Perturbation Analysis of the GI/GI/1 Queue

Michael A. Zazanis
Dept. of Industrial Engineering
and Operations Research
University of Massachusetts
Amherst, MA 01003

Rajan Suri
Dept. of Industrial Engineering
University of Wisconsin
1513 University Avenue
Madison, WI 53706

Abstract

We examine a family of GI/GI/1 queueing processes generated by a parametric family of service time distributions, $F(x, \theta)$, and we show that under suitable conditions the corresponding customer stationary expectation of the system time is twice continuously differentiable with respect to θ. Expressions for the derivatives are given which are suitable for single run derivative estimation. These results are extended to parameters of the interarrival time distribution and expressions for the corresponding second derivatives (as well as partial second derivatives involving both interarrival and service time parameters) are also obtained. Finally we present perturbation analysis algorithms based on these expressions along with simulation results demonstrating their performance.

KEYWORDS: SENSITIVITY ANALYSIS, SECOND DERIVATIVES ESTIMATION.

1 Introduction

Consider a GI/GI/1 queue with service time distribution belonging to a parametric family $F(x, \theta)$, $\theta \in [a, b]$. Assume the system to be stable for all θ in $[a, b]$ and denote by $T(\theta)$ a random variable distributed according to the steady state distribution of the system time of a customer. We present a sample path construction of two such queueing processes on the same probability space, one with service time distribution $F(x, \theta)$ and the other with $F(x, \theta + \Delta\theta)$, starting both with the arrival of a customer to an empty system, and obtain an exact expression for the limit of the difference $\frac{1}{n} \sum_{i=1}^{n} [T_i(\theta + \Delta\theta) - T_i(\theta)]$ as $n \to \infty$ where $T_i(\theta)$ is the system time of the ith customer in the system with service time distribution $F(x, \theta)$.

Analyzing the limit of that difference we establish the existence of $\frac{d}{d\theta}ET$, obtain for it an expression suitable for estimation from a single simulation experiment, and show that the corresponding Perturbation Analysis (PA) estimates are strongly consistent. Under additional assumptions we also show that the second derivative $\frac{d^2}{d\theta^2}ET$ exists and obtain strongly consistent PA estimates for it as well. Extension of these results to vector parameters leads to efficient PA algorithms for estimating the gradient and Hessian matrix of the response time from a *single* experiment without the use of finite differences. We also present analogous results for parameters of the interarrival time distribution as well as for families of queues parametrized with respect to both the interarrival and the service time distribution. In this latter case we give expressions for the corresponding mixed partial derivatives.

The algorithm for first derivatives was originally given in Suri and Zazanis [29] and its convergence properties were established there for M/G/1 queues. (See also Zazanis, [32]). Since this paper was first submitted for publication, a number of related results have appeared in the literature. Hu [15] established the differentiability of $ET(\theta)$ and the strong consistency of the first derivative estimates under the assumption that the service time distribution has the SSCX property (Strong Stochastic Convexity). This result for the first derivative is also implicit in Glasserman, Hu, and Strickland [9]. Finally, Konstantopoulos and Zazanis [4, 5], and Brémaud and Lasgouttes [3] examine PA algorithms in a stationary and ergodic framework.

Our proof of the differentiability of $ET(\theta)$ is based on direct sample path arguments without relying on restrictive convexity assumptions and our strong consistency results for the corresponding PA estimates are obtained under natural moment conditions. Our sample path approach involving an auxiliary queueing system is specifically tailored to queueing systems and thus allows us to establish the first derivative result under weaker moment conditions compared to more general

[1]Acknowledgements: We would like to thank an anonymous referee for detailed comments. This work was partly supported by the U.S. Office of Naval Research Contracts N00014-75-C-0648 and N00014-79-C-0776, and by N.S.F. Grant ENG82-13680.

GSMP results (Glasserman, Hu, and Strickland [9]). More important, it provides a way of obtaining similar results for higher derivatives which in this paper is exploited for the purpose of establishing the existence of second derivatives and obtaining strongly consistent estimators for them from a single sample path. For a review of PA techniques the reader is referred to Suri [27], Glasserman [7], Ho and Cao [14], and the references therein.

The corresponding finite difference derivative estimates involve *two* or *three* experiments on the system: If $\hat{T}(\theta)$ is an estimate of $ET(\theta)$ obtained from an experiment with parameter value θ, the simplest finite difference estimates are $\hat{T}' = \frac{\hat{T}(\theta+\Delta\theta)-\hat{T}(\theta-\Delta\theta)}{2\Delta\theta}$ and $\hat{T}'' = \frac{\hat{T}(\theta+\Delta\theta)-2\hat{T}(\theta)+\hat{T}(\theta-\Delta\theta)}{\Delta\theta^2}$. The advantage in the number of experiments becomes significant when θ is an N-dimensional vector. Estimating the entries of the $N \times N$ Hessian matrix by means of finite differences would require $2N^2 + 1$ simulation experiments whereas the PA algorithm we propose still requires only one. An equally important, though less obvious, advantage of PA estimates is that they are significantly less noisy than their finite difference counterparts (Zazanis and Suri [31]).

One could also obtain second derivative estimates using Likelihood Ratio (LR) methods (Glynn [10], Reiman and Weiss [20], Rubinstein [23]). However it is noted in Reiman and Weiss [20] that second derivatives estimated by LR methods are likely to be noisy. As the experimental results in part II of this paper indicate, our estimates have surprisingly low variance and thus could be used very effectively for optimization purposes. So far, the first derivative information from PA algorithms has been used to obtain efficient algorithms for multiparameter optimization of the performance of complex discrete event systems (Ho and Cao [13]) including fast optimization during a *single* simulation run (Suri and Zazanis [29], Suri and Leung [28], L'Ecuyer [18]). However, in optimization of deterministic systems, Newton algorithms are known to be superior to algorithms that only use first derivatives. The availability of low variance estimates for the Hessian could make possible the development of stochastic approximation algorithms using this information to achieve improved convergence rates. Finally we point out that Reiman and Simon [21] have proposed an alternative way for estimating second (and higher) derivatives for systems in light traffic driven by a Poisson process.

