Optimal Pricing and Admission Control in a Queueing System with Periodically Varying Parameters

Seunghwan Yoon and Mark E. Lewis[1]

Department of Industrial and Operations Engineering
University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117
syoon@engin.umich.edu
melewis@engin.umich.edu
(734) 763-0519 (Office)
(734) 764-3451 (Fax)

[1]corresponding author

Abstract

We consider congestion control in a non-stationary queueing system. Assuming that the arrival and service rates are bounded, periodic functions of time, a Markov decision process (MDP) formulation is developed. We show under the infinite horizon discounted and average reward optimality criteria, for each fixed time, optimal pricing and admission control strategies are non-decreasing in the number of customers in the system. This extends stationary results to the non-stationary setting. Despite this result, the problem still seems intractable. We propose an easily implementable *pointwise stationary approximation* (PSA) to approximate the optimal policies, suggest a heuristic to improve the implementation of the PSA and verify its usefulness via a numerical study.

1   Introduction

Advances in telecommunications technology have sparked recent interest in real-time revenue management for multi-class queueing systems. When the system capacity is finite, a decision-maker must dynamically allocate resources while maintaining revenue maximization considerations. By and large this is achieved by two congestion control mechanisms; pricing control and admission control. Pricing control is used when a customer's priority or class cannot be discerned upon arrival. As the system becomes more congested, the price for admittance can be raised so that the number of arriving customers is reduced. By contrast, when a customer's class is revealed upon arrival, the revenue maximization and congestion control dilemmas can be alleviated via admission control. As in the pricing problem, lower priority customers are more likely to be admitted when the system is not operating close to capacity. The difference in the two scenarios, while subtle, is distinct in that the amount of information available to the decision-maker and the reward received is quite different. The pricing controller sets a price without knowing the class of an arriving customer and may only receive that price while the admission controller receives the exact reward offered by admitted customers.

The vast majority of the literature on congestion control assumes that the arrival and service processes are stationary; independent of the current time. Under this assumption, a decision-maker making pricing or admission control decisions needs only know the current number of customers in the system. With the added complication of non-stationary arrival and/or service processes, when a customer arrives, the decision-maker needs to know the current number of customers in the system *and* the current time. This makes the problem of finding the optimal decision intractable. The main theoretical contribution of this paper is that, under reasonable assumptions, many of the structural results of the stationary case can be extended to the non-stationary problem. In particular, when the arrival and service rates vary in a known periodic fashion, an intuitive result continues to hold; as the congestion increases in the pricing problem the optimal price increases and in the admission control problem, the admittance requirements become more strict.

This (unfortunately) does not complete the picture. In contrast to the infinite horizon stationary problem, the optimal decisions in either scenario are not static for each congestion level. Thus, while the structural

results suggest for each fixed time, we can restrict attention to monotone policies, this restriction does not make the search for optimal policies simple. To exploit the connection between the stationary and non-stationary cases, we suggest a *pointwise stationary approximation* (PSA) scheme that allows for the decisions to be monotone in the congestion level, but dynamic in time. Since we use results from the stationary problem to implement the approximation, the procedure is intuitive and easily implementable. We then compare several refinements of the PSA to the average reward stationary policies that are used in practice.

With the growth of Internet service providers and mobile network service providers, the study of congestion control in queueing systems with non-stationary arrivals have centered around the telecommunication industries. The focus of most of the work on non-stationary queueing systems has been on obtaining insights for fixed polices. For example, Massey and Whitt [13] obtain the mean number of busy servers and the time lag between the maximum arrival time and the actual peak congestion time for a system without control. The same authors discuss stationary approximations to non-stationary problems in [12]. Green and Kolesar [3, 4] use a pointwise stationary approximation for the congestion in a non-stationary system. Although, in the present study we use the same term, "PSA", to relate to a dynamic control scheme rather than the estimation of congestion levels, the concept is quite similar.

Monotone policies in admission control were originally discussed by Yadin and Naor [18], Heyman [6], Lippman and Ross [9] and Miller [15]. Monotone policies in pricing control was considered by Low [11] and more recently by Paschalidis and Tsitsiklis [16], Feng and Xiao [1] and Fu, Markus, and Wang [2]. None of the models deal with the periodic non-stationary extension presently considered.

The remainder of the paper is organized as follows. In Section 2 we formulate the problems as Markov decision processes (MDPs) and introduce the optimality criteria of interest (discounted and average reward). We prove the existence of monotone optimal policies under each optimality criterion in Section 3. In Section 4 we compare the average reward pricing and admission control models numerically and introduce the pointwise stationary approximation heuristic. The concluding remarks follow in Section 5.

2   Model Formulation

Consider a single commodity queueing system with finite capacity m over an infinite planning horizon. Customers arrive according to a non-homogeneous Poisson process with rate $\Lambda(t) > 0$ for all t. Each customer belongs to one of several classes $j \in \{1, \ldots, \ell\}$. Class j customers have a reservation price $p_j < \infty$, and the probability that an arrival at time t is a class j customer is $q_j(t) \in (0, 1)$, such that $\sum_{j=1}^{\ell} q_j(t) = 1$ for each t. In the pricing scenario (from now on called the *pricing control problem*), if an arriving customer sees a price lower than the reservation price, the price is paid and the customer enters the system. Otherwise the customer is lost forever. In the *admission control problem*, upon arrival the customer reveals the reservation price and is either accepted or rejected; rejected customers are lost forever. Let the set $\mathbb{A} \equiv \{p_0, p_1, \ldots, p_\ell\}$ be the set of possible reservation prices (p_0 will be used to model a full system), and without loss of generality, assume that $p_0 > p_1 \cdots > p_\ell > 0$. It should be clear that the decision-maker in the pricing control problem would never set a price outside of the set $\mathbb{A}$ since it would only result in lost revenue.

Let the arrival rate of class j customers be denoted $\beta_j(t) \equiv q_j(t)\Lambda(t)$, $j \in \{1, \ldots, \ell\}$. When price a is set, the customers that actually enter the buffer form a Poisson process with rate $\lambda_a(t) \equiv \sum_{j=1}^{w(a)} \beta_j(t)$, where $w(p_j) \equiv j$. The service process is assumed to be exponential with rate $\mu_i(t)$ when there are i customers in the system at time t. We assume that $\mu_0(t) = 0$, $\mu_1(t) > 0$ and $\mu_i(t)$ is non-decreasing and concave in i. Since $\mu_i(t)$ is not assumed to be strictly non-decreasing in i, buffer space can be modelled by setting $\mu_{i+1}(t) = \mu_i(t)$ for all $i \geq i'$ for some i'. Assume that the functions $\Lambda(t)$, $q_j(t)$ and $\mu_i(t)$ are measurable, bounded, periodic functions with period T for all $j \in \{1, 2, \ldots, \ell\}$ and $i \in \{0, 1, \ldots, m\}$. Define the maximal transition rate $\Psi \equiv \sup_{t \in [0, T)}\{\Lambda(t) + \mu_m(t)\}$. We apply *uniformization* (see Lippman [8]) with uniformization constant Ψ, so that we may consider discrete-time problems instead of the continuous-time problems described in Section 1. This has the effect of scaling the actual rewards by a constant, but has no effect on the optimal policies (in either case). In addition to arrival and departure times, we embed the times corresponding to the beginning of a new period; $0, T, 2T$, etc. Thus, if $\{\sigma_t, t = 0, 1, \ldots\}$ represents

the sequence of decision epochs, $\sigma_{t+1} - \sigma_t$ is the minimum of an exponential with rate Ψ and the residual time left in the current period.

In each problem, when the system is in a state x the decision-maker chooses from a set of available actions; a set of admission prices in the pricing control problem and either accept or reject in the admission control problem. A decision rule is a function, d, from the state space to the action space that describes what action will be taken for each state of the system at a particular time. A policy π is a sequence of decision rules, i.e., $\pi = \{d_0, d_1, \ldots\} \in \Pi$ where Π is the set of all non-anticipating policies. Each policy generates a sequence of random variables $\{(X_t,\ d_t(X_t)) : t = 0, 1, \ldots\}$ where X_t denotes the state of the system and $d_t(X_t)$ denotes the action chosen by decision rule d_t in state X_t at decision epoch t. Let $r(x, a)$ be the expected reward when the system is in state x and action a is chosen. The $n-$stage, $\alpha-$discounted expected reward of the policy π given that the initial state is x is given by

$$V_{n,\alpha}^{\pi}(x) \equiv \mathbb{E}_x^{\pi}\left[\sum_{t=0}^{n-1} e^{-\alpha\sigma_t} r(X_t, d_t(X_t))\right],$$

where $\mathbb{E}_x^{\pi}$ denotes expectation with respect to the probability measure determined by the initial state x and the policy π, $\{\sigma_t, t = 0, 1, \ldots\}$ represents the sequence of event times and $\alpha \in \mathbb{R}^+$. Without loss of generality, assume that $\alpha + \Psi = 1$. Note that although policies are defined for the infinite horizon, $V_{n,\alpha}^{\pi}(x)$ is only defined for a finite number of steps. In this case, only that portion of the policy that corresponds to the appropriate time horizon is used. Since the rewards we consider are bounded and non-negative the above expectation is guaranteed to exist. Define the $\alpha-$discounted expected reward to be $V_{\alpha}^{\pi}(x) \equiv \lim_{n\to\infty} V_{n,\alpha}^{\pi}(x)$. Similarly, when $\alpha = 0$ define the long-run average expected reward (or *gain*) to be $G^{\pi}(x) \equiv \liminf_{n\to\infty} V_{n,0}^{\pi}(x)/n$. Consider the value functions and criterion given in Table 1.

