Sensitivity Analysis for Stationary and Ergodic Queues

P. Konstantopoulos *
INRIA, Sophia Antipolis
2004 Route des Lucioles
06565 Valbonne Cedex
FRANCE

M. Zazanis †
Dept. of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL 60208

Abstract

Starting with some mild assumptions on the parametrization of the service process, perturbation analysis (PA) estimates are obtained for stationary and ergodic single server queues. Besides relaxing the stochastic assumptions, our approach solves some problems associated with the traditional regenerative approach taken in most of the previous work in this area. First, it avoids problems caused by perturbations interfering with the regenerative structure of the system. Second, given that the major interest is in steady-state performance measures, it examines directly the stationary version of the system, instead of considering performance measures expressed as Cesaro limits. Finally, it provides new estimators for general (possibly discontinuous) functions of the workload and other steady-state quantities.

KEYWORDS: STATIONARY PROCESSES, PERFORMANCE EVALUATION AND QUEUEING, NON-MARKOVIAN PROCESSES ESTIMATION.

1 Introduction

The increasing importance of sensitivity analysis in communications networks and manufacturing systems makes it desirable to construct reasonable estimators whose asymptotic behavior does not depend on the renewal character of the arrival or service processes. More specifically, consider a node in a queueing network that is assumed to be stable. Consider also a parameter, say θ, of the distribution of the service process at this node. The sensitivity of the mean delay is defined as the derivative of the mean delay with respect to θ. The question is whether, by operating (or

*Current address: EECS Department, University of California, Berkeley CA 94720

†Supported in part by N.S.F. Grants ECS-88110033 and DDM-8905638

simulating) the system at the nominal value θ and by real time measurements, one can obtain an estimate for the sensitivity at θ.

In this paper we obtain strongly consistent perturbation analysis estimators for general (average) performance criteria for a queue with stationary and ergodic input. Also, we give some new proofs, distinct from the existing ones, that not only simplify the underlying ideas, but also generalize them to the stationary and ergodic context. For an overview of the area and further references see [11] and [7].

Consider a queue with stationary and ergodic arrival $\{t_n\}$ and service $\{\sigma_n\}$ processes. Let θ be a parameter (in a sense to be made precise later) of the service process such that the queue under consideration is stable for all allowable values of θ. Let $J(\theta)$ be an average steady-state performance measure of the system. For instance, $J(\theta)$ can be the variance of the workload in steady-state, with respect to the parameter θ.

We will develop direct estimators for $\frac{\partial}{\partial\theta}J(\theta)$ of the form

$$\frac{1}{t}\int_0^t Y_t(\theta)dt, \tag{1}$$

or

$$\frac{1}{n}\sum_{k=1}^{n} Y_{t_k}(\theta), \tag{2}$$

where $Y_t(\theta)$ is a process constructed on the same probability space on which our original data $\{t_n\}$ and $\{\sigma_n\}$ are defined.

Strong consistency of an estimator of the above form means convergence, as $t \to \infty$ or $n \to \infty$, to $\frac{\partial}{\partial\theta}J(\theta)$. Our method for constructing these estimators is the following: We will assume that the system is in steady state (meaning that a suitable process like the queue length is stationary) and construct a stationary and ergodic process Y_t such that $\frac{\partial}{\partial\theta}J(\theta) = EY_0(\theta)$, where E denotes expectation under the stationary measure P or a Palm transformation of it (e.g., conditioning). Indeed, we shall be freely using the ideas and formulas for Palm measures. The fact that we assume that the system is in steady state is only for reasons of constructing the estimators. Strong consistency is guaranteed by the ergodic theorem for any initial condition.

In a renewal context, i.e. when interarrival times are i.i.d. and independent of the service times, also i.i.d. with common distribution G_θ, one can construct the service time processes on the same probability space by doing the obvious thing; namely by choosing i.i.d. numbers ξ_n uniformly distributed on $[0,1]$ and by letting $\sigma_n(\theta) = G_\theta^{-1}(\xi_n)$, where G_θ^{-1} is a version of the inverse of the function G_θ. This way, one has constructed all processes associated with the evolution of the queue on the same probability space. Consider, for instance, the resulting workload process, denoted by $W_t(\theta)$. It has been shown (see [15] and [17]) that a perturbation analysis estimator for the derivative of the expected waiting time in steady-state is given by (2) where $Y_t(\theta)$ is the pathwise right derivative of $W_t(\theta)$ with respect to θ. These steady-state results have been obtained by considering Cesaro limits of the form

$$\lim_{\delta\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t \delta^{-1}\left[W_s(\theta+\delta) - W_s(\theta)\right] ds \ , \tag{3}$$

and showing that it is indeed possible to interchange the order in which the two limits are taken in (3). This was also the approach taken in [5] and [6] for Markovian networks and in [8] for GSMP's with a special type of regenerative structure. A problem that arises in this case is that the regeneration points are in general not the same for the two processes, $W_t(\theta)$ and $W_t(\theta+\delta)$. This was pointed out in [10]. An alternative approach requiring convexity of the sample performance measures with respect to θ was taken in [12].

