## IV. Discrete LTR via the Kalman Filter Design

We now use the results of the previous section to establish a discrete time LQG/LTR result through singular Kalman filtering.

*Lemma 1:* In a SISO system, given $\lambda = \rho$, then $\mathcal{G}_{KP} = H(zI-F)^{-1}M^P$ is equal to $\mathcal{G}_{LQ} = -K(zI-F)^{-1}G$.

*Proof:* Choose a balanced state space realization. Assume $\lambda = \rho = \alpha$. Then

$$\begin{aligned}\mathcal{G}_{LQ} &= -K_\alpha(zI-F)^{-1}G \\ &= (M_\alpha^P)^T\Sigma(zI-\Sigma F^T\Sigma)^{-1}\Sigma H^T \\ &= (M_\alpha^P)^T(zI-F^T)^{-1}H^T \\ &= (\mathcal{G}_{KP})^T \\ &= \mathcal{G}_{KP}.\end{aligned}$$ □

*Theorem 2:* In a SISO system, provided $\det(HG) \neq 0$ and $\mathcal{P}(z)$ is minimum phase, then given $\rho = 0$

$$\mathcal{G}_{KF}^{LQ}(z) = \mathcal{G}_{LQ}(z). \quad (29)$$

*Proof:* It is shown in [5] that provided $\det(HG) \neq 0$ and $\mathcal{P}(z)$ is minimum phase then

$$\mathcal{G}_{KF}^{LQ}(z) = \mathcal{G}_{KP}(z), \quad \text{if } \lambda = 0. \quad (30)$$

Equation (30) is true $\forall\rho$. Choose $\rho = \rho^*$. From Theorem 1 the left-hand side of (30), which is equal to $\mathcal{P}(z)\mathcal{C}_{KF}^{LQ}(z)$, is the same if we interchange $\lambda$ and $\rho$. That is if $\lambda = \rho^*$ and $\rho = 0$. From the previous lemma, the RHS of (30) is equivalent to $\mathcal{G}_{LQ}(z)$ if $\rho = 0$. The theorem result follows. □

## V. Conclusion

We have demonstrated that in the special case of a SISO plant, a relationship exists between the discrete time LQ design and Kalman filter design. As a result of this, loop transfer recovery is possible through the Kalman filter design.

## References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Control—Linear Quadratic Methods.* Englewood Cliffs, NJ: Prentice-Hall, 1990.

[2] K. V. Fernando and H. Nicholson, "On the structure of balanced and other principle representations of SISO systems," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 229–231, 1983.

[3] T. Ishihara and H. Takeda, "Loop transfer recovery techniques for discrete-time optimal regulators using prediction estimators," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 1149–1151, 1986.

[4] R. E. Kalman, "When is a linear control system optimal?" *Trans. A.S.M.E. Ser. D:J. Basic Eng.*, vol. 86, pp. 1–10, 1964.

[5] J. M. Maciejowski, "Asymptotic recovery for discrete-time systems," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 602–605, 1985.

[6] R. J. Ober, "Topology of the set of asymptotically stable minimal systems," *Int. J. Contr.*, vol. 46, no. 1, pp. 263–280, 1987.

# Optimal Control of Arrivals to Queues with Delayed Queue Length Information

Joy Kuri and Anurag Kumar

***Abstract*—We consider discrete-time versions of two classical problems in the optimal control of admission to a queueing system: i) optimal routing of arrivals to two parallel queues and ii) optimal acceptance/rejection of arrivals to a single queue. We extend the formulation of these problems to permit a $k$ step delay in the observation of the queue lengths by the controller. For geometric inter-arrival times and geometric service times the problems are formulated as controlled Markov chains with expected total discounted cost as the minimization objective.**

**For problem i) we show that when $k = 1$, the optimal policy is to allocate an arrival to the queue with the smaller expected queue length (JSEQ: Join the Shortest Expected Queue). We also show that for this problem, for $k \geq 2$, JSEQ is not optimal.**

**For problem ii) we show that when $k = 1$, the optimal policy is a threshold policy. There are, however, two thresholds $m_0 \geq m_1 > 0$, such that $m_0$ is used when the previous action was to reject, and $m_1$ is used when the previous action was to accept.**

## I. Introduction

Problems in the control of queueing systems often arise in communication networks, computer systems, and manufacturing systems. Explicit structural results for optimal policies in such control problems are usually very difficult to obtain and have only been derived for the simplest of problems (see, e.g., [16], [15], [3], [4], [10], and [12]). All of these formulations assume that at the decision epochs the instantaneous queue length information is available to the control algorithm. In practice, however, the controller may only be able to observe old queue lengths. For example, in a packet communication network or a distributed computer system, the source of traffic (that is to be controlled) and the sink of traffic (at which congestion is of concern) are connected by communication links, and consequently there are propagation delays.

We consider discrete-time and delayed queue-length information versions of two classical problems for which explicit structural results have been obtained for the zero delay case in the above mentioned references.

The first problem is that of optimally allocating arriving customers to one of two parallel queues so as to minimize the expected total discounted number in the system. For exponential service times the optimality of the join the shortest queue (JSQ) policy is the well-known result for this problem ([16], [15], [3]). We consider a discrete-time version of this problem, with geometric inter-arrival times and geometric service times. Further, we assume that the queue lengths are observed only after a delay of $k$ time steps, whereas, of course, all the previous control actions are known to the controller. We show that for $k = 1$, the optimal policy is for the controller to calculate the expected queue lengths conditioned on the most recently known queue lengths (i.e., for $k = 1$, the one-step old queue lengths) and the controls applied since then (i.e., for $k = 1$, the last control action), and then allocate an arrival to the queue with the smaller expected length, i.e., the policy is now join the shortest expected

Manuscript received September 8, 1994; revised January 7, 1995. Part of this paper appeared in *Proc. 31st IEEE CDC Conf. on Decision and Control* (Tucson, AZ, Dec. 16–18, 1992), pp. 991–992.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012, India.

IEEE Log Number 9412293.

0018-9286/95$04.00 © 1995 IEEE

queue (JSEQ). We also prove that for $k \geq 2$, JSEQ is no longer optimal by giving an example in which a different policy does strictly better (for a suitable value of the discount factor).

