

Submitted to IEEE Trans.
on Autom. Control

February, 1976

ESL-P-648

OPTIMAL CONTROL OF NOISY FINITE-STATE MARKOV PROCESSES

Adrian Segall

The author is with the Electronic Systems Laboratory, Research Laboratory of Electronics and the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge Mass. 02139. This research was supported by the Advanced Research Project Agency of the Department of Defense (monitored by ONR) under contract Nr. N00014-75-C-1183 and by NSF under Grant ENG75-14103.

ABSTRACT

The paper treats the problem of optimal control of finite-state Markov processes observed in noise. Two types of noisy observations are considered: additive white Gaussian noise and jump-type observations. Sufficient conditions for the optimality of a control law are obtained similar to the stochastic Hamilton-Jacobi equation for perfectly observed Markov processes. An illustrative example concludes the paper.

1. Introduction

The theory of optimal control of perfectly observed Markov processes is by now well understood and quite general optimality conditions are available (see e.g. [1]). When the observations are perturbed by noise however, the problem becomes much more difficult and various models have to be considered separately. For the linear Gaussian model with additive white Gaussian observation noise, sufficient conditions for optimality of a causal control law have been obtained in [2]. The purpose of the present paper is to consider another class of models arising often in practical situations: the class of noisy observations of a controlled finite-state Markov process. Two types of noise will be considered: first we treat observations of the system state mixed with additive white Gaussian noise and secondly the situation in which the finite-state Markov process modulates the rate of a point process [3].

Estimation schemes for such processes have been provided in several previous works [4]-[7]; here we show how these estimates are to be incorporated in the optimal control scheme. It was shown in those works that for finite-state Markov processes, a recursive explicit finite-dimensional expression can be obtained for the estimate of the state, which moreover completely characterizes the posterior distribution. It is natural to expect therefore that this estimate is a sufficient statistic for the optimal control, and the main point of the present paper is to prove this. The main result is a differential equation providing a sufficient condition simultaneously for the

optimality of the control law and for the separation of control and estimation. The equation is similar to the stochastic Hamilton-Jacobi equation for perfectly observed Markov processes [1]. The latter is restated below for easy reference and comparison to our results in Sec. III.

Theorem 1 [1]

Consider a Markov process $x(\cdot)$ with backward Kolmogorov operator $\mathcal{L}(u)$ dependent on the control law $u(\cdot) \in \mathcal{U}$. Then a sufficient condition for the control u^* to minimize the criterion

$$C = E \left[\int_0^T L(t, x(t), u(t)) dt + \phi(x(T)) \right] \quad (1.1)$$

is provided by the stochastic Hamilton-Jacobi equation

$$-\frac{\partial V^*(t, x)}{\partial t} = \min_{u \in \mathcal{U}} \left\{ L(t, x, u) + \mathcal{L}(u) \left[V^*(t, x) \right] \right\} \quad (1.2a)$$

$$V^*(T, x) = \phi(x) \quad (1.2b)$$

Here $V^*(t, x)$ is the optimal expected cost-to-go given that $x(t)=x$, and u^* is obtained by the minimization in (2a).

The motivation for the present study has been provided by an earlier work [8], where the problem of dynamic file allocation in a computer network gave rise to a controlled finite-state Markov model. The Markov process had two components, one which was perfectly observable and one

which could be observed only through a point process, the rate of which was modulated by the Markov process. For this model a separation property was shown to hold and a differential - difference equation providing a sufficient condition for optimality has been derived. The present paper provides a generalization of the results in [8] to arbitrary finite-state Markov processes.

We may mention here that the discrete-time analog of our results can be proven in a straightforward manner using dynamic programming. This has been done in [9], where efficient algorithms are also proposed for solving the dynamic programming equations.

In Section II we present the models for the state and observations processes and state the optimal control problem. Section III presents the recursive estimation schemes for these models and Sections IV, V contain the sufficient conditions for optimality of the control law and for the separation of control and state estimation. A simple example is described in Section VI.

II. The Basic Models

Consider a controlled continuous-time finite-state Markov process $z(t)$ with states

$$\rho^{(1)} < \rho^{(2)} < \dots < \rho^{(n)}, \quad (2.1)$$

initial occupancy probabilities

$$\pi_i = \text{Prob} \{z(0) = \rho^{(i)}\} \quad (2.2)$$

and transition probabilities

$$P \left\{ z(t+dt) = \rho^{(j)} \mid z(t) = \rho^{(i)} \right\} = \begin{cases} q_{ij}(t)dt + o(dt) & j \neq i \\ 1 + q_{ii}(t)dt + o(dt) & j = i \end{cases} \quad (2.3)$$

where

$$q_{ii}(t) = - \sum_{j \neq i} q_{ij}(t) \quad (2.3a)$$

The control $\underline{u}(\cdot)$ taking values in $U \subset R^m$ is affecting the signal $z(\cdot)$ by controlling its transition probabilities, so that in fact

$$q_{ij}(t) = q_{ij} \left(t, \underline{u}(t) \right) \quad i, j = 1, \dots, n \quad (2.4)$$

The Markov process $z(t)$ and all other processes with jumps are taken throughout the paper to be right-continuous. As in [6], it is useful to introduce the notations

$$x_i(t) = \begin{cases} 1 & \text{if } z(t) = \rho^{(i)} \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

$$\underline{x}(t) = \left[x_1(t), \dots, x_n(t) \right]^T \quad (2.6)$$

$$\underline{Q}(t, \underline{u}(t)) = \left\{ q_{ij}(t, \underline{u}(t)) \right\} \quad i, j=1, \dots, n \quad (2.7)$$

Throughout this paper, the controlled Markov process $z(t)$, or equivalently the vector $\underline{x}(t)$, will be the model for the signal process. The problem is to find conditions for optimality of the control law, when observations of the process $z(t)$ are corrupted by noise. Two types of observations processes are considered, and the models are described below.

