation taken with respect to the possible values of $y(t_o)$. Thus the players seek saddle-point strategies with respect to this J. An immediate question is whether the strategies obtained for (10) and the saddle-point condition (2) also provide a solution to the nonzero-sum game equilibrium condition (9ab) which holds during play.

Fortunately, the answer is in the affirmative and can be easily established via contradiction. Suppose from time t onwards based on a given $\mathfrak{z}^p(t)$, a strategy $U^1$ exists such that $J_p(t, U^1, V^o) < J_p(t, U^o, V^o)$. Then we can immediately construct another strategy

$$U^2 = \begin{cases} U^1 & \text{for } \tau > t \text{ and } \mathfrak{z}^p(\tau) \text{ as given} \\ U^o & \text{otherwise} \end{cases} \tag{11}$$

and write for (10)

$$J = E_{y(t_o)}\{\mathcal{G}\} \triangleq E_{y(t_o),\mathfrak{z}^p(t)}[E\{\mathcal{G}/\mathfrak{z}^p(t)\}] \triangleq E_{y(t_o),\mathfrak{z}^p(t)}[J_p] \tag{12}$$

which implies the contradiction

$$J(U^2, V^o) < J(U^o, V^o) \tag{13}$$

unless $V^1$ is valid only on a set of measure zero. This means any solution based on (10) automatically satisfies (9) except on a set of measure zero.

For the deterministic game, i.e. $\mathfrak{z}^p(t) = y(t) = \mathfrak{z}^e(t)$ the optimal strategies were found to be feedback ones based on the state vector[2].

$$U^o: \quad u(t) = C_p(t)y(t) \tag{14}$$

$$V^o: \quad v(t) = C_e(t)y(t) \tag{15}$$

For the particular stochastic game considered by the authors, one player retains perfect knowledge ($Q_p(t) = 0$) while the other makes only noisy measurements, i.e. $z^p(t) = y(t)$. If the evader is the player with the noisy information set, the optimal strategies are of the form

$$U^o: \quad u(t) = C_p(t)y(t) + D_p(t)\tilde{y}(t) \tag{16}$$

$$V^o: \quad v(t) = C_e(t)\hat{y}(t) \tag{17}$$

where $\hat{y}(t)$ is the evader's optimal estimate of the state $y(t)$ and $\tilde{y}(t)$ is the error of this estimate.

$$\tilde{y}(t) = y(t) - \hat{y}(t) \tag{18}$$

Under certain conditions, the pursuer can calculate this error[1,5]. If the information sets of the two players are exchanged, the strategy forms are also, though as discussed in Section IV below the values of the feedback gains are not interchangeable.

III. Estimation Problem in Stochastic Games.

For the stochastic game, discussed in section II the exact values of the feedback gains are

$$C_p(t) = -R_p^{-1}(t)G_p^T(t)K^{-1}(t_f,t), \tag{19}$$

$$C_e(t) = -R_e^{-1}(t)G_e^T(t)K^{-1}(t_f,t), \tag{20}$$

$$D_p(t) = -R_p^{-1}(t)G_p^T(t)\Gamma(t) \tag{21}$$

$K^{-1}(t_f,t)$ is given by

$$K^{-1}(t_f,t) = \left[\frac{I}{a^2} + M_p(t_f,t) - M_e(t_f,t)\right]^{-1} \tag{22}$$

where

$$M_p(t_f,t) = \int_t^{t_f} G_p(\tau)R_p^{-1}(\tau)G_p(\tau)d\tau \tag{23}$$

and a similar definition applies for $M_e(t_f,t)$. Equation (22) is the solution to the differential equation

$$\frac{d}{dt}K^{-1}(t_f,t) = K^{-1}(t_f,t)\left[G_p(t)R_p^{-1}(t)G_p^T(t) - G_e(t)R_e^{-1}(t)G_e^T(t)\right]K^{-1}(t_f,t),$$

$$K^{-1}(t_f,t_f) = a^2I. \tag{24}$$

$\Gamma(t)$ is obtained by solving a two-point boundary-value problem since the differential equation for $\Gamma$ is coupled to the one for $P(t)$, the covariance matrix of $\tilde{y}(t)$.

$$\dot{\Gamma} = \Gamma G_pR_p^{-1}G_p^T\Gamma + \Gamma\left[G_pR_p^{-1}G_p^TK^{-1} + PH^TQ_e^{-1}H\right] + \left[K^{-1}G_pR_p^{-1}G_p^T + H^TQ^{-1}HP\right]\Gamma + K^{-1}G_eR_e^{-1}G_eK^{-1}, \quad \Gamma(t_f) = 0 \tag{25}$$

$$\dot{P} = -G_pR_p^{-1}G_p^T(K^{-1}+\Gamma)P + P(\Gamma+K^{-1})G_pR_p^{-1}G_p^T - PH^TQ_e^{-1}HP, \quad P(t_o) = P_o \tag{26}$$

$P(t)$ is used by the evader in his Kalman-Bucy filter[3] which from $z(t)$

produces his optimal estimate of y(t).

$$\dot{\hat{y}} = -[G_p R_p^{-1} G_p^T - G_e R_e^{-1} G_e^T] K^{-1} \hat{y} + P H^T Q_e^{-1} [z - H\hat{y}], \quad \hat{y}(t_o) = \hat{y}_o \tag{27}$$

This solution was previously obtained and published by the authors[1]. Note that the Kalman-Bucy filter (26) and (27) is more complicated than usual requiring the solution of a T.P.B.V.P. (25) and (26). In the deterministic game it is only necessary to solve a differential equation with terminal conditions, (24). When determining the measurement correction-gain in the Kalman-Bucy filter, $P(t)H^T(t)Q_e(t)$, for ordinary, estimation problems, P(t) is calculated from a differential equation with only initial conditions[3]. Then why is it necessary to now solve a two-point boundary-value problem--certainly a computational headache--to determine these gains?

The answer to this question is a result of the fact that neither player can select his own strategy without giving thought to what strategy would be optimal for his opponent to employ. The pursuer selects his strategy $U^o$ by using the (9b), but in doing so he must check the validity of (9a). If one inequality is not satisfied, the other inequality is meaningless.

