ON STOCHASTIC OPTIMAL CONTROL PROBLEMS
WITH SELECTION AMONG DIFFERENT COSTLY OBSERVATIONS

by
Stéphane Lafortune

Memorandum No. UCB/ERL M85/99
5 December 1985
ON STOCHASTIC OPTIMAL CONTROL PROBLEMS
WITH SELECTION AMONG DIFFERENT COSTLY OBSERVATIONS

by

Stéphane Lafortune

Memorandum No. UCB/ERL M85/99
5 December 1985

ELECTRONICS RESEARCH LABORATORY
College of Engineering
University of California, Berkeley
94720
ON STOCHASTIC OPTIMAL CONTROL PROBLEMS
WITH SELECTION AMONG DIFFERENT COSTLY OBSERVATIONS

by
Stéphane Lafortune

Memorandum No. UCB/ERL M85/99
5 December 1985

ELECTRONICS RESEARCH LABORATORY
College of Engineering
University of California, Berkeley
94720
On Stochastic Optimal Control Problems

with Selection Among Different Costly Observations

Stéphane Lafortune

Department of Electrical Engineering and Computer Sciences
and Electronics Research Laboratory
University of California
Berkeley, CA 94720

Abstract

Optimizing observations is an important issue in the control of systems whose state is partially observed. This paper presents a general theorem for the computation of the optimal solution to discrete-time stochastic control problems, when the decision makers have the additional possibility of choosing at each step among different sets of observations on the system, each set incurring a different cost. Dynamic programming is employed to determine the optimal observations and controls. The result is applied to the special cases of: (i) finite-state controlled Markov chains, and (ii) linear Gaussian systems with a cost function quadratic in the states and controls.

Keywords: stochastic optimal control, observation selection, partial state information, dynamic programming, controlled Markov chains, LQG.

1. Problem statement

We consider the problem of the optimization of discrete-time stochastic systems, where at each step two consecutive decisions must be taken: (i) a decision on what type of observation to make on the system, and (ii) a decision on what control action to exert. The cost criterion depends on the state and these two control actions. We shall only consider finite-horizon problems.

Our aim is to present a general theorem for systems of the form:

\[ x_{k+1} = f_k(x_k, m_k, u_k, w_k), \]  
\[ y_k = h_k(x_k, m_k, v_k). \]  

for \( k \geq 0 \), with initial condition \( x_0, x_0, w_0, \ldots, v_0, \ldots \) are mutually-independent

* Research supported by the U.S. Army Research Office contract DAAG29-82-K-0091, and by a scholarship from the Natural Sciences and Engineering Research Council of Canada.
random variables defined on an underlying probability space. Their probability distributions on \( R^n \), \( R^n \), and \( R^m \), respectively, are known. \( x_k \in R^n \) is the state. \( m_k \) and \( u_k \) are control variables taking values in \( M \subset R^1 \) and \( U \subset R^m \). In particular, \( m_k \) parametrizes the observation equation (1.1b), where \( y_k \in R^p \) is the observed process.

Let
\[
I_k := \{ y_0, \ldots, y_k, m_0, \ldots, m_k, u_0, \ldots, u_k \}
\]
denote the information available to the decision maker at step \( k \). At each step \( k \), the values of the control actions \( m_k \) and \( u_k \) are determined by feedback in the following way:
\[
m_k = g^1_k(I_{k-1}) \in M,
\]
\[
u_k = g^2_k(I_{k-1}, m_k, y_k) \in U.
\]

Let the control strategy be denoted by \( g = (g^1, g^2) \), where \( g^i = \{g^i_0, \ldots, g^i_{N-1}\} \), and let \( G \) denote the set of all admissible strategies. We define
\[
J(g) := \mathbb{E}^g \left[ \sum_{k=0}^{N-1} c_k(x_k, m_k, u_k) + c_N(x_N) \right]
\]

To be the cost function associated with the control strategy \( g \). The superscript \( g \) in the expectation emphasizes the fact that the stochastic processes \( x, m, u, \) and \( y \) become well-defined only when \( g \) is given.\(^1\) We want to find an optimal strategy \( g^* \in G \), i.e., a strategy satisfying (a.s.)
\[
J(g^*) = J^* := \inf_{g \in G} J(g).
\]