A policy whose expected reward achieves the supremum in one of these criteria (for all states) is called *optimal* under that criterion. Moreover, although we view the finite horizon case as interesting, we are actually most interested in the infinite horizon discounted and average reward cases. For the remainder of the paper we differentiate quantities identified with the pricing control problem from those with the admission control problem by using superscripts P and C, respectively.

Value Function	Criterion
$V_{n,\alpha}(x) \equiv \sup_{\pi \in \Pi} V_{n,\alpha}^{\pi}(x)$	[Finite Horizon]
$V_{\alpha}(x) \equiv \sup_{\pi \in \Pi} V_{\alpha}^{\pi}(x)$	[Infinite Horizon Discounted]
$G(x) \equiv \sup_{\pi \in \Pi} G^{\pi}(x)$	[Long-run Average]

Table 1: OPTIMALITY CRITERION CONSIDERED.

2.1 The Pricing Control Problem

Since the arrival and service rates are periodic functions of time, from the decision-maker's perspective under the infinite horizon discounted and average reward criteria the actual period within which events occur is not of importance. This fact is easily seen from the *principle of optimality* since a decision-maker that sees i customers s time units into the current period is faced with precisely the same dilemma if there are i customers in the system s time units into the next period. In the pricing control problem this means that only the current time since the last period, the number of customers currently in the system, and the type of event that has just occurred (departure, arrival, or dummy transition due to uniformization) need to be included in the state space. Let $\mathbb{X}^P \equiv \{(i, j, z) | i \in \{0, 1, \ldots, m\}, j \in \{-1, 0, 1\}, \text{ and } z \in [0, T)\}$ be the state space, where at relative time z (the current time modulo T), i represents the number of customers in the system, $j = 1$ denotes a customer's arrival, $j = -1$ a departure, and $j = 0$ a "dummy" transition due to uniformization. Note that one may conjecture that the second element of the state space is superfluous. However, recall that the model we propose only requires pricing decisions be made at arrival times. This is facilitated by inclusion of the "transaction type" j. The set of available actions in state $(i, 1, z) \in \mathbb{X}^P$, denoted A_i^P, is

$$A_i^P = \begin{cases} \{p_1, \ldots, p_\ell\} & \text{if } i < m; \\ \{p_0\} & \text{if } i = m. \end{cases}$$

$Q_a(t) \equiv \sum_{j=1}^{w(a)} q_j(t) = \lambda_a(t)/\Lambda(t)$ is the probability that given an arrival has just occurred at time t and price a is set, the arrival has a class higher than a (recall class 1 is the highest class). For a function f

on the state space, define the mapping U_α^P

$$(U_\alpha^P f)(i, z) \equiv \int_z^T [\Lambda(t)f(i, 1, t) + \mu_i(t)f(i, -1, t)$$

$$+ (\Psi - \Lambda(t) - \mu_i(t)f(i, 0, t)]e^{-(t-z)}dt + e^{-(T-z)}f(i, 0, 0).$$

Intuitively, $(U_\alpha^P f)(i, z)$ represents the total expected reward as measured by f starting from the next decision epoch, given that there are currently i customers in the system at z time units into the period. Let $v_{0,\alpha}^P \equiv 0$ and consider the following systems of equations for $(i, j, z) \in \mathbb{X}^P$,

$$v_{n,\alpha}^P(i, j, z) = \begin{cases} \max_{a \in A_i^P}\{Q_a(z)(a + (U_\alpha^P v_{n-1,\alpha}^P)(i+1, z)) \\ \qquad\qquad + (1 - Q_a(z))(U_\alpha^P v_{n-1,\alpha}^P)(i, z)\} & \text{if } 0 \le i < m,\ j = 1; \\ (U_\alpha^P v_{n-1,\alpha}^P)(i-1, z) & \text{if } 0 < i \le m,\ j = -1; \\ (U_\alpha^P v_{n-1,\alpha}^P)(i, z) & \text{otherwise,} \end{cases} \tag{2.1}$$

$$g^P + h^P(i, j, z) = \begin{cases} \max_{a \in A_i^P}\{Q_a(z)(a + (U_0^P h^P)(i+1, z)) \\ \qquad\qquad + (1 - Q_a(z))(U_0^P h^P)(i, z)\} & \text{if } 0 \le i < m,\ j = 1; \\ (U_0^P h^P)(i-1, z) & \text{if } 0 < i \le m,\ j = -1; \\ (U_0^P h^P)(i, z) & \text{otherwise,} \end{cases} \tag{2.2}$$

The first system, (2.1), are the *finite horizon optimality equations* (FHOE) and (2.2) the *average optimality equations* (AOE). If we replace $v_{n,\alpha}^P$ and $v_{n-1,\alpha}^P$ with v_α^P in (2.1) we have the *discount optimality equations* (DOE). If there exists a solution to the FHOE then $v_{n,\alpha}^P(x) = V_{n,\alpha}^P(x)$. Similarly, for the DOE (when $\alpha > 0$) $v_\alpha^P(x) = V_\alpha^P(x)$. The existence in either case is guaranteed since the rewards are bounded and the action space is finite (cf. Theorem 11.3.2 of [17]). A solution to the AOE is a little more subtle since the recurrent class structure must be considered. However, for the problem we consider a solution (g^P, h^P) implies $g^P = G^P(x)$ (for all x) while h^P is unique up to an additive constant and is called a *relative value function*.

2.2 The Admission Control Problem

In an analogous manner to the previous section, since we are interested in the infinite horizon problems, the decision-maker is indifferent to which period decisions are being made. The state and action spaces for the admission control problem are

$$\mathbb{X}^C \equiv \{(i,j,z)\mid i \in \{0,1,\ldots,m\}, j \in \{-1,0,1,\ldots\ell\}, \text{ and } z \in [0,T)\}$$

$$A_i^C = \begin{cases} \{\text{accept, reject}\} & \text{if } i < m \text{ and } j > 0; \\ \{\text{reject}\} & \text{if } i = m \text{ and } j > 0. \end{cases}$$

where $i \leq m$, $j \leq 0$ and $z \in [0,T)$ are as in the pricing control model and $j > 0$ denotes an arrival of class j. Define the mapping U_α^C for a function, f, on the state space to be

$$(U_\alpha^C f)(i,z) \equiv \int_z^T [\sum_{j=1}^{\ell} \beta_j(t)f(i,j,t) + \mu_i(t)f(i,-1,t) + (\Psi - \Lambda(t) - \mu_i(t)f(i,0,t)]e^{-(t-z)}dt$$

$$+ e^{-(T-z)}f(i,0,0).$$

The analogous systems to (2.1) and (2.2) follow ($v_{0,\alpha}^C = 0$)

$$v_{n,\alpha}^C(i,j,z) = \begin{cases} \max\{p_j + (U_\alpha^C v_{n-1,\alpha}^C)(i+1,z), \\ \quad (U_\alpha^C v_{n-1,\alpha}^C)(i,z)\} & \text{if } 0 \leq i < m, \ j \geq 1; \\ (U_\alpha^C v_{n-1,\alpha}^C)(i-1,z) & \text{if } 0 < i \leq m, \ j = -1; \\ (U_\alpha^C v_{n-1,\alpha}^C)(i,z) & \text{otherwise,} \end{cases} \tag{2.3}$$

$$g^C + h^C(i,j,z) = \begin{cases} \max\{p_j + (U_0^C h^C)(i+1,z), \\ \quad (U_0^C h^C)(i,z)\} & \text{if } 0 \leq i < m, \ j \geq 1; \\ (U_0^C h^C)(i-1,z) & \text{if } 0 < i \leq m, \ j = -1; \\ (U_0^C h^C)(i,z) & \text{otherwise.} \end{cases} \tag{2.4}$$

Again replacing $v_{n,\alpha}^C$ and $v_{n-1,\alpha}^C$ with v_α^C yields the DOE. The first element in each maximum refers to accepting the arriving customer and the second corresponds to rejection.

3   Monotone Optimal Policies

In this section we show the existence of non-decreasing optimal pricing strategies and optimal control levels. To this end, we show that the result holds in the finite horizon case and then extend to the infinite horizon models. To ease notation, let the set of non-increasing, concave functions (the domain will be clear from the context) be denoted by DC. We find the following results from Lippman and Stidham [10] useful.

