DISCRETE TIME PARTIALLY OBSERVED CONTROL

ROBERT J. ELLIOTT
Department of Statistics and Applied Probability
University of Alberta
Edmonton, Alberta
Canada T6G 2G1
AND
JOHN B. MOORE
Department of Systems Engineering
Research School of Physical Sciences
Australian National University, G.P.O. Box 4
Canberra, A.C.T. 2601 Australia

Acknowledgments: Research partially supported by NSERC Grant A7964 and Boeing BCAC. The hospitality of the Department of Systems Engineering at the Australian National University is gratefully acknowledged.

Key Words: Stochastic control, reference probability, unnormalized density, separated control, dynamic programming, minimum principle, adjoint process, dual control.

Abstract: A discrete time, partially observed control problem is discussed by explicitly constructing a reference probability under which the observations are independent. Using the unnormalized conditional probabilities as information states the problem is treated in separated form. Dynamic programming and minimum principle results are obtained.

1. Introduction.

Much effort has been expended on discussing the optimal control of partially observed diffusions in continuous time, but the results are still not entirely satisfactory. Discrete time control problems are treated in the books of Kumar and Varaiya [5] and Caines [1]. In this paper we discuss the discrete time, partially observed control problem using the reference probability. This idea is described in references [3] and [4]; the reference probability is constructed explicitly, and the role of the dynamics in the separated problem is clarified. The unnormalized conditional probabilities, which describe the state of the process given the observations, play the role of 'information states', and the control problem can be re-cast as a fully observed optimal control problem. A dynamic programming result and minimum principle are obtained, in terms of separated controls, and an adjoint process is described. Finally, when some of the parameters of the model are unknown it is shown how the methods extend to dual control problems.

2. Dynamics.

We shall consider a finite time horizon control problem and, for simplicity, we suppose noise is additive in the state and observation processes. All processes are defined initially on a probability space $(\Omega, \mathcal{F}, P)$.

The state process $\{x_k\}$, $k = 0, 1, \ldots, M$, take values in R^d and has dynamics

$$x_{k+1} = F_k(x_k, u_k) + w_{k+1}. \tag{2.1}$$

We suppose the initial density $\pi_0(z)$ of x is known.

The observation process $\{y_k\}$, $k = 0, 1, \ldots, M$ takes values in R^m and has dynamics

$$y_{k+1} = H_k(x_k) + b_{k+1}. \tag{2.2}$$

We suppose $y_0 = 0 \in R^m$. For $0 \leq k \leq M$ write $y^k = \{y_0, y_1, \ldots, y_k\}$. $\{G_k\}$ is the complete filtration generated by x and y. $\{\mathcal{Y}_k\}$ is the complete filtraton generated by y.

The noise in the state process is a sequence $\{w_k\}$, $1 \leq k \leq M$, of independent R^d valued random variables having densities ψ_k.

The noise in the observation process is a sequence $\{b_k\}$, $1 \leq k \leq M$, of independent R^m valued random variables having positive densities ϕ_k, $\phi_k(b) > 0$ for all $b \in R^m$. The parameter u_k in (2.1) represents the control variable, and takes values in a set $U \subset R^p$. At time k, u_k is $\mathcal{Y}_k$ measurable, that is, u_k is a function of y^k. For $0 \leq k < M$ write $\underline{U}(k)$ for the set of such control functions and

$$\underline{U}(k, k+\ell) = \underline{U}(k) \cup \underline{U}(k+1) \cup \cdots \cup \underline{U}(k+\ell).$$

For $u \in \underline{U}(0, M-1)$, x^u will denote the trajectory $(x_0, x_1^u, x_2^u, \ldots, x_M^u)$ determined by (2.1).

3. Unnormalized Densities.

We review the recurrence relation for the unnormalized conditional density of the state given the observations. For details see [3] and [4].