The pursuer's strategy consists of selecting two feedback gains to apply to the vectors y and $\tilde{y}$. In determining these gains, the pursuer cannot ignore how $\hat{y}$--and thus $\tilde{y}$ are determined.*

His opponent's strategy involves selecting an optimal estimate, $\hat{y}$, and the feedback gain to be used in conjunction with that estimate. The evader, therefore, cannot ignore how his opponent is using $\tilde{y}$.

Now note that the equation for P(t) cannot be derived from the equation for $\hat{y}(t)$ alone. The usual form of the Kalman-Bucy filter is

$$\dot{\hat{y}} = F\hat{y} + Gv + PH^T Q^{-1}(z - H\hat{y}), \quad \hat{y}(t_o) = \hat{y}_o \tag{28}$$

where

$$\dot{P} = FP + PF^T - PH^T Q^{-1} HP, \quad P(t_o) = P_o \tag{29}$$

This is not the relationship between (27) and (26). Comparing (27) to (28), $G_p R_p^{-1} G_p^T K^{-1}$ corresponds to F, $-G_e$ to G, and $-R_e^{-1} G_e^T K^{-1} y$ to v. However, using these correspondences, (26) does not compare directly with (29). This is because the estimation error, $\tilde{y}$, enters the system equation for which (27) is the estimation equation; the form of the system equation is really

$$\dot{y} = Fy + Gv + B\tilde{y}, \quad y(t_o) = y_o. \tag{30}$$

Consequently the differential equation for $\tilde{y}$ is

$$\dot{\tilde{y}} = (F + B)\tilde{y} + PH^T Q^{-1}(H\tilde{y} + w), \quad \tilde{y}(t_o) = \tilde{y}_o. \tag{31}$$

---

*It should be noted that of y, $\hat{y}$ and $\tilde{y}$ only two can be selected independently. See (18). The pursuer could use any pair of these three in his strategy: y and $\hat{y}$, y and $\tilde{y}$, or $\hat{y}$ and $\tilde{y}$.

and thus the one for P(t) is

$$\dot{P} = (F + B)P + P(F^T + B^T) - PH^TQ^{-1}HP, \qquad P(t_o) = P_o. \tag{32}$$

When B (which is $G_pD_p$) is optimized the result is dependent on P, and when P is optimized the result is dependent on B; thus it is only natural that the differential equations for the two, (25) and (26), are effectively coupled. That a two-point boundary-value problem is involved results from the fact that the known value of P(t) is the initial one, $P(t_o)$, while the fixed value of $\Gamma(t)$ is the terminal one, zero.*

Another counter-intuitive question concerning the results that might be raised in this: Suppose the pursuer knowing the filter employed by the evader decides to use different $C_p$ and $D_p$ feedback matrices on (16). Then clearly the "estimate" $\hat{y}(t)$ by the evader as calculated by (27) will not be optimal and can possibly be meaningless. Doesn't the pursuer stand to gain more this way? The reason that this possibility need not be of concern is due to the fact that the saddle point condition (2) has been satisfied for this game[1]. In fact by virtue of (2), we know the pursuer cannot gain anything and only stands to lose by the above suggestion.

In considering how the information set effects the performance of the games and the strategies of the two players, the chart in Figure 2 is helpful.# The stochastic game defined by (3), (10) and (6) with optimal strategies (16) and (17) is appropriate for the top row of the chart. If the pursuer is making the noisy measurements, the left column is appropriate.

I. B. Rhodes has considered the game where either or both of the players make no measurements[4]. The authors are aware that several colleagues including Rhodes have worked and are working on the problem where both players make noisy measurements and partial solutions have been obtained. The major difficulty with stochastic differential games where both sides receive imperfect information is the problem of "closure"[1]. Each side is faced with the vicious cycle of second guessing the action of the other side. One must continuously build estimators to estimate the errors of the estimators of ....... leading to an infinite state controller. In fact indication so far seems to say that this is the inevitable price one has to pay to solve this problem.

The solution to the stochastic game is also applicable to the deterministic game where both players have perfect information. This can be seen by taking the limiting process as $Q_e(t)$ approaches zero. P(t)

------------

*The terminal value of $\Gamma(t)$ is zero since this term does not enter explicitly into the criterion.

#To the authors' knowledge, this chart was first used by I. B. Rhodes approaches zero and thus $\tilde{y}(t)$ does likewise; $\hat{y}(t)$ approaches the correct value, $y(t)$. Thus the stochastic strategies, (16) and (17), approach the deterministic ones, (14) and (15) respectively.

The limiting process as $Q_e(t)$ approaches infinity produces the game and solution for the upper-right box in Figure 2. This means that the evader's estimate at time t is based entirely on the initial estimate, since no inflight information is obtained. Thus for $Q_e$ equal to infinity, the evader is employing a strictly open-loop strategy, while for $Q_e$ equal to zero his optimal strategy was closed-loop. In a sense then, the stochastic strategy represents a bridge between open-loop and closed-loop strategies.

IV. Performance as a Function of the Information Set

The optimized value of the criterion in terms of the game parameters is

$$J^{\circ} = \mathrm{Tr}\{\frac{1}{2} K^{-1}(t_f,t_o)Y(t_o) + \frac{1}{2}\Gamma(t_o)P(t_o) + \frac{1}{2}\int_{t_o}^{t_f} P(t)H^T(t)Q^{-1}(t)H(t)P(t)\Gamma(t)dt\} \quad (33)$$

where $Y(t)$ is the covariance of $y(t)$. The first term corresponds to the criterion for the deterministic game; it is a function of the initial separation but involves no stochastic parameters. Thus the second two terms can be called the relative criterion; they indicate the effect of the information parameters, $Q_e(t)$ and $P_o$, on the outcome.

For a particular (constant parameter) scalar game, the relative criterion is plotted in Figure 3. The quality $Q_e(t)/P_o$ is the independent variable, this ratio being the only relevant quantity defining the effect of the information. The relative J is always negative indicating a reduction in the evader's measurement capabilities--he being the maximizing player--and the pursuer's ability to capitalize on his opponent's errors. Note that as $Q_e(t)/P_o$ approaches zero the relative criterion approaches zero and thus the total criterion approaches the deterministic value.