Lemma 3.1 (Lippman and Stidham (1977)) *The following hold.*

1. *If $H(i) \equiv \mu_i f(i-1) + (\tilde{\mu} - \mu_i)f(i)$, where $f(i) \in DC$ and $\tilde{\mu} = \max_i\{\mu_i\}$, then $H(i) \in DC$.*

2. *Let $r \geq 0$ be fixed and let $g(i) \equiv \max\{r + f(i+1), f(i)\}$. Then g is concave if f is concave.*

We note that it is simple to show that $\tilde{\mu}$ can be replaced with any upper bound of $\tilde{\mu}$. For a function f define the first and second differences, $\Delta f(i) \equiv f(i+1) - f(i)$ and $\Delta^2 f(i) \equiv \Delta f(i+1) - \Delta f(i)$, respectively. Although these mappings will be applied to functions of more than one variable, they will always correspond to the differences of the first element.

Lemma 3.2 *In the discounted finite horizon problem, $V_{n,\alpha}^P(i,j,z), (U_\alpha^P v_{n,\alpha}^P)(i,z) \in DC$ as a function of i for all $n \geq 0$, for each fixed j, z and α.*

Proof.   Since $v_{n,\alpha}^P = V_{n,\alpha}^P$, it suffices to show the result holds for solutions to the FHOE. By induction. For $n = 0$ the results hold trivially. Suppose they hold for $n - 1$. We first show that $(U_\alpha^P v_{n-1,\alpha}^P)(i,z) \in DC$ implies $v_{n,\alpha}^P(i,j,z) \in DC$. It should be clear that this holds for $j = -1, 0$ since

$$v_{n,\alpha}^P(i,0,z) = (U_\alpha^P v_{n-1,\alpha}^P)(i,z) \qquad\qquad v_{n,\alpha}^P(i,-1,z) = (U_\alpha^P v_{n-1,\alpha}^P)(i-1,z).$$

Consider the case with $j = 1$, suppress j and z (since they are fixed) and recall $d_n^P(i)$ represents the optimal price when in state $(i,1,z)$. Suppose the decision-maker uses the potentially sub-optimal action $d_n^P(i+1)$ in state $(i,1,z)$. Then

$$v_{n,\alpha}^P(i,1,z) \geq Q_{d_n^P(i+1)}(U_\alpha^P v_{n-1,\alpha}^P)(i+1,z) + (1 - Q_{d_n^P(i+1)})(U_\alpha^P v_{n-1,\alpha}^P)(i,z), \qquad (3.1)$$

while

$$v_{n,\alpha}^P(i+1,1,z) = Q_{d_n^P(i+1)}(U_\alpha^P v_{n-1,\alpha}^P)(i+2,z) + (1 - Q_{d_n^P(i+1)})(U_\alpha^P v_{n-1,\alpha}^P)(i+1,z), \qquad (3.2)$$

Subtracting (3.1) from (3.2) yields

$$\Delta v_{n,\alpha}^P(i,1,z) \le Q_{d_n^P(i+1)}(\Delta U_\alpha^P v_{n-1,\alpha}^P)(i+1,z) + (1 - Q_{d_n^P(i+1)})\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i,z) \qquad (3.3)$$

Thus, since $\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i,z) \le 0$ for i and z arbitrary, $v_{n,\alpha}^P$ is non-increasing in i for all n as desired.

On the other hand, to get the concavity suppose $d_n^P(i)$ is employed in $(i+1,1,z)$. The FHOE yields

$$\Delta v_{n,\alpha}^P(i,1,z) \ge Q_{d_n^P(i)}\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i+1,z) + (1 - Q_{d_n^P(i)})\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i,z). \qquad (3.4)$$

Combining (3.3) (for $(i+1,1,z)$) and (3.4)

$$\Delta^2 v_{n,\alpha}^P(i,1,z) \le Q_{d_n^P(i+2)}\Delta^2(U_\alpha^P v_{n-1,\alpha}^P)(i+1,z) + (1 - Q_{d_n^P(i)})\Delta^2(U_\alpha^P v_{n-1,\alpha}^P)(i,z). \qquad (3.5)$$

Thus, $\Delta^2 v_{n,\alpha}^P \le 0$ and we have $v_{n,\alpha}^P \in DC$.

It remains to show that $v_{n,\alpha}^P \in DC$ implies $(U_\alpha^P v_{n,\alpha}^P)(i-1,z) \in DC$. Recall

$$(U_\alpha^P v_{n,\alpha}^P)(i,z) \equiv \int_z^T \left[\Lambda(t)v_{n,\alpha}^P(i,1,t) + \mu_i(t)v_{n,\alpha}^P(i,-1,t) \right.$$

$$\left. + (\Psi - \Lambda(t) - \mu_i(t))v_{n,\alpha}^P(i,0,t) \right] e^{-(t-z)}dt + e^{-(T-z)}v_{n,\alpha}^P(i,0,0).$$

The result now follows from the linearity of the difference operator and part 1 of Lemma 3.1. ∎

The next result asserts the existence of monotone optimal policies in the finite horizon case.

Theorem 3.3 *Under the finite horizon optimality criterion, for each fixed time $z \in [0,T)$, the optimal price at stage n when there are i customers in the system, denoted $d_n^P(i)$ (since j and z are fixed), is non-decreasing in i.*

Proof. Fix $z \in [0, T)$ and $\alpha \geq 0$. Clearly $d_n^P(m) > d_n^P(m-1)$ since $d_n^P(m) = p_0$. Thus, we must show $d_n^P(i+1) \geq d_n^P(i)$, $0 \leq i \leq m-2$. Let $d_n^P(i)$ be the optimal price when in state $(i, 1, z)$. Suppose the decision-maker uses the potentially sub-optimal action $d_n^P(i+1)$ in state $(i, 1, z)$. Then

$$v_{n,\alpha}^P(i, 1, z) \geq Q_{d_n^P(i+1)}(z)(d_n^P(i+1) + (U_\alpha^P v_{n-1,\alpha}^P)(i+1, z))$$

$$+ (1 - Q_{d_n^P(i+1)}(z))(U_\alpha^P v_{n-1,\alpha}^P)(i, z),$$

(3.6)

while

$$v_{n,\alpha}^P(i+1, 1, z) = Q_{d_n^P(i+1)}(z)(d_n^P(i+1) + (U_\alpha^P v_{n-1,\alpha}^P)(i+2, z))$$

$$+ (1 - Q_{d_n^P(i+1)}(z))(U_\alpha^P v_{n-1,\alpha}^P)(i+1, z),$$

(3.7)

Subtracting (3.6) from (3.7) yields

$$\Delta v_{n,\alpha}^P(i, 1, z) \leq Q_{d_n^P(i+1)}(z)\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i+1, z)$$

$$+ (1 - Q_{d_n^P(i+1)}(z))\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i, z)$$

(3.8)

Similarly, if we use action $d_n^P(i)$ in state $(i+1, 1, z)$. The FHOE imply

$$\Delta v_{n,\alpha}^P(i, 1, z) \geq Q_{d_n^P(i)}(z)\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i+1, z)$$

$$+ (1 - Q_{d_n^P(i)}(z))\Delta(U_\alpha^P v_{n-1,\alpha}^P)(i, z)$$

(3.9)

$(i, 1, z)$. Combining (3.8) and (3.9) we obtain

$$(Q_{d_n^P(i+1)}(z) - Q_{d_n^P(i)}(z))\Delta^2(U_\alpha^P v_{n-1,\alpha}^P)(i, z) \geq 0.$$

Lemma 3.2 implies $\Delta^2(U_\alpha^P v_{n-1,\alpha}^P)(i, z) \leq 0$. There are two possibilities to consider. First, suppose that $\Delta^2(U_\alpha^P v_{n-1,\alpha}^P)(i, z) = 0$. Note that the FHOE for $j = 1$ may be rewritten,

$$v_{n,\alpha}^P(i, 1, z) = (U_\alpha^P v_{n-1,\alpha}^P)(i, z) + \max_{a \in A(x)}\{Q_a(z)(a + \Delta(U_\alpha^P v_{n-1,\alpha}^P)(i, z))\}$$

Our assumption on $\Delta^2(U_\alpha^P v_{n-1,\alpha}^P)(i,z) = 0$ implies

$$\max_{a \in A(x)} \{Q_a(z)(a + \Delta(U_\alpha^P v_{n-1,\alpha}^P)(i,z))\} = \max_{a \in A(x)} \{Q_a(z)(a + \Delta(U_\alpha^P v_{n-1,\alpha}^P)(i+1,z))\}$$

and there exists an optimal policy such that $d_n^P(i) = d_n^P(i+1)$.