Suppose we have an equivalent probability measure $\overline{P}$ on (Ω, G_M) such that under $\overline{P}$:

1) $\{y_k\}$ is a sequence of independent random variables having positive densities ϕ_k,
2) for any $u \in \underline{U}(0, M-1)$, $x_{k+1}^u \in R^d$ satisfies the dynamics

$$x_{k+1}^u = F_k(x_k^u, u_k) + w_k, \qquad k \in Z^+,$$

where w_k is a sequence of independent random variables having densities ψ_k.

Suppose $u \in \underline{U}(0, M-1)$. Define $\overline{\gamma}_\ell^u = \phi_\ell(y_\ell - H_{\ell-1}(x_{\ell-1}^u))/\phi_\ell(y_\ell)$ and $\overline{\Lambda}_n^u = \prod_{\ell=1}^n \overline{\gamma}_\ell^u$. Then a probability P^u can be defined by setting the restriction of $\frac{dP^u}{d\overline{P}}$ to G_M equal to $\overline{\Lambda}_M^u$. It is under P^u that the state and observation processes have the form (2.1) and (2.2).

Suppose Φ_ℓ is any G-adapted process. Then a version of Bayes' theorem states that

$$E^u[\Phi_n \mid \mathcal{Y}_n] = \overline{E}[\overline{\Lambda}_n^u \Phi_n \mid \mathcal{Y}_n] / \overline{E}[\overline{\Lambda}_n^u \mid \mathcal{Y}_n]. \tag{3.1}$$

Write $q_n^u(z)$ for the unnormalized conditional density such that

$$\overline{E}[\overline{\Lambda}_n^u I(x_n^u \in dz) \mid \mathcal{Y}_n] = q_n^u(z)dz.$$

The equation (3.1) indicates why we consider $q_n^u(z)$, because the normalized conditional density $p_n^u(z)$ is then given by:

$$p_n^u(z) = q_n^u(z) / \int_{R^d} q_n^u(\xi)d\xi,$$

and for any Borel test function f

$$E^u[f(x_n^u) \mid \mathcal{Y}_n] = \int_{R^d} f(z)p_n^u(z)dz.$$

Now for $u \in \underline{U}(0, M-1)$ and any Borel test function f consider

$$\begin{aligned} \overline{E}[f(x_n^u)\overline{\Lambda}_n^u \mid \mathcal{Y}_n] &= \int_{R^d} f(z)q_n^u(z)dz = \langle f(z) q_n^u(z)\rangle \qquad (3.2) \\ &= \overline{E}[f(F_{n-1}(x_{n-1}^u, u_{n-1}) + w_n)\overline{\Lambda}_{n-1}^u \phi_n(y_n - H_{n-1}(x_{n-1}^u)) \mid \mathcal{Y}_n]/\phi_n(y_n) \\ &= \phi_n(y_n)^{-1}\overline{E}\Big[\int_{R^d} f(F_{n-1}(x_{n-1}^u, u_{n-1}) + w)\psi_n(w)dw\overline{\Lambda}_{n-1}^u \phi_n(y_n - H_{n-1}(x_{n-1}^u)) \mid \mathcal{Y}_n\Big] \\ &= \phi_n(y_n)^{-1} \int\int f(F_{n-1}(\xi, u_{n-1}) + w)\psi_n(w)\phi_n(y_n - H_{n-1}(\xi))q_{n-1}^u(\xi)dwd\xi. \end{aligned}$$

Substituting $z = F_{n-1}(\xi, u_{n-1}) + w$ this is

$$= \phi_n(y_n)^{-1} \int\int f(z)\psi_n(z - F_{n-1}(\xi, u_{n-1}))\phi_n(y_n - H_{n-1}(\xi))q^u_{n-1}(\xi)d\xi dz. \tag{3.3}$$

The equality of (3.2) and (3.3) holds for all Borel test functions f, so we have the following recurrence relation for q^u_n:

THEOREM 3.1.

$$q^u_n(z) = \phi_n(y_n)^{-1} \int_{R^d} \psi_n(z - F_{n-1}(\xi, u_{n-1}))\phi_n(y_n - H_{n-1}(\xi))q^u_{n-1}(\xi)d\xi. \tag{3.4}$$

REMARKS 3.2 This equation describes the observable dynamics of a separated problem. $q^u_n(\cdot)$ is an "information state" in the sense of Kumar and Varaiya [5]. That is, if we know $q^u_{n-1}(\cdot)$, y^n and u_{n-1}, equation (3.4) enables us to determine $q^u_n(\cdot)$.