\( m_k \) is an additional control variable parametrizing the observation equation. The choice among different sets of observations can be as simple as deciding whether or not to observe, in which case \( \text{card}(M) = 2 \). For the sake of generality, we shall also allow the

\(^1\) In the following, we will use the two notations \( \mathbb{E}^g c_k(x_k, m_k, u_k) \) and \( \mathbb{E} c_k(x_k, m_k, u_k) \) interchangeably.
possibility that the state equation depends on $m_k$. For example, in section 5, we consider linear Gaussian systems where the matrix $C_k$ in $y_k = C_k x_k + v_k$ and the variances of the processes $w$ and $v$ depend on $m_k$. Another point is that at each step $k$, the decisions on $m_k$ and $u_k$ are made sequentially, and therefore $u_k$ is allowed to depend on $m_k$, whereas the converse is not true. In short, this problem corresponds to optimizing the trade-off between the increased performance resulting from better observations (via better state estimates) and the higher cost of making better observations.

2. Information state for the system

We are dealing with a stochastic control problem with partial state information. We want to determine a suitable information state (in the terminology of Kumar and Varaiya [1]), or sufficient statistic (in that of Bertsekas [2]), for the system (1.1), i.e., a function of $I_k$ that possesses the Markov property. For simplicity, we assume that densities exist. Let $p_0(x_0)$ denote the probability density (p.d.) of the initial condition $x_0$, and, for a given control strategy $g$, let $p_{k|k-1}(x_k \mid I_{k-1})$ and $p_{k+1|k}(x_{k+1} \mid I_{k-1}, m_k, y_k)$ denote the conditional p.d. of $x_k$, given $I_{k-1}$ and $I_{k-1} \cup \{m_k, y_k\}$, respectively.

Lemma 2.1: $p_{k+1|k}(\cdot \mid I_{k-1})$ is an information state for (1.1). It does not depend explicitly on $g$ (and therefore we can drop the superscript $g$). There exists a function $S_k$ such that

\[ P_{k+1|k}(\cdot \mid I_k) = S_k [ P_{k|k-1}(\cdot \mid I_{k-1}), m_k, y_k, u_k ] . \] (2.1)

with initial condition $P_{0|1-1} = P_0$. $S_k$ can be broken into two functions $\Phi_k$ and $\Psi_k$:

\[ p_{k|k-1}(\cdot \mid I_{k-1}, m_k, y_k) = \Phi_k [ p_{k|k-1}(\cdot \mid I_{k-1}), m_k, y_k ] ; \] (2.2)

\[ P_{k+1|k}(\cdot \mid I_k) = \Psi_k [ p_{k|k}(\cdot \mid I_{k-1}, m_k, y_k), m_k, u_k ] . \] (2.3)

\[ 2 \text{ The organization of this paper and the proofs it contains were inspired by the treatment of standard stochastic systems (no } m_k \text{ in (1.1)) in [1], chapters 2 to 6.} \]
Proof: Given in the Appendix. □

The dynamics of the information state are illustrated in Figure 2.1.

\[ p_{k+1|k-1}(\cdot I_{k-1}) \quad p_{k|k}(\cdot I_{k-1}, m_k, y_k) \quad p_{k+1|k}(\cdot I_k) \]

\[ \Phi \quad \Psi \]

\[ m_k \quad y_k \quad m_k \quad u_k \]

Figure 2.1 - Information State

For simplicity, we shall often denote \( p_{k|k-1}(\cdot I_{k-1}) \) by \( p_{k|k-1} \), and similarly, \( p_{k|k}(\cdot I_{k-1}, m_k, y_k) \) by \( p_{k|k} \). Observe that although the functions \( p_{k|k-1} \) and \( p_{k|k} \) do not depend explicitly on \( g \), the processes \( x, m, u, \) and \( y \), and consequently \( I_k \), do depend on \( g \). For this reason, it will sometimes be necessary to write \( I_k \) in the arguments of \( p_{k|k-1} \) and \( p_{k|k} \) to emphasize the strategy considered. Observe also that (2.2) and (2.3) imply that

\[ p_{k+1|k+1}(\cdot I_k, m_{k+1}, y_{k+1}) = T_{k+1}[ p_{k|k}(\cdot I_{k-1}, m_k, y_k), m_k, u_k, m_{k+1}, y_{k+1} ] \quad (2.4) \]

and so \( p_{k+1|k+1} \) is not an information state, because, due to the explicit dependence of \( x_k \) on \( m_k \) via (1.1a), \( m_k \) has to appear as an argument in \( T_{k+1} \), even though it is already in \( p_{k|k} \).