Suppose now that $\Delta^2(U_\alpha^P v_{n-1,\alpha}^P)(i,z) < 0$. Then, $Q_{d_n^P(i+1)}(z) - Q_{d_n^P(i)}(z) \leq 0$. Since $Q_a(z)$ is (strictly) decreasing in a, this implies that $d_n^P(i+1) \geq d_n^P(i)$ for each fixed time. ∎

Lemma 3.4 *In the discounted finite horizon admission control problem, $V_{n,\alpha}^C(i,j,z), (U_\alpha^C v_{n,\alpha}^C)(i,z) \in DC$ as a function of i for all n, for each fixed j, z and α.*

Proof. By induction. The case for $n = 0$ is trivial. Assume that the results hold for $n-1$. The fact that $(U_\alpha^C v_{n-1,\alpha}^C)(i,z) \in DC$ implies $v_{n,\alpha}^C(i,j,z) \in DC$ follows in the same manner as Lemma 3.2 for $j = -1, 0$. When $j > 0$ using the optimal admission control decision for state $(i+1,j,z)$ in state (i,j,z) implies

$$\Delta v_{n,\alpha}^C(i,j,z) \leq \begin{cases} \Delta(U_\alpha^C v_{n-1,\alpha}^C)(i+1,z) & \text{if } d_n^C(i+1,j,z) = \text{accept}, \\ \Delta(U_\alpha^C v_{n-1,\alpha}^C)(i,z) & \text{if } d_n^C(i+1,j,z) = \text{reject}. \end{cases} \tag{3.10}$$

Thus, $v_{n,\alpha}^C(i,j,z)$ is non-increasing. Part 2 of Lemma 3.1 implies the concavity. To show $v_{n,\alpha}^C(i,j,z) \in DC$ implies $(U_\alpha^C v_{n,\alpha}^C)(i,z) \in DC$, apply the same argument as Lemma 3.2. ∎

Theorem 3.5 *For each fixed time $z \in [0,T)$, the optimal admission control policy when there are i customers in the system at stage n, denoted $d_n^C(i,j,z)$, is non-decreasing in i under the finite horizon optimality criterion. That is, there exists control limits $\eta_n(j,z)$ such that*

$$d_n^C(i,j,z) = \begin{cases} \text{accept } (0) & \text{if } i \leq \eta_n(j,z); \\ \text{reject } (1) & \text{if } i > \eta_n(j,z). \end{cases} \tag{3.11}$$

Furthermore, $\eta_n(j,z)$ is non-increasing in j for each fixed time z.

Proof. Note that it is optimal to accept (reject) at stage n if $p_j + \Delta(U_\alpha^C v_{n-1,\alpha}^C)(i, z) \geq (<) \, 0$. Let i' be the

smallest i such that $p_j + \Delta(U_\alpha^C v_{n-1,\alpha}^C)(i, z) < 0$ (letting $i' = m$ if none exists). Since Lemma 3.4 implies

$\Delta(U_\alpha^C v_{n-1,\alpha}^C)(i', z) \geq \Delta(U_\alpha^C v_{n-1,\alpha}^C)(i' + 1, z)$ we have $p_j + \Delta(U_\alpha^C v_{n-1,\alpha}^C)(i, z) < 0$ for all $i \geq i'$ and

reject is optimal for each such i. Hence, if we set $\eta_n(i, z) = i'$ the proof is complete.

The last claim follows from the fact that the acceptance (rejection) criterion when in state (i, j, z) for

$j > 0$ is $p_j + \Delta(U_\alpha^C v_{n-1,\alpha}^C)(i, z) \geq (<) \, 0$ and the assumption that $p_1 > p_2 > \cdots > p_\ell$. ∎

3.1 Infinite Horizon Monotone Optimal Policies

The main theoretical result of the paper is stated in the following theorem.

Theorem 3.6 *For each fixed time z, under both the infinite horizon discounted and average reward optimality criteria the following hold*

1. *In the pricing control problem there exists an optimal policy such that the admission price, when there are i customers in the system, is non-decreasing in i.*

2. *In the admission control problem there exists an optimal policy such that for each class j there are control limits $\eta(j, z)$ such that the decision-maker accepts (rejects) class j customers if $i \leq (>) \, \eta(j, z)$. Furthermore, $\eta(j, z)$ is non-increasing in j.*

The results of the previous sections suggest that to prove Theorem 3.6 it suffices to show that the value

functions for each infinite horizon model are concave in i. To this end, we note that the concavity in

the infinite horizon discounted case is achieved virtually immediately by letting the horizon length go to

infinity in Lemmas 3.2 and 3.4. The concavity of the relative value functions in the average reward case

requires several technical assertions. We then let the horizon length go to infinity in the undiscounted (and

normalized) finite horizon problem to get the result. This is detailed in the next proposition.

Proposition 3.7 *For the pricing and admission control models, $g_n^k(x) \equiv V_{n,0}^k(x) - V_{n-1,0}^k(x) \to g^k$ and*

$h_n^k(x) \equiv V_{n,0}^k(x) - V_{n,0}^k(y) \to h^k(x)$, $\forall x \in \mathbb{X}^k$, *where y is an arbitrary (fixed) state in $\mathbb{X}^k$, (g^k, h^k)*

satisfies the AOE, and $k = P, C$ for the pricing and admission control models, respectively. Furthermore, this convergence is uniform on compact subsets of $\mathbb{X}^k$.

Proof. See Appendix ∎

The following lemma is an immediate consequence of Lemmas 3.2, 3.4, Proposition 3.7 and the observation that $\Delta V_{n,1}^P(i,z) = \Delta h_n^P(i,z)$. It completes the proof of Theorem 3.6.

Lemma 3.8 *In the pricing and admission control models, the value functions for the infinite horizon discounted reward problem and the relative value function for the average reward problem are non-increasing and concave in i. Similarly for $(U_\alpha^k v_\alpha^k)(i,z)$ and $(U_0^k h^k)(i,z)$, where $k = P, C$.*

4 Pointwise Stationary Approximations

The results of the previous section imply a close relationship between the non-stationary and stationary versions of congestion control problems. In fact, for each fixed time, we have shown that the intuition for the optimal policy is precisely the same; the more congested the network becomes, the more strict should be the admission structure. This intuition does not carry over in all non-stationary models as an example in Lewis, Ayhan, and Foley [7] exhibits. Furthermore, this observation does not stand to make the solution of the non-stationary problems simple since the problems remain uncountable state space MDPs. In fact, even approximating these problems with a discretization of the time horizon quickly becomes computationally intensive for problems of reasonable size.

With this in mind, we present a pointwise stationary approximation to the non-stationary problem and verify that it achieves an average reward that is quite close to optimal. Our numerical study shows that it can be a considerable improvement over the stationary policy that is often used in practice. Note that although in this section we only consider the average reward case, a similar analysis could be performed for the discounted case. A brief description of the PSA follows.

Choose a set of γ points, say $S \equiv \{\tau_1, \tau_2, \ldots, \tau_\gamma\}$, where $\tau_1 = 0$ and $\tau_\gamma = T$ in the period $[0, T]$ and consider a sequence of stationary problems. For the nth problem, $n \in \{1, 2, \ldots, \gamma\}$, $\lambda(t) = \lambda(\tau_n)$,

$\mu_i(t) = \mu_i(\tau_n)$ for $i = 0, 1, \ldots, m$, and $q_j(t) = q_j(\tau_n)$ for $j \in \{1, 2, \ldots, \ell\}$ for all t. That is to say, we solve the *stationary* pricing (or admission) problem with the parameters for each fixed time τ_n and obtain the optimal (stationary) policy denoted $\{f_{\tau_n}^k(i), i = 0, 1, \ldots, m\}$ for $k = P, C$.

Definition 4.1 *The pointwise stationary approximation for the set S is defined to be the non-stationary policy $d^P(i, t) = f_{\tau_n}^P(i)$ for pricing control and $d^C(i, j, t) = f_{\tau_n}^C(i, j)$ for admission control, for $t \in [\tau_n, \tau_{n+1})$ and $n = 1, 2, \ldots, \gamma - 1$.*

In essence the PSA is the piecewise linear policy that uses the optimal policy for the stationary problem at several points during the period in hopes that the optimal non-stationary policy can be closely approximated in this manner. This exploits the fact that the structure of the optimal policy is the same for the stationary and non-stationary problems and is a quite intuitive way of congestion control in practice. Equally important is the observation that the stationary problem may be solved via solution of the optimality equations from one of the well-known solution techniques for MDPs, for example policy iteration or value iteration.

One might notice that we have made no mention of how the set S is chosen. An obvious implementation would be to choose the number of points one would like to include in S and space them equally on the interval. This has the drawback that it ignores the observation that congestion (not time) is the driving force behind changes in the optimal policy. As an alternative, we propose a PSA that attempts to track times of peak congestion and adjust the policy accordingly. Define the *instantaneous congestion*

$$\rho(t) = \frac{\Lambda(t)}{\mu_m(t)}$$

for $t \in [0, T)$. For the remainder of the section, we assume that $\rho(t)$ is continuous and differentiable for all $t \in [0, T)$. We contend that this is not much of a restriction since we can replace the derivative condition contained in the third step of the algorithm below with a similar condition on left-hand derivatives, while discontinuous functions can be approximated by continuous functions. We next define the "distance" between two stationary policies at τ_{n_1} and τ_{n_2} for each problem considered. In the pricing control problem let

$$D^P(f_{\tau_{n_1}}^P, f_{\tau_{n_2}}^P) \equiv \max_{0 \leq i \leq m-1} |w(f_{\tau_{n_1}}^P(i)) - w(f_{\tau_{n_2}}^P(i))|.$$

This represents the largest change in the action chosen (in each state) for different policies in the stationary problem. Similarly, in the admission control problem, assume that $f^C_{\tau_{n_1}}$ and $f^C_{\tau_{n_2}}$ are stationary policies that are of the form of the optimal policy (i.e. have control limits) and let

$$D^C(f^C_{\tau_{n_1}}, f^C_{\tau_{n_2}}) \equiv \max_{1 \leq j \leq \ell} |\eta(j, \tau_{n_1})) - \eta(j, \tau_{n_2})|.$$

Note that this represents the maximum difference in the admission control limits for the stationary policies $f^C_{\tau_{n_1}}$ and $f^C_{\tau_{n_2}}$.