As $Q_e(t)P_o$ approaches infinity, the relative criterion asympototically approaches a finite value. For large values of $Q_e(t)/P_o$, the evader's estimate is dependent more on his initial estimate than on his measurements, for the former is far more accurate. However, at such large values of $Q_e(t)/P_o$, further increases of this quantity do not permit the pursuer to take greater advantage of the evader's errors. Once the evader is effectively operating open-loop, the pursuer cannot further reduce the value of the criterion.

V. The Asymmetry of the Roles of Pursuer and Evader

The differing abilities to use open-loop strategies, as discussed in Section I above, is the most obvious asymmetry characteristic in the roles of the pursuer and evader. This phenomena has also been considered for the deterministic pursuit-evasion game presented in Section II. Specifically, it has been shown that the evader's open-loop strategy is non-optimal when $[\frac{I}{a^2} + M_p(t_f,t)]^{-1}$ fails to exist; this is obviously never true. The pursuer's open-loop strategy is non-optimal when $[\frac{I}{a^2} - M_e(t_f,t)]^{-1}$ fails to be finite--a distinct possibility. These results were obtained from the conjugate point condition for the opposing strategies[5].

For the stochastic game where the pursuer is the player making the noisy measurements, the forms of the optimal strategies are the symmetric images of those given in (16) and (17)

$$U^o: \quad u(t) = C_p(t)\hat{y}(t) \tag{34}$$

$$V^o: \quad v(t) = C_e(t)y(t) + D_e(t)\tilde{y}(t) \tag{35}$$

$C_p(t)$ and $C_e(t)$ are exactly as given in (19) and (20), while $D_e(t)$ is given by

$$D_e(t) = -R_e^{-1}(t)G_e^T(t)\Gamma_\rho(t) \tag{36}$$

This is again symmetric with the form for $D_p(t)$ given in (21).

However, the equations which determine $\Gamma_\rho(t)$ and $P_\rho(t)$* for this problem

$$\begin{aligned}\dot{\Gamma}_\rho &= -\Gamma_e G_e R_e^{-1} G_e^T \Gamma_e + \Gamma_\rho (P_\rho H^T Q_e^{-1} H - G_e R_e^{-1} G_e^T K^{-1}) \\ &\quad + (H^T Q_e^{-1} H P_\rho - K^{-1} G_e R_e^{-1} G_e^T - K^{-1} (G_p R_p^{-1} G_p^T) K^{-1}, \; \Gamma_\rho(t_f) = 0,\end{aligned} \tag{37}$$

$$\begin{aligned}\dot{P}_\rho &= G_e R_e^{-1} G_e^T (K^{-1} + \Gamma_\rho) P_\rho + P_\rho (K^{-1} + \Gamma_\rho) G_e R_e^{-1} G_e^T \\ &\quad - P_\rho H^T Q_e^{-1} H P_\rho, \qquad P_\rho(t_o) = P_{\rho o}\end{aligned} \tag{38}$$

have asymmetric properties when compared with (25) and (26). The roles of $G_p R_p^{-1} G_p^T$ and $G_e R_e^{-1} G_e^T$ are completely reversed and some--but not all-- of the signs are reversed.

Consequently, the characteristics of the families of curves produced by (25) and (26) are quite different from those of (37) and (38). This can be seen by comparing, Figures 4 and 5 which display curves of $\Gamma(t)$ and $\Gamma_\rho(t)$.

Specifically note that $\Gamma(t)$ is negative semi-definite in (25); the driving term, $+K^{-1}G_e R_e^{-1} G_e^T K^{-1}$, is destabilizing--when integrating backwards in time from $t_f$--and is balanced by the quadratic stabilizing term, $\Gamma G_p R_p^{-1} G_p^T \Gamma$. However, $\Gamma_\rho(t)$ in (37) is positive semi-definite and both the quadratic and the driving term are destabilizing.

Now as $Q_e(t)$ and $Q_p(t)$ approach infinity, take this limit in (25) and (37). $\Gamma(t)$ approaches $[\frac{I}{a^2} + M_p(t_f,t)]^{-1} - K^{-1}(t_f,t)$ while $\Gamma_p(t)$ approaches

----------

*The subscripts $\rho$ have been employed here to denote that these are the equations appropriate for the game where the pursuer makes noisy measurements.

$[\frac{I}{a^2} - M_e(t_f,t)]^{-1} - K^{-1}(t_f,t)$. Obviously in this case $\Gamma(t)$ cannot become infinite if $K^{-1}$ remains finite, while $\Gamma_p(t)$ certainly can. Note that here $\Gamma(t)$ approaches an expression, one term of which gives the open-loop conjugate point condition for the evader in the deterministic game; $\Gamma_p(t)$ approaches the corresponding expression for the pursuer. This means if a deterministic open-loop conjugate point exists for the pursuer then the solution for the corresponding T.P.B.V.P. for the stochastic case will be numerically difficult. This fact is clearly illustrated in Figure 5. This discussion again indicates the role of the strategies for the stochastic game as a bridge between closed-loop and open-loop strategies.

A graphic description of this phenomenon on the time axis is perhaps worth the proverbial thousand words. In Figure 6 it can be seen exactly where valid solutions can be found. Here it is assumed that the relative controllability condition

$$M_p(t_f,t) > M_e(t_f,t) \tag{39}$$

is satisfied, so that $K^{-1}(t_f,t)$ is always finite.

References

1. R. D. Behn and Y. C. Ho, "On a Class of Linear Stochastic Differential Games," IEEE Transactions on Automatic Control, AC-13, No. 3, June 1968.
2. Y. C. Ho, A. E. Bryson, Jr. and S. Baron, "Differential Games and Optimal Pursuit-Evasion Strategies," IEEE Transactions on Automatic Control, AC-10, No. 4, October 1965, pp. 385-389.
3. R. E. Kalman and R. Bucy, "New Results in Linear Filtering and Prediction," Journal of Basic Engineering, 1961, Vol. 83D, Transactions of the ASME, pp. 95-108.
4. I. B. Rhodes, Optimal Control of Dynamic Systems by Two Controllers with Conflicting Objectives, Ph.D. Thesis, Stanford University, Palo Alto, California, 1967.
5. R. D. Behn, Linear Stochastic Differential Games, Ph.D. Thesis, Harvard University, Cambridge, Massachusetts, 1968.