2 Parametric families of service time distributions and stochastic service functions

Consider a GI/GI/1 queueing system with interarrival distribution $G(x)$ and service time distribution $F(x,\theta)$ depending on a parameter $\theta \in [a,b]$. Let $F^{-1}(u,\theta)$ be defined by

$$F^{-1}(u,\theta) = \inf\{x : F(x,\theta) > u\} \ . \tag{1}$$

Let U be a random variable uniformly distributed in $[0,1]$ and $X(\theta) = F^{-1}(U,\theta)$. In this fashion we have determined a family of random variables indexed by θ, $\{X(\theta) \ ; \ \theta \in [a,b]\}$, satisfying $P(X(\theta) \le x) = F(x,\theta)$ for all $\theta \in [a,b]$. We will call $X(\theta)$ constructed in this fashion a *stochastic service function*. Suppose that the following condition is satisfied:

Condition C.1 *The derivative* $\frac{dX}{d\theta}(\theta) = \lim_{\Delta\theta\to 0} \frac{X(\theta+\Delta\theta)-X(\theta)}{\Delta\theta}$ *exists and is a continuous function of* $\theta \in [a,b]$ *w.p.1.*

Additionally, for the purpose of estimating second derivatives, we require the following

Condition C.2 *The second derivative,* $\frac{d^2X}{d\theta^2}(\theta) = \lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta}[\frac{dX}{d\theta}(\theta+\Delta\theta) - \frac{dX}{d\theta}(\theta)]$, *exists and is continuous for all* $\theta \in [a,b]$ *w.p.1.*

While in general $\{X(\theta); \theta \in [a,b]\}$ may not satisfy **C.1** and **C.2**, simple sufficient conditions to ensure this can be given in terms of $F(x,\theta)$: Suppose that $F(x,\theta)$ is absolutely continuous for all θ with density $f(x,\theta) = D_1 F \stackrel{\text{def}}{=} \frac{\partial F(x,\theta)}{\partial x}$ and that $D_2 F \stackrel{\text{def}}{=} \frac{\partial F}{\partial \theta}$ exists for all $(x,\theta) \in \mathbf{R}^+ \times [a,b]$. Let $\Lambda_\theta = \{\omega : f(X(\theta,\omega),\theta) = 0\}$. Then $P(\Lambda_\theta) = 0$ and from the chain rule together with (1), it follows that $\frac{dX}{d\theta}$ exists on $\Omega \setminus \Lambda_\theta$ and is given by $-\frac{D_2F(X,\theta)}{D_1F(X,\theta)}$. On Λ_θ we can arbitrarily set $\frac{dX}{d\theta} = 0$ and thus we have

$$\frac{dX}{d\theta} = -\frac{D_2F(X,\theta)}{D_1F(X,\theta)} \quad \text{w.p.1.} \tag{2}$$

Similarly, assuming that $F(x,\theta)$ is twice continuously differentiable w.r.t. $(x,\theta) \in \mathbf{R}^+ \times [a,b]$, we have

$$\frac{d^2X}{d\theta^2} = -\frac{D_{22}\,F(D_1F)^2 + D_{11}F\,(D_2F)^2 - 2D_{12}F\,D_1F\,D_2F}{(D_1F)^3} \qquad \text{w.p. } 1, \tag{3}$$

where $D_{11}F \stackrel{\text{def}}{=} \frac{\partial^2 F}{\partial x^2}$ and similarly for the other partial derivatives. The above expressions for $\frac{dX}{d\theta}$ and $\frac{d^2X}{d\theta^2}$ are necessary for the implementation of the PA algorithm we propose in §12. Despite the fact that in general they can be complicated, they assume simple forms for two cases particularly important in applications as shown in the following two examples.

Example 2.1: Suppose that θ is a location parameter of the service time distribution. Then the distribution of $X - \theta$ does not depend on θ and thus, for all ω, $X(\omega,\theta)$ has the form $X(\omega,\theta) = \theta + \zeta_1(\omega)$ which in turn implies that $\frac{dX}{d\theta} = 1$ and $\frac{d^2X}{d\theta^2} = 0$.

Example 2.2: Suppose that θ is a scale parameter, i.e. that the distribution of $\frac{X}{\theta}$ is independent of θ. Then, $X(\omega,\theta)$ has the form $X(\omega,\theta) = \theta\zeta_2(\omega)$, and thus $\frac{dX}{d\theta} = \zeta_2 = \frac{X}{\theta}$ and $\frac{d^2X}{d\theta^2} = 0$.

In the final example, θ is neither a location nor a scale parameter:

Example 2.3: Consider the service time distribution

$$F(x,\theta) = \begin{cases} x(x+\theta) & \text{if} \quad 0 \le x < \frac{1}{2}(\sqrt{4+\theta^2} - \theta) \\ 1 & \text{if} \quad \frac{1}{2}(\sqrt{4+\theta^2} - \theta) < x \;, \end{cases}$$

with $\theta > 0$. From (2) and (3), $\frac{dX}{d\theta} = -\frac{X}{2X+\theta}$ and $\frac{d^2X}{d\theta^2} = \frac{X}{(2X+\theta)^2}$. (Notice that $2X + \theta > 0$ w.p.1.)

3 Assumptions

The following assumptions define the class of systems within which we shall confine ourselves. They are divided into two groups, the second group containing the additional assumptions required for obtaining expressions for second derivatives.

We will assume that the parametric family of service time distributions $F(x,\theta)$ is such that the corresponding stochastic service function $X(\theta)$ satisfies condition **C.1** and furthermore that the following assumptions hold:

Assumption A.1 *Let* $\chi(\omega) = \sup_{\theta\in[a,b]} X(\theta,\omega)$ *for all* ω*. Then* $E\chi < EA < \infty$ *where* A *is a random variable distributed according to the interarrival time distribution* G*.*

Assumption A.2 $E[\chi^2] < \infty$ *and* $E[\xi^3] < \infty$ *with* $\xi(\omega) = \sup_{\theta\in[a,b]} |\frac{dX}{d\theta}(\theta,\omega)|$.

If there exists $\theta^* \in [a,b]$ such that $X(\theta) \le X(\theta^*)$ w.p.1 for all $\theta \in [a,b]$ then assumption **A.1** simply states that the family of queueing systems with service distribution $F(x,\theta)$ are stable for all values on θ in that interval. Similarly $E\chi^2 < \infty$ in **A.2** guarantees that $ET(\theta) < \infty$ for all $\theta \in [a,b]$ (e.g. see Asmussen [1]).