Additive white Gaussian noise

Suppose the (m -dimensional) observations process $\underline{y}(t)$ satisfies

$$d\underline{y}(t) = \underline{h}(t, z(t)) dt + d\underline{w}(t) \quad (2.8)$$

where $\underline{w}(t)$ is a vector-valued Wiener process with

$$E\underline{w}(t) = 0 \quad E\underline{w}(t)\underline{w}^T(t) = \underline{I} \cdot t \quad (2.9)$$

and such that the "future" increments of $\underline{w}(\cdot)$

$$\underline{w}(\tau) - \underline{w}(t) \quad \tau \geq t \quad (2.10)$$

are independent of the "past"

$$\mathcal{B}_t = \sigma \left\{ \underline{z}(s), \underline{y}(s), \quad s \leq t \right\} \quad (2.11)$$

Clearly we can write (2.8) in the form

$$\begin{aligned} d\underline{y}(t) &= \sum_{i=1}^n \underline{h} \left(t, \rho^{(i)} \right) x_i(t) dt + d\underline{w}(t) \\ &= \underline{H}(t)\underline{x}(t)dt + d\underline{w}(t) \end{aligned} \quad (2.12)$$

where $\underline{H}(t)$ is an $(m \times n)$ -dimensional matrix with columns $\left\{ \underline{h} \left(t, \rho^{(i)} \right), i=1, \dots, n \right\}$.

We may also note that (2.8) can be significantly generalized by allowing \underline{h} and the covariance of \underline{w} at time t to be dependent of $\underline{y}(t)$, but this will only complicate the expressions below, without adding much to their contents.

We can also observe that with (2.11) and (2.5), we can write the signal evolution (2.3) as

$$\begin{aligned}
 E \left\{ x_j(t+dt) \mid \mathcal{B}_t \right\} &= \text{Prob} \left\{ z(t+dt) = \rho^{(j)} \mid \mathcal{B}_t \right\} = \\
 &= \sum_{i \neq j} q_{ij}(t, \underline{u}) x_i(t) dt + \left(1 + q_{jj}(t, \underline{u}) dt \right) x_j(t) \quad (2.13)
 \end{aligned}$$

or in vector form

$$E \left\{ d\underline{x}(t) - \underline{Q}^T(t, \underline{u}) \underline{x}(t) dt \mid \mathcal{B}_t \right\} = 0. \quad (2.14)^1$$

This says that the signal equation can be written as

$$d\underline{x}(t) = \underline{Q}^T(t, \underline{u}) \underline{x}(t) dt + d\underline{v}(t) \quad (2.15)$$

where \underline{v} is a martingale w.r.t. \mathcal{B} . (see [10]). Observe that since \underline{w} is a continuous martingale and \underline{v} is a compensated sum of jumps martingale their conditional covariance process ([10], formulas (63)-(69)) is

$$\langle \underline{v}, \underline{w} \rangle = 0 \quad (2.16)$$

Counting observations

Suppose the finite-state Markov process $z(\cdot)$ modulates the rate of a counting process $N(\cdot)$. This model is discussed at length in [3], [11],

[6 , Sec. VIII], but we may mention here only that this situation is quite similar to the model (2.8), where the signal $z(\cdot)$ modulates the mean of a Wiener process (because conditioned on $z(\cdot)$, we can regard the observations process $y(\cdot)$ as being a Wiener process with mean $g(\cdot, z(\cdot))$ and variance \underline{I}).

In general, the jumps of the signal $z(\cdot)$ and of the observations process $N(\cdot)$ may not be independent, so that one has to consider the following transition probabilities

$$\text{Prob} \left\{ N(t+dt) - N(t) = k, z(t+dt) = \rho^{(j)} \mid z(t) = \rho^{(i)} \right\} =$$

$$= \begin{cases} s_{ij}(t) dt & j \neq i, k = 1 \\ \left[q_{ij}(t, \underline{u}) - s_{ij}(t) \right] dt & j \neq i, k = 0 \\ \left[h(t, \rho^{(i)}) + s_{ii}(t) \right] dt & j = i, k = 1 \\ 1 - \left(h(t, \rho^{(i)}) - q_{ii}(t, \underline{u}) - s_{ii}(t) \right) dt & j = i, k = 0 \end{cases} \quad (2.17)$$

where

$$q_{ii}(t, \underline{u}) = - \sum_{j \neq i} q_{ij}(t, \underline{u}) \quad ; \quad s_{ii}(t) = - \sum_{j \neq i} s_{ij}(t) \quad (2.17b)$$

If we define now

$$\mathcal{B}_t = \sigma \left\{ N(s), z(s), \quad s \leq t \right\}, \quad (2.18)$$

it can be seen as in (2.15) (see also [6 , Sec. VII]), that (2.17) can be written

$$d\underline{x}(t) = \underline{Q}^T(t, u) \underline{x}(t) dt + d\underline{v}(t) \quad (2.19a)$$

$$dN(t) = \underline{h}(t) \underline{x}(t) dt + dw(t) \quad (2.19b)$$

where \underline{v} , and w are \mathcal{B} -martingales, with conditional covariance process [10, formula (63)]

$$\langle \underline{v}, w \rangle_t = \int_0^t \underline{S}^T(\tau) \underline{x}(\tau) d\tau \quad (2.20)$$