The *instantaneous congestion based PSA* (ICPSA) is described for the pricing control problem by the following algorithm (the admission control problem would be approximated analogously):

1. Set $n = 1$.

2. Compute $f^P_{\tau_n}$ from the stationary problem (recall $\tau_1 = 0$).

3. Let $\tau^* = \inf\{t > \tau_n | \rho'(t) = 0\}$, where the infimum of the empty set if taken to be T.

4. Compute $f^P_{\tau^*}$ from the stationary problem and set $n' = n + 1 + D^P(f^P_{\tau_n}, f^P_{\tau^*})$. If $\tau^* = T$ use the left hand limits of $\lambda_a(t)$ and $\mu_i(t)$ for each a and i as t approaches T.

5. Choose $\{\tau_{n+1}, \ldots, \tau_{n'-1}\}$ equally spaced between τ_n and τ^* and compute $f^P_{\tau_k}$ for $\tau_k \in \{\tau_{n+1}, \ldots, \tau_{n'-1}\}$.

6. Let $n = n'$. If $\tau_n = T$. Stop. Otherwise return to step 2.

Steps 3, 4, and 5 make up the crux of the ICPSA algorithm. The construction includes all times that the derivative of the instantaneous congestion is zero as well as the beginning and end of the period. Between these times, the number of points used is set equal to the largest change in the pricing action. This policy has the desirable property that more time points are placed between times when the optimal stationary approximation makes a significant change in the policy. When the algorithm is complete, the set $\{f^P_{\tau_n}\}$ is defined and thus, so is the corresponding ICPSA policy.

In order to compare the equally spaced PSA and the ICPSA to what is currently done in practice we define the nominal arrival and service rates

$$\bar{\beta}_j \equiv \frac{\int_0^T \beta_j(t)dt}{T} \quad \text{and} \quad \bar{\mu}_i \equiv \frac{\int_0^T \mu_i(t)dt}{T}$$

for $j = 1, 2, ...\ell$ and $i = 0, 1, ..., m$. Define the *average stationary approximation* (ASA) policy to be the policy that maximizes the reward of the stationary (pricing or admission control) problem when the vectors $\bar{\beta}$ and $\bar{\mu}$ are used as constant arrival and service rate vectors, respectively. Thus, when estimating the arrival and service rate vectors the decision-maker simply uses the average of each rate over the period.

We discretize the time horizon and approximate the gains of the optimal policy and the heuristics (equally spaced PSA, ICPSA, and ASA) that have been discussed. Suppose we divide the period into n equally spaced time segments. Denote the time elapsed between adjacent decision epochs by Δt so that $\Delta t = T/n$. At each decision epoch, the decision-maker chooses the "optimal" action. Since transitions follow a Poisson process with rate Ψ, an event occurs at every Δt with probability $1 - \Psi e^{-\Psi \Delta t}$ whereas nothing occurs (a dummy transition) during time period Δt with probability $\Psi e^{-\Psi \Delta t}$. Recall that the crucial quantity in defining the AOE, (2.2), was $(U_0^P f)(i, z)$ for a function f on the state space. Let

$$(\hat{U}_0^P f)(i, z) \equiv (1 - e^{-\Psi \Delta t}) \Big[\frac{\Lambda(z + \Delta t)}{\Psi} f(i, 1, z + \Delta t) + \frac{\mu_i(z + \Delta t)}{\Psi} f(i, -1, z + \Delta t)$$

$$+ (1 - \frac{\Lambda(z + \Delta t)}{\Psi} - \frac{\mu_i(z + \Delta t)}{\Psi}) f(i, 0, z + \Delta t) \Big] + e^{-\Psi \Delta t} f(i, 0, z + \Delta t), \tag{4.1}$$

for $z = 0, \Delta t, \ldots, T - 2\Delta t$. Similarly,

$$(\hat{U}_0^P f)(i, T - \Delta t) \equiv (1 - e^{-\Psi \Delta t}) \Big[\frac{\Lambda(T)}{\Psi} f(i, 1, 0) + \frac{\mu_i(T)}{\Psi} f(i, -1, 0)$$

$$+ (1 - \frac{\Lambda(T)}{\Psi} - \frac{\mu_i(T)}{\Psi}) f(i, 0, 0) \Big] + e^{-\Psi \Delta t} f(i, 0, 0), \tag{4.2}$$

where the rate functions are replaced with their left hand limits in the case of a discontinuity at time T. The "discretized" AOE is the analogue of (2.2) using $\hat{U}_0^P$ instead of U_0^P. The result produces an estimate of the
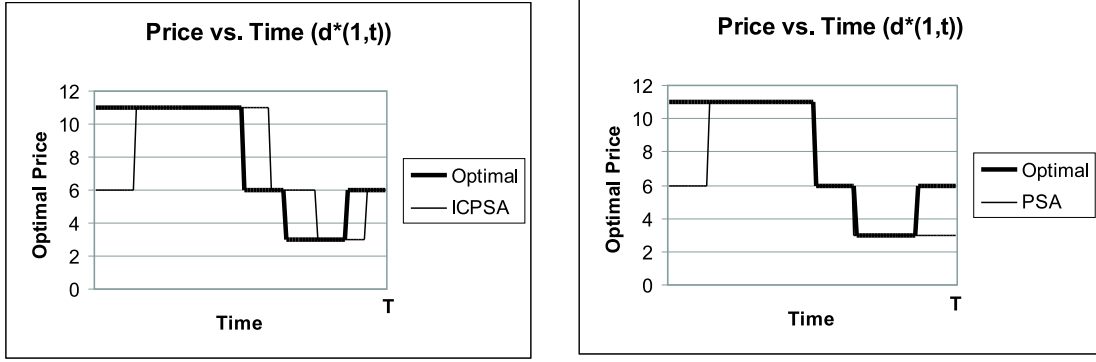
Figure 1: THE COMPARISON OF OPTIMAL PRICES FOR $T = \pi$.

relative value function and the optimal gain for the pricing control problem. An analogous system is used in the admission control problem.

We present an example that captures two important observations; the first is that the ASA can perform quite poorly while the second highlights the importance of knowing the reservation price to the decision-maker in the admission control model.

Example 4.2

Let $\ell = 3$, $m = 3$, and $\Delta t = \frac{T}{100}$, and the set of reservation prices $\{11, 6, 3\}$. We consider the cases in which $T = n\pi/4$, $n = 1, 2, 3$, and 4, so that the nonstationarity is captured by the whole sine wave (bimodal) in one case and various portions of the sine function in others. The arrival rates and service rates are

$$\lambda_{p_j}(t) = \begin{cases} \frac{1}{\Psi}(10\sin(2t) + 11(2^{j-1})) & \text{if } j = 1, 2, 3; \\ 0 & \text{otherwise}, \end{cases}$$

$$\mu_i(t) = \begin{cases} \frac{1}{\Psi}(30 + 10(i-1)) & \text{for } i = 1, 2, 3; \\ 0 & \text{otherwise}, \end{cases}$$

respectively, where $\Psi = 104$ and $\Lambda(t) = \lambda_{p_3}(t)$.

Figure 1 shows how the heuristic policies compare to the (discretized) optimal policy when 1 customer is in the system and $T = \pi$ (other congestion levels are similar). The PSA with equally spaced time points is
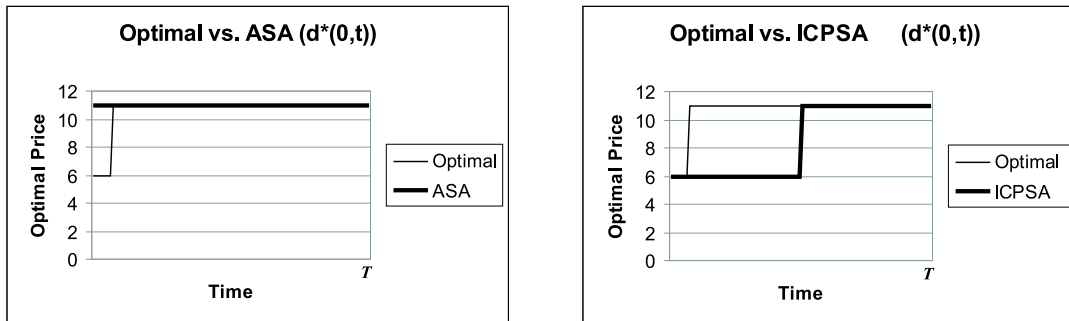
Figure 2: THE COMPARISON OF OPTIMAL PRICES FOR $T = \pi/4$.

denoted "PSA". The number of time points were set equal to the number of points used in ICPSA. The ASA policy for this example uses the static price 6 for all t when there is 1 customer in the system. One might note that both the ICPSA and the PSA closely track the optimal policy, but that the ICPSA overcompensates. That is to say that there is a shift to the left in the ICPSA away from the optimal policy. We believe this is due to the fact that the pointwise stationary approximations do not take into account the decision-maker's ability to anticipate an increase in congestion. The shift is particularly apparent in Figure 2 and somewhat justifies the performance of the ASA when $T = \pi/4, \pi/2$ (the compensation can have a negative effect).