In this paper we extend the results on strong consistency of PA estimators to stationary and ergodic queues. This extension achieves the following goals: First, it shows that renewal assumptions and the existence of special regenerative structure are not necessary. Second, it allows one to apply our results to any node of a stationary and ergodic network, provided the node does not belong to a feedback loop. Third, it presents a new construction which allows us to analyze the system directly in steady-state thus eliminating the need to examine Cesaro limits and to justify the interchange of two limits as in (3). Finally, in this paper we consider general (possibly discontinuous) functions of the workload and obtain interesting new expressions.

Some assumptions on the parametrization of the service times are given in the sequel. Consider two jointly stationary and ergodic sequences $\{\tau_n\}$ and $\{\xi_n\}$ under a probability measure that will be denoted by P^0 in order to conform with the notation used in the rest of the paper. The first one is the sequence of interarrival times while the latter is the sequence of random variables from which the service times are generated in the following way: There is a deterministic function $(\xi, \theta) \to h(\xi, \theta)$ such that $\sigma_n(\theta) = h(\xi_n, \theta)$. The parameter θ ranges over a finite interval $[a, b]$ and $E^0 \sigma_n(\theta) < E^0 \tau_n := \lambda^{-1}$ for all θ. Moreover:

(i) $\theta \to h(\xi, \theta)$ is differentiable and Lipschitz, i.e., $|h(\xi, \theta_1) - h(\xi, \theta_2)| \leq K(\xi)|\theta_1 - \theta_2|$. (Of course, differentiability automatically implies that the function is Lipschitz on $[a, b]$.)

(ii) $h(\xi, \theta) \leq h(\xi, b)$.

(iii) $\xi \to h(\xi, \theta)$ is one-to-one.

We note at this point that the problem of putting the data on a common probability space, so that θ is a parameter of the random variables and not of the probability measure, is an important one that can be crucial in the behavior of the estimators. Our approach here simply mimics the usual one. One can also see that the above conditions are not too restrictive. This is best seen by an example: Let σ_n be i.i.d. with common exponential distribution with rate θ. Consider the function $h(\xi, \theta) = -\theta^{-1} \log \xi$. Let ξ_n be i.i.d. and uniform in the interval $(0, 1)$. Then $\sigma_n = h(\xi_n, \theta)$ has the required exponential distribution. Since we are interested in the derivative of a performance measure at a specific value, say θ_0, of the parameter, it suffices to consider the function $\theta \to h(\xi, \theta)$ defined on an interval $[a, b]$ that contains θ_0 in its interior. In this case clearly $h(\xi, \theta) = -\theta^{-1} \log \xi$ satisfies the above conditions.

In Section 2 we present an outline of the method and the main ideas behind it. Under conditions

(i), **(ii)** and **(iii)** (which are not necessarily the minimal ones) plus some extra moment conditions we construct estimators for the derivatives of functions of the workload and prove their strong consistency in Sections 3 and 4. A sketch of these proofs can also be found in [13]. In Section 5 we consider other performance measures. Finally in Section 6 we discuss what happens in the classical cases and also some simulation applications.

2 Sketch of the method

In this section we will give an outline of the basic ideas. Let $\tilde{W}(\theta)$ be a random variable distributed like the steady-state workload when the parameter of the service process has the value θ. Let f be a function of bounded variation such that $Ef(\tilde{W}(\theta)) < \infty$. The questions that we want to deal with are:

- Does the derivative of $Ef(\tilde{W}(\theta))$ exist?
- If yes, is there a perturbation analysis estimator for this derivative?