To obtain second derivative estimates, we need additionally condition **C.2** and the following assumptions:

Assumption A.3 *The interarrival distribution* G *is absolutely continuous with density* g *which we will assume right continuous, and has a bounded hazard function,* $\frac{g(x)}{1-G(x)} \le \alpha < \infty$*, for all* $x \in [0,\infty)$*.*

Assumption A.4 *There exists* $\epsilon > 0$ *such that* $E[e^{\epsilon\chi}] < \infty$ *and* $E[e^{\epsilon\xi}] < \infty$*. Furthermore, if* $\psi(\omega) = \sup_{\theta\in[a,b]} |\frac{d^2X}{d\theta^2}(\theta,\omega)|$, $E|\psi|^3 < \infty$.

When **C.1** holds, $X(\theta,\omega)$ and $\frac{dX}{d\theta}(\theta,\omega)$ are w.p.1 continuous functions of θ on $[a,b]$ and hence χ and ξ are well defined random variables. When **C.2** holds as well, then $\frac{d^2X}{d\theta^2}(\theta,\omega)$ is also a continuous function of θ w.p.1 and ψ is well defined.

The following remarks will be useful in the sequel.

Remark 1: Let $\Delta X = X(\theta+\Delta\theta) - X(\theta)$. An immediate consequence of **A.2** is that $|\Delta X| \le \xi|\Delta\theta|$.

Remark 2: In view of remark 1, **A.4** implies that $E[e^{\epsilon\frac{\Delta X}{\Delta\theta}}] < \infty$, and since $X(\theta) \le \chi$, $E[e^{\epsilon X(\theta)}] < \infty$.

Most of the analysis in this paper will be carried out under an additional monotonicity assumption:

Assumption M.1 *For* $\Delta\theta > 0$*,* $\Delta X = X(\theta + \Delta\theta) - X(\theta) \geq 0$ *w.p.1. Furthermore, if* A *is a random variable distributed according to the interarrival distribution,* $EX(b) < EA < \infty$*.*

This monotonicity assumption (which implies the weaker **A.1**) is introduced in order to simplify the sample path analysis that follows. In §8 we show how it can be replaced by **A.1**.

4 Sample path analysis of the system

For $i = 1, 2, \ldots$, let $\Omega_i = [0,1]$, $\mathcal{F}_i = \mathcal{B}_{[0,1]}$ the Borel σ-field on $[0,1]$, and P_i the Lebesgue measure on $[0,1]$. Our probability space will be the product space $(\Omega, \mathcal{F}, P) = (\prod_{i=1}^{\infty} \Omega_i, \prod_{i=1}^{\infty} \mathcal{F}_i, \prod_{i=1}^{\infty} P_i)$. Let $\omega = (\omega_1, \omega_2, \ldots)$ be an element of Ω and $U_i : \Omega \to [0,1]$ the projection mapping $U_i(\omega) = \omega_i$, $i = 1, 2, \ldots$. Thus $\{U_i\}$ is a sequence of i.i.d. random variables, uniformly distributed on $[0,1]$. Define

$$X_i(\theta) = F^{-1}(U_{2i-1}, \theta) \quad \text{and} \quad A_i = G^{-1}(U_{2i}), \quad i = 1, 2, \ldots \,. \tag{4}$$

$X_i(\theta)$ is the service time of the ith customer, C_i, and A_i the interarrival time between C_i and C_{i+1}. Assume that the first customer that arrives to the system, C_1, finds it empty. Denote by $T_i(\theta)$ the system time (waiting plus service time) of the ith customer. Under assumption **A.1** the family of queueing processes defined in this fashion is stable for all $\theta \in [a, b]$ and the corresponding sequence of system times $\{T_i(\theta)\}$ converges in distribution for each θ to a random variable $T(\theta)$.

The analysis in this section is based on considering two sample paths, the *nominal* with parameter value θ, and the *perturbed* with parameter value $\theta + \Delta\theta$. It involves various functionals of these sample paths, such as the number of customers in the ith busy period, the length of the kth idle period, etc. Of course, these functionals depend on θ. Whenever the dependence on θ is dropped this will *always* signify that they are computed from the nominal sample path (with parameter value θ).

Consider the sample path depicted in Fig.1. consisting of m busy periods which we will label BP_1 , $BP_2, \ldots, BP_m$. Let N_k be the number of customers in BP_k, Y_k be the length of BP_k, and I_k be the length of the idle period following BP_k. Let M_k, $k = 0, 1, \ldots$ be the discrete time renewal process defined by $M_k = N_1 + \cdots + N_k$, for $k = 1, 2, \ldots$, and $M_0 = 0$. Thus C_1 initiates BP_1 and, in general, $C_{M_{k-1}+1}$ initiates BP_k. Let us also define

$$S_k(\theta) = \sum_{i=M_{k-1}+1}^{M_k} T_i(\theta) \ \ , \tag{5}$$

and

$$S_k(\theta + \Delta\theta) = \sum_{i=M_{k-1}+1}^{M_k} T_i(\theta + \Delta\theta) \ \ . \tag{6}$$

Notice that we use the same subsequence $\{M_k\}$ *corresponding to parameter value* θ *in both (5) and (6).* Thus, while $S_k(\theta)$ is the area under BP_k in the nominal sample path, $S_k(\theta + \Delta\theta)$ is not necessarily the area under the kth busy period in the perturbed path since busy periods may have coalesced.