3. Optimal control

The sequentiality assumption for the decisions on \( m_k \) and \( u_k \) suggest that the optimal strategy \( g \) could be determined by a dynamic programming algorithm, where the dynamic programming equation (d.p.e.) would contain two nested minimizations. From Lemma 2.1, we expect that restricting attention to separated strategies is sufficient. Such strategies are of the form:

\[ m_k = g_k(I_{k-1}) = g_k(p_{k|k-1}) \in M \quad (3.1) \]
\[ u_k = g_k^2(I_{k-1}, m_k, y_k) = g_k^2(p_{k | k-1}, m_k, y_k) = g_k^2(p_{k | k}, m_k) \in U. \] (3.2)

The following theorem shows that these claims are true. \( P \) denotes the set of all probability densities on \( R^n \). We define the cost-to-go from step \( k \)

\[ J_k(g) := E\left[ \sum_{j=k}^{N-1} c_k(x_k, m_k, u_k) + c_N(x_N) \mid I_{k-1} \right]. \] (3.3)

**Theorem 3.1:** Define recursively the functions \( V_k(g), 0 \leq k \leq N, \) and \( p \in P, \) by:

\[ V_N(p) := E\left[ c_N(x_N) \mid p_{N|N-1} = p \right]; \] (3.4)

\[ V_k(p) := \inf_{m \in M} \left\{ E\left[ \inf_{u \in U} \left( c_k(x_k, m, u) + V_{k+1}(p_{k+1}, m, u) \mid p_{k+1}, m \right) \mid p_{k|k-1} = p \right] \right\}. \] (3.5)

(a) Consider any \( g \in G \). Then

\[ V_k(p_{k|k-1}(\cdot \mid I_{k-1})) \leq J_k(g) \text{ a.s., } 0 \leq k \leq N. \] (3.6)

(b) Let \( g^* \) be a separated policy such that for all \( 0 \leq k \leq N-1 \) and for all \( p \in P, g_k^*(p) \) achieves the infima in (3.5). Then

\[ V_k(p_{k|k-1}(\cdot \mid I_{k-1})) = J_k(g^*) \text{ a.s., } 0 \leq k \leq N. \] (3.7)

and \( g^* \) is optimal. In particular, \( V_0(p_0) = J^* \) a.s.

**Proof:** (a) The proof is by induction. Consider any \( g \in G \). (3.6) is true with equality for \( k = N \), because

\[ J_N(g) = E[c_N(x_N) \mid I_{N-1}] \]

\[ = \int c_N(x)p_{N|N-1}(x \mid I_{N-1})dx \]

\[ = V_N(p_{N|N-1}(\cdot \mid I_{N-1})). \] (3.8)

by definition of \( p_{N|N-1} \), and from (3.4). Now, suppose that (3.6) is true for \( k+1 \). We show that it is true for \( k \), thus proving (a). Using successively the smoothing property of conditional expectations, (3.3), and the induction hypothesis, we get
\[ J_k(g) = E^g[c_k(x_k, m_k, u_k) + E^g \left\{ \sum_{j=k+1}^{N-1} c_j(x_j, m_j, u_j) + c_N(x_N) \mid I_k \right\} \mid I_{k-1}] \text{ a.s.} \]

\[ = E^g[c_k(x_k, m_k, u_k) + J_{k+1}(g) \mid I_{k-1}] \]

\[ \geq E^g[c_k(x_k, m_k, u_k) + V_{k+1}(pk+1k(\cdot \mid I_k)) \mid I_{k-1}] \text{ a.s.} \]

\[ = E^g[E^g[c_k(x_k, m_k, u_k) + V_{k+1}(pk+1k(\cdot \mid I_k) \mid I_k) \mid I_{k-1}] \text{ a.s.} \] (3.9)