To construct a perturbation analysis estimator one should start by constructing stationary versions of the processes $W_{\cdot}(\theta), \theta \in [a, b]$ on a common probability space. Recall that we work in a non-renewal/regenerative framework and so the initial condition may play a role. Let us first introduce some terminology. We denote by P the distribution of the stationary version of the marked point process $[t_n, \xi_n]$ (the n-th point t_n is the n-th arrival time and has ξ_n as a mark, with the convention $t_0 \leq 0 < t_1$—this convention will be used throughout the paper). A superscript or a subscript to P will denote a Palm transformation of P with respect to some stationary point process (i.e., loosely speaking, conditioning on the event that there is a point on the origin of time). For instance, P^0 is in fact the Palm transformation of P with respect to the stationary arrival process. Since $E^0\sigma_n(b) < E^0\tau_n$, there is a stationary workload process $W_t(b)$ under the probability measure P. For more details see [3] and [16], Chapter 7. Because of assumption **(iii)** of Section 1, knowledge of the sequence $\sigma_n(b)$ implies knowledge of $\sigma_n(\theta)$ for all $\theta \in [a, b]$. Now assumption **(ii)** allows us to construct each realization of $W_t(\theta)$ from the realization of $W_t(b)$ by keeping the so-called construction points (times of arrival of customers that find the b-system empty) and then by replacing the $\sigma_n(b)$'s by the $\sigma_n(\theta)$'s in each busy cycle of the b-system. This may introduce additional construction points for the θ-system resulting from the possible breaking of the busy cycles of the b-system. Denote by $T_n(\theta)$ the arrival times of customers that find the θ-system empty. We just said that $\{T_n(\theta)\} \supseteq \{T_n(b)\}$.

Suppose for the moment that the function f is smooth and nice. One would then like to prove that a PA estimator of $\partial EW_0(\theta)/\partial\theta$ is given by (1) where $Y_t = f'(W_t(\theta))W'(\theta)$, with $W'(\theta)$ being the right derivative of $W(\theta)$ with respect to θ (see below for the definition). This is tantamount to showing that

$$\frac{\partial}{\partial\theta} Ef(W_0(\theta)) = E\frac{\partial}{\partial\theta} f(W_0(\theta)). \tag{4}$$

The standard approach that would mimic the one of [17] would consist in looking at busy cycles of the θ-system and essentially trying to establish that

$$\frac{\partial}{\partial\theta}\frac{1}{E^*_\theta(T_1(\theta)-T_0(\theta))}E^*_\theta\int_{T_0(\theta)}^{T_1(\theta)} f(W_t(\theta))dt = \frac{1}{E^*_\theta(T_1(\theta)-T_0(\theta))}E^*_\theta\int_{T_0(\theta)}^{T_1(\theta)}\frac{\partial}{\partial\theta}f(W_t(\theta))dt \tag{5}$$

(Here P^*_θ denotes the Palm transformation of P with respect to the point process $\{T_n(\theta)\}$ and E^*_θ the corresponding expectation.) Indeed, the left hand sides and the right hand sides of (4) and (5) are the same. Observe now that we might as well write

$$Ef(W_0(\theta)) = \frac{1}{E^*_b(T_1(b)-T_0(b))}\; E^*_b\int_{T_0(b)}^{T_1(b)} f(W_t(\theta))dt, \tag{6}$$

since by construction $W_t(\theta)$ and $T_n(b)$ are jointly stationary. This way the integration limits no longer depend on θ. In other words, we avoid problems caused by the fact that when the value of θ is changed to $\theta+\delta$, some of the busy periods in the original sample path will split into smaller ones and others will coalesce. The remaining of the argument, i.e. the justification of differentiating inside the expectation is purely analytical and depends on the way the service processes have been defined on the common probability space.

Our goal is to consider more general functions f, for instance $f(w) = 1(w > x)$. In that case there is no obvious guess for the process Y_t that would lead to the construction of the PA estimator. The following section deals with this problem in the spirit of the above discussion.

3 Derivatives for indicator functions

Consider $f(w) = 1(w > x)$ and the problem of finding a PA estimator for $Ef(W_0(\theta)) = P(W_0(\theta) > x)$, the time stationary probability that the workload exceeds a level x. Before proceeding further let us define the process $W_t'(\theta)$, the derivative of the workload at nominal value θ. It is not difficult to see that this is a right continuous, piecewise constant process given by

$$W_t'(\theta) = (W_{t_n-}'(\theta) + \sigma_n'(\theta))1(W_t(\theta) > 0) \quad , \quad t_n \le t < t_{n+1} \; . \tag{7}$$

Thus it jumps by an amount equal to $\sigma_n'(\theta)$ at each arrival t_n and is set to zero as soon as the system empties. In the analysis that follows it is important to bear in mind that $W_t(\theta)$ has right continuous paths.