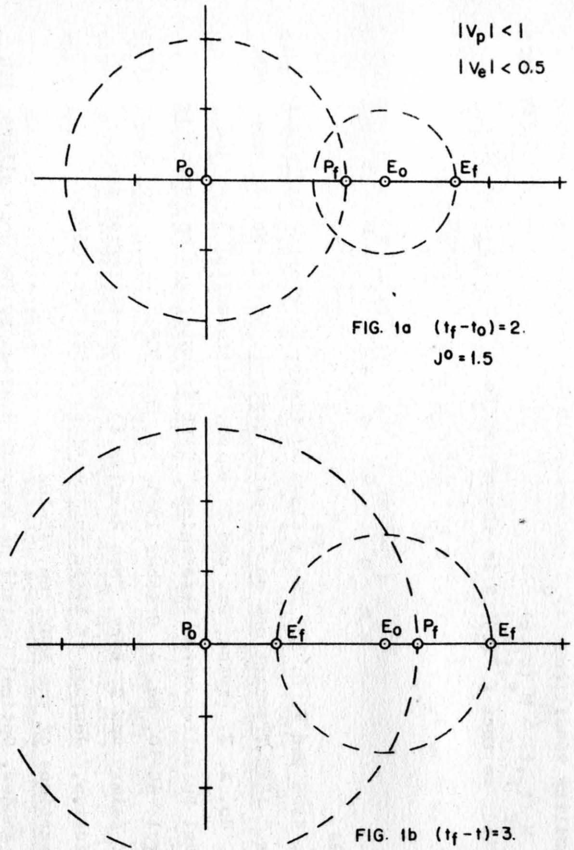



FIG. 1a $(t_f - t_0) = 2.$
$J^0 = 1.5$

FIG. 1b $(t_f - t) = 3.$
$J^0 = 1.0$
J = 2.0 FOR PURSUER OPEN-LOOP

FIG. 1 A VELOCITY LIMITED GAME

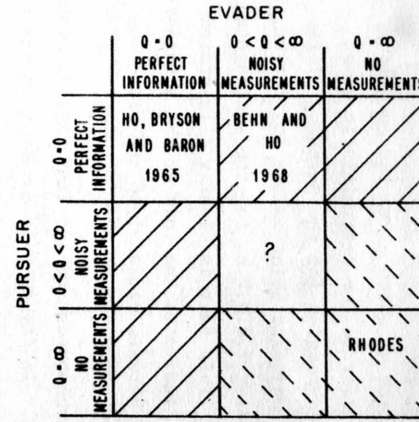



FIG. 2 A CHART OF STOCHASTIC GAMES WITH IMPERFECT INFORMATION

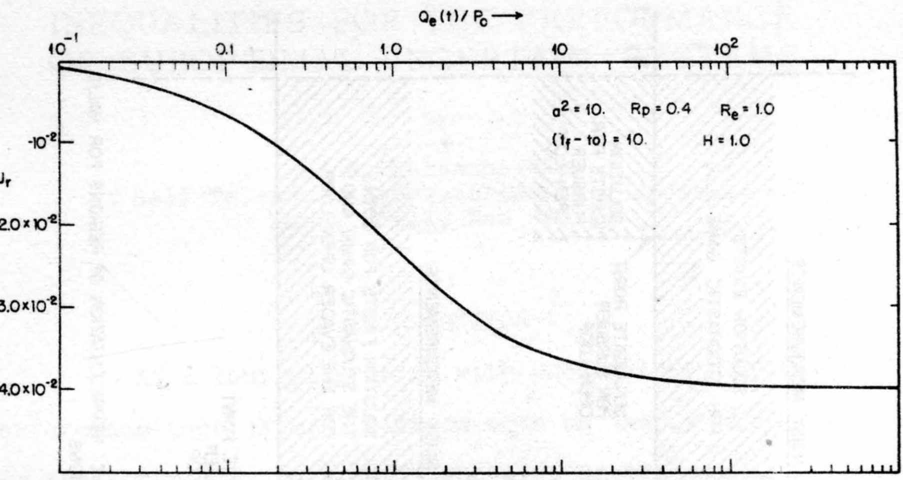



FIG. 3 THE RELATIVE CRITERION

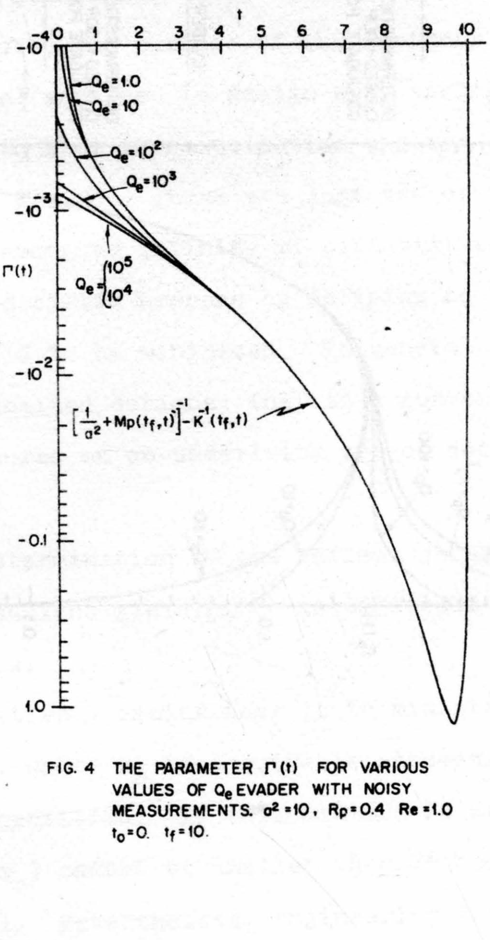



FIG. 4 THE PARAMETER $\Gamma(t)$ FOR VARIOUS VALUES OF $Q_e$ EVADER WITH NOISY MEASUREMENTS. $a^2 = 10$, $R_p = 0.4$ $Re = 1.0$ $t_o = 0.$ $t_f = 10.$