But, by Lemma 2.1, we can replace the information sets by information states in (3.9):

\[ J_k(g) \geq E^g[E^g[c_k(x_k, m_k, u_k) + V_{k+1}(psi_k[pk+1k(\cdot \mid I_k-1, m_k, y_k), m_k, u_k] \mid pk+1k, m_k, y_k) \mid I_k-1] \text{ a.s.} \]

\[ = E^g[E^g[c_k(x_k, m_k, u_k) + V_{k+1}(psi_k[pk+1k, m_k, y_k, m_k, u_k] \mid pk+1k, m_k, y_k) \mid pk+1k, m_k, y_k) \mid pk+1k-1] \]

\[ \geq V_k(pk+1k-1(\cdot \mid If_k-1)). \] (3.10)

the last inequality holding by (3.5).

(b) Again, we use induction to prove (3.7). First, we observe that (3.8) implies that (3.7) is true for \( k = N \). Next, we repeat the development in (a), but with the given \( g^* \) in place of \( g \). However, the two inequalities in (a) now become equalities: (3.9), by the induction hypothesis, and (3.10), because, by assumption, \( g^*_k \) achieves the infima in (3.5) for all \( p \in P \). This proves (3.7). To show the optimality of \( g^* \), we set \( k = 0 \) in (3.7) and (3.6) to get

\[ J(g^*) = V_0(p(x_0)) \leq J(g) \text{ a.s., for all } g \in G. \] (3.11)

\[ \Box \]

Remarks: (i) Observe that the \( V_{k+1} \) term could be removed from the inner conditional expectation in the d.p.e. (3.5).

(ii) (2.2) implies that, for each fixed \( m \), the outer conditional expectation in (3.5) is an integral over \( y_k \).
The argument of the value function $V_k$ in (3.5) is a function, meaning that finding the optimal $g^*$ is computationally difficult. In the next two sections, we consider two special cases where the problem is more amenable because the information state is finite-dimensional.

4. Special case I: finite-state controlled Markov chains

Consider a Markov chain whose state process $x$ takes values in a finite set $S = \{1, 2, \ldots, S\}$, and whose transition-probability matrix $P(m, u)$ can depend on two different controls $m$ and $u$:

$$[P(m, u)]_{i,j} = P_{ij}(m, u) := \text{Prob}(x_{k+1} = j \mid x_k = i, m_k = m, u_k = u). \quad (4.1)$$

Let the observed process $y \in S$ be described by the output probability

$$P_j(i, m) := \text{Prob}(y_k = j \mid x_k = i, m_k = m). \quad (4.2)$$

These probabilities do not depend on $k$. It is convenient to define the $S \times S$ matrix $D(m, j)$ by

$$D(m, j) := \text{diag}[P_j(i, m)]_{i=1}^{S}. \quad (4.3)$$

Let $\text{Prob}_{k \mid k}^{-1}(i \mid I_{k-1})$ and $\text{Prob}_{k \mid k}(i \mid I_{k-1}, m_k, y_k)$ be the probabilities that $x_k = i$, given the respective information sets. Since the state space is finite, these probabilities are completely described by the $1 \times S$ row-vectors:

$$\pi_{k \mid k-1}(I_{k-1}) := [\text{Prob}_{k \mid k-1}(1 \mid I_{k-1}), \ldots, \text{Prob}_{k \mid k-1}(S \mid I_{k-1})]; \quad (4.4)$$

$$\pi_{k \mid k}(I_{k-1}, m_k, y_k) := [\text{Prob}_{k \mid k}(1 \mid I_{k-1}, m_k, y_k), \ldots, \text{Prob}_{k \mid k}(S \mid I_{k-1}, m_k, y_k)]. \quad (4.5)$$

To simplify the notation, we shall often omit writing the arguments of these two probabilities. Also, $\pi_{k \mid k-1}(j)$ will denote the $j$th component of $\pi_{k \mid k-1}$.