We need some extra notation: We will denote by P_x the Palm transformation of P with respect to the point process of the downcrossings of $W_t(\theta)$ at level x. Informally speaking, this is P conditional on having an x-downcrossing at the origin of time. Note that P_x depends on θ. Let also λ_x be the rate of the x-downcrossings by $W_t(\theta)$. Finally let $A(I)$ denote the number of arrivals in a set I.

Theorem 1 *If $E^0 A[T_0(b), T_1(b))^2 < \infty$ and $E^0 K(\xi_0)^2 < \infty$ then*

$$\frac{\partial}{\partial\theta} P(W_0(\theta) > x) = \lambda_x E_x W_0'(\theta). \tag{8}$$

Proof Using the Palm inversion formula (e.g. see [3]) we get

$$P(W_0(\theta) > x) = \lambda_b^* E_b^* \int_{T_0(b)}^{T_1(b)} 1(W_t(\theta) > x)dt, \tag{9}$$

where λ_b^* is the rate of the point process $\{T_n(b)\}$. As mentioned earlier, this particular point process plays a key role in our proof. The basic observation here is that $W_t(\theta) \le W_t(b)$ for all $\theta \in [a, b]$ (due to the assumption that $\sigma_n(b)$ dominates $\sigma_n(\theta)$ for all θ) and hence the customers that find the b-system empty also find the θ-system empty, for all $\theta \in [a, b]$. Thus in the right hand side of (9) the dependence on θ is only inside the integrand and not in the integration limits. We now have

$$\frac{1}{\delta}\{P(W_0(\theta+\delta) > x) - P(W_0(\theta) > x)\} = \lambda_b^* E_b^* \int_{T_0(b)}^{T_1(b)} \frac{1}{\delta}\{1(W_t(\theta+\delta) > x) - 1(W_t(\theta) > x)\}dt. \quad (10)$$

Observe that the quantity inside the expectation of (10) converges pathwise as $\delta \to 0$ to

$$\sum_{T_0(b) \le d_n < T_1(b)} W'_{d_n}(\theta),$$

where d_n is the sequence of downcrossings of the level x by the process $W_t(\theta)$ (and hence depends on θ). This is a direct consequence of the definition of $W'_t(\theta)$ (see formula (7)). It can also be easily seen that the quantity inside the expectation of (10) is bounded by

$$\frac{1}{\delta} \sum_{T_0(b) \le t_n < T_1(b)} \sum_{i=0}^{n} |\sigma_i(\theta+\delta) - \sigma_i(\theta)| =: Z(\delta).$$

To show existence of the derivative of $P(W_0(\theta) > x)$ it now suffices to show that $Z(\delta)$ is bounded above by an E_b^*-integrable random variable (dominated convergence theorem). To this end, use the Lipschitz property to get

$$Z(\delta) \le \sum_{T_0(b) \le t_n < T_1(b)} \sum_{i=0}^{n} K(\xi_i).$$

This is further bounded above by

$$N \sum_{i=0}^{N-1} K(\xi_i),$$

where $N = A[T_0(b), T_1(b))$. It remains to show that $E_b^* N \sum_{i=0}^{N-1} K(\xi_i) < \infty$. The cycle formula (see [14]—for a short proof see [4]) between P_b^* and P^0 gives

$$E_b^* N \sum_{i=0}^{N-1} K(\xi_i) = \frac{\lambda}{\lambda^*} E^0 N K(\xi_0).$$

A straightforward application of the Cauchy-Schwarz inequality shows that

$$E^0 NK(\xi_0) \leq \frac{\lambda}{\lambda_b^*} \sqrt{E^0 N^2} \sqrt{E^0 K(\xi_0)^2}, \tag{11}$$

which is finite by the assumptions. Thus the conditions for the dominated convergence theorem hold and so we can interchange limit and expectation in (10). We conclude that $\frac{\partial}{\partial\theta} P(W_0(\theta) > x)$ exists and

$$\frac{\partial}{\partial\theta} P(W_0(\theta) > x) = \lambda_b^* E_b^* \sum_{T_0(b) \leq d_n < T_1(b)} W'_{d_n}(\theta) = \lambda_x E_x W'_0(\theta).$$

The latter equality above is obtained by another application of the cycle formula. This concludes the proof of the theorem. □

Corollary Under the conditions of Theorem 1 we also have

$$\frac{\partial}{\partial\theta} P(W_0(\theta) > x) = \lambda E^0 W'_0(\theta)[1(W_0(\theta) > x) - 1(W_{t_1-}(\theta) > x)]. \tag{12}$$

