

OPTIMAL CONTROL OF AN IMPERFECTLY OBSERVED QUEUEING SYSTEM

Wakuta, Kazuyoshi
Nagaoka Technical College

<https://doi.org/10.5109/13430>

出版情報 : Bulletin of informatics and cybernetics. 25 (3/4), pp.181-188, 1993-03. Research
Association of Statistical Sciences

バージョン :

権利関係 :

OPTIMAL CONTROL OF AN IMPERFECTLY OBSERVED QUEUEING SYSTEM

By

Kazuyoshi WAKUTA*

Abstract

A control problem of an $M/G/1$ queue with imperfectly observed queue length is considered. It is shown that our control problem can be formulated as an imperfect state information semi-Markov decision process with the countable state space and unbounded costs and that there exists an optimal stationary I-policy when the set of control parameters is finite.

1. Introduction and the control model

Recently, a control problem of queues with imperfectly observed queue length has been investigated by Wakuta [7]. The system treated there has the finite input source. In this paper, a control problem of queues with imperfectly observed queue length having an infinite input source is considered: the system concerned with is based on an $M/G/1$ queue. The infinite input source leads to the infinite state space of the system and hence to unbounded costs. The control problem is described precisely in this section at first. Next, in section 2, it is shown to be formulated as an imperfect state information semi-Markov decision process with the countable state space and unbounded costs. Finally, in Section 3, it is transformed to a perfect state information semi-Markov decision process, and the existence of an optimal stationary I-policy is demonstrated if the set of control parameters is finite.

Consider a service facility with a single server, where customers arrive to be served in accordance with a Poisson process. The service facility separates into two parts: a waiting room and a service room. When one customer is in service, the others have to wait in the waiting room. If no one is waiting in the service room and no one is in service when a customer arrive there, then he can enter the service room at once and is served according to a general distribution. Simultaneously with the accomplishment of the service, a customer in the waiting room enters the service room in turn, and the server chooses a control parameter which determines a way to serve among many ways. The server, however, cannot know the number of customers in the waiting room precisely. But he can get some information on it, for example, (i) many customers are waiting; (ii) few customers are waiting; (iii) no customer is waiting. If no customer enters the service room at the time of a departure, then the server can assume that no one is waiting. So he can know whether there is no customer or there is at least one

* Nagaoka Technical College, Nagaoka, Niigata, 940, Japan

customer at the time of a departure. The costs are incurred on the server: a holding cost and a service cost. Thus, the server has to find a rule to choose control parameters which minimize the expected cost.

The system we consider here is based on an M/G/1 queue, i.e., a single-server queueing system subject to a Poisson process of arrivals and independent service times with general distributions. The arrival rate λ is constant, and each service time has a probability distribution $F_a(t)$, $a \in A$, $t \in R_+$, where A is a Borel set, the set of control parameters and $R_+ = [0, \infty)$. We assume that $F_a(t)$ has a density function $f_a(t)$ which is Borel measurable on $A \times R_+$. The state of the system is the queue length, i.e., the number of customers in the service facility and is assumed to be only imperfectly observed through the observations generated by an observation system. The observations are generated at the initial time and the times of successive departures, and then, control parameters are chosen, based on the observable histories. We refer to these times as decision epochs. The observation system is characterized by a stochastic kernel g on M given S , where $S = \{0, 1, 2, \dots\}$, the set of states, and $M = \{0, 1, 2, \dots\}$, the set of observations (For the definition of a stochastic kernel, see, for example, Bertsekas and Shreve [1]). We assume that we are given an initial distribution $p \in P(S)$ for the initial state of the system, where $P(S)$ is the set of all probability measures on S . Moreover, since it is known whether there is no customer or there is at least one customer in the service facility at the decision epochs, we assume that g satisfies the following condition: $g(m|i) = 1$, $i = 0$, $m = \mu_0$ and $g(m|i) = 0$, $i \neq 0$, $m = \mu_0$, where μ_0 means that there is no customer in the service facility. If there is no customer in the service facility at a decision epoch, then there is no sense in choosing control parameters. Nevertheless, we shall choose a control parameter, but decision yields nothing. Then, the arrival time of a new customer is the next decision epoch, and the time until then is regarded as the service time for convenience. An observation is generated also at that time, and then, a control parameter is chosen. But we can then assume that there is only one customer in the service facility regardless of the observation. Denote by i_n , m_n , a_n , and t_n the state of the system at the n -th decision epoch, the n -th observation, the n -th control parameter, and the n -th service time, respectively. Let $T_0 = 0$ and $T_n = t_1 + \dots + t_n$. Then, T_n denotes the n -th decision epoch. Two kinds of costs are incurred: a service cost at rate c_a , $a \in A$ and a holding cost at rate h ($h > 0$) per person. We assume that c_a is bounded and Borel measurable on A (We do not assume here the nonnegativity of c_a).