It is well known (Asmussen [1, p.182]) that assumption **A.1** implies $EN_1 < \infty$ and furthermore that the renewal process $\{M_k\}$ is aperiodic. Hence,

$$ET(\theta) \; = \; \lim_{m\to\infty} \frac{\sum_{k=1}^{m} S_k(\theta)}{\sum_{k=1}^{m} N_k} \; = \; \lim_{m\to\infty} \frac{\sum_{k=1}^{m} \sum_{i=M_{k-1}+1}^{M_k} T_i(\theta)}{\sum_{k=1}^{m} N_k} \quad \text{w.p.1} \; .$$

Suppose now that the value of θ is increased to $\theta + \Delta\theta \le b$. We also have

$$ET(\theta+\Delta\theta) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} T_i(\theta+\Delta\theta) \; = \; \lim_{m\to\infty} \frac{1}{M_m} \sum_{i=1}^{M_m} T_i(\theta+\Delta\theta) \quad \text{w.p.1} \; , \tag{7}$$

the first equality following from the ergodicity of the system. The second equality in (7) is less obvious since $\{M_k\}$ is a random subsequence not corresponding in general to the regenerative cycles of the system at parameter $\theta+\Delta\theta$. It can be justified using Neuveu's cycle formula (e.g. see Baccelli and Brémaud [2]) and the ergodicity of the system. In Lemma 7 of the Appendix an elementary renewal-theoretic proof of (7) is provided.

If we let

$$\Delta S_k = S_k(\theta+\Delta\theta) - S_k(\theta) \; = \sum_{i=M_{k-1}+1}^{M_k} T_i(\theta+\Delta\theta) - T_i(\theta) \; \ge 0 \; ,$$

we can rewrite (7) as

$$\begin{aligned} ET(\theta+\Delta\theta) \quad &= \quad \lim_{m\to\infty} \frac{\sum_{k=1}^{m} S_k(\theta+\Delta\theta)}{\sum_{k=1}^{m} N_k} \; = \; \lim_{m\to\infty} \frac{\sum_{k=1}^{m} S_k(\theta)}{\sum_{k=1}^{m} N_k} \; + \; \lim_{m\to\infty} \frac{\sum_{k=1}^{m} \Delta S_k}{\sum_{k=1}^{m} N_k} \\ &= \quad ET(\theta) \; + \; \lim_{m\to\infty} \frac{\sum_{k=1}^{m} \Delta S_k}{\sum_{k=1}^{m} N_k} \quad \text{w.p.1} \; . \end{aligned} \tag{8}$$

Now we calculate the change in the total system time for all customers in these m busy periods when the value of the parameter θ changes to $\theta + \Delta\theta$. To this effect consider customer C_i who belongs to the kth busy period, BP_k, (i.e. $M_{k-1} < i \le M_k$). When the value of θ is increased to $\theta + \Delta\theta$ there are in general three sources of delay for C_i:

(i) The delay $\Delta X_i = X_i(\theta+\Delta\theta) - X_i(\theta)$ in C_i's own service time.

(ii) The delay caused by changes in the service times of customers $C_{M_{k-1}+1}, \ldots, C_{i-1}$ i.e. of all the preceding customers belonging to the same busy period when the parameter value is θ.

(iii) A delay arising from the possibility that accumulated perturbations, introduced in BP_{k-1} and the busy periods prior to it, may cause it to coalesce with BP_k. In this case, $C_{M_{k-1}+1}$ will not find the system empty when the value of the parameter is increased to $\theta + \Delta\theta$ and as a result all the customers belonging to BP_k will have to wait for an additional amount of time. It will be useful to think of **(i)** and **(ii)** as the *local* effects of the perturbations introduced in the sample path while **(iii)** represents the *global* effects. We proceed to examine these effects next.

The effect of the first two sources of delay in the system time of C_i as described above is easily seen to be a change in $T_i(\theta)$ equal to $\Delta X_{M_{k-1}+1} + \Delta X_{M_{k-1}+2} + \cdots + \Delta X_i$. Next, let us evaluate the effect of the third source (which for customers in BP_1 is of course 0). As a result of the perturbations in the service times of the customers in BP_1, its length is increased by

$$\Delta Y_1 \;=\; \Delta X_1 + \cdots + \Delta X_{N_1} \;\; .$$

Obviously, as long as $\Delta Y_1 < I_1$, this has no effect on the system time of the customers in the next busy period. However, if $\Delta Y_1 \geq I_1$, then the two busy periods coalesce and as a result the system time of every customer in BP_2 is increased by $\Delta Y_1 - I_1$. Using the notation $x^+ = \max(0, x)$ for the positive part of a real number x, the effect of coalescence of the two busy periods on $T_i(\theta)$, $M_1 < i \leq M_2$, is given by $(\Delta Y_1 - I_1)^+$. Thus

$$\Delta T_i \;=\; T_i(\theta + \Delta\theta) - T_i(\theta) \;\;=\;\; (\Delta Y_1 - I_1)^+ \;\;+\;\; \sum_{j=M_1+1}^{i} \Delta X_j \quad \text{for} \quad M_1 < i \leq M_2 \;\; . \qquad (9)$$

Summing (9) from $i = M_1 + 1$ to M_2 we get the total change in the system time of all the customers of BP_2 as

$$\Delta S_2 \;\;=\;\; N_2 \, (\Delta Y_1 - I_1)^+ \;+\; \sum_{i=M_1+1}^{M_2} \sum_{j=M_1+1}^{i} \Delta X_j \;\; .$$

There is only one way for perturbations introduced in BP_k to cause a delay in customers of BP_l (with $k < l$), namely that the intervening $l - k$ idle periods disappear that all the $l - k + 1$ busy periods coalesce into a single busy period as a result of these perturbations. To be specific, consider Fig.1 which depicts a sample path consisting of three busy periods, BP_1, BP_2, BP_3. Let

$$\Delta Y_k \;=\; \sum_{i=M_{k-1}+1}^{M_k} \Delta X_i \; . \tag{10}$$

BP_1 and BP_2 will coalesce if and only if $\Delta Y_1 \geq I_1$. It is also clear that BP_2 and BP_3 will coalesce and BP_1 and BP_2 will not if and only if $\Delta Y_2 \geq I_2$ and $\Delta Y_1 < I_1$. Finally, all three will coalesce into one busy period if and only if $\Delta Y_1 \geq I_1$ and $\Delta Y_2 + \Delta Y_1 \geq I_2 + I_1$. Hence, the delay for the customers of BP_3 caused by perturbations in BP_2 and BP_1 is given by

$$\max(0,\ \Delta Y_2 - I_2,\ \Delta Y_2 + \Delta Y_1 - I_2 - I_1) \;\; .$$

Generally, for k busy periods, we can easily check that the effect of perturbations introduced in all the busy periods prior to BP_k on the customers of BP_k is $V_0 \equiv 0$ when $k = 1$, and

$$\begin{aligned} V_{k-1} \;&=\; \max(0,\ \Delta Y_{k-1} - I_{k-1},\ \Delta Y_{k-1} + \Delta Y_{k-2} - I_{k-1} - I_{k-2},\ \ldots, \\ &\qquad \Delta Y_{k-1} + \Delta Y_{k-2} + \cdots + \Delta Y_1 \; - \; I_{k-1} - I_{k-2} - \cdots - I_1) \; , \quad k \geq 2 \; . \end{aligned} \tag{11}$$