We write recursive relations for $\pi_{k \mid k}$ and $\pi_{k+1 \mid k}$. The initial condition is $\pi_{0 \mid -1} = \pi_0$, the given law of the initial state. It can be shown (cf: proof of Lemma 2.1, (A.8) and (A.11)) that the functions $\Phi_k$ and $\Psi_k$ in (2.2) and (2.3) have the following
expression:

\[ \pi_{k+1}(I_{k-1}, m_k, y_k) = \frac{\pi_{k+1}(I_{k-1}) D(m_k, y_k)}{\pi_{k+1}(I_{k-1}) D(m_k, y_k) \mathbb{1}} ; \]  

\[ \pi_{k+1}(I_k) = \pi_{k+1}(I_{k-1}, m_k, y_k) P(m_k, u_k) . \]  

\( \mathbb{1} \) in (4.6) is the \( S \times 1 \) column-vector \((1, \ldots, 1)^T\).

We now write the complete expression of the d.p.e. (3.5). Consider \( \pi \in \Pi \), the set of all \( 1 \times S \) probability row-vectors. Then, (3.4) and (3.5) become:

\[ V_N(\pi) = \sum_{i \in S} c_N(i) \pi(i) , \]  

\[ V_k(\pi) = \inf_{m \in M} E \left[ \inf_{u \in U} \left\{ \sum_{i \in S} c_k(i, m, u) \pi_{k+1}(i) + V_{k+1}(\pi_{k+1} P(m, u)) \right\} \mid \pi_{k+1} = \pi \right] . \]  

Let \( u_k^* = g_k^2(\pi_{k+1}, m) = g_k^2(\pi_{k+1-1}, m, y_k) \) achieve the inner infimum. We evaluate the conditional expectation in (4.9):

\[ V_k(\pi) = \inf_{m \in M} E \left[ \sum_{i \in S} c_k(i, m, u_k^*) \pi_{k+1}(i) + V_{k+1}(\pi_{k+1} P(m, u_k^*)) \mid \pi_{k+1} = \pi \right] \]

\[ = \inf_{m \in M} \int_{y_k} \left[ \sum_{i \in S} c_k(i, m, u_k^*) \frac{\pi D(m, y_k)}{\pi D(m, y_k) \mathbb{1}}(i) \right. \]

\[ + \left. V_{k+1}(\pi D(m, y_k) \mathbb{1} P(m, u_k^*)) \right] \pi D(m, y_k) \mathbb{1} \, dy_k . \]  

since

\[ \text{Prob}(y_k \mid \pi_{k+1-1} = \pi, m) = \sum_{i \in S} \text{Prob}(y_k \mid i, m) \text{Prob}(x_k = i \mid \pi_{k+1-1} = \pi) \]  

\[ = \sum_{i \in S} \mathcal{P}_{y_k}(i, m) \pi(i) \]

\[ = \pi D(m, y_k) \mathbb{1} . \]  

where in (4.11) we have used (3.1).
5. Special case II: linear Gaussian systems

Consider the case where (1.1) is of the form:

\[ x_{k+1} = A_k x_k + B_k u_k + w_k , \]  
(5.1a)

\[ y_k = C_k(m_k)x_k + v_k . \]  
(5.1b)

with \( x_0 \sim N(\bar{x}_0, \Sigma_0) \), \( w_k \sim N(0, Q_k(m_k)) \), and \( v_k \sim N(0, R_k(m_k)) \). Consider a cost-function quadratic in the states and in the controls \( u \) :

\[ J(g) := E \left[ \sum_{k=0}^{N-1} (x_k^T M_k x_k + u_k^T N_k u_k + c_k(m_k)) + x_k^T M x_k \right] . \]  
(5.2)

(Here, we make the usual symmetry and positive (semi-)definiteness assumptions on \( M_k, N_k, Q_k \), and \( R_k \).) We mention at this point that Aoki and Li [3] have studied a version of this problem where the decision makers select the total number and the spacings between each observation, which is quite different from our formulation. Also, their model has no \( w \) term in (5.1a).

The derivation of the Kalman filter remains valid when the matrices \( A_k, B_k \), and \( C_k \) are random, provided they are measured at time \( k \), i.e., that they are in \( I_k \), and that they are independent of the noise variables (e.g., [1]). In our case, once \( m_k \) is chosen, all the parameters in (5.1) that depend on it can simply be regarded as time-varying, with the important difference that their time variation can be altered. However, that decision is based on past information, namely, \( I_{k-1} \). It follows that the p.d. \( p_{k+1|k} \) and \( p_{k|1k} \) defined in section 2 are Gaussian, and therefore the information state \( p_{k+1|k} \) is two-dimensional. In fact, this remains true if \( A_k \) and \( B_k \) also depend on \( m_k \).