Now, suppose that the system is in state i ($i \geq 1$) at a decision epoch and that control parameter $a \in A$ is chosen. If the current service is accomplished after t unit time, and until then, $(j - i + 1)$ new customers arrive at the service facility, then the expected discounted cost incurred is given by

$$\int_0^t c_a e^{-\alpha s} ds + i \int_0^t h e^{-\alpha s} ds + \sum_{k=1}^{j-i+1} \int_{\tau_k}^t h e^{-\alpha s} ds, \quad i \geq 1, \quad (1.1)$$

where α is a positive number, the discount factor, and τ_k is the k -th arrival time during t unit time. Conditional on the event that $(j - i + 1)$ customers arrive at the service facility until t unit time, each arrival time has the uniform distribution independent of

the others (cf. Cinlar [2, p. 78]). Then, given that $(j - i + 1)$ new customers arrive at the service facility during t unit time, the conditional expected discounted cost is as follows:

$$\ell(i, a, t, j) = \int_0^t c_a e^{-\alpha s} ds + i \int_0^t h e^{-\alpha s} ds + (j - i + 1) \int_0^t \int_\tau^t h e^{-\alpha s} ds dG(\tau), \quad i \geq 1, \tag{1.2}$$

where G is the uniform distribution over the interval $[0, t]$. If $i = 0$, the cost incurred is given by $\ell(i, a, j, t) = 0$. Then, our problem is to find a rule to choose control parameters which minimize the expected total discounted cost:

$$E \left[\sum_{n=0}^{\infty} e^{-\alpha T_n} \ell(i_n, a_n, t_{n+1}, i_{n+1}) \right]. \tag{1.3}$$

2. Formulation as an imperfect state information semi-Markov decision process

In this section we show that the control problem introduced in Section 1 can be formulated as an imperfect state information semi-Markov decision process (ISISMDP) with the countable state space and unbounded costs.

From the definition of the control problem, it can be formulated as ISISMDP($S, M, A, q, g, \ell, \alpha$), where S is the state space; M is the observation space; A is the action space; q is the transition law (which is defined afterwards); g is the observation system; ℓ is the cost function; α is the discount factor. An initial distribution $p \in P(S)$ is given at the initial time. Then, an observation is generated by g , and an action is chosen. If the process is in state $i \in S$, and action $a \in A$ is chosen, then:

- (i) the current state changes to state $j \in S$ until t unit time according to a stochastic kernel q on $S \times R_+$ given $S \times A$, provided that q has a density function k such that $k(j, t|i, a) = f_a(t)e^{-\lambda t}(\lambda t)^{j-i+1}/(j - i + 1)!, i \geq 1, j \geq i - 1$, and $k(j, t|i, a) = 0, i \geq 1, j < i - 1; k(j, t|0, a) = \lambda e^{-\lambda t}, j = 1$, and $k(j, t|0, a) = 0, j > 1$;
- (ii) the cost given by $\ell(i, a, t, j)$ is incurred;
- (iii) the state $j \in S$ can be only imperfectly observed through the observation $m \in M$ generated by a stochastic kernel g on M given S , provided that g satisfies the condition that $g(m|i) = 1, i = 0, m = \mu_0$ and $g(m|i) = 0, i \neq 0, m = \mu_0$, where μ_0 means that no customer is waiting.

A policy ω is a rule to choose control parameters and is a sequence $\{\omega_0, \omega_1, \dots\}$, where ω_n is a Borel measurable stochastic kernel on A given $H_n = P(S) \times M \times (A \times R_+ \times M)^n$. H_n means the set of all observable histories when the n -th action must be chosen. For a policy ω , the expected total discounted cost J_ω is defined as:

$$J_\omega(p) = E_{\omega,p} \left[\sum_{n=0}^{\infty} e^{-\alpha T_n} \ell(i_n, a_n, t_{n+1}, i_{n+1}) \right], \quad p \in P(S), \tag{2.1}$$

where $E_{\omega,p}$ denotes the expectation by the unique probability measure induced by policy ω (cf. Bertsekas and Shreve [1, Proposition 7.28]). A policy ω^* is called the

optimal policy if $J_{\omega^*}(p) \leq J_{\omega}(p)$ for all policies ω and all initial distributions p .