Hence, in general, the change in system time for C_i is given by

$$\Delta T_i \;=\; V_{k-1} \;+ \sum_{j=M_{k-1}+1}^{i} \Delta X_j \; , \qquad M_{k-1} < i \leq M_k \; , \tag{12}$$

and consequently

$$\Delta S_k \;=\; \sum_{i=M_{k-1}+1}^{M_k} \Delta T_i \;=\; N_k V_{k-1} \;+\; \sum_{i=M_{k-1}+1}^{M_k} \; \sum_{j=M_{k-1}+1}^{i} \Delta X_j \;\; . \tag{13}$$

Notice that (11) implies that V_{k-1}, $k = 1, 2, \ldots$, can be interpreted as the waiting time of the kth customer in an *auxiliary* queueing system in which $\Delta Y_1, \Delta Y_2, \ldots$ is the service time sequence and $I_1, I_2, \ldots$ is the interarrival time sequence. In general I_j depends on ΔY_j unless the arrival process to the original system is Poisson. From (8) and (13) we have then

$$\begin{aligned} ET(\theta + \Delta\theta) - ET(\theta) \;&=\; \lim_{m\to\infty} \frac{1}{\sum_{k=1}^m N_k} \sum_{k=1}^{m} \; \sum_{i=M_{k-1}+1}^{M_k} \; \sum_{j=M_{k-1}+1}^{i} \Delta X_j \\ &\quad+\; \lim_{m\to\infty} \frac{\sum_{k=1}^{m-1} N_{k+1} V_k}{\sum_{k=1}^m N_k} \qquad \text{w.p.1} \; . \end{aligned} \tag{14}$$

From the strong law of large numbers the first term in the right hand side (rhs) of (14) is equal to

$$\frac{1}{EN_1}\, E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\Delta X_j\ ] \quad \text{w.p.1.}$$

The evaluation of the second limit on the rhs of (14) is not as straightforward since the sequence $\{V_k N_{k+1}; k = 1, 2, \ldots\}$, is not regenerative w.r.t. the renewal process $\{M_k\}$. This second term which incorporates the global ("non-infinitesimal" or nonlinear) effects of the perturbations introduced in the sample path is considered in the next section.

5 Global effects of perturbations and the auxiliary queueing system

The condition for the stability of the auxiliary queueing system is $E[\Delta Y_1 - I_1] < 0$. It is easy to show that the auxiliary system is always stable when the original system, with service time distribution $F(x, \theta + \Delta\theta)$, is stable. Indeed, using Wald's lemma repeatedly,

$$\begin{aligned} E\Delta Y_1 - EI_1 &= EN_1(\theta)[EX_1(\theta+\Delta\theta) - EX_1(\theta)] - EN_1(\theta)[EA_1 - EX_1(\theta)] \\ &= EN_1(\theta)[EX_1(\theta+\Delta\theta) - EA_1] \ < \ 0 \ . \end{aligned} \tag{15}$$

The last inequality follows from the fact that $a \le \theta + \Delta\theta \le b$ and **A.1** which postulates that $EX_1(\theta) \le EA_1$ for all $\theta \in [a, b]$. From (15) follows that the auxiliary queueing system is stable, i.e. that

$$V(\theta, \Delta\theta) \ \stackrel{\text{def}}{=} \ \sup(0,\ \Delta Y_1 - I_1,\ \Delta Y_1 + \Delta Y_2 - I_1 - I_2,\ \ldots) \tag{16}$$

is finite w.p. 1. (In the sequel, where no confusion arises, the dependence of V on θ, $\Delta\theta$, will not be made explicit.) Let

$$\Phi_i = \sum_{j=M_{i-1}+1}^{M_i} \xi_j \ , \qquad i = 1, 2, \ldots, \tag{17}$$

where $\xi_j = \sup_{[a,b]} \frac{dX_j}{d\theta}$. Then

$$\Delta Y_i \le \Phi_i \Delta\theta \ . \tag{18}$$

Since $E[\Phi_1^2] \leq \infty$ (see Lemma 9 in the Appendix), $EV < \infty$ (Asmussen [1]).

Write

$$\sum_{k=1}^{m-1} V_k N_{k+1} = \sum_{k=1}^{m-1} V_k E[N_1] + \sum_{k=1}^{m-1} V_k (N_{k+1} - E[N_1]) \ . \tag{19}$$

Notice first that

$$\lim_{m\to\infty} \frac{1}{m-1} \sum_{k=1}^{m-1} V_k E[N_1] = E[V]\, E[N_1] < \infty \quad \text{w.p.1,} \tag{20}$$

where V is the r.v. defined in (16), because of the ergodicity of the auxiliary system and the finiteness of moments assumption **A.2**. On the other hand, as it is shown in Lemma 12 of the Appendix, we can use a martingale stability theorem to prove that

$$\lim_{m\to\infty} \frac{1}{m-1} \sum_{k=1}^{m-1} V_k (N_{k+1} - E[N_1]) = 0 \quad \text{w.p.1.} \tag{21}$$

From (19), (20) and (21) follows that

$$\lim_{m\to\infty} \frac{\sum_{k=1}^{m-1} V_k N_{k+1}}{\sum_{k=1}^{m} N_k} = E[V] \quad \text{w.p.1.} \tag{22}$$

Combining (14) and (22) gives

$$ET(\theta + \Delta\theta) - ET(\theta) = \frac{1}{E[N_1]} E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} \Delta X_j] + E[V] \ . \tag{23}$$

In (23) we established that the contribution of the global effects of perturbations is given by the term $E[V]$. Now we state three lemmas that characterize the behavior of $E[V]$ as $\Delta\theta \downarrow 0$.