The initial information state q_0 is just π_0, the (normalized) density of x_0. Note that, even if π_0 is a unit mass at a particular x_0, $q^u_1(z) = \phi_1(y_1)^{-1}\psi_1(z - F_0(x_0, u_0))\phi_1(y_1 - H_0(x_0))$, and the consequent terms $q^u_2, q^u_3, \ldots$ follow from equation (3.4).

4. Cost.

Suppose, given x_0 and $u \in \underline{U}(0, M-1)$, the cost function associated with the problem is of the form

$$C(x_0, u) = \sum_{k=0}^{M-1} c_k(x^u_k, u_k) + c_M(x^u_M).$$

Then the expected cost, if control u is used and the density of x_0 is $\pi_0(\cdot)$, is

$$V_0(\pi_0, u) = E[C(\overline{x}_0, u)].$$

This can be expressed

$$V_0(\pi_0, u) = \overline{E}\Big[\overline{\Lambda}^u_M\Big(\sum_{k=0}^{M-1} c_k(x^u_k, u_k) + c_M(x^u_M)\Big)\Big]$$

$$= \sum_{k=0}^{M-1} \overline{E}[\overline{\Lambda}^u_k c_k(x^u_k, u_k)] + \overline{E}[\overline{\Lambda}^u_M c_M(x^u_M)]$$

$$= \overline{E}\Big[\sum_{k=0}^{M-1} \langle c_k(z, u_k), q^u_k(z)\rangle + \langle c_M(z), q^u_M(z)\rangle\Big]$$

$$= \overline{E}\Big[\overline{E}\Big[\sum_{k=0}^{M-1} \langle c_k(z, u_k), q^u_k(z)\rangle + \langle c_M(z), q^u_M(z)\rangle \mid \mathcal{Y}_M\Big]\Big]$$

where, for example, we write

$$\langle c_k(z,u_k), q_k^u(z)\rangle = \int_{R^d} c_k(z,u_k) q_k^u(z)dz$$
$$= \overline{E}[\overline{\Lambda}_k^u c_k(x_k^u, u_k) \mid \mathcal{Y}_k].$$

REMARKS 4.1. We have seen the information state at time k belongs to the set S of positive measures $q(\cdot)$ on R^d. Note the probability measures are a subset of S.

S is an infinite dimensional space. A metric can be defined on S using the L^1 norm, so that for $q^1(\cdot), q^2(\cdot) \in S$

$$d(q^1, q^2) = \|q^1 - q^2\| = \int_{R^d} |q^1(z) - q^2(z)|dz.$$

Any $q \in S$ can be normalized to give a probability measure $\pi(q) = q(\cdot)/\|q\|$.

Consider the process starting from some intermediate time k, $0 \leq k \leq M$, from some state $q(\cdot) \in S$. Then, for $u \in \underline{U}(k, M-1)$

$$q_{k+1}^u(z) = \phi_{k+1}(y_{k+1})^{-1} \int_{R^d} \psi_{k+1}(z - F_k(\xi, u_k))\phi_{k+1}(y_{k+1} - H_k(\xi))q(\xi)d\xi. \qquad (4.1)$$

The remaining information states $q_n^u(\cdot)$, $k+1 < n \leq M$, are similarly obtained from (3.4).

The expected cost accumulated, starting from state $q(\cdot) \in S$ and using control $u \in \underline{U}(k, M-1)$ is, therefore

$$V_k(q,u) = \overline{E}\Big[\sum_{j=k}^{M-1} \langle c_j(z,u_j), q_j^u(z)\rangle + \langle c_M(z), q_M^u(z)\rangle \mid q_k = q\Big]. \qquad (4.2)$$

REMARKS 4.2. The problem is now in a separated form. The filtering recursively determines the unnormalized, conditional probabilities which are the information states, $q_k^u(\cdot)$. These evolve according to the dynamics (3.4), and the cost is expressed in terms of these information states.