Admissible controls

Suppose an arbitrary control law u is used and observations $y(\cdot)$ or $N(\cdot)$ are taken. We denote the corresponding σ -fields

$$\mathcal{F}_t(\underline{u}) = \sigma\{y(s), \quad s \leq t\} \quad (2.21a)$$

for Gaussian noise and

$$\mathcal{F}_t(\underline{u}) = \sigma\{N(s), \quad s \leq t\} \quad (2.21b)$$

for counting observations. Clearly these fields are dependent on the control law \underline{u} and this is made explicit in (2.21a) and (2.21b). A control law \underline{u} will be said to be admissible if its value at time t is dependent only on the observations up to time t^- or more rigorously if $u(t)$ is $\mathcal{F}_t(u)$ -predictable [10] for all t . The set of all admissible control laws will be denoted by \mathcal{U} , namely

$$\mathcal{U} = \{ u \mid u(t) \in U \text{ and } u(t) \text{ is } \mathcal{F}_t(u)\text{-predictable for all } t \} \quad (2.22)$$

Cost functional

We consider a cost functional of the form

$$\begin{aligned} J(\underline{u}) &= E \left[\int_0^T L(t, z(t), \underline{u}(t)) dt + \phi(z(T)) \right] \\ &= E \left[\int_0^T \sum_{i=1}^n L(t, \rho^{(i)}, \underline{u}(t)) x_i(t) dt + \sum_{i=1}^n \phi(\rho^{(i)}) x_i(T) \right] \end{aligned} \quad (2.23)$$

or in vector form

$$J(\underline{u}) = E \left[\int_0^T \underline{L}^T(t, \underline{u}(t)) \underline{x}(t) dt + \underline{\phi}^T \underline{x}(T) \right] \quad (2.24)$$

Statement of the problem

The problem is to find sufficient conditions for the optimality of a control law $u^* \in \mathcal{U}$ controlling the signal via (2.15) or (2.19a), with observations

(2.12) or (2.19b) respectively and cost functional (2.24).

III Recursive estimation formulas.

In sections IV and V we present Bellman-type sufficient conditions for a control law \underline{u}^* to be optimal in the class of admissible controls. It is also shown that if these conditions are satisfied, the optimal control $\underline{u}^* \in \mathcal{U}$ is of the form

$$\underline{u}^* \left(t, \hat{\underline{x}}^*(t-) \right), \quad (3.1)$$

where

$$\hat{\underline{x}}^*(t) = E \left[\underline{x}(t) \mid \mathcal{F}_t(\underline{u}^*) \right] \quad (3.2)$$

is the minimum mean-squared error estimate of $\underline{x}(t)$ given the observations $\mathcal{F}_t(\underline{u}^*)$ and $\hat{\underline{x}}^*(t-)$ is the value of this estimate at $t-$. Filtering equations for finite-state Markov processes observed in white Gaussian noise and in counting observations have been obtained in [4 , formula (5)] and [6 , Sec. VII] respectively. With the notation introduced in Sec. II and for an arbitrary admissible control law u , the estimate

$$\hat{\underline{x}}(t) = E [\underline{x}(t) \mid \mathcal{F}_t(\underline{u})] \quad (3.3)$$

is given by the recursive formula

$$d\hat{\underline{x}}(t) = \underline{Q}^T(t, \underline{u}) \hat{\underline{x}}(t) dt + \underline{K} \left(t, \hat{\underline{x}}(t-) \right) d\underline{v}(t) \quad (3.4)$$

where for Gaussian noise (2.12) the coefficients are [4]

$$\underline{K} \left(t, \hat{\underline{x}}(t-) \right) = \underline{P} \left(\hat{\underline{x}}(t-) \right) \underline{H}^T(t) \quad (3.5)$$

$$d \underline{v}(t) = d\underline{y}(t) - \underline{H}(t) \hat{\underline{x}}(t) dt \quad (3.6)$$

and for counting observations (2.19), they are [6]

$$\underline{K} \left(t, \hat{\underline{x}}(t-) \right) = \left[\underline{P} \left(\hat{\underline{x}}(t-) \right) \underline{h}^T(t) + \underline{C} \left(\hat{\underline{x}}(t-) \right) \right] \cdot \frac{1}{\underline{h}(t) \hat{\underline{x}}(t-)} \quad (3.7)$$

$$d \underline{v}(t) = dN(t) - \underline{h}(t) \hat{\underline{x}}(t) dt \quad (3.8)$$

$$\underline{C} \left(\hat{\underline{x}}(t) \right) = \underline{S}^T(t) \hat{\underline{x}}(t) \quad (3.9)$$

In both cases, $\underline{P} \left(\hat{\underline{x}}(t) \right)$ is the conditional covariance matrix of the estimate

$$\underline{P} \left(\hat{\underline{x}}(t) \right) = E \left\{ \underline{\tilde{x}}(t) \underline{\tilde{x}}^T(t) \mid \mathcal{F}_t(\underline{u}) \right\} \quad (3.10)$$

where

$$\underline{\tilde{x}}(t) = \underline{x}(t) - \hat{\underline{x}}(t) \quad (3.10a)$$

and is given by

$$\underline{P} \left(\hat{\underline{x}}(t) \right) = \text{diag} \left(\hat{x}_i(t) \right) - \hat{\underline{x}}(t) \hat{\underline{x}}^T(t) \quad (3.11)$$

Remarks

1. In the case with Gaussian noise, the innovations process $\underline{v}(t)$ is a Wiener process with covariance $\underline{I} \cdot t$ and the two \mathcal{B} -martingales $\underline{v}, \underline{w}$

have zero conditional covariance process (see (2.16)). For the problem with counting observations, the innovations process $\underline{v}(t)$ is a $\mathcal{F}_t(\underline{u})$ -martingale with covariance process $\int \underline{h}(\tau) \hat{\underline{x}}(\tau) d\tau$ and the two \mathcal{B} -martingales \underline{v} , w have covariance process given in (2.20). It is because of these facts that the two estimation formulas are slightly different. But except for $\underline{c}(t)$ and the denominator in (3.7), they are identical.