Consider a fixed feedback strategy \( g \) and the corresponding processes \( x, m, u, \) and \( y \).\(^3\) Then, using the notation

\[ p_{k+1|k} (x_{k+1} \mid I_k) \sim N(\hat{x}_{k+1|k}, \Sigma_{k+1|k}) \]  
(5.3)

\(^3\) For simplicity, we omit writing the superscript \( g \) for these processes.
the Kalman filter equations corresponding to (2.2) and (2.3) are:

\[
\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k (y_k - C_k (m_k) \hat{x}_{k|k-1})
\]

\[
\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + B_k u_k
\]

\[
\Sigma_{k|k} = \Sigma_{k|k-1} - L_k C_k (m_k) \Sigma_{k|k-1}
\]

\[
\Sigma_{k+1|k} = A_k \Sigma_{k|k} A_k^T + Q_k (m_k)
\]

where

\[
L_k := \Sigma_{k|k-1} C_k (m_k) [C_k (m_k) \Sigma_{k|k-1} C_k (m_k)]^T + R_k (m_k)^{-1}
\]

The interesting feature of this special case is that, due to the quadratic form of the cost and the fact that only \( R_k \) in (5.1a) depends on \( m_k \), the certainty-equivalence principle still holds and the value function has a partially-closed form. This is not true for linear Gaussian systems in general, and this was our motivation for these extra assumptions. More precisely, it can be shown, by substituting (5.10) in (3.5), that

\[
V_k (\hat{x}_{k|k-1}, \Sigma_{k|k-1}) = \hat{x}_{k|k-1}^T P_k \hat{x}_{k|k-1} + W_k (\Sigma_{k|k-1})
\]

\[0 \leq k \leq N. \ P_k \] is determined by solving the standard backward Riccati equation

\[
P_k = M_k + A_k^T P_{k+1} A_k - K_k^T (N_k + B_k^T P_{k+1} B_k) K_k
\]

\[0 \leq k < N, \] with final condition \( P_N = M_N \). \( K_k \) is the deterministic optimal control gain

\[
K_k := -[N_k + B_k^T P_{k+1} B_k]^{-1} B_k^T P_{k+1} A_k
\]

\[i.e., u_k = K_k \hat{x}_{k|k} = g_k^T (\hat{x}_{k|k}). \ P_k \text{ and } K_k \text{ do not depend on } m \text{ and that they can be completely determined beforehand. The other part of } V_k \text{ has no closed-form solution and must be solved recursively as follows:}
\]

\[
W_k (\Sigma) = \inf_{m \in M} \left[ c_k (m) + \text{Trace} \{ M_k \Sigma + l_k (\Sigma, m) (P_k - M_k) \} \right]
\]

\[+ W_{k+1} (A_k \Sigma A_k^T + Q_k (m) - A_k l_k (\Sigma, m) A_k^T) \].

(5.13)
with final condition $W_N(\Sigma) = \text{Trace}(M_N \Sigma)$, where we have defined

$$l_k(\Sigma, m) := \Sigma C_k(m)^T[C_k(m)\Sigma C_k(m)^T + R_k(m)]^{-1}C_k(m)\Sigma.$$  \hspace{1cm} (5.14)

The optimal sequence $m^*$ can be determined beforehand, but it depends on $P_k$, and consequently the Riccati equation must be solved first. If $M$ is finite with $\text{card}(M) = n$, then at step $k$, the domain of $\Sigma$ in (5.13) can contain up to $n^k$ values.

As in the standard LQG problem, the control $u$ has no learning role, but the control $m$ has one, since it can influence the estimation covariance of the state. Clearly, if $A_k$ or $B_k$ were dependent on $m_k$, $V_k$ would possess no separation property as (5.10) exhibits, even with a quadratic cost. Thus, $u_k^*$ would in general also depend on $\Sigma_k|k-1$, meaning that it too would have a learning function.