In Table 2 for $T = \pi, 3\pi/4$ the equally spaced PSA and ICPSA outperform the ASA considerably (over 8% savings in one case). When the non-stationarity is more mild ($T = \pi/2, \pi/4$) the ASA dominates the PSA however not by a significant mount. Another notable phenomenon is that the percentage difference in the average reward is shown to vary widely for the ASA (over 10%) while the equally spaced PSA and the ICPSA are consistently within 4% of optimal.

One should also note that there is quite a large difference in the gain for the pricing and admission control models (over 28%). Since the arrival and service rates are the same for each model, this can be attributed to the fact that the decision-maker has complete knowledge of the class of a customer upon arrival.

5 Conclusions

We have introduced non-stationary congestion control problems with periodic rates for both the arrival and service processes. By embedding the current time (modulo the period) in the state space, we are able to

Cases			Opt. Gain	% Diff. from g^P
$T = \pi$	Admission Control	Opt. Admission (g^C)	1.75857	+36.06
	Pricing Control	Opt. Pricing (g^P)	1.29246	–
		ASA (g^P_{ASA})	1.15075	-10.96
		PSA (g^P_{PSA})	1.26951	-1.78
		ICPSA (g^P_{ICPSA})	1.25754	-2.70
$T = \frac{3\pi}{4}$	Admission Control	Opt. Admission (g^C)	1.79780	+33.14
	Pricing Control	Opt. Pricing (g^P)	1.35028	–
		ASA (g^P_{ASA})	1.21622	-9.92
		PSA (g^P_{PSA})	1.30139	-3.62
		ICPSA (g^P_{ICPSA})	1.30935	-3.03
$T = \frac{\pi}{2}$	Admission Control	Opt. Admission (g^C)	1.81467	+29.02
	Pricing Control	Opt. Pricing (g^P)	1.40647	–
		ASA (g^P_{ASA})	1.40475	-0.12
		PSA (g^P_{PSA})	1.38299	-1.67
		ICPSA (g^P_{ICPSA})	1.38299	-1.67
$T = \frac{\pi}{4}$	Admission Control	Opt. Admission (g^C)	1.26077	+29.23
	Pricing Control	Opt. Pricing (g^P)	0.97558	–
		ASA (g^P_{ASA})	0.97436	-0.13
		PSA (g^P_{PSA})	0.94429	-3.21
		ICPSA (g^P_{ICPSA})	0.94429	-3.21

Table 2: Comparison of the average optimal rewards for Example 4.2.

model each problem as a discrete-time stationary Markov decision process and compare the average rewards of pricing and admission control models. We showed that under the discounted and average reward optimality criteria, the intuition that the control structure should become more strict as the number of customers increases continues to hold under very light regularity conditions. Unfortunately, since the state space is uncountable, this does not stand to make the problems tractable. On the other hand, it does lead us to the conclusion that we should be able to adapt some of the results for the stationary problem, to the more general case. This link is exploited by the use of a the pointwise stationary approximation. The PSA uses the knowledge of the arrival and service rates, coupled with the results presented in Theorem 3.6, to get an easily implementable and intuitive approximation to the optimal policies.

There are several avenues for further study. Note that the equally spaced PSA and possibly non-uniformly spaced ICPSA both performed quite well in Example 4.2. This may be due to the fact that the arrival rates are assumed to be a sinusoid. We conjecture that when this does not hold, the equally spaced

PSA will not perform as well since as more elaborate congestion rates arise, the equally spaced PSA will not adapt.

Furthermore, we have presented only two possibilities for the PSAs. Intuition tells us that as we increase the number of time points for the PSA the accuracy of the approximation should increase. This was noticed in the numerical examples, but not pursued to its fullest extent. For example one might ask the question, "How many time points is enough?" Since each added time point adds another stationary MDP to be solved, continuing to add points ad infinitum is not desirable. Furthermore, since it is the congestion that drives the decision-maker's policy, placing more points on the time interval cannot alone guarantee improved performance. We intend to pursue the study of where these points should be optimally placed.

Another consideration is exhibited by Figure 1. As has previously been alluded to, both the equally spaced PSA and the ICPSA "overestimate" the optimal policy. That is, it appears that a better approximation could be achieved by shifting the points to the left. We conjecture that this is due to the fact that the instantaneous congestion does not "anticipate" future congestion while a decision-maker pricing optimally would certainly do so. The idea that a PSA has this shift in the optimal policy is consistent with the shift in the estimation of congestion levels discussed in [4]. We believe that each of these observations offer particularly interesting challenges and hope to explore them further in future research.

Acknowledgements

References

[1] Y. Feng and B. Xiao. Optimal policies of yield management with multiple predetermined prices. *Operations Research*, 48(2):332–343, 2000.

[2] M. C. Fu, S. I. Marcus, and I. Wang. Monotone optimal policies for a transient queueing staffing problem. *Operations Research*, 48(2):327–331, 2000.

[3] L. V. Green and P. J. Kolesar. On the accuracy of the simple peak hour approximation for Markovian queues. *Management Science*, 41(8):1353–1370, August 1995.

[4] L. V. Green and P. J. Kolesar. The lagged psa for estimating peak congestion in multiserver Markovian queues with periodic arrival rates. *Management Science*, 43(1):80–87, January 1997.

[5] O. Hernández-Lerma and J. B. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Applications of Mathematics: Stochastic Modelling and Applied Probability. Springer-Verlag Inc., New York, 1996.

[6] D. P. Heyman. Optimal operating policies for M/G/1 queueing systems. *Operations Research*, 16:362–382, 1968.

[7] M. E. Lewis, H. Ayhan, and R. D. Foley. Bias optimal admission control policies for a multi-class non-stationary queueing system. *Journal of Applied Probability*, 39(1):20–37, 2002.

[8] S. A. Lippman. Applying a new device in the optimization of exponential queueing systems. *Operations Research*, 23(4):687–712, 1975.

[9] S. A. Lippman and S. M. Ross. The streetwalker's dilemma: A job shop model. *SIAM Journal of Applied Mathematics*, 20(3):336–342, 1971.

[10] S. A. Lippman and J. Stidham, Shaler. Individual versus social optimization in exponential congestion systems. *Operations Research*, 25(2):233–247, Mar.-Apr. 1977.

[11] D. W. Low. Optimal dynamic pricing policies for an M/M/s queue. *Operations Research*, 22(3):545–561, 1974.

[12] W. A. Massey and W. Whitt. Stationary-process approximations for the nonstationary erlang loss model. *Operations Research*, 44(6):976–983, 1996.

[13] W. A. Massey and W. Whitt. Peak congestion in multi-server service systems with slowly varying arrival rates. *Operations Research*, 25:157–172, 1997.

[14] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag Inc., New York, 1993.

[15] B. Miller. A queueing reward system with several customer classes. *Management Science*, 16(3):234–245, 1969.

[16] I. C. Paschalidis and J. N. Tsitsiklis. Congestion-dependent pricing of on-line Internet services. In *Proceedings of the 38th Conferences on Decision and Control*, Phoenix, Arizona, December 1999.

[17] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., New York, 1994.

[18] M. Yadin and P. Naor. Queueing systems with a removable service station. *Operations Research Quarterly*, 14(4):393–405, 1963.

Appendix

This section is dedicated to the proof of Proposition 3.7. After some slight changes to the assumptions since we are maximizing and not minimizing this is a direct application of Theorem 5.6.3 of Hernandez-Lerma and Lasserre [5]. In essence the first two results (Lemmas 5.3 and 5.4 below) yield the existence of a solution to the AOE (2.2) and (2.4). The third result (Lemma 5.6) then leads to the convergence of the normalized value iterates. Denote the state and action spaces of a generic MDP by $\mathbb{X}$ and $\mathbb{A}$, respectively. When speaking of the pricing and admission models we continue to use the superscripts P and C.

We begin with several definitions. Recall, in the pricing control problem, the function $w^P(p_j) = j$ for $j = 0, 1, \ldots, \ell$. Similarly in the admission control model, let $w^C(accept(0)) = 0$ and $w^C(reject(1)) = 1$. Define the graph of $\mathbb{A}$ by $Gr(\mathbb{A}) \equiv \{(x, a) \mid x \in \mathbb{X}, a \in \mathbb{A}(x)\}$, where $\mathbb{A}(x)$ is the set of available actions when in state x.

Definition 5.1 *For $v = ((i, j, z), a)$, $v' = ((i', j', z'), a') \in Gr(\mathbb{A})$ let the* **Euclidean distance** *between the*

elements v and v' be

$$\|v - v'\| \equiv \sqrt{(i - i')^2 + (j - j')^2 + (z - z')^2 + (w^k(a) - w^k(a'))^2},$$

where k is P or C when considering the pricing or admission control model, respectively.