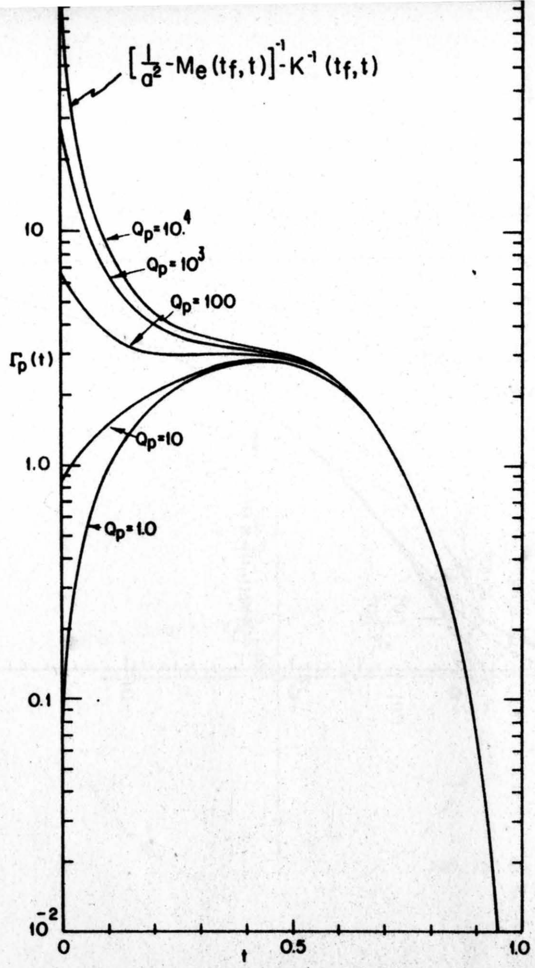



FIG. 5 THE PARAMETER $\Gamma_p(t)$ FOR VARIOUS VALUES OF Q. PURSUER WITH NOISY MEASUREMENTS. $a^2 = 10.$, $R_p = 0.4$, $R_e = 1.0$, $t_0 = 0.$, $t_f = 1.0$

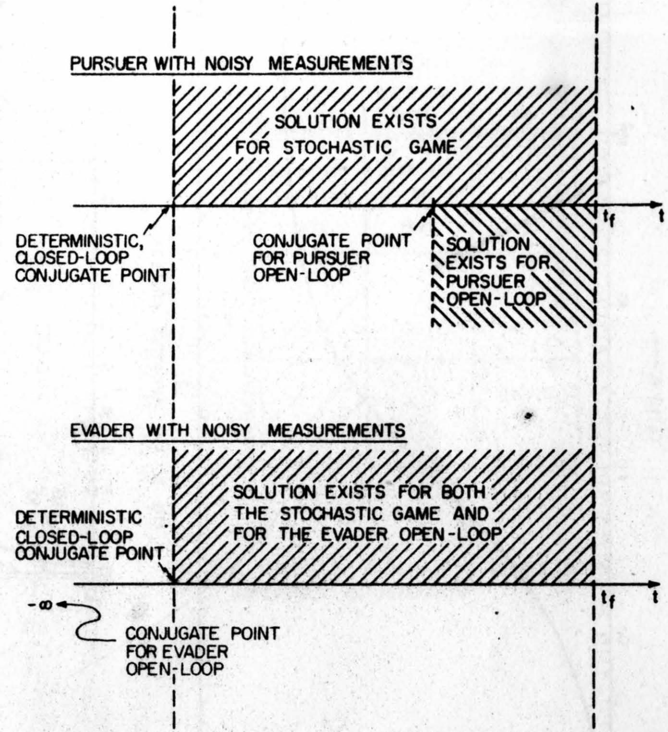



FIG. 6 GRAPHIC REPRESENTATION OF REGIONS FOR VALID SOLUTIONS

# INEQUALITIES FOR THE PERFORMANCE OF SUBOPTIMAL UNCERTAIN SYSTEMS

by

H. S. Witsenhausen
Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey
U. S. A.

## INTRODUCTION

In a control problem with uncertainty the index of performance (cost) $K(\alpha,\beta)$ depends both on the design $\alpha$ chosen for the controller in a set A and also on the values $\beta$ taken by the uncertain quantities in a set B.

Frequently the purpose of the designer is to minimize a number $J(\alpha)$ assigned to design $\alpha$ by taking the supremum over $\beta$ in B or the expectation under a probability measure on B, of $K(\alpha,\beta)$. These are just two of the most important ways, among an infinity of possibilities, of formalizing the decision process by defining on the set A a "supercriterion" J to be minimized. In general A could even be a set of randomized designs, that is a convex set of probability measures on an underlying set of deterministic designs.

The determination of the infimum $J^*$ of J over A and of optimal designs yielding $J^*$ is difficult outside a few special cases.

A relatively easier task is to minimize over A the function $K(\alpha,\beta_o)$ where $\beta_o$ is a suitable assumed value for the uncertain quantities. If this minimum is attained for $\alpha_o$ then $J_o \equiv J(\alpha_o)$ cannot be smaller than $J^*$ and will be large in general. Nevertheless, engineering intuition suggests that, under reasonable assumptions, $\alpha_o$ cannot be an extremely bad design. Inequalities are needed to shed light on the validity of this belief.

We consider problems in which

$$K(\alpha,\beta) \equiv \|f(\alpha) - g(\beta)\| \tag{1}$$

where $\|\cdot\|$ is a norm on a real linear space L into which the functions f,g map the sets A and B. In terms of $u = f(\alpha)$, $q = g(\beta)$ one redefines

$$K(u,q) = \|u-q\| \tag{2}$$

where u is to be selected in a set $U = f(A)$ and q is subject to the probability measure $\mu$ induced by $g(\beta)$ in the space L, or else is only known to belong to the set $Q = g(B)$.