The second problem is that of optimally accepting or rejecting customers arriving to a single queue so as to minimize the expected total discounted cost, where there is a fixed cost per unit time for each queued customer and there is a reward for each accepted customer. For general i.i.d. arrivals and exponential service times it has been shown that the optimal policy is of threshold type, i.e., there is an integer $m$ such that arrivals are accepted so long as the queue length is less than $m$ and rejected otherwise. We consider a discrete-time version of this problem, with geometric inter-arrival times and geometric service times. Further, we assume that the queue length is observed at the controller only after a delay of $k$ units. We show that for $k = 1$, the optimal policy is again of threshold type. There are two thresholds $m_0 \geq m_1 > 0$; when the previous action was to accept (respectively, reject) the policy is to accept an arrival if the queue length one step back was less than $m_1$ (respectively, $m_0$).

In each case the approach is to formulate the problem as a completely observed controlled Markov chain, whose state is the queue length(s) $k$ steps back, and the previous $k$ control actions. In each problem, the optimal value function is shown to possess certain natural properties. These when used in the Dynamic Programming equation yield the desired structural properties (see [4], [14], and [7]).

The problem of control of discrete-time systems with delayed information has been considered before in the contexts of delayed information sharing patterns in decentralized control ([17], [13], [6]) and routing under imperfect information ([2]). Our problems can be cast into the framework of [13] if we consider two "controllers," one of which only makes noiseless observations (the full history) and exercises no control, and the other only controls and has no observations of its own. The information is shared after a delay of $k$. But since our problems can be formulated as Completely Observed Controlled Markov Chains with the state being the queue length(s) $k$ steps back, and the past $k$ actions at the controller, the Conjecture in [13, Section I] holds true for every $k$ in our problems.

In [2], Beutler and Teneketzis develop a general approach for showing the optimality of switch-type policies for the routing problem under imperfect information. They provide conditions for the value function to possess the property of submodularity, which implies that a switching-type policy is optimal. The formulation in [2] is different from ours, however, in the following respects: a) arrivals are perfectly observed by the controller, b) it is assumed that a finite number of arrivals occurs, and c) the feedback delay is random. With our approach, we are able to provide a complete characterization of the optimal policy that is of switching type, without recourse to submodularity.

Results that are closely related to our results for the flow control problem presented in this correspondence have been arrived at independently by Altman and Nain [1]. We note that a summary of our results reported here has already appeared in [8].

Our note is organized as follows. In Section II we present the results of the problem of optimal routing to parallel queues. In Section III, the problem of optimal acceptance/rejection of arrivals is presented. We conclude in Section IV.

## II. Optimal Customer Allocation to Two Parallel Queues

In this section, we consider the classical problem of optimal customer allocation to two queues in parallel, with the additional feature that the queue lengths are available at the scheduler not instantaneously, but only after some delay.

In our model, we assume that time $t$ is discrete. Let $(q_1(t), q_2(t))$ denote the discrete-time queue length process, where by "queue length" we mean the total number in the queue, including the service position. At time $t, t \in \{0, 1, 2, \cdots\}$ the controller must decide on a control action $u(t) \in \{1, 2\}$, and is allowed only to observe the queue lengths until time $t-k$ and, of course, knows all control actions until time $t-1$. In particular, we assume that at time zero the process has already been evolving since time $-k$ and the controller is given $(q_1(-k), q_2(-k))$ and $((u(-k), u(-k+1), \cdots, u(-1))$; this is the given initial condition. The problem is to choose $\{u(0), u(1), \cdots\}$ so as to optimize a cost function.

We remark here that the controller does not use any information about the arrival process. This models situations where the controller has access to the queue lengths (delayed) but cannot see the arrivals. For example, if the control action has to be computed at the queues and sent to the router, then a decision made by the controller impacts the queue lengths after a round-trip delay, and the arrival information during this period is unavailable to it.

The arrivals and departures occur as follows. An arrival occurs to the system with probability $\lambda$ at $t = n+, n \geq -k$, and a departure occurs from a nonempty queue with probability $\mu$ at $t = n-, n \geq -k+1$. The control action at $t = n, n \geq -k$, decides to which queue an arrival at $n+$ must be routed. If no customer arrives at $n+$ then the decision has no effect.

Thus the scheduler, at time $t$, has the information

$$\{\{q_i(t-l)\}_{l=k}^{t+k}, i = 1, 2\}$$

and

$$\{\{u(t-l)\}_{l=1}^{t+k}\}.$$

We need a policy $\pi$ for choosing $\{u(0), u(1), u(2)\cdots\}$ so as to minimize the cost function

$$E_{s(0)}^{\pi}\left[\sum_{n=0}^{\infty} \beta^n(\lambda + q_1(n) + q_2(n))\right]$$

where

$$s(0) = \{\{q_i(-k)\}, i = 1, 2, \{u(-k), \cdots, u(-1)\}\}$$

and $\beta \in (0, 1)$ is a discount factor. It is clear that $(\lambda + q_1(n) + q_2(n))$ is the expected holding cost of customers in the interval $n \in \{0, 1, 2, \cdots\}$ if the holding cost per customer per time step is 1.

The problem posed above can be formulated as a partially observed controlled Markov chain (PO-CMC), and then converted into a completely observed controlled Markov chain (CO-CMC) with the state being the "information state" (see, for example, [5]). It is, however, quite natural to formulate this problem directly as a CO-CMC, which we proceed to do here; the formulation as a PO-CMC and the conversion into the corresponding equivalent CO-CMC are not shown owing to lack of space (see [9]). We show the formulation for $k = 1$. This is for ease of notation and clarity. It should be clear what the formulation for $k \geq 2$ is.

We list the elements of the CO-CMC for $k = 1$ as follows

a) State at time $n$

$$s(n) = (q_1(n-1), q_2(n-1), u(n-1)) \quad \forall n \in \mathcal{N}.$$

So the state space is $\mathcal{N} \times \mathcal{N} \times \{1, 2\}$.

b) Action at time $n$

$$u(n) \in \{1, 2\} \; \forall n \in \mathcal{N}.$$

So the action space is $\{1, 2\}$.

c) Transition Probabilities: Let

$$i = (i_1, i_2, v)$$
$$j = (j_1, j_2, w), \quad i_1, i_2, j_1, j_2 \in \mathcal{X}, \; v, w \in \{1, 2\}.$$

Then

$$\begin{aligned}\text{Prob}\,(s(n+1) = j | s(n) = i, u(n) = d) &= I\{w = d\} \\ &\times \text{Prob}\,(q_1(n) = j_1, q_2(n) = j_2 | q_1(n-1) = i_1, \\ &\quad q_2(n-1) = i_2, u(n-1) = v).\end{aligned}$$

We denote by $P_v$ the $\mathcal{X}^2 \times \mathcal{X}^2$ matrix with elements

$$\begin{aligned}\text{Prob}\,(q_1(n) = j_1, q_2(n) = j_2 | q_1(n-1) \\ = i_1, q_2(n-1) = i_2, u(n-1) = v).\end{aligned}$$

Consider a function $f\colon \mathcal{X}^2 \to \Re$ and think of it as a column vector on $\mathcal{X}^2$, whose $(x_1, x_2)$th element is $f(x_1, x_2)$. Now for an $\mathcal{X}^2 \times \mathcal{X}^2$ matrix (say $Q$), denote by $Qf$ the column vector on $\mathcal{X}^2$, whose $(x_1, x_2)$th element (i.e., $(Qf)(x_1, x_2)$) is the product of the $(x_1, x_2)$th row of $Q$ and the column vector $f$.