2. The estimates in (3.4) provide a complete description of the posterior distribution of the signal. This is because the i -th component of $\hat{\underline{x}}(t)$ is

$$\hat{x}_i(t) = \text{Prob} \left\{ z(t) = \rho^{(i)} \mid \mathcal{F}_t(\underline{u}) \right\} \quad (3.12)$$

3. It is easy to generalize equation (3.4) to the situation when the observations process has random jump sizes [3]. If we denote by $Y(t)$ the observation process and by $N(t, dy)$ the number of jumps of size $[y, y+dy]$ up to time t , then the estimate $\hat{\underline{x}}(t)$ is given by

$$d\hat{\underline{x}}(t) = \underline{Q}^T(t, \underline{u}) \hat{\underline{x}}(t) dt + \int_{y=-\infty}^{\infty} \underline{K} \left(t, \hat{\underline{x}}(t-), y \right) \underline{v} (dt, dy) \quad (3.13)$$

where

$$\underline{v} (dt, dy) = N(dt, dy) - \underline{h}(t, dy) \hat{\underline{x}}(t) dt \quad (3.14)$$

$$\underline{K} \left(t, \hat{\underline{x}}(t-), y \right) = \left[\underline{p}(t-) \underline{h}^T(t, dy) + \underline{c}(t-, dy) \right] \frac{1}{\underline{h}(t, dy) \hat{\underline{x}}(t-)} \quad (3.15)$$

$$\underline{c}(t, dy) = \underline{s}^T(t, dy) \hat{\underline{x}}(t). \quad (3.16)$$

The quantities $h(t, \rho^{(i)}, dy)$ and $s_{ij}(t, dy)$ will be defined exactly as in (2.17) with $N(t)$ replaced by $N(t, dy)$. In particular, if $Y(\cdot)$ can have only a countable or finite number of jump sizes, the integral in (3.13) will reduce to the appropriate sum. This latter case is sometimes called a partially observable finite-state Markov process, since the joint process $\{z(t), Y(t)\}$ is a Markov process and only one of its components is observable.

4. In the case with Gaussian noise, the estimate is continuous so that we can in fact change $\hat{\underline{x}}(t-)$ with $\hat{\underline{x}}(t)$ in (3.5). For counting or general jump observations, $\hat{\underline{x}}(t)$ jumps at the times the observations jump, so that it is important to keep $\hat{\underline{x}}(t-)$ in (3.7) (and the corresponding (3.4)).

5. It will be useful to observe for future reference that the cost functional (2.24) can be written in terms of the estimates $\hat{\underline{x}}(t)$ as

$$\begin{aligned} J(\underline{u}) &= \int_0^T E \left\{ E \left[\underline{L}^T(t, \underline{u}(t)) \underline{x}(t) \mid \mathcal{F}_t(\underline{u}) \right] \right\} dt + E \left\{ E \left[\underline{\phi}^T \underline{x}(T) \mid \mathcal{F}_T(\underline{u}) \right] \right\} \\ &= E \left\{ \int_0^T \underline{L}^T(t, \underline{u}(t)) \hat{\underline{x}}(t) dt + \underline{\phi}^T \hat{\underline{x}}(T) \right\} \end{aligned} \quad (3.17)$$

The last equality follows from the fact that $\underline{u}(t)$ is $\mathcal{F}_t(\underline{u})$ -measurable (see(2.22)).

IV. Optimality and control-estimation separation for Gaussian noise

We are now ready to investigate conditions for optimality for the model (2.12), (2.15) of finite-state Markov process observed in additive white Gaussian noise. In preparation for the statement and proof of the Bellman-type sufficient conditions, consider an arbitrary twice differentiable function of $(n+1)$ variables $V(t, \underline{x})$ where \underline{x} is n -dimensional. Later we shall interpret V as the optimal conditional cost-to-go. If $\hat{\underline{x}}$ is the estimate with Gaussian additive noise given by (3.4) with (3.5), (3.6), (3.10), (3.11), then the Itô differential rule gives

$$\begin{aligned} dV(t, \hat{\underline{x}}(t)) &= \frac{\partial V}{\partial t} dt + \left(\frac{\partial V}{\partial \underline{x}} \right)^T d\hat{\underline{x}}(t) + \frac{1}{2} (d\hat{\underline{x}})^T \frac{\partial^2 V}{\partial \underline{x}^2} (d\hat{\underline{x}}) = \\ &= \left\{ \frac{\partial V}{\partial t} + \left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{Q}^T(t, u) \hat{\underline{x}}(t) + \frac{1}{2} \text{tr} \left[\left(\frac{\partial^2 V}{\partial \underline{x}^2} \right) \cdot \underline{K}(t, \hat{\underline{x}}(t)) \underline{K}^T(t, \hat{\underline{x}}(t)) \right] \right\} dt \\ &\quad + \left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{K}(t, \hat{\underline{x}}(t)) d\underline{v}(t) \end{aligned} \quad (4.1)$$

Here $\partial V / \partial \underline{x}$ is the gradient of V and $\partial^2 V / \partial \underline{x}^2$ is a matrix with elements $\partial^2 V / \partial x_i \partial x_j$ both evaluated at $\underline{x} = \hat{\underline{x}}(t)$. We also consider the operator

$$\mathcal{L}(u)V(t, \underline{x}) = \left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{Q}^T(t, u) \underline{x} + \frac{1}{2} \text{tr} \left[\left(\frac{\partial^2 V}{\partial \underline{x}^2} \right) \underline{K}(t, \underline{x}) \underline{K}^T(t, \underline{x}) \right] \quad (4.2)$$

and are ready to state the optimality criterion for this problem. Both the statement and the proof are similar to Wonham's separation theorem for linear

Gaussian signals in white Gaussian noise.