This most elementary model applies to open loop control of perturbed linear systems with norm type criteria and of distinct nonlinear systems only coupled through the criteria as per (1). As an example consider the dynamic system

$$\dot{x}(t) = A(t)x(t) + B(t)\alpha(t) + C(t)\beta(t)$$
$$x(0) \text{ given}$$

here $\beta$ is a random process of known statistics and $\alpha$ is to be elected in the set of measurable time functions with values n a constraint set $\Omega$. The objective is to minimize the kpectation of the cost functional

$$K = \left[\int_0^1 \Big(x'(t)H(t)x(t) + \alpha'(t)M(t)\alpha(t)\Big)dt + x'(1)Qx(1)\right]^{1/2}$$

The space L is the space of pairs $(x,\alpha)$ where x is continuous and $\alpha$ measurable with $\int_0^1 \alpha'(t)M(t)\alpha(t)dt < \infty$. In this space K defines either a norm or a pseudo-norm. (In the latter case K defines a norm on a quotient space.) By the principle of superposition this criterion has the form (1) with f and g linear-plus-constant operators. (For the minimax problem it is assumed that $\beta$ belongs to a given set S of integrable function and the supremum over this set replaces the expectation.)

If the data concerning $\beta$ is invariant under sign change there is symmetry about the typical function $\beta \equiv 0$. If in addition $\Omega$ is convex then theorem 5 below will apply.

THE STOCHASTIC CASE

Let $\Sigma$ be a $\sigma$-algebra of subsets of the real normed space L, such that the linear operations and the norm are measurable. The uncertain point q is subject to a probability measure $\mu$ on $\Sigma$, with

$$E\{\|q\|\} < \infty \tag{3}$$

Then

$$J(x) = E\{\|x-q\|\} \tag{4}$$

defines J as a real valued function on L, Lipschitz continuous with the constant 1.

Define

$$J^* = \inf\{J(u)|u \in U\} \tag{5}$$

The natural choice for $q_0$ is the expectation of q, which we assume exists in the weak Dunford-Pettis sense, that is, for any continuous linear functional p on L

$$E\{\langle p,q\rangle\} = \langle p,q_o\rangle \tag{6}$$

where $\langle p,q\rangle$ denotes the value at q of linear functional p. Then because of the representation

$$\|x\| = \sup\{\langle p,x\rangle \mid p \in L^*, \|p\| \leq 1\} \tag{7}$$

(6) implies that for all x in L

$$\|x-q_o\| \leq J(x) \tag{8}$$

Now suppose that an element $u_o$ of U has been found such that for all u in U

$$\|u_o-q_o\| \leq \|u-q_o\| \tag{9}$$

and let $J_o = J(u_o) \geq J^*$. Bounds of the form $J_o \leq kJ^*$ are sought where k is as small as possible under the basic assumptions above and various combinations of additional assumptions. Three additional assumptions will be considered:

Assumption C: The set U is convex.

Assumption S: $q_o$ is a center of symmetry for the measure $\mu$, that is $G \in \Sigma$ implies

$$\mu(G) = \mu(2q_o-G) \tag{10}$$

from which

$$J(x) = J(2q_o-x) \tag{11}$$

Assumption P: The norm satisfies the parallelogram law

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{12}$$

which implies that L is a prehilbert space with the inner product.

$$x \cdot y = \frac{\|x+y\|^2 - \|x-y\|^2}{4} \tag{13}$$

<u>Theorem 1:</u> Under the basic assumptions alone or augmented by C or augmented by P (but not by both) the smallest number k such that $J_o \leq kJ^*$ is three.

Proof: By the triangle inequality, for $u \in U$, $q \in L$

$$\|u_o - q\| \leq \|u_o - q_o\| + \|u - q_o\| + \|u - q\|$$

$$\leq 2\|u - q_o\| + \|u - q\| \quad \text{by (9)}$$

$$\leq 2J(u) + \|u - q\| \quad \text{by (8)}$$

taking the expectation over q

$$J_o \leq 2J(u) + J(u) = 3J(u)$$

and the infimum over u in U

$$J_o \leq 3J^*$$

Hence the bound holds. To see that it is the best with assumption C let L be $R^2$ with the $\ell_1$ norm. Let $U = \{(x,y) | y - x = 1\}$, let $q = (-1,0)$ with probability $1 - \varepsilon$ and $q = \left(\frac{1}{\varepsilon} - 1, 0\right)$ with probability $\varepsilon$. Let $q_o = (0,0)$, $u_o = (0,1)$. Then $J^* = 1$ and $J_o = 3 - 2\varepsilon$ while C holds. With assumption P let $L = R^1$, $q = -1$ with probability $1 - \varepsilon$ and $\frac{1}{\varepsilon} - 1$ with probability $\varepsilon$. Let $U = \{-1, +1\}$, $q_o = 0$, $u_o = +1$, then $J^* = 1$ and $J_o = 3 - 4\varepsilon$ which completes the proof.

Note that if P and C both hold then (9) implies

$$(u - u_o) \cdot (q_o - u_o) \leq 0 \tag{14}$$

Theorem 2: With assumptions P and C the smallest number k such that $J_o \leq kJ^*$ is two.

Proof: For $u \in U$,

$$\|u-q_o\|^2 = \|(u-u_o) + (u_o-q_o)\|^2$$

$$= \|u-u_o\|^2 + \|u_o-q_o\|^2 + 2(u-u_o)\cdot(u_o-q_o)$$

$$\geq \|u-u_o\|^2 \quad \text{by (14)}$$

Hence

$$\|u-u_o\| \leq \|u-q_o\| \leq J(u) \quad \text{by (8)}$$

Thus

$$\|q-u_o\| \leq \|q-u\| + \|u-u_o\| \leq \|q-u\| + J(u)$$

Taking expectation over q

$$J_o \leq 2J(u)$$

and infimum over u

$$J_o \leq 2J^*$$

This bound is sharp: take $L = U = R^1$, $q = -1$ with probability $1 - \varepsilon$ and $q = \frac{1}{\varepsilon} - 1$ with probability $\varepsilon$, $q_o = u_o = 0$. Then $J^* = 1$, $J_o = 2 - 2\varepsilon$ and the theorem is proved.