The following lemma is straightforward from the definition of ℓ and k and the boundedness of c_a .

LEMMA 2.1. *There exist positive numbers β and γ such that*

$$\sum_{j \in S} \int_0^{\infty} |\ell(i, a, t, j)| k(j, t | i, a) dt \leq \beta i + \gamma, \quad i \in S, a \in A \quad (2.2)$$

The following assumptions are imposed throughout the paper.

Assumption (I). There exist $\varepsilon > 0$ and $\delta > 0$ such that

$$F_a(\delta) < 1 - \varepsilon, \quad a \in A. \quad (2.3)$$

Assumption (II). Any initial distribution has a finite expectation.

LEMMA 2.2. *There exists a number ρ ($0 < \rho < 1$) such that*

$$\sum_{j \in S} \int_0^{\infty} e^{-\alpha t} w(j) k(j, t | i, a) dt \leq \rho w(i), \quad i \in S, a \in A, \quad (2.4)$$

where

$$w(i) = \beta i + \gamma, \quad i \in S. \quad (2.5)$$

PROOF. From Assumption (I), we have

$$\int_0^{\infty} e^{-\alpha t} f_a(t) dt \leq \mu < 1, \quad a \in A, \quad (2.6)$$

where $\mu = 1 - \varepsilon + \varepsilon e^{-\alpha \delta}$ (cf. Ross [5, Theorem 7.3]). Then, from the definition of k , we have

$$\begin{aligned} & \sum_{j \in S} \int_0^{\infty} e^{-\alpha t} w(j) k(j, t | i, a) dt \\ &= \int_0^{\infty} e^{-\alpha t} f_a(t) \left[\sum_{j \geq i-1} (\beta i + \gamma) e^{-\lambda t} (\lambda)^{j-i+1} / (j - i + 1)! \right] dt \\ &= \int_0^{\infty} e^{-\alpha t} f_a(t) [\beta \lambda t + \beta i - \beta + \gamma] dt \\ &\leq \mu \beta i + \beta \lambda / (\alpha \varepsilon) + \mu (\gamma - \beta), \quad i \geq 1 \end{aligned}$$

(Note that $t e^{-\alpha t} \leq 1/(\alpha \varepsilon)$, $t \geq 0$). For $i = 0$, we have

$$\begin{aligned} & \sum_{j \in S} \int_0^{\infty} e^{-\alpha t} w(j) k(j, t | 0, a) dt \\ &= \int_0^{\infty} e^{-\alpha t} w(1) k(1, t | 0, a) dt \\ &= [\lambda / (\alpha + \lambda)] (\beta + \gamma). \end{aligned}$$

If we choose an appropriately large γ in Lemma 2.1, then

$$\mu\beta + \beta\lambda/(\alpha e) + \mu(\gamma - \beta) < \beta + \gamma$$

and

$$[\lambda/(\alpha + \lambda)] (\beta + \lambda) < \gamma,$$

which imply that (2.4) holds for $i = 0$ and 1 . Since $0 < \mu < 1$, there exists a number ρ ($0 < \rho < 1$) such that

$$\mu\beta i + \beta\lambda/(\alpha e) + \mu(\gamma - \beta) \leq \rho(\beta i + \gamma), i \geq 1$$

and

$$[\lambda/(\alpha + \lambda)] (\beta + \gamma) \leq \rho\gamma$$

from which the result follows.

REMARK 2.1. The condition (2.4) is equivalent to the condition given by van Nunen and Wessels [3] except that the weighting function w is linear in this paper.

Let

$$P = \left\{ p \in P(S) : \sum_{i \in S} ip(i) < \infty \right\} \tag{2.7}$$

and

$$\bar{w}(p) = \sum_{i \in S} w(i)p(i), p \in P. \tag{2.8}$$

Then, for any $p \in P$, $0 < \gamma \leq \bar{w}(p) < \infty$. For any Borel measurable function v on P , set

$$\|v\|_w = \sup_{p \in P} |v(p)|\bar{w}(p)^{-1} \tag{2.9}$$

and define B^w as the Banach space of all such v for which $\|v\|_w < \infty$. Then, we have the following proposition.