Definition 5.2 *Let $Q(B|x, a)$ represent the probability of entering the set of states $B \subseteq \mathbb{X}$ given that action a is chosen in state $x \in \mathbb{X}$ (the transition kernal).*

1. *$Q(\cdot|x, a)$, is called **strongly continuous** if the function $F(x, a) \equiv \int_{y \in \mathbb{X}} f(y) Q(dy|x, a)$ is continuous and bounded whenever f is measurable and bounded.*

2. *A function f is said to be **inf-compact** if, for every $x \in \mathbb{X}$ and $r \in \mathbb{R}$, the set $\{a \in \mathbb{A}(x)| \ f(x, a) \leq p_0 - r\}$ is compact.*

Lemma 5.3 *The following hold for the non-stationary pricing and admission control models;*

1. *The one-stage reward function $r(x, a)$ is nonnegative, upper-semicontinuous in a and inf-compact.*

2. *The transition kernel Q is strongly continuous.*

Proof. Since the set of available actions A_i^k for $k = P, C$ is finite, the first assertion holds trivially. We show that the second assertion holds for the pricing control model. The admission control scenario is similar. Fix $\epsilon > 0$ and let $x = (i, j, z)$ and $x' = (i', j', z')$. Without loss of generality, assume that $z < z'$. Since $|Q_a(z)| \leq 1$ for all $z \in [0, T)$ it suffices to show the strong continuity of

$$F(i, z) \equiv \int_z^T \left[\Lambda(t) f(i, 1, t) + \mu_i(t) f(i, -1, t) + (1 - \Lambda(t) - \mu_i(t) f(i, 0, t) \right] e^{-(t-z)} dt. \qquad (5.1)$$

To ease the notation, let

$$f'(i, t) \equiv \Lambda(t) f(i, 1, t) + \mu_i(t) f(i, -1, t) + (1 - \Lambda(t) - \mu_i(t)) f(i, 0, t)$$

for $0 \leq i \leq m$. Assume that $\delta < 1$ so that $||x - x'|| < \delta$ implies $i = i'$, $j = j'$, and $a = a'$. Consider

$$|F(i, z) - F(i, z')| = \left| \int_z^T f'(i, t)e^{-(t-z)}dt - \int_{z'}^T f'(i, t)e^{-(t-z')}dt \right|$$

$$= \left| \int_z^{z'} f'(i, t)e^{-(t-z)}dt + (e^{-(z'-z)} - 1) \int_{z'}^T f'(i, t)e^{-(t-z')}dt \right|.$$

Let M denote the bound of f. Since $|p + q|^2 \leq 2p^2 + 2q^2$, we have

$$|F(i, z) - F(i', z')| \leq \sqrt{2 \left(\int_z^{z'} f'(i, t)e^{-(t-z)}dt \right)^2 + 2 \left((e^{-(z'-z)} - 1) \int_{z'}^T f'(i, t)e^{-(t-z')}dt \right)^2}$$

$$\leq 2M(1 - e^{-(z'-z)}).$$

Thus, if we choose $z' - z = \delta$ so that $2M(1 - e^{-(z'-z)}) \leq \epsilon$, the proof is complete. ∎

Recall that the discounted reward problems include a continuous discount fact $e^{-\alpha\sigma_t}$. However, most of the MDP theory is developed for the discrete-time discounted reward. To this end, for $\xi \in (0, 1)$ define

$$W_{n,\xi}^\pi(x) = \mathbb{E}_x^\pi \left[\sum_{t=0}^{n-1} \xi^t r(X_t, d(X_t)) \right].$$

to be the $n-$stage, discrete-time, $\xi-$discounted reward of policy π. Moreover, $W_\xi^\pi(x) \equiv \liminf_{n \to \infty} W_{n,\xi}^\pi(x)$ is the infinite horizon, discrete-time discounted reward. The optimal n-stage, discrete-time $\xi-$discounted reward and the optimal discrete-time, $\xi-$discounted reward are respectively defined

$$W_{n,\xi}(x) \equiv \sup_{\pi \in \Pi} W_{n,\xi}^\pi(x)$$

$$W_\xi(x) \equiv \sup_{\pi \in \Pi} W_\xi^\pi(x),$$

Lemma 5.4 *For the nonstationary pricing and admission control models, the following hold with $M = p_1 \Psi$ and $N = b(x) = p_1[1 + \Psi(T + \mathbb{E}_{(m,0,0)}^{\pi^*} \tau_0)]$, where τ_0 is the first time the queue length of a process started in state $(m, 0, 0)$ using the (discounted reward) optimal policy π^* reaches zero;*

1. *There exist a state $s \in \mathbb{X}$ and numbers $\omega \in (0,1)$ and $M \geq 0$ such that $(1-\xi)W_\xi(s) \leq M$ for all $\xi \in [\omega, 1)$.*

2. *Suppose we fix a distinguished state $s \in \mathbb{X}$. There exist $N \geq 0$ and a nonnegative function $b(\cdot)$ on $\mathbb{X}$ such that, $-N \leq h_\xi(x) \leq b(x)$ for all $x \in \mathbb{X}$ and $\xi \in [\omega, 1)$, where $h_\xi(x) \equiv W_\xi(x) - W_\xi(s)$.*

3. *Furthermore, the function $b(\cdot)$ is measurable and such that, for every $x \in \mathbb{X}$ and $a \in \mathbb{A}(x)$, $\int_{y \in \mathbb{X}} b(y)Q(dy|x,a) < \infty$.*

4. *Let $s \in \mathbb{X}$ be a fixed state. The sequence $\{h_{\xi(n)}\}$ is equicontinuous, where $h_{\xi(n)}(x) \equiv W_{\xi(n)}(x) - W_{\xi(n)}(s)$, $x \in \mathbb{X}$.*

Proof. The first assertion holds since the expected reward is bounded by $M = p_1\Psi$. To prove the second assertion, note that the difference in the expected discrete-time discounted reward between starting in state $x = (i,j,z)$ and starting in state $s = (0,0,0)$ is bounded. In doing so, let $S_0 \equiv \{(i,j,z) \in \mathbb{X} | z = 0\}$ and let N_{T-z} be a Poisson random variable of rate $\Psi(T-z)$ representing the expected number of decision epochs seen by a process starting in state (i,j,z) before reaching time T. Thus, a process starting in state (i,j,z) satisfies $X_{N_{T-z}+1} \in S_0$. Suppose $\{I_t, \ t \geq 0\}$ and $\{J_t, \ t \geq 0\}$ represent the number of customers in the system and the type of event that just occurred at the tth decision epoch. For any stationary policy π,

$$W_\xi^\pi(i,j,z) = \mathbb{E}_{(i,j,z)}^\pi \left[\sum_{t=0}^{N_{T-z}} \xi^t r(X_t, d(X_t)) \right] + \mathbb{E}_{(i,j,z)}^\pi W_\xi^\pi(I_{N_{T-z}+1}, J_{N_{T-z}+1}, 0)$$

$$\leq p_1\Psi T + \mathbb{E}_{(i,j,z)}^\pi W_\xi^\pi(I_{N_{T-z}+1}, J_{N_{T-z}+1}, 0). \tag{5.2}$$

It remains to show that the expected difference in the discounted rewards from two processes starting in different states in S_0 is also bounded. Suppose we start two processes on the same probability space each using the policy π. Process 1 starts in state $(i,j,0)$ and Process 2 in state $(i',j',0)$. Note that, if a decision must be made ($j = 1$), the difference in the rewards is at most p_1. Consider the processes immediately following this decision and without loss of generality assume that process 1 has a higher queue length.

Since each of the queue length processes is *skip-free*, the time that Process 1 first has queue length zero, say τ_0, bounds the coupling time of the two processes. Moreover, since the capacity is finite (and the service rate is strictly positive unless the system is empty), the first passage time τ_0 is finite almost surely with finite expectation. Thus,

$$|W_\xi^\pi(i,j,0) - W_\xi^\pi(i',j',0)| \leq p_1[1 + \Psi \, \mathbb{E}_{(i,j,0)}^\pi \, \tau_0] < \infty. \tag{5.3}$$

To complete the proof of the second assertion, combine (5.2) and (5.3) to obtain

$$|W_\xi^\pi(i,j,z) - W_\xi^\pi(0,0,0)| \leq p_1 \Psi T + |\,\mathbb{E}_{(i,j,z)}^\pi \, W_\xi^\pi(I_{N_{T-z}+1}, J_{N_{T-z}+1}, 0) - W_\xi^\pi(0,0,0)|$$

$$\leq p_1[1 + \Psi(T + \mathbb{E}_{(m,0,0)}^\pi \, \tau_0)] < \infty. \tag{5.4}$$

Since π was arbitrary, the inequality holds for the optimal policy as well. Thus if π^* is an optimal policy, the second assertion holds with $N = b(x) = p_1[1 + \Psi(T + \mathbb{E}_{(m,0,0)}^{\pi^*} \, \tau_0)]$. The third assertion clearly holds: since $b(\cdot)$ is constant. For the last assertion, let $x = (i,j,z)$ and $x' = (i',j',z')$ in $\mathbb{X}$ and assume $\epsilon > 0$. Without loss of generality, assume $z < z'$. We want to show that $|h_{\xi(n)}(x) - h_{\xi(n)}(x')| \leq \epsilon$ whenever $\|x - x'\| < \delta$ for some $\delta > 0$ where ϵ is independent of $\xi(n)$, x, and x'.