The symmetry assumption S is naturally a powerful one as the following theorems show.

Theorem 3: With assumption S or with S and C the smallest number k such that $J_o \leq kJ^*$ is two.

Proof: Define $F_u(q) = \|u-q\| + \|u-2q_o+q\|$. Then for q in L, u in U

$$2\|q-q_o\| = \|(u-q) - (u-2q_o+q)\| \leq F_u(q)$$

$$2\|u-q_o\| = \|(u-q) + (u-2q_o+q)\| \leq F_u(q)$$

Hence

$$F_u(q) \geq 2 \max(\|u-q_o\|, \|q-q_o\|) \tag{15}$$

On the other hand

$$F_{u_o}(q) = \|(u_o-q_o) + (q_o-q)\| + \|(u_o-q_o) - (q_o-q)\|$$

$$\leq 2\|u_o-q_o\| + 2\|q_o-q\|$$

$$\leq 4 \max(\|u_o-q_o\|, \|q_o-q\|)$$

and, using (9) and (15)

$$F_{u_o}(q) \leq 2F_u(q) \tag{16}$$

Now

$$J_o = E\{\|u_o-q\|\}$$

$$= E\{\|u_o-2q_o+q\|\}, \quad \text{by S}$$

$$= \frac{1}{2} E\left\{F_{u_o}(q)\right\}$$

$$\leq E\{F_u(q)\} \quad \text{by (16)}$$

$$= 2E\{\|u-q\|\} \quad \text{by S}$$

$$= 2J(u)$$

Taking the infimum over u in U yields $J_o \leq 2J^*$. Now let $L = R^2$ with $\ell_1$ norm. Let $U = \{(x,y)|x + y = 1\}$, $q = (-1,0)$ or $(1,0)$ with probability 1/2, $q_o = (0,0)$, $u_o = (0,1)$. Then $J_o = 2$, $J^* = 1$ while assumptions S and C are satisfied.

Theorem 4: Under assumptions S and P the smallest number k such that $J_o \leq kJ^*$ is $\sqrt{2}$.

Proof:

$$F^2_{u_o}(q) = (\|u_o - q\| + \|u_o - 2q_o + q\|)^2$$

$$\leq 2\|u_o - q\|^2 + 2\|u_o - 2q_o + q\|^2$$

$$= \|2(u_o - q_o)\|^2 + \|2(q_o - q)\|^2 \quad \text{by P}$$

Hence

$$F_{u_o}(q) \leq 2(\|u_o - q_o\|^2 + \|q_o - q\|^2)^{\frac{1}{2}}$$

$$\leq 2\left(2\ \max(\|u_o - q_o\|^2, \|q_o - q\|^2)\right)^{1/2}$$

$$= 2\sqrt{2}\ \max(\|u_o - q_o\|, \|q_o - q\|)$$

$$\leq \sqrt{2}\ F_u(q) \quad \text{by (9) and (15)}$$

And this implies $J_o \leq \sqrt{2}\ J^*$ in exactly the same way as in theorem 3. Now let $L = E^2$, $U = \{(x,y) | x^2 + y^2 = 1\}$ $q = (-1,0)$ or $(1,0)$ with probability 1/2, $q_o = (0,0)$, $u_o = (0,1)$. Then $J^* = 1$ and $J_o = \sqrt{2}$. Thus the bound is sharp.

Theorem 5: Under assumptions S, P and C the smallest number k such that $J_o \leq kJ^*$ is $2/\sqrt{3}$.

Proof: For an Euclidean triangle ABC whose median AM makes an angle no less than 90° with AB (or AC) the sides satisfy

$$b + c \leq \frac{2}{\sqrt{3}}\, a$$

as is easily shown by calculus. Therefore $x \cdot (y - x) \geq 0$ implies

$$\|x+z\| + \|x-z\| \leq \frac{2}{\sqrt{3}} (\|y+z\| + \|y-z\|) \qquad (17)$$

for x,y,z in $E^2$.

Orthogonal projection of y on a plane containing x and z extends this result to $E^3$ and thus to any prehilbert space.

Letting $x = u_o - q_o$, $y = u - q_o$, $z = q - q_o$, with $u \in U$, the relation $x \cdot (y-x) \geq 0$ holds by (14). This implies (17) which can be written

$$F_{u_o}(q) \leq \frac{2}{\sqrt{3}} F_u(q)$$

from which $J_o \leq \frac{2}{\sqrt{3}} J^*$ follows just as in theorem 3. To show that the bound is sharp let $L = E^2$, $q = (0,0)$ or $(2, 2\sqrt{2})$ with probability 1/2. $U = \{(a,0) | a \in R\}$, $q_o = (1, \sqrt{2})$, $u_o = (1,0)$ then $J_o = 4$ and $J^* = 2\sqrt{3}$ which completes the proof.

Note that with assumption P the problem of minimizing $E\{\|u-q\|^2\}$ leads to $k = 1$. This differs from the results obtained here because squaring does not commute with expectation.

THE MINIMAX CASE

Instead of a probability measure for q, assume that q is only known to belong to a set Q in L. Then (3) is replaced by the assumption that Q is bounded, and

$$J(x) = \sup\{\|x-q\| \mid q \in Q\} \qquad (18)$$

replaces (4). J is again Lipschitz continuous with constant 1. J* is defined by (5). The basic assumption about $q_o$ is simply that it be an element of Q (or of the convex hull of Q) which implies (8). With $u_o$ assumed to satisfy (9) one now seeks the smallest k such that $J_o = J(u_o) \leq kJ^*$ under the basic assumptions and their strengthening by C, P and the revised form of S: $q_o$ is a center of symmetry for Q (or for the convex hull of Q).

The results are summarized in:

Theorem 6: The sharp bounds are given for each combination of assumptions by the following table

| Assumptions: | - | C | P | S | PC | SC | SP | SPC |
|---|---|---|---|---|---|---|---|---|
| Bound: | 3 | 3 | 3 | 2 | 2 | 2 | $\sqrt{2}$ | $2/\sqrt{3}$ |

These results have been established earlier[6] with as elementary proofs as for theorems 1 to 5.