Theorem 4.1 (Optimality and separation criterion).

Suppose there exists a control law $u^* \in \mathcal{U}$ and a function $V(t, \underline{x})$, such that for all $t \in [0, T]$, all \underline{x} with components $\{0 \leq x_i \leq 1\}$ and all control laws $u \in \mathcal{U}$ the following holds

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} V(t, \underline{x}) + \mathcal{L}(u^*)V(t, \underline{x}) + \underline{L}^T(t, u^*) \underline{x} \leq \\ &\leq \frac{\partial}{\partial t} V(t, \underline{x}) + \mathcal{L}(u)V(t, \underline{x}) + \underline{L}^T(t, u) \underline{x} \end{aligned} \quad (4.3a)$$

with terminal condition

$$V(T, \underline{x}) = \underline{\phi}^T \cdot \underline{x} \quad (4.3b)$$

Then the control law

$$u^* \left(t, \hat{\underline{x}}^*(t) \right) \quad (4.4)$$

where $\hat{\underline{x}}^*$ is given by (3.4) - (3.6) and (3.10) - (3.11) with u^* replacing u , is optimal in the sense that it minimizes the criterion (2.24) in the class \mathcal{U} of admissible controls.

The minimum cost given $\hat{\underline{x}}(0)$ is then

$$\min_{\underline{u} \in \mathcal{U}} J(\underline{u}) = V(0, \hat{\underline{x}}(0)) \quad (4.5)$$

Proof

As said before the proof will be similar to Wonham's proof for linear Gaussian models [2]. Although here the observation fields \mathcal{F}_t are dependent

on the control, whereas for the linear Gaussian model they are not, it turns out that in fact Wonham's proof in [2] does not need this property. This can be seen by following essentially the same argument as below.

For an arbitrary control law \underline{u} , the corresponding fields $\mathcal{F}_t(\underline{u})$ of (2.21a) and the corresponding estimate $\hat{\underline{x}}$ given by (3.4) - (3.6) and (3.10) - (3.11), the following hold

$$\begin{aligned}
 & E \left\{ v \left(T, \hat{\underline{x}}(T) \right) - \int_t^T \left[\frac{\partial}{\partial s} v \left(s, \hat{\underline{x}}(s) \right) + \mathcal{L}(\underline{u}) v \left(s, \hat{\underline{x}}(s) \right) \right] ds \middle| \mathcal{F}_t(\underline{u}) \right\} \\
 &= E \left\{ v \left(T, \hat{\underline{x}}(T) \right) - \int_t^T dV \left(s, \hat{\underline{x}}(s) \right) + \right. \\
 &+ \left. \int_t^T \left(\frac{\partial v(s, \hat{\underline{x}}(s))}{\partial \hat{\underline{x}}(s)} \right)^T \underline{K} \left(s, \hat{\underline{x}}(s) \right) d\underline{v}(s) \middle| \mathcal{F}_t(\underline{u}) \right\} \\
 &= E \left\{ v \left(t, \hat{\underline{x}}(t) \right) \middle| \mathcal{F}_t(\underline{u}) \right\} = v \left(t, \hat{\underline{x}}(t) \right) \tag{4.6}
 \end{aligned}$$

The first equality above follows from (4.2),(4.1) and the second equality follows from the fact that $\underline{v}(t)$ is an $\mathcal{F}_t(\underline{u})$ -measurable Wiener process and therefore the appropriate stochastic integral has zero conditional expectation w.r.t. $\mathcal{F}_t(\underline{u})$. The last equality follows from the fact that $\hat{\underline{x}}(t)$ is $\mathcal{F}_t(\underline{u})$ measurable.

Equation (4.6) holds for arbitrary control laws, and in particular for our candidate for optimal law \underline{u}^* , so that, also using the first part of (4.3a) and (4.3b) we have

$$\begin{aligned}
 & \mathbb{E} \left\{ \underline{\phi}^T \hat{\underline{x}}^*(T) + \int_t^T \underline{L}^T(t, \underline{u}^*(t)) \hat{\underline{x}}^*(t) dt \middle| \mathcal{F}_t(\underline{u}^*) \right\} = \\
 & = \mathbb{E} \left\{ v(T, \hat{\underline{x}}^*(T)) - \int_t^T \left[\frac{\partial}{\partial s} v(s, \hat{\underline{x}}^*(s)) + \mathcal{L}(\underline{u}^*) v(s, \hat{\underline{x}}^*(s)) \right] ds \middle| \mathcal{F}_t(\underline{u}^*) \right\} \\
 & = v(t, \hat{\underline{x}}^*(t)) \tag{4.7}
 \end{aligned}$$