Consider two independent Markov chains, $X \equiv \{X_t,\ t \geq 0\}$ starting in (i,j,z) and $X' \equiv \{X_t',\ t \geq 0\}$ starting in (i',j',z') on the same probability space. Assume both chains use the discrete-time discounted reward optimal policy π^* and that $\delta < 1$ so that $i = i'$ and $j = j'$, i.e., $\|x - x'\| = z' - z$. Consider

$$|W_{\xi(n)}^*(i,j,z) - W_{\xi(n)}^*(i',j',z')| = |W_{\xi(n)}^*(i,j,z) - W_{\xi(n)}^*(i,j,z')|$$

$$\leq \mathbb{E}_{(i,j,z),(i,j,z')}^{\pi^*} \left[\sum_{t=0}^\infty \left| r(\mathbb{X}_t, d(X_t)) - r(\mathbb{X}_t', d(X_t')) \right| \right]. \tag{5.5}$$

Let $Z \equiv \sum_{t=0}^\infty |r(\mathbb{X}_t, d(X_t)) - r(\mathbb{X}_t', d(X_t'))|$ and E be the event that no decision epochs occur between

times z and z'. Thus,

$$\mathbb{E}^{\pi^*}_{(i,j,z),(i,j,z')}[Z] = \mathbb{E}^{\pi^*}_{(i,j,z),(i,j,z')}[Z|E]\,\mathbb{P}\{E\} + \mathbb{E}^{\pi^*}_{(i,j,z),(i,j,z')}[Z|E^c]\,\mathbb{P}\{E^c\}$$

$$\leq p_1[1 + \Psi(T + \mathbb{E}^{\pi^*}_{(m,0,z)}\,\tau_0)](1 - e^{-\Psi(z'-z)}) \tag{5.6}$$

where the inequality follows from the proof of the second assertion. Thus, choosing δ such that $\delta < 1$ and

$|z - z'| < \delta$ implies $p_1[1 + \Psi(T + \mathbb{E}^{\pi^*}_{(m,0,0)}\,\tau_0)](1 - e^{-\Psi(z'-z)}) < \epsilon$ yields the result. ∎

Applying Theorem 5.5.4 of Hernandez-Lerma and Lasserre [5] the results of Lemmas 5.3 and 5.4 yield the existence of a solution to the AOE and an average optimal policy. It remains to show that the undiscounted value functions may be normalized such that they approach a solution to the AOE in the limit. Recall that probability measures $\{\mu_n,\ n \geq 1\}$ on $\mathbb{X}$ are said to *converge weakly* to a probability measure μ, written $\mu_n \xrightarrow{w} \mu$, if $\int f d\mu_n \to \int f d\mu$ as $n \to \infty$ for every continuous and bounded function f on $\mathbb{X}$. Moreover, a Markov chain with transition kernal Q is called $\psi-$irreducible if for all $x \in \mathbb{X}$, whenever $\psi(B) > 0$, there exists some $n > 0$, possibly depending on both B and x, such that $Q^n(B \mid x) > 0$. The following lemma appears as Proposition 10.1.2 part (ii) of Meyn and Tweedie [14].

Lemma 5.5 *Suppose that $\{X_t,\ t \geq 0\}$ is ψ-irreducible. If μ is any subinvariant measure with $\mu(B) < \infty$ for some $B \in \mathcal{B}^+(\mathbb{X})$, the set of Borel subsets of $\mathbb{X}$ with positive ψ measure, then $\mu \succ \psi$, i.e., the measure ψ is absolutely continuous with respect to μ.*

Lemma 5.6 *For the nonstationary pricing and admission control models, the following hold;*

1. *The sequence $\{W_{n,\xi}\}$ is equicontinuous.*

2. *There is a probability measure ϕ on $\mathbb{X}^k$, independent of the initial state $x \in \mathbb{X}^k$, such that for all $x \in \mathbb{X}^k$ and each decision rule d, $Q^n(\cdot|x, d(x)) \xrightarrow{w} \phi$ as $n \to \infty$ and $\phi(G) > 0$ for every open set G and $k = P, C$.*

3. *There is a function $L : \mathbb{X}^k \to \mathbb{R}$ such that*

$$\int_{y \in \mathbb{X}^k} b(y) Q^n(dy | x, (d_n(x), d_{n-1}(x), \ldots, d_1(x))) \leq L(x), \text{ for all } x \in \mathbb{X}^k, \ n \geq 1,$$

where $b(\cdot)$ is the function in Lemma 5.4 and $k = P, C$.

Proof. We prove the result for the pricing control problem. The admission control problem is analogous. The proof of the first assertion follows in precisely the same manner as that used in part 4 of Lemma 5.4. To prove that the second assertion holds, let the *total variation norm* of the difference of two probability measures $Q(B \mid x, a)$ and $Q(B \mid x', a')$ be defined

$$\|Q(B \mid x, a) - Q(B \mid x', a')\|_W \equiv 2 \sup_{B \in \mathcal{B}(\mathbb{X})} |Q(B \mid x, a) - Q(B \mid x', a')|, \tag{5.7}$$

where $\mathcal{B}(\mathbb{X})$ is the Borel σ-algebra of $\mathbb{X}$. Remark 5.6.2 of Hernandez-Lerma and Lasserre [5] explains that $Q^n(B|x, d) \xrightarrow{w} \phi$ as $n \to \infty$ if for some $0 < \omega < 1$

$$\|Q(B \mid x, a) - Q(B \mid x', a')\|_W \leq 2\omega \qquad \text{for all } (x, a), (x', a') \in Gr(\mathbb{A}). \tag{5.8}$$

Note that (5.8) does not hold when $\sup_{B \in \mathcal{B}(\mathbb{X})} |Q(B \mid x, a) - Q(B \mid x', a')| = 1$. Thus, we will show in both models that $0 < Q(B \mid x, a) < 1 - \theta$ for $0 < \theta < 1$, independent of B. Let

$$B_{i,j}(z_1, z_2) = \{(i, j, z) | z \in (z_1, z_2), \ 0 \leq z_1 < z_2 \leq T\}$$

be an arbitrary open set in $\mathbb{X}^P$ for fixed $i \in \{0, 1, \ldots, m\}$ and $j \in \{-1, 0, 1\}$. For $0 \leq i < m$

$$0 < Q(B_{i+1,j}(z_1, z_2) \mid (i, j, z), a) \leq \int_{z_1}^{z_2} \Lambda(t) \Psi e^{-\Psi t} dt \leq 1 - e^{-\Psi T}.$$

Applying similar algebra to the other cases we conclude that $0 < Q(B \mid x, a) \leq 1 - e^{-\Psi T}$ for all open sets B and hence for all Borel sets. Since $0 < e^{-\Psi T} < 1$, substituting θ with $e^{-\Psi T}$ leads to

$$0 < \sup_{B \in \mathcal{B}(\mathbb{X})} |Q(B \mid x, a) - Q(B \mid x', a')| \leq 1 - \theta < 1, \tag{5.9}$$

which satisfies (5.8).

To complete the proof of the second assertion, it remains to show that $\phi(G) > 0$ for any open set G. Let γ be the counting measure defined on $\mathbb{Z}^+$ and κ the Lebesgue measure on $\mathbb{R}^+$. For any set $G \subseteq \mathbb{X}^P$, G can be written as a triplet (G_1, G_2, G_3) where G_i, $i = 1, 2, 3$ represents a set of elements for each dimension of the state space. Define the product measure ψ on $\mathbb{X}^P$ as $\psi(G) \equiv \gamma(G_1) \times \gamma(G_2) \times \kappa(G_3)$.

Since the first two elements of the state space are discrete, open sets under the measure ψ correspond to open sets in $[0, T)$. Recall that class 1 customers are always allowed entry into the system as long as it is not full. Next note that for $B \subset \mathbb{X}^P$, such that $\psi(B) > 0$, $Q^2(B \mid x, a) > 0$ for all $(x, a) \in Gr(\mathbb{A})$ (we may require two steps since we embed at the end of the horizon). That is to say, the Markov chain generated by any deterministic, stationary policy is ψ-irreducible. Given that ϕ is a limiting probability measure, the following two facts are evident: (i) the probability measure ϕ on $\mathbb{X}^P$ is subinvariant (in fact, invariant) since $\phi^n = \phi$ for all $n \in \mathbb{Z}^+$ and (ii) $\phi(B) \leq 1$ since ϕ is a probability measure. Therefore, we can apply Lemma 5.5 and conclude that $\phi \succ \psi$. By the definition of absolute continuity, $\phi(G) = 0$ implies $\psi(G) = 0$. Thus, the contrapositive of the definition leads to the conclusion that any set of positive ψ measure has positive ϕ measure. In particular, this fact holds for any open set G. The last assertion holds since $b(\cdot) = p_1[1 + \Psi(T + \mathbb{E}^{\pi^*}_{(m,0,0)} \tau_0)]$ is constant. ∎

Lemmas 5.3, 5.4 and 5.6 now imply that Theorem 5.6.3 of Hernandez-Lerma and Lasserre [5] can be applied to prove Proposition 3.7 directly.