Notice the equality between the stochastic and minimax bounds, in all eight cases. This is one form of a duality which holds in a wider field, as will be reported elsewhere.

THE TWO-STAGE CASE

To introduce feedback into the situation, consider the problem of minimizing the expectation of

$$\|u_1 - q_1 + \gamma(u_1 - q_1) - q_2\| .$$

where $q_1, q_2$ are independent random elements of L with probability measures $\mu_1, \mu_2$ and means $q_{10}, q_{20}$. The element $u_1$ is to be selected from a set $U_1$ and the function $\gamma$ from the set of all measurable functions mapping L into a set $U_2$.

Suppose $u_{10}$ and $\gamma_o$ are such that

$$\|u_{10}+u_{20}-q_{10}-q_{20}\| = \min\{\|u_1+u_2-q_{10}-q_{20}\| \mid u_1 \in U_1,\ u_2 \in U_2\} \tag{19}$$

where

$$u_{20} \equiv \gamma_o(u_{10}-q_{10}) \tag{20}$$

and such that for all $x \in L$ one has $\gamma_o(x) \in U_2$ and

$$\|x + \gamma_o(x) - q_{20}\| = \min\{\|x+u-q_{20}\| \mid u \in U_2\} \tag{21}$$

Then the naive open-loop policy is the pair $(u_{10}, \gamma_{oo})$ where $\gamma_{oo}(x) = u_{20}$ for all x. If this policy is used it will result in an expected cost

$$J_o = E\{\|u_{10}+u_{20}-q_1-q_2\|\} \tag{22}$$

The naive closed-loop policy, also known as the synthesis of optimal control in feedback form, is the pair $(u_{10}, \gamma_o)$ which leads to the expected cost.

$$J_f = E\{\|u_{10} - q_1 + \gamma_o(u_{10}-q_1) - q_2\|\} \tag{23}$$

One of the most entrenched beliefs of control engineers is that $J_f \leq J_o$. Though this is often true it is known[5] not to hold generally. The smallest numer $\varphi$ such that $J_f \leq \varphi J_o$ under some combination of assumptions will be called the "fooling factor" for these assumptions.

An upper bound on the fooling factor can be obtained from the single-stage result by means of the following "conversion theorem."

<u>Theorem 7:</u> Suppose $U_2$, $\mu_2$, $q_{20}$ and the norm satisfy one of the combinations of assumptions required for the single-stage

data in one of theorems 1 through 5. Let k be the corresponding bound. Then $J_f \leq kJ_o$ in the two-stage problem, that is, $\varphi \leq k$.

Proof: For x arbitrary and fixed in L consider the single-stage problem of selecting $v \in V = x + U_2$ to minimize the expectation of $\|v-q_2\|$, where $q_2$ has distribution $\mu_2$ with mean $q_{20}$. Because of relation (21) the element $v_o = x + \gamma_o(x) \in V$ is suboptimal in the sense of (9). Since V differs from $U_2$ only by a translation the single-stage theorem applies and gives the inequality

$$\underset{q_2}{E} \{\|v_o-q_2\|\} \leq k \underset{q_2}{E} \{\|v-q_2\|\}$$

for all v in V.

Since x was arbitrary and $V = x + U_2$ one has for all x in L and all $u_2$ in $U_2$

$$\underset{q_2}{E} \{\|x + \delta_o(x) - q_2\|\} \leq k \underset{q_2}{E} \{\|x+u_2-q_2\|\}$$

Letting $x = u_{10} - q_1$ and $u_2 = u_{20}$ gives

$$\underset{q_2}{E} \{\|u_{10} - q_1 + \gamma_o(u_{10}-q_1) - q_2\|\} \leq k \underset{q_2}{E} \{\|u_{10}-q_1+u_{20}-q_2\|\}$$

Taking the expectation over $q_1$ yields, by (22) and (23). $J_f \leq kJ_o$ as claimed.

Note that no assumptions on the first stage data are used in the above proof. One would therefore expect that with such assumptions the inequality $\varphi \leq k$ is strict. But in fact equality holds in the most interesting case.

Theorem 8: Suppose the first and second stage data satisfy assumption S, P and C. Then the fooling factor $\varphi$ is $2/\sqrt{3}$.

Proof: The bound $\varphi \leq 2/\sqrt{3}$ holds by theorem 5 via theorem 7. That equality holds is shown by an example published elsewhere.[5]

The minimax counterparts of theorems 7 and 8 are true and admit of similar proofs.[6]

If one considers the expectation of the square of a norm satisfying P, that is "quadratic criteria," then $\varphi = 1$ by certainty equivalence.[1]

## CONCLUSIONS

The subject of inequalities for suboptimal performance has received attention only recently.[2-4] Because of the limitations of automatic computers, this subject is of lasting practical interest, at least to obtain a feeling for the implications of various properties and the reliance that can be put on suboptimal designs. Nearly everything remains to be done in this area. The equality between the stochastic and minimax bounds is an interesting and not entirely expected phenomenon.

## REFERENCES


1. Dreyfus, S. E., "Some types of optimal control of stochastic systems," J. SIAM Control v. 2, p. 120, 1962.
2. Durbeck, R. D., "An Approximation Technique for Suboptimal Control," IEEE Trans. AC-10, p. 144, 1965.
3. McClamroch, N. H., "A Result on the Performance Deterioration of Optimum Systems," IEEE Trans. AC-12, p. 209, 1967.

4. Rissanen, J. J., "Performance Deterioration of Optimum Systems," IEEE Trans. AC-11, p. 530, 1966.

5. Thau, F. E., and Witsenhausen, H. S., "A Comparison of Closed-Loop and Open-Loop Optimum Systems," IEEE Trans. AC-11, p. 619, 1966.

6. Witsenhausen, H. S., "Minimax Control of Uncertain Systems," M.I.T. Report ESL-R-269, May 1966; (N.A.S.A. CR-77050; Accession number N66-33441).