Lemma 1 *When the original system satisfies* **C.1, M.1,** *and* **A.2***, for the auxiliary system defined above,* $\lim_{\Delta\theta\downarrow 0} \frac{1}{\Delta\theta} E[V] = 0$.

In view of this lemma and (23) we do not expect $E[V]$ to contribute to the value of the right derivative $D^+ET(\theta)$.

Proof: We need to establish that

$$\lim_{\Delta\theta \downarrow 0} E[\sup(0, \frac{\Delta Y_1}{\Delta\theta} - \frac{I_1}{\Delta\theta}, \frac{\Delta Y_1}{\Delta\theta} + \frac{\Delta Y_2}{\Delta\theta} - \frac{I_1}{\Delta\theta} - \frac{I_2}{\Delta\theta}, \ldots)] = 0 .$$

From (18) we have $\frac{\Delta Y_i}{\Delta\theta} \leq \Phi_i$ and thus

$$0 \leq E[\sup(0, \frac{\Delta Y_1}{\Delta\theta} - \frac{I_1}{\Delta\theta}, \ldots)] \leq E[\sup(0, \Phi_1 - \frac{I_1}{\Delta\theta}, \ldots)] . \tag{24}$$

Since $E[\Phi_1^2] < \infty$ (see Lemma 9 in the Appendix), the expectation on the last term of (24) is finite (Asmussen [1]). Letting $\Delta\theta \downarrow 0$, the last term of the above inequality goes to 0 by monotone convergence and the lemma follows. □

The behavior of $E[V]$ as $\Delta\theta \downarrow 0$ is characterized more precisely in the following

Lemma 2 *Under assumptions* **C.1, M.1, A.2, A.3,** *and* **A.4,** *for sufficiently small* $\Delta\theta > 0$, *there is a positive* $L < \infty$ *such that*

$$0 < E[V] - E[(\Delta Y_1 - I_1)^+] \leq L\Delta\theta^3 .$$

This suggests that, for the purpose of establishing the existence of $\frac{d^2}{d\theta^2}ET(\theta)$ and obtaining an expression for it, we can substitute $E[V]$ with $E(\Delta Y_1 - I_1)^+$, a considerably more tractable expression.

Proof: Let $K^*(\gamma, \Delta\theta) = E[\exp\{\gamma(\frac{\Delta Y_1}{\Delta\theta} - \frac{I_1}{\Delta\theta})\}]$. Let Z_1 be the age of the arrival process at the end of the first busy period. Then

$$K^*(\gamma, \Delta\theta) = E[e^{\gamma \frac{\Delta Y_1}{\Delta\theta}} E[e^{-\gamma \frac{I_1}{\Delta\theta}} \mid Z_1]] , \tag{25}$$

since ΔY_1 and I_1 are conditionally independent given Z_1. We next show that

$$E[\exp\{-\gamma \frac{I_1}{\Delta\theta}\} \mid Z_1] \leq \frac{\alpha}{\gamma/\Delta\theta + \alpha} = \frac{\Delta\theta\alpha}{\gamma + \Delta\theta\alpha} . \tag{26}$$

Since the conditional distribution of I_1 given Z_1 has density $\frac{g(Z_1+x)}{1-G(Z_1)}$, the conditional expectation on the lhs of (26) is $\int_0^\infty e^{-x\gamma/\Delta\theta} \frac{g(Z_1+x)}{1-G(Z_1)} dx$ which, after integration by parts, gives $1 -$

$\int_0^\infty \frac{\gamma}{\Delta\theta} e^{-x\gamma/\Delta\theta} \frac{1-G(Z_1+x)}{1-G(Z_1)} dx$. This, together with the inequality $\frac{1-G(Z_1+x)}{1-G(Z_1)} = \exp(-\int_{Z_1}^{Z_1+x} \frac{g(\xi)}{1-G(\xi)} d\xi) \geq \exp(-\int_{Z_1}^{Z_1+x} \alpha d\xi = \exp(-\alpha x)$, gives inequality (26). Combining (25) and (26) we obtain

$$K^*(\gamma, \Delta\theta) \ \leq \ E[\exp\{\gamma \frac{\Delta Y_1}{\Delta\theta}\}] \frac{\Delta\theta\alpha}{\gamma + \Delta\theta\alpha} \ .$$

In the Appendix (Lemma 11) we show that there exists $\epsilon > 0$ such that $E[\exp\{\gamma \frac{\Delta Y_1}{\Delta\theta}\}] \ \leq \ \overline{K}$ for $0 < \gamma < \epsilon$. Hence, for $0 < \gamma < \epsilon$

$$K^*(\gamma, \Delta\theta) \ \leq \ \overline{K} \frac{\Delta\theta\alpha}{\gamma + \Delta\theta\alpha}$$

From this follows that $K^*(\gamma, \Delta\theta) < 1/2$ for $0 < \gamma < \epsilon$ and $\Delta\theta < \gamma/(2\alpha\overline{K})$. But then from Kingman's inequality (Kingman [16]),

$$0 \ \leq \ E[\sup(0\,, \ \frac{\Delta Y_1}{\Delta\theta} - \frac{I_1}{\Delta\theta}, \ldots)] \ - \ E(\frac{\Delta Y_1}{\Delta\theta} - \frac{I_1}{\Delta\theta})^+ \ \leq \ \frac{[K^*(\gamma, \Delta\theta)]^2}{2e\gamma[1 - K^*(\gamma, \Delta\theta)]} \ .$$

Fix $\gamma < \epsilon$ and let $\Delta\theta < \gamma/(2\alpha\overline{K})$. Then,

$$0 \ \leq \ \frac{1}{\Delta\theta} EV \ - \ E(\frac{\Delta Y_1}{\Delta\theta} - \frac{I_1}{\Delta\theta})^+ \ < \ \frac{\overline{K}^2 \alpha^2}{e\gamma^3} \Delta\theta^2 \ . \tag{27}$$

Multiplying (27) by $\Delta\theta$ we obtain the inequality of Lemma 2 with $L = \frac{\overline{K}^2\alpha^2}{e\gamma^3}$. □

Finally, we examine the limiting behavior of $E(\Delta Y_1 - I_1)^+$ when $\Delta\theta \downarrow 0$.