Proof This follows by yet another application of the cycle formula, this time between measures P_x and P^0. Indeed, observe that the difference of the two indicator functions of (12) is one if and only if there is a downcrossing of the level x by the process $W_t(\theta)$ on the interarrival interval $[t_0, t_1)$. □

Note The Cauchy-Schwarz inequality and the conditions of Theorem 1 in (11) were used to show that $E^0 NK(\xi_0)$ is finite. Using Hölder's inequality we could trade off the existence of lower moments of N for higher moments of $K(\xi_0)$. In particular, if $K(\xi_0)$ is constant which is the case if for instance θ is a location parameter of the service times (i.e. if $\sigma_n(\theta) = \xi_n + \theta$), only $E^0 N$ needs to be finite.

4 Derivatives for general functions

We now extend the result of Theorem 1 to more general functions f. That is, we want to obtain an estimator for $\frac{\partial}{\partial\theta} Ef(W_0(\theta))$. Recall that for f = indicator function we have two formulas (8) and (12).

We confine ourselves to the class of functions that are of bounded variation on bounded intervals (*locally bounded variation*). In this general case we will need the following set of assumptions:

A1 $E^0 K(\xi_0)^4 < \infty$,

A2 $E^0 A[T_0(b), T_1(b))^4 < \infty$,

A3 $E^0 \sup_\theta f(W_0(\theta))^2 < \infty$,

A4 $E^0 \sup_\theta f(W_{0-}(\theta))^2 < \infty$.

Note that these are stronger than the assumptions of Theorem 1, but this is not surprising in view of the generality of the function f. We should also note that these assumptions can be relaxed depending on the specific nature of f. When f is increasing, both **A3** and **A4** can be replaced by:

A3′ $E^0[f(W_0(b))^2] < \infty$.

See also the remarks on special cases at the end of the section and Section 6 for the renewal case. We are now ready to state and prove the following theorem:

Theorem 2 *Consider a function f that is locally of bounded variation. Then under the assumptions* **A1**–**A4** *we have*

$$\frac{\partial}{\partial \theta} E f(W_0(\theta)) = \int_0^\infty \lambda_x E_x W_0'(\theta) f(dx) \tag{13}$$

$$= \lambda E^0 W_0'(\theta)[f(W_0(\theta)) - f(W_{t_1-}(\theta)]. \tag{14}$$

Proof Suppose first that f is an elementary function, that is, a finite linear combination of indicator functions: $f(w) = \sum_{i=1}^k \alpha_i 1(w > x_i)$. Then (13) follows directly from (8) of Theorem 1 and (14) from (12) of the Corollary.

For a general f of locally bounded variation, there is no loss of generality if we assume that it is nonnegative and increasing. We can then approximate f from below by an increasing sequence of elementary functions f_n that converge to f uniformly over compact sets. We shall establish (14) first and then (13) will follow form an application of the cycle formula. Let

$$g_n(\theta) = Ef_n(W_0(\theta)),\ g(\theta) = Ef(W_0(\theta)).$$

The derivative of g_n exists and

$$g_n'(\theta) = \lambda E^0 W_0'(\theta)[f_n(W_0(\theta)) - f_n(W_{t_1-}(\theta))].$$

To show that the derivative of g exists and is equal to (14) it suffices to show that g_n converges to g for some θ_0 in $[a, b]$, and that g_n' converges uniformly in θ (e.g., see [2]). In other words, it suffices to show that

$$\sup_\theta |E^0 W_0'(\theta)[f(W_0(\theta)) - f_n(W_0(\theta))]| \to_{n\to\infty} 0 \tag{15}$$

and that

$$\sup_\theta |E^0 W_0'(\theta)[f(W_{t_1-}(\theta)) - f_n(W_{t_1-}(\theta))]| \to_{n\to\infty} 0. \tag{16}$$

We show (15) and then (16) will follow by the same token. The expression in (15) is bounded above by

$$E^0 \sup_\theta |W_0'(\theta)| \sup_\theta |f(W_0(\theta)) - f_n(W_0(\theta))|. \tag{17}$$

Since $0 \leq \ldots \leq f_n \leq f_{n+1} \leq \ldots \leq f$, the quantity inside the expectation of (17) is bounded above by $\sup_\theta |W_0'(\theta)| \sup_\theta |f(W_0(\theta)) - f_1(W_0(\theta))|$ We next show that this has finite expectation. Clearly,

$$E^0 \sup_\theta |W_0'(\theta)| \sup_\theta |f(W_0(\theta)) - f_1(W_0(\theta))| \leq \sqrt{E^0 \sup_\theta W_0'(\theta)^2} \sqrt{E^0 \sup_\theta f(W_0(\theta))^2}.$$

The latter term is finite by **A3**. For the proof of the finiteness of the first term we refer the reader to the Appendix.