Define $\sigma\colon \mathcal{X}^2 \to \Re$, with $\sigma(x_1, x_2) = x_1 + x_2$. Also define "arrival" and "departure" operators $\alpha$ and $\delta$ as follows

$$\alpha_1(x_1, x_2) = (x_1 + 1, x_2),$$
$$\alpha_2(x_1, x_2) = (x_1, x_2 + 1),$$
$$\delta_1(x_1, x_2) = ((x_1 - 1)^+, x_2)$$

etc. Then $(P_v\sigma)(y_1, y_2)$ is the expected total population in the two queues given that the queue lengths one step back were $(y_1, y_2)$ and the previous control was $v$.

Defining

$$\begin{aligned}(P\sigma)(\underline{y}) = \mu\overline{\mu}\{\sigma(\delta_1(\underline{y})) + \sigma(\delta_2(\underline{y}))\} \\ + \mu^2\sigma(\delta_1(\delta_2(\underline{y}))) + \overline{\mu}^2\sigma(\underline{y})\end{aligned}$$

we see that

$$(P_v\sigma)(\underline{y}) = \overline{\lambda}(P\sigma)(\underline{y}) + \lambda(P\sigma)(\alpha_v\underline{y}).$$

d) The one-step cost is defined to be

$$\begin{aligned}c(s(n), u(n)) &= E[a(n) + q_1(n) + q_2(n) | s(n)] \\ &= \lambda + (P_{u(n-1)}\sigma)(q_1(n-1), q_2(n-1))\end{aligned}$$

where $a(n) \in \{0, 1\}$ indicates whether there was an arrival at time $n$ or not (one if there was an arrival and zero if not). We note that $c(s(n), u(n))$ does not depend on $u(n)$, since the arrival at $n$, if any, must be routed to one of the two queues.

e) Performance Criterion: We use the discounted cost criterion

$$J^\beta(\pi, s(0)) = E^\pi_{s(0)}\left[\sum_{n=0}^{\infty} \beta^n c(s(n), u(n))\right]$$

where $s(0)$ is the initial state, $\pi$ is the policy, and $\beta \in (0, 1)$ is the discount factor, and $E^\pi_{s(0)}$ denotes expectation under the law of the controlled process with policy $\pi$ and initial state $s(0)$.

We observe that $J^\beta(\pi, s(0))$ exists, since $c(s(n), u(n))$ can grow at most linearly with $n$ while the discount factor $\beta^n$ decreases exponentially, and thus dominates. Therefore, applying the Bounded Convergence Theorem, we can take the expectation inside the summation, and substituting for $c(s(n), u(n))$, we find that

$$J^\beta(\pi, s(0)) = E^\pi_{s(0)}\left[\sum_{n=0}^{\infty} \beta^n(\lambda + q_1(n) + q_2(n))\right].$$

Thus the performance criterion is nothing but the expected total discounted population in the two queues; this is the same as the criterion presented in Section II.

### A. Optimality of JSEQ for One-Step Delay

Let the initial state be $s(0) = (x_1, x_2, d)$; the control actions are to be chosen from this epoch onwards. Define, for $d \in \{1, 2\}$

$$V^*(x_1, x_2, d) = \min_\pi E^\pi_{(x_1, x_2, d)}\left[\sum_{n=0}^{\infty} \beta^n c(s(n), u(n))\right].$$

For ease of notation and readability, we shall write $V^*(x_1, x_2, d)$ as $V^*_d(x_1, x_2)$. Then the dynamic programming equations are

$$\begin{aligned}V^*_1(x_1, x_2) &= \min\{c((x_1, x_2, 1), 1) + \beta(P_1V^*_1)(x_1, x_2), \\ &\qquad c((x_1, x_2, 1), 2) + \beta(P_1V^*_2)(x_1, x_2)\} \\ &= \lambda + (P_1\sigma)(x_1, x_2) \\ &\qquad + \beta\min\{(P_1V^*_1)(x_1, x_2), (P_1V^*_2)(x_1, x_2)\}. \quad (1)\end{aligned}$$

Similarly

$$\begin{aligned}V^*_2(x_1, x_2) &= \lambda + (P_2\sigma)(x_1, x_2) \\ &\qquad + \beta\min\{(P_2V^*_1)(x_1, x_2), (P_2V^*_2)(x_1, x_2)\}. \quad (2)\end{aligned}$$

Formally, the function $V^*(x_1, x_2, d)$ can be viewed as the unique fixed point of the dynamic programming operator $T$ (defined below). Let $S := \mathcal{X} \times \mathcal{X} \times \{1, 2\}$; $S$ is the state space of the problem. Define the norm of a real-valued function $v$ defined on $S$ as

$$\|v\| = \sup_{(x_1, x_2, d) \in S} \frac{|v(x_1, x_2, d)|}{(\|x_1, x_2, d\|_2 \vee 1)^r}$$

where $\|\cdot\|_2$ is the $l_2$ norm, $\vee$ is the max operator, and $r > 0$ is an integer (see [11] and [9, the Appendix] for details). Consider the space of real-valued functions

$$\mathcal{V} = \{v\colon S \to \Re \text{ such that } \|v\| < \infty\}$$

metrized by the metric $\rho(v, w) := \|v - w\|$ for $v, w \in \mathcal{V}$. Thus $\mathcal{V}$ is the space of real-valued functions on $S$ that increase at most polynomially in the $l_2$ norm of $(x_1, x_2, d) \in S$. It can be seen without difficulty that the optimal value function in our problem belongs to the space $\mathcal{V}$. Define the operator $T$ that picks a function $f \in \mathcal{V}$ and transforms it to another function $Tf$ as follows

$$\begin{aligned}(Tf)_1(x_1, x_2) &= \lambda + (P_1\sigma)(x_1, x_2) \\ &\qquad + \beta\min\{(P_1f_1)(x_1, x_2), (P_1f_2)(x_1, x_2)\} \\ (Tf)_2(x_1, x_2) &= \lambda + (P_2\sigma)(x_1, x_2) \\ &\qquad + \beta\min\{(P_2f_1)(x_1, x_2), (P_2f_2)(x_1, x_2)\}.\end{aligned}$$