Similarly, for an arbitrary control law $\underline{u} \in \mathcal{U}$, the second part of (4.3a) and (4.3b) give together with (4.6)

$$\begin{aligned}
 v(t, \hat{\underline{x}}(t)) & = \mathbb{E} \left\{ v(T, \hat{\underline{x}}(T)) - \int_t^T \left[\frac{\partial}{\partial s} v(s, \hat{\underline{x}}(s)) + \right. \right. \\
 & \quad \left. \left. + \mathcal{L}(\underline{u}) v(s, \hat{\underline{x}}(s)) \right] ds \middle| \mathcal{F}_t(\underline{u}) \right\} \leq \\
 & \leq \mathbb{E} \left\{ \underline{\phi}^T \hat{\underline{x}}(T) + \int_t^T \underline{L}^T(t, \underline{u}(t)) \hat{\underline{x}}(t) dt \middle| \mathcal{F}_t(\underline{u}) \right\} \tag{4.8}
 \end{aligned}$$

Now comparing (4.7) and (4.8) evaluated at $t=0$ and observing that

$$\mathcal{F}_0(\underline{u}) = \mathcal{F}_0(\underline{u}^*) = \text{trivial } \sigma\text{-field} \tag{4.9}$$

$$\hat{\underline{x}}(0) = \hat{\underline{x}}^*(0) = \mathbb{E} \underline{x}(0) = \underline{\pi} \tag{4.9a}$$

we have from (3.17)

$$J(\underline{u}^*) = v\left(0, \underline{\hat{x}}^*(0)\right) = v\left(0, \underline{\hat{x}}(0)\right) \leq J(\underline{u}) \quad (4.10)$$

for all $\underline{u} \in \mathcal{U}$. This proves simultaneously (4.4) and (4.5).

V. Optimality and control - estimation separation for counting processes.

As already seen before, the problem with counting observations is quite similar to the one with additive Gaussian noise, and therefore one should not be surprised that similar methods and results will hold. We consider again a differentiable function $V(t, \underline{x})$, and if $\hat{\underline{x}}$ is given by (3.4) with (3.7)-(3.11), the Doleans Dade-Meyer differential rule [10, formula (101)] gives

$$dV \left(t, \hat{\underline{x}}(t) \right) = \frac{\partial V}{\partial t} dt + \left(\frac{\partial V}{\partial \underline{x}} \right)^T d\underline{\hat{x}}^C + v \left(t, \hat{\underline{x}}(t) \right) - v \left(t, \hat{\underline{x}}(t-) \right) \quad (5.1)$$

where $(\partial V / \partial \underline{x})$ is the gradient of V evaluated at $\hat{\underline{x}}(t)$ and $\underline{\hat{x}}^C$ is the continuous part of $\hat{\underline{x}}$ given by (see (3.4))

$$d\underline{\hat{x}}^C = \left[\underline{Q}^T(t, \underline{u}) \hat{\underline{x}}(t) - \underline{K} \left(t, \hat{\underline{x}}(t) \right) \underline{h}(t) \hat{\underline{x}}(t) \right] dt \quad (5.1a)$$

$$\triangleq \underline{A} \left(t, \underline{u}, \hat{\underline{x}}(t) \right) \hat{\underline{x}}(t) dt$$

Now the last part of (5.1) can be calculated as

$$v \left(t, \hat{\underline{x}}(t) \right) - v \left(t, \hat{\underline{x}}(t-) \right) = \left[v \left(t, \hat{\underline{x}}(t-) + \underline{K} \left(t, \hat{\underline{x}}(t-) \right) \right) - v \left(t, \hat{\underline{x}}(t-) \right) \right] dN(t)$$

$$\triangleq \left[\Delta_{\underline{K}} V \left(t, \hat{\underline{x}}(t-) \right) \right] dN(t) \quad (5.2)$$

(to see this, compare the two sides when $dN(t)=0$ and when $dN(t)=1$).

Therefore (5.1) becomes

$$\begin{aligned} dV(t, \hat{\underline{x}}(t)) = & \left\{ \frac{\partial V}{\partial t} + \left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{A}(t, u, \hat{\underline{x}}(t)) \hat{\underline{x}}(t) + \left[\Delta_K V(t, \hat{\underline{x}}(t)) \right] \underline{h}(t) \hat{\underline{x}}(t) \right\} dt \\ & + \left[\Delta_K V(t, \hat{\underline{x}}(t-)) \right] d\underline{v}(t) \end{aligned} \quad (5.3)$$

The Kolmogoroff operator will be now

$$\mathcal{L}(u)V(t, \underline{x}) = \left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{A}(t, u, \underline{x}) \underline{x} + [\Delta_K V(t, \underline{x})] \underline{h}(t) \underline{x} \quad (5.4)$$

and the sufficient condition for optimality and separation will be the same as before.

Theorem 5.1 (Optimality and separation criterion)

Suppose that for $\mathcal{L}(\cdot)$ defined in (5.4), there exists control law u^* $\in \mathcal{U}$ and a function $V(t, \underline{x})$, such that for all $t \in [0, T]$, all \underline{x} with components $\{0 \leq x_i \leq 1\}$ and all control laws $\underline{u} \in \mathcal{U}$, the following holds

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} V(t, \underline{x}) + \mathcal{L}(\underline{u}^*)V(t, \underline{x}) + \underline{L}^T(t, u^*) \underline{x} \leq \\ &\leq \frac{\partial}{\partial t} V(t, \underline{x}) + \mathcal{L}(\underline{u})V(t, \underline{x}) + \underline{L}^T(t, u) \underline{x} \end{aligned} \quad (5.5a)$$

with terminal condition

$$V(T, \underline{x}) = \underline{\phi}^T \cdot \underline{x} \quad (5.5b)$$

Then the control law

$$\underline{u}^* \left(t, \hat{\underline{x}}^*(t-) \right) \quad (5.6)$$

where $\hat{\underline{x}}^*$ is given by (3.4) and (3.7)-(3.11) with u^* replacing u , is optimal in the sense that it minimizes the criterion (2.23) in the class \mathcal{U} of admissible controls. The minimum cost given $\hat{\underline{x}}(0)$ is then

$$\min_{\underline{u} \in \mathcal{U}} J(\underline{u}) = v \left(0, \hat{\underline{x}}(0) \right) \quad (5.7)$$

Note

We may note that for the case of counting observations treated here, the optimal predictable control is a function of $\hat{\underline{x}}^*(t-)$. This is in contradistinction of the problem with Gaussian noise treated in Sec. IV where $\hat{\underline{x}}(t)$ is continuous and therefore $\hat{\underline{x}}(t-)$ can be replaced by $\hat{\underline{x}}(t)$ (see (4.4)).