Let now $X_n = \sup_\theta |f(W_0(\theta)) - f_n(W_0(\theta))|$. If we show that X_n converges to zero P^0-almost surely then we are done (dominated convergence theorem). Equivalently, it suffices to show that $E^0 X_n \to 0$.

To this end, given an $\epsilon > 0$, choose a constant K such that

$$E^0 f(W_0(b)) 1(W_0(b) > K) \leq \epsilon.$$

As mentioned above, the approximating sequence f_n can be chosen so that it converges uniformly over the interval $[0, K]$. Write

$$E^0 X_n = E^0 X_n 1(W_0(b) \leq K) + E^0 X_n 1(W_0(b) > K). \tag{18}$$

For the first term of (18) we have the bound

$$\sup_{0 \leq w \leq K} |f(w) - f_n(w)| P^0(W_0(b) \leq K),$$

which goes to zero by the uniform convergence of f_n. Using the inequalities $f_n \leq f$ and $W_0(\theta) \leq W_0(b)$ we see that the second term of (18) is bounded above by

$$2E^0 f(W_0(b)) 1(W_0(b) > K) \leq 2\epsilon.$$

This readily shows that $E^0 X_n \to 0$ and hence (15). The proof of (16) is similar. □

Let us now examine some special cases and see how assumptions **A1**–**A4** can be relaxed.

The case of a bounded f.

From the proof of Theorem 2 it is clear that in this case we only need to have assumptions that ensure $E^0 \sup_\theta |W_0'(\theta)| < \infty$. Repeating the proof of the Appendix, we can readily see that assumptions **A1**–**A4** can be replaced by the assumptions of Theorem 1.

The case of a smooth f.

Here one is typically able to differentiate not only inside the expectation of the right hand side of (6), but also inside the integral, to obtain

$$\frac{\partial}{\partial \theta} E[f(W_0(\theta))] = E[f'(W_0(\theta)) W_0'(\theta)] \ , \tag{19}$$

or equivalently (5). The moment conditions necessary for this depend of course on f. When f is a polynomial of degree r, an analysis similar to that in the Appendix shows that **A1**,**A2**, and **A3′** (that is, the same conditions for an increasing function) are sufficient to guarantee (5). We point out that, in the case of smooth f, 19 is equivalent to 14. Indeed, applying the Palm inversion formula to the right hand side of 19 we get

$$E[f'(W_0(\theta))W_0'(\theta)] \;=\; \lambda E^0 \int_0^{t_1} f'(W_s(\theta))W_s'(\theta)ds$$

which, since $W_s'(\theta)$ is constant and equal to $W_0'(\theta)$ on $[0, t_1)$, can be written as

$$\lambda E^0 W_0'(\theta) \int_0^{t_1} f'(W_s(\theta))ds.$$

This is easily seen to be equal to the right hand side of 12.

5 Other performance measures

5.1 Queue length

We now consider briefly the queue length process $Q_t(\theta)$. In contrast to the workload process, the pointwise derivative of $Q_t(\theta)$ with respect to θ is equal to 0 for almost all t. Let $D(\theta)$ be the departure process, $S_n(\theta)$ the n'th departure epoch, and $Y_t(\theta)$ the derivative of the last departure time before t with respect to θ. Dropping the dependence on θ for simplicity, Y_t is constant in $[S_i, S_{i+1})$, and given by the following recursion on departure epochs:

$$Y_{S_{n+1}} = \sigma_{n+1}' + 1(Q_{S_n} > 0)Y_{S_n}.$$

As usual, these processes are defined to be right continuous. Finally N_k is the point process of the k-downcrossings of $Q_t(\theta)$ (downcrossings from k to $k-1$, i.e. departures that leave the system with $k-1$ customers), λ_k is the intensity of N_k, and E_k is the expectation with respect to the Palm probability of k-downcrossings. Then an analysis similar in every respect to the one in the proof of Theorem 1 gives

$$\frac{\partial}{\partial\theta} P(Q_0 \leq k) = \lambda_k E_k[Y_0]. \tag{20}$$

To obtain the counterpart of (13) consider a function $f : N \to R$. Write f as $\sum_{i=1}^{\infty} \alpha_i 1(k \geq i)$ and approximate it by below by $f_n(k) = 1(k \leq n) f(k)$. Then, by (20) and the linearity of expectation,

$$\frac{\partial}{\partial \theta} E[f_n(Q_0)] = \sum_{i=1}^{n} \alpha_i \lambda_i E_i[Y_0].$$