DEFINITION 4.3. A control $u \in \underline{U}(0, M-1)$ is said to be separated if u_k depends on y^k only through the information state $q_k^u(\cdot)$. Write $\underline{U}_S(0, M-1)$ for the set of separated controls.

DEFINITION 4.4. For $0 \leq k \leq M-1$ the cost process is defined by:

$$V(k,q) = \bigwedge_{u \in \underline{U}(k,M-1)} V_k(q,u).$$

Here $V_k(q,u)$ is given by (4.2). Also, set $V(M,q) = \langle c_M(z), q(z)\rangle$. We now establish the dynamic programming identity.

THEOREM 4.5. *For* $0 \le k \le M-1$ *and* $q \in S$

$$V(k,q) = \bigwedge_{u \in \underline{U}(k)} \overline{E}[\langle c_k(z,u_k), q(z)\rangle + V(k+1, q^u_{k+1}) \mid q_k = q]. \tag{4.3}$$

Proof.

$$\begin{aligned}
V(k,q) &= \bigwedge_{u \in \underline{U}(k,M-1)} V_k(q,u) = \bigwedge_{u \in \underline{U}(k)} \bigwedge_{v \in \underline{U}(k+1,M-1)} V_k(q,u) \\
&= \bigwedge_{u \in \underline{U}(k)} \bigwedge_{v \in \underline{U}(k+1,M-1)} \overline{E}\Big[\overline{E}\Big[\langle c_k(z,u_k), q(z)\rangle \\
&\quad + \sum_{j=k+1}^{M-1} \langle c_j(z,v_j), q^v_j(z)\rangle + \langle c_M(z), q^v_M(z)\rangle \mid \mathcal{Y}_{k+1}\Big] q_k = q\Big] \\
&= \bigwedge_{u \in \underline{U}(k)} \Big\{\overline{E}[\langle c_k(z,u_k), q(z)\rangle \mid q_k = q] \\
&\quad + \bigwedge_{v \in \underline{U}(k+1,M-1)} \overline{E}\Big[\overline{E}\Big[\sum_{j=k+1}^{M-1} \langle c_j(z,v_j), q^v_j(z)\rangle + \langle c_M(z), q^v_M(z)\rangle \mid \mathcal{Y}_{k+1}\Big] q_k = q\Big]\Big\}.
\end{aligned}$$

Using the Lattice property for the controls, (see Lemma 16.14 of [2]), the inner minimization and first expectation can be interchanged, so this is

$$\begin{aligned}
&= \bigwedge_{u \in \underline{U}(k)} \Big\{\overline{E}[\langle c_k(z,u_k), q(z)\rangle \mid q_k = q] \\
&\quad + \overline{E}\Big[\bigwedge_{v \in \underline{U}(k+1,M-1)} \overline{E}\Big[\sum_{j=k+1}^{M-1} \langle c_j(z,v_j), q^v_j(z)\rangle + \langle c_M(z), q^v_M(z)\rangle \mid \mathcal{Y}_{k+1}\Big] q_k = q\Big\} \\
&= \bigwedge_{u \in \underline{U}(k)} \overline{E}\big[\langle c_k(z,u_k), q(z)\rangle + V(k+1, q^u_{k+1}) \mid q_k = q\big],
\end{aligned}$$

and the result follows.