Proof

For an arbitrary control law \underline{u} , the corresponding fields $\mathcal{F}_t(\underline{u})$ of (2.21b) and the corresponding estimate $\hat{\underline{x}}$ given by (3.4) with (3.7)-(3.11), the following hold

$$\begin{aligned} & E \left\{ v \left(T, \hat{\underline{x}}(T) \right) - \int_t^T \left[\frac{\partial}{\partial s} v \left(s, \hat{\underline{x}}(s) \right) + \mathcal{L}(\underline{u}) v \left(s, \hat{\underline{x}}(s) \right) \right] ds \Big| \mathcal{F}_t(\underline{u}) \right\} = \\ & = E \left\{ v \left(T, \hat{\underline{x}}(T) \right) - \int_t^T dv \left(s, \hat{\underline{x}}(s) \right) + \int_t^T \left[\Delta_{\underline{K}} v \left(s, \hat{\underline{x}}(s-) \right) \right] d \underline{v}(s) \Big| \mathcal{F}_t(\underline{u}) \right\} = \end{aligned}$$

$$= E \left\{ v \left(t, \hat{\underline{x}}(t) \right) \mid \mathcal{F}_t(\underline{u}) \right\} = v \left(t, \hat{\underline{x}}(t) \right) \quad (5.8)$$

The second equality here follows from the fact that $\underline{v}(t)$ is a $\mathcal{F}_t(\underline{u})$ -martingale and $\left[\Delta_K v \left(t, \hat{\underline{x}}(t-) \right) \right]$ is $\mathcal{F}_t(\underline{u})$ -predictable (since it is left-continuous) and hence

$$M_t(t) = \int_0^t \left[\Delta_K v \left(s, \hat{\underline{x}}(s-) \right) \right] d \underline{v}(s) \quad (5.9)$$

is also a $\mathcal{F}_t(\underline{u})$ -martingale (see [10, formula (75)]) which says

$$E \left[M(T) - M(t) \mid \mathcal{F}_t(\underline{u}) \right] = 0$$

The first equality in (5.8) follows from (5.3). Now (5.8) is similar to (4.6) and the rest of the proof is identical to the proof of Theorem 4.1.

Remark

The above can be easily extended to the situation when the observation process has jumps of random size (see (3.13) - (3.16)). The only change will be to redefine $\mathcal{L}(\underline{u})$ as

$$\begin{aligned} \mathcal{L}(\underline{u})v(t, \underline{x}) &= \left(\frac{\partial v}{\partial \underline{x}} \right)^T \underline{A}(t, \underline{u}, \underline{x}) \underline{x} + \\ &+ \int_{-\infty}^{\infty} \left[v \left(t, \underline{x} + K(t, \underline{x}, y) \right) - v(t, \underline{x}) \right] \underline{h}(t, dy) \underline{x} \end{aligned} \quad (5.8)$$

with

$$\underline{A}(t, \underline{u}, \underline{x}) = \underline{Q}^T(t, \underline{u}) - \int_{-\infty}^{\infty} \underline{K}(t, \underline{x}, y) \underline{h}(t, dy) \quad (5.9)$$

VI. Example

Consider a machine producing components according to a Poisson process with constant rate λ . Suppose the machine can be in one of two possible states: a normal state in which the probability that a finished component is defective is small, ρ' say, and a breakdown state in which this probability is $\rho'' > \rho'$. We denote

$$x(t) \equiv x_1(t) = 1 - x_2(t) = \begin{cases} 1 & \text{if machine is in breakdown state} \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

Suppose the transition probabilities are given as in (2.3), (2.4), (2.7), where the dependence on the control is given by

$$\underline{Q}(t, u) = (1-u)\underline{Q}_0(t) + u \underline{Q}_1(t) \quad (6.2)$$

Here $\underline{Q}_0(t)$, $\underline{Q}_1(t)$ are given matrices whose rows add up to zero and u represents the amount of attention and preventive maintenance the machine receives. If $u = 0$ the machine is not checked or repaired, and the machine is operated according to the transition matrix $\underline{Q}_0(t)$, which has a strong bias towards the breakdown state $x_1(t)$. If $u = 1$, maximum attention and maintenance is given, and the machine will operate according to $\underline{Q}_1(t)$, which has a bias towards the normal state $x_2(t)$. For simplicity, we assume that the jumps in the state \underline{x} are conditionally independent of the production process, so that $\underline{S}(t)$ in (2.20) is zero. A reasonable cost functional for this problem will be a weighted sum of the expected defective components $\int (\rho'\lambda x_1 + \rho''\lambda x_2) dt$ and the total effort $\int u dt$, so that in (2.23), (2.24)

$$L(t, u, \underline{x}) = \alpha u + (\rho' \lambda_{x_1} + \rho'' \lambda_{x_2}) \quad (6.3)$$

where $\alpha \geq 0$ is a normalizing constant. If we let

$$N(t) = \text{number of defective components produced within the interval } [0, t], \quad (6.4)$$

the problem is to find the optimal control u^* based on

$$\mathcal{F}_t(u^*) = \sigma\{N(s), s \leq t\} \quad (6.5)$$

to minimize the cost functional

$$J(u) = E \int_0^T \left[\alpha u(t) + \left(\rho' \lambda_{x_1}(t) + \rho'' \lambda_{x_2}(t) \right) \right] dt \quad (6.6)$$