Lemma 3 *Let Z_1 be the age of the arrival process at the end of the busy period. Under* **C.1, M.1, A.2,** *and* **A.3,**

$$\lim_{\Delta\theta \downarrow 0} \frac{1}{\Delta\theta^2} E(\Delta Y_1 - I_1)^+ = \frac{1}{2} E[\frac{g(Z_1)}{1 - G(Z_1)} \ (\sum_{i=1}^{N_1} \frac{dX_i}{d\theta})^2] \ . \tag{28}$$

Taken together, Lemmas 2 and 3 suggest that the contribution of $E[V]$ to the second derivative of ET is given by (28).

Proof: We start by conditioning on Z_1 and ΔY_1. Then, w.p. 1,

$$E(\Delta Y_1 - I_1)^+ \ = \ E[\ E[(\Delta Y_1 - I_1)^+ | \Delta Y_1, Z_1]\ ] \ , \tag{29}$$

where Z_1 is the time that has elapsed since the last arrival, at the instant when BP_1 ends (which in a FCFS system coincides with the system time of the last customer in the busy period). Since the interarrival time distribution is absolutely continuous with density $g(\cdot)$ (by Assumption **A.3**), the conditional density of the length of the idle period I_1 given Z_1 is $\frac{g(Z_1+x)}{1-G(Z_1)}$ for $x \geq 0$ and the conditional expectation in (29) can be written as

$$\begin{aligned} E[(\Delta Y_1 - I_1)^+ | \Delta Y_1, Z_1] &= \int_0^\infty (\Delta Y_1 - x)^+ \, \tfrac{g(Z_1+x)}{1-G(Z_1)} \, dx \quad \text{w.p. } 1 \\ &= \Delta\theta^2 \int_0^\infty (\tfrac{\Delta Y_1}{\Delta\theta} - y)^+ \, \tfrac{g(Z_1+y\Delta\theta)}{1-G(Z_1)} \, dy \ , \end{aligned} \tag{30}$$

where in the second integral we have made the change of variables $x = y\Delta\theta$. Notice that $\frac{g(Z_1+x)}{1-G(Z_1)} = \frac{g(Z_1+x)}{1-G(Z_1+x)} \frac{1-G(Z_1+x)}{1-G(Z_1)} \leq \alpha$ since by **A.3** the interarrival distribution has hazard rate bounded above by α. Hence the quantity inside the integral in the last term of (30) is dominated by $\alpha(\frac{\Delta Y_1}{\Delta\theta} - y)^+$. Since

$$E[\int_0^\infty (\frac{\Delta Y_1}{\Delta\theta} - y)^+ \, \alpha dy \,] \ = \ \frac{\alpha}{2} E(\frac{\Delta Y_1}{\Delta\theta})^2 \ < \ \frac{\alpha}{2} E\Phi_1^2 \ < \ \infty \ ,$$

(the last inequality following from $\frac{\Delta Y_1}{\Delta\theta} \leq \Phi_1$ and Lemma 9 in the Appendix) we can apply the Dominated Convergence Theorem to obtain

$$\begin{aligned} \lim_{\Delta\theta\downarrow 0} E[\, \textstyle\int_0^\infty (\frac{\Delta Y_1}{\Delta\theta} - y)^+ \, \frac{g(Z_1+y\Delta\theta)}{1-G(Z_1)} dy \,] \ &= \ E[\, \textstyle\int_0^\infty \lim_{\Delta\theta\downarrow 0} (\frac{\Delta Y_1}{\Delta\theta} - y)^+ \, \frac{g(Z_1+y\Delta\theta)}{1-G(Z_1)} dy \,] \\ = \ E[\, \tfrac{g(Z_1)}{1-G(Z_1)} \textstyle\int_0^\infty (\frac{dY_1}{d\theta} - y)^+ dy \,] \ &= \ \tfrac{1}{2} E[\, \tfrac{g(Z_1)}{1-G(Z_1)} \, \left(\textstyle\sum_{i=1}^{N_1} \frac{dX_i}{d\theta} \right)^2 \,] \ . \end{aligned} \tag{31}$$

From (29) and (31) Lemma 3 follows. □

6 Differentiability and expressions for the steady state derivatives

In this section we establish the existence of $\frac{d}{d\theta}ET$ and (under additional conditions) of $\frac{d^2}{d\theta^2}ET$ and obtain expressions for them in terms of expectations that can be estimated from a single simulation run. As a first step we show that, under assumptions **M.1** and **A.2**, $ET(\theta)$ is continuous on $[a, b]$.

The proof of differentiability, presented in Theorem 2, is based on the results of the sample path analysis of §4 and §5 which relied on the monotonicity assumption **M.1** and the positivity of the increment $\Delta\theta$. We show that, under the same assumptions, the right derivative $D^+ET(\theta)$ exists, obtain an expression for it, and show that it is a continuous function of θ. Since a continuous function with a continuous right derivative must be differentiable (a special case of Lemma 14) the differentiability of $ET(\theta)$ is established. This indirect argument is necessary in order to avoid a complicated sample path analysis involving negative perturbations. Second derivatives are dealt with similarly.

Throughout this section the monotonicity assumption **M.1** is maintained. It will be relaxed in §8.

Theorem 1 *For a GI/GI/1 queue satisfying* **C.1, M.1,** *and* **A.2***, the expected system time in steady state, $ET(\theta)$, is continuous on $[a, b]$.*

Proof: Under **C.1, A.1, A.2,** and **M.1**, we have shown that

$$ET(\theta+\Delta\theta)-ET(\theta) = \frac{1}{E[N_1(\theta)]}E[\sum_{i=1}^{N_1(\theta)}\sum_{j=1}^{i}X_j(\theta+\Delta\theta)-X_j(\theta)]+EV(\theta,\Delta\theta) \quad \text{for } \Delta\theta \geq 0\ . \tag{32}$$

(This is equation 23.) Combining this with inequality (97) established in Lemma 14 we obtain for all $\Delta\theta$, positive or negative, such that $|\Delta\theta| \leq \min(\theta - a, b - \theta)$,

$$|ET(\theta + \Delta\theta) - ET(\theta)| \leq \frac{1}{E[N_1(\theta)]}E[\sum_{i=1}^{N_1(\theta)}\sum_{j=1}^{i}|X_j(\theta + \Delta\theta) - X_j(\theta)|] + EV(\theta, |\Delta\theta|)\ . \tag{33}$$