A uniform convergence argument allows us to pass in the limit:

$$\frac{\partial}{\partial \theta} E[f(Q_0)] = \sum_{i=1}^{\infty} \alpha_i \lambda_i E_i[Y_0]. \tag{21}$$

Let E^1 denote the expectation with respect to the Palm probability corresponding to departures. We now use the cycle formula in order to get another expression for (21):

$$\lambda_i E_i[Y_0] = \lambda E^1 \int_{[S_0, S_1)} Y_s N_i(ds) = \lambda E^1 Y_{S_0} 1(Q_{S_0} = k-1),$$

where we have used the fact that Y_s is constant on $[S_0, S_1)$ and that there is a k-downcrossing at S_0 if and only if $Q_{S_0} = k-1$. Writing $1(Q_{S_0} = k-1) = 1(Q_{S_0} \geq k-1) - 1(Q_{S_0} \geq k)$ we obtain

$$\frac{\partial}{\partial \theta} Ef(Q_0) = \lambda E^1 Y_{S_0} [f(Q_{S_0 -}) - f(Q_{S_0})].$$

In the above expression $Q_{S_0 -} = Q_{S_0} + 1$ is the number of customers in the system just before a departure.

5.2 Sojourn time

Up to this point we have considered only "time stationary" performance measures (to use standard queueing theoretic terminology). "Customer stationary" performance measures (such as the expected sojourn time in steady state) can be treated as in Section 4, with the exception of additional smoothness requirements on f. One starts by using a formula similar to (6) in discrete time:

$$E^0 f(W_0(\theta)) = \frac{1}{E_b^* N} \sum_{T_0(b) \leq t_n < T_1(b)} f(W_{t_n}(\theta)).$$

By using arguments similar to those of Theorems 1 and 2, one can show that, under appropriate conditions,

$$\frac{\partial}{\partial \theta} E^0 f(W_0(\theta)) = \frac{1}{E_b^* N} \sum_{T_0(b) \leq t_n < T_1(b)} \frac{\partial}{\partial \theta} f(W_{t_n}(\theta))$$

$$= E^0 f'(W_0(\theta))W_0'(\theta).$$

It is exactly this case that has been studied in [17] under renewal assumptions on the arrival and service processes.

5.3 Joint distributions

Fix an $s > 0$ and consider $P(W_0(\theta) > x, W_s(\theta) > y)$. A procedure similar to that of Theorem (1) will give us the formula

$$\frac{\partial}{\partial\theta} P(W_0(\theta) > x, W_s(\theta) > y) = \lambda_x E_x[W_0'(\theta)1(W_s(\theta) > y)] + \lambda_y E_y[W_0'(\theta)1(W_{-s}(\theta) > x)],$$

which can also be easily translated into a formula of the form (12).

6 The renewal case—final remarks

Explicit conditions involving only the interarrival and service distributions can be given in the renewal case. In fact it is well known (e.g. see [9]) that the existence of $E^0[\sigma(b)^r]$ implies that of $E^0 A[T_0(b), T_1(b))^r$ for $r \geq 1$. For example, **A2** can be replaced with $E^0\sigma(b)^4 < \infty$. If f is polynomial of degree r, then it is enough to show that **A3′** holds and for this $E^0\sigma(b)^{2r+1} < \infty$ is sufficient (see [1]).

From a simulation point of view, the results in this paper are important since they justify isolating a single server node not belonging to any feedback loop and using the estimators developed above. The proposed estimators would be strongly consistent for performance criteria of the type described here. Theorem 2 is particularly useful in this context. If $\{t_n\}$ is the sequence of customer arrival times with the convention $t_0 \leq 0 < t_1$ and we simulate (or observe) the system in the interval $(0, t)$, Theorem 2 suggests

$$\frac{1}{t} \sum_{\{n: 0 < t_{n+1} < t\}} W_{t_n}' [f(W_{t_n}) - f(W_{t_{n+1}})],$$

as a strongly consistent estimator for $\frac{\partial}{\partial\theta}E[f(W_0)]$ which can be used whether f is differentiable or not. Unlike the estimator based on level crossings the above estimator does not require the introduction of crossing events which presents an advantage in simulation.