PROPOSITION 2.1. For any policy ω

$$\|J_\omega\|_w \leq 1/(1 - \rho). \tag{2.10}$$

PROOF. By virtue of Lemma 2.2, we can easily show by induction that for any policy ω and any initial distribution p ,

$$E_{\omega,p} [e^{-\alpha T_n} w(s_n)] \leq \rho^n \bar{w}(p), n \geq 0, \tag{2.11}$$

Then, the result follows directly from Lemma 2.1 and the property of the conditional expectation.

REMARK 2.2. From (2.11), we have $E_{\omega,p} [\gamma e^{-\alpha T_n}] \leq \rho^n \bar{w}(p)$. Letting $n \rightarrow \infty$, we have $E_{\omega,p} [e^{-\alpha L}] = 0$, where $L = \sup_{n \geq 0} T_n$. Hence, L is infinite almost surely for any policy ω and any initial distribution p .

REMARK 2.3. By the property of the conditional expectation, the expected total discounted cost can be also written as:

$$J_\omega(p) = E_{\omega,p} \left[\sum_{n=0}^{\infty} e^{-\alpha T_n} c(i_n, a_n) \right], \quad (2.12)$$

where

$$c(i, a) = \sum_{j \in S} \int_0^{\infty} \ell(i, a, t, j) k(j, t | i, a) dt. \quad (2.13)$$

3. Transformation to a Perfect state Information Semi-Markov Decision Process and the Existence of an Optimal Stationary I-policy

In this section we construct a perfect state information semi-Markov decision process (SMDP), which is sufficient for ISISMDP defined in Section 2. Then, we show that there exists an optimal stationary I-policy when A is finite.

Let p_n denote the conditional distribution of i_n given $h_n = (p, m_0, a_0, t_1, m_1, \dots, a_{n-1}, t_n, m_n)$. Then, $p_n, n \geq 0$, can be obtained recursively from the Bayes' theorem (see Wakuta [7, p. 214]).

LEMMA 3.1. *Any conditional distribution $p_n, n \geq 0$, has a finite expectation, i.e., $p_n \in P, n \geq 0$.*

PROOF. The claim is established by induction. By Assumption (II), p_0 has a finite expectation. Assume as an induction hypothesis that p_n has a finite expectation. Then, we have

$$\begin{aligned} & \sum_{j \in S} j p_{n+1}(j) \\ &= \sum_{j \in S} j \left\{ \sum_{i \in S} g(m|j) k(j, t | i, a) p_n(i) \right\} / z \\ &= \sum_{i \in S} \sum_{j \in S} j g(m|j) k(j, t | i, a) p_n(i) / z \\ &\leq \sum_{i \in S} \sum_{j \in S} j k(j, t | i, a) p_n(i) / z \\ &= \left[\lambda e^{-\lambda t} p_n(0) + \sum_{i \geq 1} f_a(t) \{ \lambda t + (i - 1) \} p_n(i) \right] / z < \infty, \end{aligned}$$

where $z = \sum_{i \in S} \sum_{j \in S} g(m|j) k(j, t | i, a) p_n(i)$. Hence, the induction argument is completed.

We introduce the discrete topology on S . Then, $P(S)$ endowed with the weak topology is a complete separable metric space, and $p(i), i \in S$, is a continuous function of p (cf. Parthasarathy [4, Theorem 6.1, p. 40]). Then, P defined by (2.7) is a Borel subset of $P(S)$.

Now, we construct a SMDP $(P, A, \bar{q}, \bar{c}, \alpha)$, where P is the state space; A is the action space; \bar{q} is the law of motion of the system; \bar{c} is the cost function; α is the

discount factor (see Wakuta [7, p. 215] for the definition of \bar{q} and \bar{c}). A policy for the SMDP is a sequence $\{\pi_0, \pi_1, \dots\}$, where each π_n is a Borel measurable stochastic kernel on A given $I_n = P \times (A \times R_+ \times P)^n$, provided that $I_0 = P$. Such a policy is called the information policy (I-policy) according to Sawaragi and Yoshikawa [6]. Then, the expected total discounted cost for the SMDP is defined as:

$$I_\pi(p_0) = E_\pi \left[\sum_{n=0}^{\infty} e^{-\alpha T_n} \bar{c}(p_n, a_n) | p_0 \right], p_0 \in P, \tag{3.1}$$

where $E_\pi[\cdot | p_0]$ denotes the expectation by the unique probability measure induced by I-policy π .