It can be shown (see [9, the Appendix] for details) that $\langle \mathcal{V}, \rho \rangle$ is a complete metric space, that $Tf \in \mathcal{V}$ and moreover, the operator $T$ has a unique fixed point in $\mathcal{V}$. Noting that (1) and (2) can be written as $V^* = TV^*$, it is clear that $V^*$ is the fixed point of $T$.

We shall show that the optimal value function $V^*_d(x_1, x_2)), d \in \{1, 2\}$, has the following properties:

P1) $V^*_i(\alpha_j\underline{x}) \geq V^*_i(\underline{x}), i, j \in \{1, 2\}$, i.e., $V^*_1(\cdot)$ and $V^*_2(\cdot)$ are coordinate-wise increasing.

P2) $V^*_1(\underline{x}) = V^*_2(\underline{x}')$, which is a consequence of the symmetry of the system (if $\underline{x} = (x_1, x_2)$, then $\underline{x}' = (x_2, x_1)$).

P3) $\forall \underline{x}$ with $x_1 < x_2, V^*_1(\underline{x}) \leq V^*_2(\underline{x})$. P3) says that if the initial state is more "unbalanced" then the cost associated with it is more. If $x_1 < x_2$ then the initial state $(x_1, x_2, 2)$ is more unbalanced than $(x_1, x_2, 1)$ in the sense that if we start with $(x_1, x_2, 2)$, it is more likely that one server will starve while the other has a queue of customers waiting for it.

P4) $\forall \underline{x}$ with $x_1 \leq x_2, V^*_i(\delta_1(\alpha_2(\underline{x}))) \geq V^*_j(\underline{x}), i, j \in \{1, 2\}$. P4) also says that the cost associated with a more unbalanced state is more.

The above properties are established in the following lemmas and theorems. We remark here that the difficult part is to obtain a closed set of properties (e.g., P1)–P4) above); the proof that the optimal value function has these properties is a matter of writing down the expressions and checking through various cases. In the proofs of the results that follow, this theme occurs repeatedly; we shall not give the details for want of space—see [9] for details.

*Lemma II-A.1:* Let $v_1: \mathcal{N}^2 \to \Re, v_2: \mathcal{N}^2 \to \Re$ be two functions. If $v_1(\underline{x})$ and $v_2(\underline{x})$ satisfy the relations in P1)–P4), then $(Pv_1)(\underline{x})$ and $(Pv_2)(\underline{x})$ also satisfy the same relations.

*Proof:* Several cases arise that need to be routinely checked. □

*Lemma II-A.2:* If $v \in \mathcal{V}$ has P1)–P4), then $Tv \in \mathcal{V}$ has P1)–P4); i.e., the dynamic programming operator $T$ preserves P1)–P4).

*Proof:* The fact that $Tv \in \mathcal{V}$ is proved in [9, Appendix]. To see that $Tv$ has P1)–P4), we write down the expressions for $T\boldsymbol{v}$ and use Lemma II-A.1 (details in [9]). □

*Theorem II-A.3:* $V^*(x_1, x_2, d)$ has properties P1)–P4).

*Proof:* We first observe that the function $v(\underline{x}) = 0, \forall \underline{x} \in S$ has properties P1)–P4) trivially. Next, consider the set of functions $\boldsymbol{H} = \{v \in \mathcal{V}: v \text{ has P1)–P4)}\}$. Using the fact that convergence under $\rho$ implies pointwise convergence (see [9, Appendix] for details), it can be proved that $\boldsymbol{H}$ is a closed set. Now the claim follows from Lemma II-A.2, and the facts that $V^*(\cdot)$ is a fixed point of $T$ and $\boldsymbol{H}$ is a closed set. □

The following lemma helps to characterize the optimal policy:

*Lemma II-A.4:* Let $e_1(\underline{x}) = x_1$ and $e_2(\underline{x}) = x_2$. For $0 < \lambda < 1$ and $0 < \mu < 1$

$$x_1 < x_2 \Leftrightarrow (P_1e_1)(\underline{x}) < (P_1e_2)(\underline{x})$$
$$x_1 \geq x_2 \Leftrightarrow (P_1e_1)(\underline{x}) > (P_1e_2)(\underline{x})$$

and similarly

$$x_1 \leq x_2 \Leftrightarrow (P_2e_1)(\underline{x}) < (P_2e_2)(\underline{x})$$
$$x_1 > x_2 \Leftrightarrow (P_2e_1)(\underline{x}) > (P_2e_2)(\underline{x}).$$

*(Remark:* Note that, for example, $(P_1e_1)(\underline{x})$ is the expected queue length in queue 1 given that the state one step back was $\underline{x}$ and the previous control was one).

*Proof:* This is a matter of checking the expressions. □

Let $s(n) = (\underline{x}, i)$ where $i = 1$ or 2. Define the join the shortest expected queue (JSEQ) policy as that which chooses $u(n) = 1$ if $(P_ie_1)(\underline{x}) < (P_ie_2)(\underline{x})$, and $u(n) = 2$, otherwise. The name follows, because

$$E[q_1(n)|s(n)] = (P_ie_1)(\underline{x}), E[q_2(n)|s(n)] = (P_ie_2)(\underline{x}).$$

*Theorem II-A.5:* For unit delay, i.e., $k = 1$, the JSEQ policy is optimal.

*Proof:* Suppose $s(n) = (\underline{x}, 1)$ and we find

$$(P_1e_1)(\underline{x}) < (P_1e_2)(\underline{x}).$$

Then Lemma II-A.4 implies $x_1 < x_2$. Further, P3) shows that $V_1^*(\underline{x}) \leq V_2^*(\underline{x})$. Now consider (1) and (2). Expanding the terms within $\min\{.\}$, and using Lemma II-A.1, it can be seen that the first term is $\leq$ the second term. This proves that the optimal decision is 1, which is what the JSEQ policy indicates. Other cases can be shown similarly. □

### B. Nonoptimality of JSEQ for $\geq$ 2 Steps Delay

In this section, we show that the JSEQ policy which was optimal for delay $k = 1$, ceases to remain so for $k \geq 2$.