COROLLARY 4.6. *Write* $\underline{U}_S(k, M-1)$ *for the set of separated controls on* $\{k, k+1, \dots, M-1\}$. *Then for* $q \in S$

$$V(k,q) = \bigwedge_{u \in \underline{U}(k,M-1)} V_k(q,u) = \bigwedge_{u \in \underline{U}_S(k,M-1)} V_k(q,u).$$

Proof. The proof will use backward induction in k. Clearly $V(M,q) = V_M(q) = \langle c_M(z), q(z)\rangle$ and the result holds for $k = M$. Suppose the result is true for $k+1$, $k+2$, $\ldots$, M. Then from Theorem 4.5

$$V(k,q) = \bigwedge_{u\in\underline{U}(k)} \overline{E}\big[\langle c_k(z,u_k), q(z)\rangle + V(k+1, q^u_{k+1}) \mid q_k = q\big].$$

It is clear that a minimizing u_k, (or a sequence of minimizing u_k), depend only on the information state $q_k = q$. Therefore,

$$\begin{aligned} V(k,q) &= \bigwedge_{u\in\underline{U}_S(k)} \overline{E}\Big[\langle c_k(z,u_k), q(z)\rangle + \bigwedge_{v\in\underline{U}_S(k+1,M-1)} V_{k+1}(q^u_{k+1}, v) \mid q_k = q\Big] \\ &= \bigwedge_{u\in\underline{U}_S(k,M-1)} V_k(q,u). \end{aligned} \tag{4.4}$$

THEOREM 4.7. *Suppose u^* is a separated control such that, for each $q \in S$, $u^*_k(q)$ achieves the minimum in (4.3). Then $V_k(q,u^*) = V(k,q)$, and u^* is an optimal control.*

Proof. We shall again prove the result by backward induction in k. Clearly

$$\begin{aligned} V_M(q,u^*) &= \langle c_M(z), q(z)\rangle \\ &= V(M,q). \end{aligned}$$

Suppose the result holds for $k+1, k+2, \ldots, M$. Then

$$\begin{aligned} V_k(q,u^*_k) &= \overline{E}[\langle c_k(z,u^*), q(z)\rangle + V_{k+1}(q^{u^*}_{k+1}, u^*) \mid q_k = q] \\ &= \overline{E}[\langle c_k(z,u^*), q(z)\rangle + V(k+1, q^{u^*}_{k+1}) \mid q_k = q] \\ &= V(k,q). \end{aligned}$$

Now for any other $u \in \underline{U}(0, M-1)$

$$V_k(q,u^*) = V(k,q) \le V_k(q,u),$$

and, in particular, $V_0(q,u^*) \le V_0(q,u)$, so u^* is optimal.

5. The Adjoint Process.

Consider any control $u \in \underline{U}(0, M-1)$. We shall suppose for simplicity of notation that the cost is purely terminal at the final time M, so

$$C(x_0, u) = c_M(x_M^u).$$

Then

$$\begin{aligned} V(\pi_0, u) &= E[c_M(x_M^u)] \\ &= \overline{E}[\langle c_M(z), q_M^u(z)\rangle]. \end{aligned}$$

THEOREM 5.1. *There is a process $\kappa_k^u(z, y^k)$, adapted to $\mathcal{Y}_k$, such that for $0 \le k \le M$*

$$\overline{E}[\langle c_M(z), q_M^u(z)\rangle \mid \mathcal{Y}_k] = \langle \kappa_k^u(z, y^k), q_k^u(z)\rangle.$$

Further, κ_k^u evolves in reverse time so that

$$\kappa_k^u(\xi, y^k) = \int_{R^d}\int_{R^m} \kappa_{k+1}^u(z, y^k, y_{k+1})\phi_{k+1}(y_{k+1} - H_k(\xi))\psi_{k+1}(z - F_k(\xi, u_k))dz\,dy_{k+1}.$$

Proof. Again we use backward induction. Define $\kappa_M^u(z, y^M) = c_M(z)$ so

$$\begin{aligned} \overline{E}[\langle c_M(z), q_M^u(z)\rangle \mid \mathcal{Y}_M] &= \langle c_M(z), q_M^u(z)\rangle \\ &= \langle \kappa_M^u(z, y^M), q_M^u(z)\rangle. \end{aligned}$$