Finally, we point out that the problem in this section was also treated by Deissenberg and Stöppler [4] for the special case when $\text{card}(M) = 2$, corresponding to the decision: observe / do not observe. However, the value function considered in that paper does not have the estimation covariance matrix $\Sigma_k|k-1$ as an argument, and therefore the solution does not have the clear recursive form of (5.10) and (5.13), which also provides for more computational efficiency.

**Acknowledgements**

The author would like to thank Professor E. Wong for his help and guidance, and Professors P. Varaiya and J. Walrand for several helpful discussions.

**Appendix - Proof of Lemma 2.1**

The independence of all the noise variables in (1.1) and the fact that the values of $m$ and $u$ are measured imply that

$$p^\xi(x_{k+1} | x_k, I_k) = p(x_{k+1} | x_k, m_k, u_k).$$  \hspace{1cm} (A.1)

$$p^\xi(y_k | x_k, m_k, I_{k-1}) = p(y_k | x_k, m_k).$$  \hspace{1cm} (A.2)
where the densities on the right-hand sides do not depend on the strategy \( g \), but only on the values of \( m \) and \( u \).

We now establish the precise form of the recursive relations (2.2) and (2.3).

\[
p_{k+1}^g(x_{k+1} \mid I_{k-1}, m_k, u_k) = \int p(y_k \mid x_{k+1}, I_{k-1}, m_k, u_k) \, p^g(x_{k+1} \mid I_{k-1}, m_k, u_k) \, dx_{k+1} - (A.3)
\]

\[
= \frac{p(y_k \mid x_k, m_k) \, p^g(x_k \mid I_{k-1}, m_k)}{\int p(x_k \mid I_{k-1}, m_k, y_k) \, dx_k} - (A.4)
\]

where (A.3) follows from Bayes' rule, and (A.4) by using (A.2). But

\[
p^g(x_k, I_{k-1}, m_k, y_k) = p(y_k \mid x_k, I_{k-1}, m_k) \, p^g(x_k \mid I_{k-1}, m_k)
\]

\[
= p(y_k \mid x_k, m_k) \, p^g(x_k \mid I_{k-1}, m_k) \, p^g(I_{k-1}, m_k)
\]

by (A.2). Substituting (A.6) in (A.4).

\[
p_{k+1}^g(x_k \mid I_{k-1}, m_k, u_k) = \int p(y_k \mid x_{k+1}, I_{k-1}, m_k, u_k) \, p^g(x_{k+1} \mid I_{k-1}, m_k, u_k) \, dx_{k+1} - (A.7)
\]

\[
= \frac{p(y_k \mid x_k, m_k) \, p^g(x_k \mid I_{k-1}, m_k)}{\int p(y_k \mid x_k, m_k) \, p^g(x_k \mid I_{k-1}, m_k) \, dx_k} - (A.8)
\]

because \( m_k \) is a function of \( I_{k-1} \) (see (1.3)) and \( x_k \) only depends on \( m_k \) via \( I_{k-1} \). (A.8) is of the form given in (2.2). Next,

\[
p_{k+1}^g(x_{k+1} \mid I_k) = \int p^g(x_{k+1} \mid x_k, I_k) \, p^g(x_k \mid I_k) \, dx_k - (A.9)
\]

\[
= \int p(x_{k+1} \mid x_k, m_k, u_k) \, p^g(x_k \mid I_{k-1}, m_k, y_k, u_k) \, dx_{k+1} - (A.10)
\]

\[
= \int p(x_{k+1} \mid x_k, m_k, u_k) \, p_{k+1}^g(x_k \mid I_{k-1}, m_k, y_k) \, dx_k - (A.11)
\]

(A.10) is a consequence of (A.1). (A.11) is true because \( x_k \) does not depend explicitly on \( u_k \) but only through \( I_{k-1} \cup \{m_k, y_k\} \), of which \( u_k \) is a function (see (1.4)). (A.11) corresponds to (2.3). \( p_{k+1}^g \) and \( p_{k+1}^{g+1} \) do not depend on \( g \) because the functions \( \Phi \) and \( \Psi \)
of (A.8) and (A.11) do not, and the initial condition is \( p_{0_{1-1}} = p_0 \) (recall that \( 1_{-1} = \emptyset \)) and is therefore independent of \( g \). □

References