We recall the cost criterion

$$J^\beta(\pi, s(0)) = E^\pi_{s(0)}\left[\sum_{n=0}^{\infty} \beta^n c(s(n), u(n))\right] = \frac{\lambda}{1-\beta} + \sum_{n=0}^{\infty} \beta^n g^\pi(n)$$

where $g^\pi(n) = E^\pi_{s(0)}(q_1(n) + q_2(n))$, with $s(0)$, the initial information state being suppressed in the notation. Observe that $g^\pi(0)$ does not depend on the policy $\pi$.

Let $A, D_1$, and $D_2$ be mutually independent Bernoulli random variables, with $A = 1$ w.p. $\lambda$, and 0, otherwise, and for $i = 1, 2, D_i = 1$ w.p. $\mu$, and 0, otherwise. Let $\pi_1$ be a policy which chooses action 1 at $t = 0$. Then

$$g^{\pi_1}(1) = E_{s(0)}[(q_1(0) + A - D_1)^+ + (q_2(0) - D_2)^+].$$

Similarly, let $\pi_2$ be a policy which chooses action 2 at $t = 0$. Then

$$g^{\pi_2}(1) = E_{s(0)}[(q_1(0) - D_1)^+ + (q_2(0) + A - D_2)^+].$$

*Lemma II-B.1:* $A, D_1, D_2, (q_1(0), q_2(0))$ are mutually independent; $A, D_1, D_2$ are Bernoulli with parameters $\lambda, \mu, \mu$, respectively. Then

$$g^{\pi_1}(1) \leq g^{\pi_2}(1) \Leftrightarrow P(q_1(0) = 0|s(0)) \geq P(q_2(0) = 0|s(0)).$$

*Proof:* Let

$$Z_1 = (q_1(0) + A - D_1)^+ + (q_2(0) - D_2)^+$$

and

$$Z_2 = (q_1(0) - D_1)^+ + (q_2(0) + A - D_2)^+.$$

Then

$$Z_2 - Z_1 = \begin{cases} 0 & \omega \in \{q_1(0) > 0, q_2(0) > 0\} \\ A - (A - D_1)^+ & \omega \in \{q_1(0) = 0, q_2(0) > 0\} \\ (A - D_2)^+ - A & \omega \in \{q_1(0) > 0, q_2(0) = 0\} \\ (A - D_2)^+ - (A - D_1)^+ & \\ & \omega \in \{q_1(0) = q_2(0) = 0\}. \end{cases}$$

Thus $Z_2 - Z_1$ takes values in $\{-1, 0, 1\}$, and we have

$$g^{\pi_2}(1) - g^{\pi_1}(1) = E_{s(0)}[Z_2 - Z_1] = \lambda\mu[P(q_1(0) = 0|s(0)) - P(q_2(0) = 0|s(0))]$$

from which the result follows. □

*Counterexample:* Let $k = 2$. The state of the CMC at $t = 0$ is given by $s(0) = ((q_1(-2), q_2(-2)), u(-2), u(-1))$. Let $s(0) = ((3, 2), 2, 2), \lambda = 0.6, \mu = 0.9$. Calculating, we find $E[q_1(0)|s(0)] < E[q_2(0)|s(0)]$. Hence, the JSEQ policy would choose action 1 at time 0. But, observe that $\text{Prob}(q_1(0) = 0|s(0)) = 0$ as there were three customers in queue 1 at $t = -2$, and only two departures can occur till $t = 0$, i.e., at $t = -1$—and $t = 0$—whereas $\text{Prob}(q_2(0) = 0|s(0)) > 0$.

Let the policy that chooses the more probably empty queue at each epoch be denoted by MPE. From Lemma II-B.1, it can be seen that $g^{\text{JSEQ}}(1) > g^{\text{MPE}}(1)$. The following result shows that this extra cost paid by the JSEQ policy at step 1 cannot be recovered in the future for some $\beta \in (0, 1)$.

*Proposition II-B.2:* There exists $\beta \in (0, 1)$ such that

$$\sum_{n=0}^{\infty} \beta^n g^{\text{JSEQ}}(n) > \sum_{n=0}^{\infty} \beta^n g^{\text{MPE}}(n).$$

*Proof:* This follows from the observation that there exists $\beta$ such that

$$\sum_{n=2}^{\infty} \beta^{n-1} g^{\mathrm{MPE}}(n) < g^{\mathrm{JSEQ}}(1) - g^{\mathrm{MPE}}(1)$$

since the left-hand side goes to zero as $\beta$ goes to zero, while the right-hand side is a positive constant. Multiplying by this particular $\beta$ we have

$$\sum_{n=2}^{\infty} \beta^{n} g^{\mathrm{MPE}}(n) < \beta(g^{\mathrm{JSEQ}}(1) - g^{\mathrm{MPE}}(1)).$$

Transposing and adding $g^{\mathrm{JSEQ}}(0) = g^{\mathrm{MPE}}(0)$ to both sides, the result follows. □

Thus we have a counterexample which shows that the JSEQ policy is not optimal for $k = 2$. It is clear that similar examples can be given for $k > 2$ as well.

## III. Optimal Customer Acceptance/Rejection at a Single Queue

In this section, we consider a model motivated by packet-switched data communication networks. A packet transmitter and a receiver in such a network are located some distance apart. The receiver controls the flow of packets from the transmitter by sending start/stop signals.

We assume that:

a) Time is slotted.
b) The round-trip propagation delay is $k$ slots, where $k$ is an integer.
c) At the beginning of each slot, the receiver sends the transmitter one of two control signals: zero (do not send) and one (send one packet).
d) At the transmitter, a packet is generated in each slot with probability $\lambda$. If the transmitter has a permit to send, the generated packet is transmitted; otherwise, the packet is immediately dropped. There is no queueing at the transmitter, and credits are not accumulated.
e) The service time $S$ of a packet is geometric $(\mu)$

$$\mathrm{Prob}\,(S = m) = (1-\mu)^{m-1}\mu, m \geq 1.$$

f) The receiver gets a reward of 1 for each packet accepted and pays a penalty of $0 < b < 1$ per queued packet per slot.