Suppose $\kappa_{k+1}^u(z, y^{k+1})$ has been defined. Then

$$\langle \kappa_{k+1}^u(z, y^{k+1}), q_{k+1}^u(z)\rangle = \int_{R^d} \kappa_{k+1}^u(z, y^{k+1}) q_{k+1}^u(z)dz$$

and

$$\begin{aligned} &\overline{E}\big[\langle \kappa_{k+1}^u(z, y^{k+1}), q_{k+1}^u(z)\rangle \mid \mathcal{Y}_k\big] \\ &= \int_{R^d}\int_{R^d}\int_{R^m} \kappa_{k+1}^u(z, y^k, y_{k+1})\phi_{k+1}(y_{k+1})^{-1}\phi_{k+1}(y_{k+1} - H_k(\xi)) \\ &\quad \times \psi_{k+1}(z - F_k(\xi, u_k)) q_k^u(\xi)\phi_{k+1}(y_{k+1})dz\,d\xi\,dy_{k+1} \\ &= \langle \kappa_k^u(\xi, y^k), q_k^u(\xi)\rangle \end{aligned}$$

where

$$\kappa_k^u(\xi, y^k) = \int_{R^d}\int_{R^m} \kappa_{k+1}^u(z, y^k, y_{k+1})\phi_{k+1}(y_{k+1} - H_k(\xi))\psi_{k+1}(z - F_k(\xi, u_k))dz dy_{k+1}.$$

REMARKS 5.2. Note in particular

$$\begin{aligned} V(\pi_0, u) &= \overline{E}[\langle c_M(z), q_M^u(z)\rangle] \\ &= \overline{E}[\langle \kappa_0^u(\xi, y_0), \pi_0(\xi)\rangle] \\ &= \overline{E}[\langle \kappa_k^u(\xi, y^k), q_k^u(\xi)\rangle]. \end{aligned}$$

6. Parameter Estimation and Dual Control.

Suppose we have a situation where the model contains unknown parameters $\theta^1, \theta^2, \theta^3$, ... , which we also wish to estimate. That is, suppose the state dynamics and observation processes are of the form:

$$\begin{aligned} x_{k+1} &= F_k(x_k, u_k, \theta^1, \theta^2,) + w_k \\ y_{k+1} &= H_k(x_k, \theta^3) + b_k, \quad 0 \le k \le M. \end{aligned}$$

Here θ^i takes values in some measure space (Θ^i, β^i, λ^i), with λ^i a probability measure. Θ^i could be a (subset of a) Euclidean space.

For example, see [4], a simple case would be (one dimensional) linear dynamics and observations of the form

$$\begin{aligned} x_{k+1} &= \theta^1 x_k + \theta^2 u_k + w_k \\ y_{k+1} &= \theta^3 x_k + b_k. \end{aligned}$$

The analysis of the previous sections goes through, taking the θ^i to be additional state variables. The unnormalized conditional density $q_n^u(z, \lambda^1, \lambda^2, \lambda^3)$ is defined by

$$\overline{E}[\overline{\Lambda}_n^u I(x_n^u \in dz) I(\theta^1 \in d\lambda^1) I(\theta^2 \in d\lambda^2) I(\theta^3 \in d\lambda^3) \mid \mathcal{Y}_n] = q_n^u(z, \lambda^1, \lambda^2, \lambda^3) dz d\lambda^1 d\lambda^2 d\lambda^3,$$

and the recursive equations (3.4) and dynamic programming results are exactly as before.

7. Conclusion.

A discrete time, partially observed control problem is discussed in separated form. Under a reference probability, which is explicitly constructed, the dynamics are given by recursive equations for the unnormalized, conditional probabilities. Dynamic programming and minimum principle results are obtained, and the extension to dual control, parameter estimation problems indicated.

References.

1. P.E. Caines, 1988. *Linear Stochastic Systems.* John Wiley, New York-Toronto.
2. R.J. Elliott, 1982. *Stochastic Calculus and Applications.* Springer–Verlag, New York-Heidelberg-Berlin.
3. R.J. Elliott. A general recursive discrete time filter. Submitted.
4. R.J. Elliott and J.B. Moore. State and parameter estimation for linear systems.
5. P.R. Kumar and P. Varaiya, 1986. *Stochastic Systems.* Prentice Hall, New Jersey.