7 Appendix

In this appendix we prove that assumptions **A1** and **A2** imply that

$$E^0 \sup_\theta W_0'(\theta)^2 < \infty.$$

From the definition of the process $W_t'(\theta)$ we have

$$W_0'(\theta) = \sum_{T_0(\theta)\le t_n<0} \sigma_n'(\theta).$$

Using the Lipschitz condition we get

$$|W_0'(\theta)| \le \sum_{T_0(\theta)\le t_n<0} K(\xi_n).$$

Since $T_0(\theta) \ge T_0(b)$ (by the way the processes have been constructed),

$$\sup_\theta |W_0'(\theta)| \le \sum_{T_0(b)\le t_n<0} K(\xi_n)$$

Let $N = A[T_0(b), T_1(b))$ and use the inequality $(x_1 + \ldots + x_n)^2 \le n(x_1^2 + \ldots + x_n^2)$ to obtain

$$E^0 \sup_\theta (W_0'(\theta))^2 \le E^0 N \sum_{T_0(b)\le t_n<0} K(\xi_n)^2.$$

Now use the cycle formula between P^0 and P_b^* to obtain the relation

$$E^0 N \sum_{T_0(b)\le t_n<0} K(\xi_n)^2 = \frac{\lambda_b^*}{\lambda} E_b^* N \sum_{T_0(b)\le t_n<T_1(b)} \sum_{i=0}^{n} K(\xi_i)^2.$$

This is smaller than

$$\le \frac{\lambda_b^*}{\lambda} E_b^* N^2 \sum_{T_0(b)\le t_n<T_1(b)} K(\xi_n)^2.$$

A second application of the cycle formula gives

$$\frac{\lambda_b^*}{\lambda} E_b^* N^2 \sum_{T_0(b) \leq t_n < T_1(b)} K(\xi_n)^2 = E^0 N^2 K(\xi_0)^2.$$

Finally, a Cauchy-Schwarz inequality applied to the latter bound gives

$$E^0 \sup_\theta (W_0'(\theta))^2 \leq \sqrt{E^0 K(\xi_0)^4} \sqrt{E^0 N^4},$$

which is finite by assumptions **A1** and **A2**. □

References

[1] S. Asmussen(1987). *Applied probability and queues.* Wiley.

[2] R.G. Bartle (1979). *The elements of real analysis.* Wiley.

[3] F. Baccelli and P. Brémaud (1987). *Palm probabilities and stationary queues.* Lect. Notes Stat., **41**, Springer-Verlag.

[4] P. Brémaud (1990). An elementary proof of Sengupta's invariance relation and a remark on Miyazawa's conservation principle. *J. Appl. Prob.*, submitted.

[5] X.R. Cao, (1987). Realization probability in closed Jackson queueing networks and its application. *Adv. Appl. Prob.*, **19**, 708-738.

[6] X.R. Cao, (1988). A sample performance function of closed Jackson queueing networks. *Operations Research*, **36**, 1, 128-136.

[7] P. Glasserman (1990). *Gradient estimation via perturbation analysis,* Kluwer Academic, Boston.

[8] P. Glasserman, J.Q. Hu, and S.G. Strickland, (1990). Strongly consistent steady-state derivative estimates, Technical Report, Division of Applied Sciences, Harvard University.

[9] A. Gut (1988). *Stopped random walks.* Springer-Verlag.

[10] Heidelberger, P., X.R. Cao, M.A. Zazanis, and R. Suri (1988). Convergence properties of Infinitesimal Perturbation Analysis, *Management Science, 34,* 11, 1281-1302.

[11] Y.C. Ho (1987). Performance evaluation and perturbation analysis of discrete event dynamic systems. *IEEE Trans. Aut. Contr.*, **32**, 563-572.

[12] J.Q. Hu, (1990) Convexity of sample path performance and strong consistency of infinitesimal perturbation analysis estimates, to appear in *IEEE Trans. on Aut. Contr.*

[13] P. Konstantopoulos and M. Zazanis (1990). Extensions of Perturbation Analysis to Stationary Queues. Allerton Conference, Oct. 1990.

[14] J. Neveu (1976). Sur les mesures de Palm de deux processus ponctuels stationnaires. *Zeit. f. Wahr.*, **34**, 199-203.

[15] R.Suri and M. Zazanis (1988). Perturbation analysis gives strongly consistent estimates for the M/G/1 queue. *Mgmt. Sci., 34*, **1**, 39-64.

[16] J. Walrand (1988) *An introduction to queueing networks.* Prentice Hall.

[17] M. Zazanis and R. Suri (1989). Perturbation analysis of the GI/G/1 queue. *QUESTA*, submitted.