Observe that the decision computed by the receiver impacts it after a delay of $k$ slots. Equivalently, the decision at any epoch is based on receiver queue-length information that is $k$ slots old. The objective is to look for control strategies for which the cost at the receiver is minimized in some sense.

It is clear that the above model yields an extension of the classical problem of optimal acceptance/rejection of arrivals to a queue (see, for example, [12]), with the additional feature that the controller is only permitted to observe queue-length information delayed by $k$ steps. Proceeding formally, let $q(t)$ denote the queue length process. At time $t, t \in \{0, 1, 2, \cdots\}$, the controller computes a control value $u(t) \in \{0, 1\}$, and is allowed only to observe the queue lengths until time $t-k$, and all controls until time $t-1$. In particular, at time zero, the controller knows $q(-k)$ and $\{u(-k), u(-k+1), \cdots, u(-1)\}$. The problem is for the controller to choose $\{u(0), u(1), \cdots\}$ to optimize a cost function.

As mentioned earlier, a decision computed by the receiver affects it after a delay of $k$ slots. Observe that the arrival information during this period cannot be used by the control algorithm, since the receiver cannot know the arrival pattern at the point in time when it has to make the decision.

Arrivals and departures occur as follows. An arrival occurs to the system with probability $\lambda$ at $t = n+, n \geq -k$. The arrival is accepted if $u(n) = 1$, otherwise it is rejected. A departure occurs from a nonempty queue with probability $\mu$ at $t = n-, n \geq -k+1$. The controller, at time $t$, has the information $\{\{q(t-l)\}_{l=k}^{t+k}$ and $\{u(t-l)\}_{l=1}^{t+k}\}$. We need a policy $\pi$ for choosing $\{u(0), u(1), \cdots\}$ so as to minimize the cost function

$$E_{s(0)}^{\pi}\left[\sum_{n=0}^{\infty} \beta^{n}(bq(n) - \lambda(1-b)u(n))\right]$$

where

$$s(0) = \{\{q(-k)\}, \{u(-k), \cdots, u(-1)\}\}$$

and $bq(n) - \lambda(1-b)u(n)$ is the expected cost in the interval $n$ if the holding cost per customer per time step is $b$, and the reward for customer acceptance is 1.

As in Section II, we formulate the problem as a CO-CMC for $k = 1$. The elements of the CO-CMC are as follows:

a) State at time $n$

$$s(n) = (q(n-1), u(n-1)), n \in \mathcal{N}.$$

So the state space is $\mathcal{N} \times \{0, 1\}$.

b) Action at time $n$: $u(n) \in \{0, 1\}$. So the action space is $\{0, 1\}$.

c) Transition Probabilities: Let

$$y = (y_1, r), l = (l_1, z), y_1, l_1 \in \mathcal{N}, r, z \in \{0, 1\}.$$

Then

$$\begin{aligned}\mathrm{Prob}\,(s(n+1) = l \mid s(n) = y, u(n) = \delta) \\ = I\{z = \delta\}P(q(n) = l_1 \mid q(n-1) = y_1, u(n-1) = r).\end{aligned}$$

We denote by $P_r$ the $\mathcal{N} \times \mathcal{N}$ matrix with elements

$$\mathrm{Prob}\,(q(n) = l_1 \mid s(n) = (y_1, r)).$$

Define two column vectors on $\mathcal{N}$ as follows: $b_0 = b(0, 1, 2, 3, \cdots)^t$ where $(\cdots)^t$ denotes transpose, and $b_1 = b_0 - \lambda(1-b)\underline{1}$, where $\underline{1}$ is the column vector of all 1's.

d) One-step expected cost: $c(s(n), u(n))$.

Let $s(n) = (x, r)$. For $u(n) = 0$, we define $c((x, r), 0) = (P_r b_0)(x)$ where we have adopted the same notation as in Section II. This quantity gives the product of the waiting cost per queued customer per slot and the expected number of customers at $n$, given $(x, r)$.

For $u(n) = 1$, we define

$$\begin{aligned}c((x, r), 1) &= (P_r b_1)(x) \\ &= (P_r b_0)(x) - \lambda(1-b).\end{aligned}$$

Note that for $u(n) = 1$, the cost is less than that for $u(n) = 0$, as one customer has been accepted.

e) Performance Criterion: We use the discounted cost criterion

$$J^{\beta}(\pi, s(0)) = E_{s(0)}^{\pi}\left[\sum_{n=0}^{\infty} \beta^{n} c(s(n), u(n))\right]$$

which can be seen to be the same as

$$J^{\beta}(\pi, s(0)) = E_{s(0)}^{\pi}\left[\sum_{n=0}^{\infty} \beta^{n}(bq(n) - \lambda(1-b)u(n))\right]$$

the cost criterion given earlier. We observe that $J^{\beta}(\pi, s(0))$ exists since $(bq(n) - \lambda(1-b)u(n))$ can increase at most linearly, while the discount factor $\beta^n$ decreases exponentially and thus dominates.

### A. Optimality of a Two-Threshold Policy for $k = 1$

Let the initial state be $s(0) = (x, r)$. Define

$$V_r^*(x) = \min_{\pi} E_{(x,r)}^{\pi}\left[\sum_{n=0}^{\infty} \beta^n c(s(n), u(n))\right].$$

The dynamic programming equations (DPE) are

$$\begin{aligned} V_0^*(x) &= \min\{(P_0 b_0)(x) + \beta(P_0 V_0^*)(x), \\ &\qquad (P_0 b_1)(x) + \beta(P_0 V_1^*)(x)\} \\ &= (P_0 b_0)(x) + \beta(P_0 V_0^*)(x) \\ &\quad + \beta \min\left\{0, (P_0(V_1^* - V_0^*))(x) - \frac{\lambda(1-b)}{\beta}\right\} \end{aligned} \tag{3}$$

and similarly

$$\begin{aligned} V_1^*(x) &= (P_1 b_0)(x) + \beta(P_1 V_0^*)(x) \\ &\quad + \beta \min\left\{0, (P_1(V_1^* - V_0^*))(x) - \frac{\lambda(1-b)}{\beta}\right\}. \end{aligned} \tag{4}$$

The state space for this problem is $\mathcal{X} \times \{0, 1\}$. As in the routing problem, it is possible to define a space of functions $\mathcal{V}$, a metric $\rho$, and an operator $T$ such that $V^* = TV^*$.

We shall show that the optimal value function $V^*$ has the following properties:

P1) For $i = 0, 1$ and $\forall x \in \mathcal{X}, V_i^*(\alpha(x)) \geq V_i^*(x)$, i.e., $V_0^*(\cdot)$ and $V_1^*(\cdot)$ are increasing with $x$.