Solution

The optimal control is given by Theorem (5.1):

$$-\frac{\partial V(t, \underline{x})}{\partial t} = \min_{u \in \mathcal{U}} \left\{ \mathcal{L}(u) V(t, \underline{x}) + \underline{L}(t, u) \underline{x} \right\} \quad (6.7)$$

$$V(T, \underline{x}) = 0 \quad (6.7a)$$

We first have to calculate the operator $\mathcal{L}(u)V(t, \underline{x})$

$$\mathcal{L}(u)V(t, \underline{x}) = \left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{A}(t, \underline{x}, u) \underline{x} + \left[\Delta_K V(t, \underline{x}) \right] \underline{h}(t) \underline{x} \quad (6.8)$$

of (5.4). From (5.1a), (5.2) and (3.7) - (3.11) we have

$$\underline{A}(t, u, \underline{x}) \underline{x} = \left[(1-u) \underline{Q}_0^T(t) + u \underline{Q}_1^T(t) \right] \underline{x} - \left\{ \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix} - \underline{x} \underline{x}^T \right\} \cdot \underline{h}^T \quad (6.9)$$

and

$$\underline{K}^V(t, \underline{x}) = V \left(t, \underline{x} + \frac{1}{\underline{h}} \underline{x} \right) \cdot \left(\begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix} - \underline{x} \underline{x}^T \right) \cdot \underline{h}^T - V(t, \underline{x}) \quad (6.10)$$

where

$$\underline{h} = [\rho^{\wedge} \lambda \quad \rho^{\vee} \lambda] \quad (6.11)$$

The only terms containing the control u in the right-hand side of (6.6) are

$$\left[\left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{Q}_1^T(t) \underline{x} - \left(\frac{\partial V}{\partial \underline{x}} \right)^T \underline{Q}_0^T(t) \underline{x} + 1 \right] u \quad (6.12)$$

so that the optimal control law is given (in terms of $V(t, \underline{x})$) by

$$u^*(t, \underline{x}) = \begin{cases} 1 & \text{if } \underline{x}^T \left(\underline{Q}_1(t) - \underline{Q}_0(t) \right) \cdot \frac{\partial V}{\partial \underline{x}} + 1 \leq 0 \\ 0 & \text{otherwise} \end{cases} \quad (6.13)$$

and the optimal control is given by

$$u^*(t, \hat{\underline{x}}^*(t-)) \quad (6.14)$$

as in (5.6).

It remains now to solve (6.7) with (6.9) - (6.14). One can somewhat simplify the algebra if one observed that always $x_1 + x_2 = 1$, so that one can define a new function

$$V_1(t, x) = V(t, x, 1-x) \quad 0 \leq x \leq 1 \quad (6.15)$$

Then (6.7) becomes a parabolic partial differential difference equation in t and x that can be solved numerically in a straightforward manner for V_1^* and then the optimal control is given by (6.14).

Footnotes

are

1. The increments in (2.14) ~~is~~ taken forward~~s~~ in time, namely $\underline{dx}(t) = \underline{x}(t+dt) - \underline{x}(t)$, where $dt > 0$.

REFERENCES

- [1]. W.M. Wonham, Stochastic Problems in Optimal Control, IEEE Int. Conv. Rec., Part II, pp. 114-124, 1963.
- [2]. W.M. Wonham, On the Separation theorem of stochastic control, SIAM J. Control, Vol. 6, No.2, pp. 312-326, 1968.
- [3]. A. Segall & T. Kailath, The modeling of randomly modulated jump processes, IEEE Trans. Infor. Theory, Vol. IT-24, No.2, pp. 64-71, March 1975.
- [4]. A.N. Shiryaev, Stochastic equations of nonlinear filtering of Markovian jump processes, Probl. Pered. Infor., Vol. 2, No.3, pp. 3-22, 1966.
- [5]. M. Fujisaki, G. Kallianpur, H. Kunita, Stochastic differential equations for the nonlinear filtering problem, Osaka J. Math., Vol. 9, pp.19-40, 1972.
- [6]. A. Segall, Recursive estimation from discrete-time point processes, IEEE Trans. Infor. Theory, Vol. IT-22, No.3, May 1976.
- [7]. A.I. Yaskin, Filtering of jump processes, Autom. Telemekh., Vol. 5, pp. 52-58, 1970.
- [8]. A. Segall, Dynamic file assignment in computer networks, IEEE Trans. on Autom. Control, Vol. AC-21, No.2, April 1976.
- [9]. R.D. Smallwood, and E.J. Sondik, The optimal control of partially observable Markov processes over a finite horizon, Oper. Res., Vol. 21, NO.5, pp. 1071-1088, 1973.
- [10]. A. Segall Stochastic processes in estimation theory, IEEE Trans. on Infor. Theory, Vol. IT-22, NO.3, May 1976.
- [11]. A. Segall, M.H.A. Davis and T. Kailath, Nonlinear filtering with counting observations, IEEE Trans. on Infor. Theory, Vol. IT-21, NO.2, pp. 143-149, March 1975.