LEMMA 3.2. For any $p \in P$ and any $a \in A$,

$$(i) \quad |\bar{c}(p, a)| \leq \bar{w}(p); \tag{3.2}$$

$$(ii) \quad \iint_{P \times R_+} e^{-\alpha t} \bar{w}(p') d\bar{q}((p', t) | p, a) \leq \rho \bar{w}(p). \tag{3.3}$$

PROOF. (i) The proof is straightforward from Lemma 2.1 and (2.13). (ii) By the definition of \bar{q} , we have

$$\begin{aligned} & \iint_{P \times R_+} e^{-\alpha t} \bar{w}(p') d\bar{q}((p', t) | p, a) \\ &= \sum_{m \in M} \int_0^\infty e^{-\alpha t} \bar{w}(u(p, a, t, m)) z(p, a, t, m) dt \\ &= \sum_{m \in M} \int_0^\infty e^{-\alpha t} \left[\sum_{j \in S} w(j) u(p, a, t, m)(j) \right] z(p, a, t, m) dt \\ &= \sum_{m \in M} \int_0^\infty e^{-\alpha t} \left[\sum_{j \in S} w(j) \left\{ \sum_{i \in S} g(m|j) k(j, t|i, a) p(i) \right\} \right] dt \\ &= \sum_{i \in S} \sum_{j \in S} \left[\int_0^\infty e^{-\alpha t} w(j) k(j, t|i, a) dt \right] p(i) \\ &\leq \sum_{i \in S} \rho w(i) p(i) = \rho \bar{w}(p) \end{aligned}$$

(see Wakuta [7, p. 214] for the definition of u). Thus, the proof is completed.

By this lemma, the following proposition can be proved in a similar way to the proof of Proposition 2.1.

PROPOSITION 3.1. For any I-policy π ,

$$\|I_\pi\|_w \leq 1/(1 - \rho). \tag{3.4}$$

Using the same method as in Wakuta [7, Proposition 3.5]), we have the following fundamental result. Note that the set of all I-policies can be regarded as a subset of policies in the original problem.

PROPOSITION 3.2. *If I-policy π^* is optimal for SMDP($P, A, \bar{q}, \bar{c}, \alpha$), then it is also optimal for ISISMDP($S, M, A, q, g, \ell, \alpha$).*

Now, we shall discuss the existence of an optimal stationary I-policy. The optimality equation for our control problem is as follows:

$$v(p) = \inf_{a \in A} \left\{ \bar{c}(p, a) + \iint_{P \times R_+} e^{-\alpha t} v(p') d\bar{q}((p', t)|p, a) \right\}, p \in P. \quad (3.5)$$

From the definition of \bar{q} , this can be written also as:

$$v(p) = \inf_{a \in A} \left\{ \bar{c}(p, a) + \sum_{m \in M} \int_0^\infty e^{-\alpha t} v(u(p, a, t, m)) z(p, a, t, m) dt \right\}, p \in P. \quad (3.6)$$

By Lemma 3.2, the operator T defined by the right-hand side of (3.5), or equivalently, (3.6) is a contraction mapping on B^w with contraction coefficient ρ and has a unique fixed point v^* . Then, from Proposition 3.2, we can prove the following theorem by the standard method (cf. Wakuta [7, Theorem 4.2]).

THEOREM 3.1. *Suppose that A is finite. Then, there exists an optimal stationary I-policy $(f^*)^\infty$, where f^* selects the action minimizing the right-hand side of (3.5), or equivalently, (3.6) with $v = v^*$.*

References

- [1] BERTSEKAS, D. P. and SHREVE, S. E.: *Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.
- [2] CINLAR, E.: *Introduction to Stochastic Processes*, Prentice-Hall, New Jersey, 1975.
- [3] VAN NUNEN, J. and WESSELS, J.: *A note on dynamic programming with unbounded rewards*. Management Sci. **24** (1978), 576–580.
- [4] PARTHASARATHY, K. R.: *Probability Measures on Metric Spaces*. Academic Press, New York, 1967.
- [5] ROSS, S. M.: *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1970.
- [6] SAWARAGI, Y. and YOSHIKAWA, T.: *Discrete-time Markovian decision process with incomplete state observation*. Ann. Math. Statist. **41** (1970), 78–86.
- [7] WAKUTA, K.: *Optimal control of an M/G/1 queue with imperfectly observed queue length when the input source is finite.*, J. Appl. Probab. **28** (1991), 210–220.

Received September 22, 1992

Communicated by N. Furukawa