P2) $\forall x \in \mathcal{X}, [V_1^*(x) - V_0^*(x)] \geq 0$, i.e., the optimal cost for the initial state $(x, 1)$ is more than that for $(x, 0)$.

P3) $\forall x \in \mathcal{X}$

$$[V_1^*(\alpha(x)) - V_0^*(\alpha(x))] \geq [V_1^*(x) - V_0^*(x)]$$

i.e., $(V_1^*(x) - V_0^*(x))$ is increasing with $x$.

P4) $\forall x \in \mathcal{X}, [V_0^*(x+1) - V_1^*(x)] \geq 0$, i.e., the optimal cost for the initial state $(x+1, 0)$ is more than that for $(x, 1)$.

P5) $\forall x \in \mathcal{X} - \{0\}$

$$[V_0^*(\alpha(x)) - V_1^*(x)] \geq [V_0^*(x) - V_1^*(\delta(x))]$$

i.e., $(V_0^*(x+1) - V_1^*(x))$ is increasing with $x$.

*Lemma III-A.1:* If $v_0(x)$ and $v_1(x)$ satisfy the relations in P1)–P5), then $\forall x \in \mathcal{X}, i, j, r \in \{0, 1\}$:

a) $(P_r v_j)(x)$ is increasing with $x$.
b) $(P_r(v_1 - v_0))(x)$ is increasing with $x$.
c) $(P_1 v_i)(x) \geq (P_0 v_i)(x)$.
d) $(P_1(v_1 - v_0))(x) \geq (P_0(v_1 - v_0))(x)$.
e) $(P_0 v_i)(x+1) \geq (P_1 v_i)(x)$.
f) $(P_0(v_1 - v_0))(x+1) \geq (P_1(v_1 - v_0))(x)$.
g) $((P_1 - P_0)(v_0))(x+1) \geq ((P_1 - P_0)(v_1))(x)$.
h) $(P_0 v_0)(x+2) - (P_0 v_1)(x+1) \geq (P_1 v_0)(x+1) - (P_1 v_1)(x)$.

*Proof:* Consider the rows of the matrix $P_r, r \in \{0, 1\}$. Each row represents the probability mass function of a random variable taking values in $\mathcal{X}$. From the structure of the matrix it is clear that in $P_r$, row $(x+1)$ is stochastically greater than row $x, x \in \mathcal{X}$. Similarly, considering $P_0$ and $P_1$, it can be seen that row $(x+1)$ of $P_0$ is stochastically greater than row $x$ of $P_1, x \in \mathcal{X}$. Now claims a)–h) above can be established using the well-known result: if $X$ and $Y$ are random variables, then $X \succeq Y \Rightarrow$ for all nondecreasing functions $f, E[f(X)] \geq E[f(Y)]$, where $\succeq$ stands for stochastic ordering. □

*Lemma III-A.2:* Let $v_0(x), v_1(x)$ be two real-valued functions on $\mathcal{X}$ having properties P3) and P5). Then $v_0(x), v_1(x)$ are convex functions.

*Proof:* This follows on applying properties P3) and P5). □

*Lemma III-A.3:* If $v$ has properties P1)–P5), then $Tv$ has P1)–P5); i.e., the dynamic programming operator preserves P1)–P5).

*Proof:* This involves writing the required expressions and checking. □

*Theorem III-A.4:* $V^*(\cdot)$ has P1)–P5).

*Proof:* The proof uses the same arguments as in Theorem II-A.3. □

*Theorem III-A.5:* For a delay of one slot, the optimal policy has the following *threshold* structure: there exist $m_0, m_1 \in \mathcal{X}, m_0 \geq m_1 > 0$ such that

if $u(n-1) = 0$ then

$$u(n) = \begin{cases} 1, & \text{if } q(n-1) < m_0 \\ 0, & \text{if } q(n-1) \geq m_0 \end{cases}$$

and, if $u(n-1) = 1$ then

$$u(n) = \begin{cases} 1, & \text{if } q(n-1) < m_1 \\ 0, & \text{if } q(n-1) \geq m_1. \end{cases}$$

Further, $m_0 \geq m_1$.

*Proof:* This is immediate from Theorem III-A.4, Lemma III-A.1 part b), and the dynamic programming equations (3) and (4). That $m_0 \geq m_1$ follows from Lemma III-A.1 part d). □

## IV. Final Remarks

We have provided explicit structural results for discrete-time infinite horizon discounted cost versions of two classical problems in the control of queues, with the additional feature that queue-length information is delayed by one time step. For the problem of customer allocation to two parallel queues, we have shown that the JSEQ policy, that is optimal for one-step delay, is not optimal for delay $\geq 2$.

In each case the approach is via a formulation as a multidimensional CO-CMC. The well-known technique of establishing the necessary properties of the value functions is used. Even for $k = 1$ a large number of properties needs to be discovered and then tediously verified. For larger $k$ this approach quickly becomes very unwieldy, and only partial results have been obtained. Further, it can be shown that for both the problems, for $k = 1$, there exist average cost optimal policies that have the same structure as the discounted cost optimal policies that we have presented in this note (see [9]).

## References


[1] E. Altman and P. Nain, "Closed-loop control with delayed information," *Performance Evaluation Rev.*, vol. 20, pp. 193–204, 1992.

[2] F. J. Beutler and D. Teneketzis, "Routing in queueing networks under imperfect information: stochastic dominance and thresholds," *Stochastics and Stochastics Reports*, vol. 26, pp. 81–100, 1989.

[3] A. Ephremides, P. Varaiya, and J. Walrand, "A simple dynamic routing problem," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 690–693, 1980.

[4] B. Hajek, "Optimal control of two interacting service stations," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 491–499, 1984.

[5] O. Hernández-Lerma, *Adaptive Markov Control Processes*. New York: Springer-Verlag, 1989.

[6] K. Hsu and S. I. Marcus, "Decentralized control of finite state Markov processes," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 426–431, 1982.

[7] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[8] J. Kuri and A. Kumar, "Optimal control of arrivals to queues with delayed queue length information," in *Proc. IEEE 31st Conf. Decis. Contr.* (Tucson, AZ, Dec. 16–18, 1992), pp. 991–992.

[9] J. Kuri, "Optimal control problems in communication networks with information delays and quality of service constraints," Ph.D. dissertation, Dept. Elec. Commun. Eng., Indian Institute of Science, Jan. 1995.

[10] W. Lin and P. R. Kumar, "Optimal control of a queueing system with two heterogeneous servers," *IEEE Trans. Automat. Contr.*, vol. AC-32, pp. 696–703, 1987.
[11] S. A. Lippman, "Semi-Markov decision processes with unbounded rewards," *Manag. Sci.*, vol. 19, no. 7, pp. 717–731, Mar. 1973.
[12] S. Stidham, "Optimal control of admission to a queueing system," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 705–713, 1985.
[13] P. Varaiya and J. Walrand, "On delayed sharing patterns," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 443–445, 1978.
[14] J. Walrand, *An Introduction to Queueing Networks.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[15] R. R. Weber, "On the optimal assignment of customers to parallel queues," *J. Appl. Probab.*, vol. 15, pp. 406–413, 1978.
[16] W. Winston, "Optimality of the shortest line discipline," *J. Appl. Probab.*, vol. 14, pp. 181–189, 1977.
[17] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proc. IEEE*, vol. 59, pp. 1557–1566, 1971.

# A New Proof of the Discrete-Time LQG Optimal Control Theorems

Mark H. A. Davis and Mihail Zervos

*Abstract*— **We present a unifying new proof for the three discrete-time linear quadratic Gaussian problems (deterministic, stochastic full information, and stochastic partial information) based on pathwise (deterministic) optimization. The essential difference between the control aspect of the three cases is that the controls should lie in different classes of "admissible controls," and we address them as constrained optimization problems using appropriate Lagrange multiplier terms.**

## I. The LQG Problems

The LQG (linear quadratic Gaussian) optimal control theorems are among the central results of linear system theory. They concern the stochastic system model

$$x_{k+1} = Ax_k + Bu_k + v_k \tag{1}$$

$$y_k = Hx_k + w_k \tag{2}$$

for $k = 0, \cdots, N-1$ where $x_k, y_k, u_k$ denote the state, observation, and control vectors at time $k$, with dimensions $n, p, m$, respectively. The noise sequence $z_k^T = (v_k^T, w_k^T)$ is Gaussian white noise, i.e., $z = (z_0, \cdots, z_{N-1})$ is a sequence of zero-mean jointly Gaussian random vectors with

$$E\left\{z_k z_j^T\right\} = \begin{bmatrix} V & U \\ U^T & W \end{bmatrix}\delta_{kj}$$

where $U, V, W$ are given matrices with $W$ strictly positive definite. The initial state $x_0$ is jointly Gaussian with, and independent of, $z$, with given mean $m_0$ and covariance matrix $\hat{P}_0$. All these random variables are defined on some probability space $(\Omega, \mathcal{F}, P)$. The optimal control problem is to determine a control process $u = (u_0, \cdots, u_{N-1})$ in some admissible class of control processes to minimize the cost

$$L(u) = E\left\{\sum_{k=0}^{N-1}(x_k^T Q x_k + 2x_k^T T^T u_k + u_k^T R u_k) + x_N^T Q_N x_N\right\} \tag{3}$$

where the matrices $Q$, $R$, and $Q_N$ are symmetric nonnegative definite with $R$ strictly positive definite. The classes of admissible control processes are subsets of the most general class $\mathcal{U}$ which contains all measurable, $L^2$-bounded control processes, i.e., all control processes $u$ such that $u_k$ is measurable and $E|u_k|^2 < \infty$ for all $k$ (note that the $L^2$ boundedness does not restrict the generality in any sense because a control which is not $L^2$-bounded has infinite cost). As is well known, the optimal control is linear feedback of the current state or state estimate; a precise formulation is given below.

Standard textbook accounts include Åström [1], Bagchi [2], Davis and Vinter [5], and Kwakernaak and Sivan [6]. In these books the full information case is treated by stochastic dynamic programming. On the other hand, using the interpretation of the output of the Kalman filter as the conditional expectation of the state given the observations whenever an admissible control is used (i.e., some variant of Theorem 1 below), the partial information problem is transformed into a full information one. Åström [1] also gives a "common" proof to the three problems by using a "completion of the squares" argument. This method (which assumes the Riccati equation) verifies the result but provides little insight.

In this paper we consider the pathwise minimization of a cost which differs from the standard one given by (3) in that an extra linear functional of the control signal (the Lagrange multiplier term) is added to it. The optimal control thus obtained lies in general in the class $\mathcal{U}$. If according to the LQG case considered, however, the Lagrange multiplier term is chosen so that both the optimal control is admissible and the Lagrange multiplier vanishes for any admissible control, then clearly this optimal control coincides with the optimal control of the corresponding LQG problem. Also note that the Lagrange multiplier term gives the "price" for perturbations outside the class of admissible controls. Related work has been done by Davis [3], [4] where the continuous-time full information LQG problem is treated. In this approach there is no stochastic optimization: all the optimization is handled by the deterministic result of Theorem 2, and the stochastic aspects are entirely concerned with the choice of the appropriate Lagrange multiplier.

There are three standard cases of the LQG problem.

*Case 1 (Deterministic):* Here $V = \hat{P}_0 = 0$, so that $v_k = 0$ and $x_0 = m_0$ a.s., whereas the current state of the system is observed exactly. In this case observations (2) are irrelevant, and the admissible controls is the class $\mathcal{U}_1$ of all sequences $u = (u_0, \cdots, u_{N-1})$ of vectors in $\mathbb{R}^m$. The optimal control sequence $u^1$ is given by

$$u_k^1 = -M_k x_k \tag{4}$$

where

$$M_k = \Theta_{k+1}^{-1}\left(T + B^T S_{k+1} A\right). \tag{5}$$

$$\Theta_k = R + B^T S_k B \tag{6}$$

and $S_N, S_{N-1}, \cdots, S_0$ is the solution of the Riccati difference equation

$$S_k = A^T S_{k+1} A + Q - M_k^T \Theta_{k+1} M_k, \qquad S_N = Q_N. \tag{7}$$

To be precise, the optimal control $u^1$ is given by $u_k^1 = -M_k x_k^1$, where $x^1$ is the solution of (1) with $u_k$ replaced by $-M_k x_k$ and

Manuscript received August 26, 1994. This work was supported by the UK Defence Research Agency under Agreement 2037/393/RAE.

The authors are with the Department of Electrical and Electronic Engineering, Imperial College, London SW7 2BT.

IEEE Log Number 9411666.

0018-9286/95$04.00 