As an immediate consequence of **A.2** (see Remark 1) $|X_j(\theta+\Delta\theta) - X_j(\theta)| \leq \xi_j |\Delta\theta|$. Furthermore,

$$E[\sum_{i=1}^{N_1}\sum_{j=1}^{i} \xi_j] \quad \leq \quad E[N_1 \sum_{i=1}^{N_1} \xi_i] \ \leq (E\Phi_1^2)^{1/2}(EN_1^2)^{1/2} \quad < \quad \infty \tag{34}$$

with $\Phi_1 = \sum_{i=1}^{N_1} \xi_i$. The first inequality above holds because $\xi_j > 0$ w.p.1, the second follows from the Cauchy-Schwartz inequality, while the last is a result of Lemmas 8 and 9. Hence

$$|ET(\theta+\Delta\theta) - ET(\theta)| \leq \frac{|\Delta\theta|}{E[N_1(\theta)]} E[\sum_{i=1}^{N_1(\theta)} \sum_{j=1}^{i} \xi_j] + EV(\theta, |\Delta\theta|) \ . \tag{35}$$

Letting $|\Delta\theta| \to 0$, the theorem follows immediately from Lemma 1. □

Theorem 2 *For a GI/GI/1 queue satisfying* **C.1, M.1,** *and* **A.2**, *the expected system time in steady state, $ET(\theta)$, is continuously differentiable for $\theta \in [a,b]$ with*

$$\frac{d}{d\theta} ET(\theta) \quad = \quad \frac{E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} \frac{dX_j}{d\theta}]}{E[N_1]} \ . \tag{36}$$

Proof: The outline of the proof is as follows:

(i) We first show that $ET(\theta)$ is a right differentiable function of θ on $[a,b)$.

(ii) We next show that the right derivative $D^+ET(\theta)$ is *continuous* on $[a,b)$ (and hence Riemann-integrable).

(iii) In the final step of the proof, we use the fact that a continuous function with Riemann-integrable right derivative must be continuously differentiable.

(i) Right differentiability of $ET(\theta)$: Dividing both sides of (32) by $\Delta\theta$ we obtain

$$\frac{1}{\Delta\theta}[ET(\theta+\Delta\theta) \ - \ ET(\theta)] \quad = \quad \frac{1}{E[N_1]} \ E[\sum_{i=1}^{N_1}\sum_{j=1}^{i} \frac{\Delta X_j}{\Delta\theta} \ ] \ + \ \frac{1}{\Delta\theta} E[V] \ . \tag{37}$$

Arguing as in the proof of Theorem 1, the expression inside the expectation in the first term on the rhs of (37) is bounded by

$$\sum_{i=1}^{N_1}\sum_{j=1}^{i} \frac{\Delta X_j}{\Delta\theta} \quad \leq \quad \sum_{i=1}^{N_1}\sum_{j=1}^{i} \xi_j \ . \tag{38}$$

In (34) we have shown that the rhs of the above inequality has finite expectation. We can therefore use the Dominated Convergence Theorem to obtain

$$\lim_{\Delta\theta\to 0} E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{\Delta X_j}{\Delta\theta}] \;=\; E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{dX_j}{d\theta}]\;. \tag{39}$$

On the other hand, in view of Lemma 1, as $\Delta\theta \downarrow 0$ the second term in the rhs of (37) converges to zero. Hence it follows that $ET(\theta)$ has a right derivative w.r.t. θ, $D^+ET \stackrel{\text{def}}{=} \lim_{\Delta\theta\downarrow 0}\frac{1}{\Delta\theta}[ET(\theta + \Delta\theta) - ET(\theta)]$, given by

$$D^+ET \;=\; \frac{E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{dX_j}{d\theta}]}{E[N_1]}\;. \tag{40}$$

Notice that to establish (40) we made no use of **C.2**, **A.3**, or **A.4**. We also note that the above dominated convergence argument insures that $D^+ET(\theta) < \infty$.

(ii) Continuity of $D^+ET(\theta)$: Since $N_1(\theta) \geq 1$ w.p.1 for all $\theta \in [a, b]$, we need only show that the numerator and the denominator on the rhs of (40) are continuous functions of θ. We start with the observation that,

$$\lim_{\delta\to 0} N_1(\theta+\delta) = N_1(\theta) \;\text{ w.p.1,} \tag{41}$$

and

$$\lim_{\delta\to 0}\sum_{i=1}^{N_1(\theta+\delta)}\sum_{j=1}^{i}\frac{dX}{d\theta}(\theta+\delta) = \sum_{i=1}^{N_1(\theta)}\sum_{j=1}^{i}\frac{dX}{d\theta}(\theta)\;\text{ w.p.1,} \tag{42}$$

where in (42) we have also used **C.1**. The above equations hold both for positive and negative δ. Once again we will use the Dominated Convergence Theorem to show that (41) and (42) imply

$$\lim_{\delta\to 0} E[N_1(\theta+\delta)] = E[N_1(\theta)]\;\;, \tag{43}$$

and

$$\lim_{\delta\to 0} E[\sum_{i=1}^{N_1(\theta+\delta)}\sum_{j=1}^{i}\frac{dX}{d\theta}(\theta+\delta)] = E\sum_{i=1}^{N_1(\theta)}\sum_{j=1}^{i}\frac{dX}{d\theta}(\theta)] \tag{44}$$

respectively, and hence establish the continuity of $D^+ET(\theta)$. Indeed, for $|\delta| < \min(\theta - a, b - \theta)$,

$$N_1(\theta + \delta) \le N(b) \ \text{ w.p.1,} \tag{45}$$

and $EN_1(b) < \infty$ as a consequence of **M.1.**

To obtain a dominating random variable for the lhs of (42) we use the fact that $|\frac{dX_i}{d\theta}(\theta)| \le \xi_i$ w.p.1 for all $\theta \in [a, b]$ (assumption **A.2**) together with (45) to get

$$| \sum_{i=1}^{N_1(\theta+\delta)} \sum_{j=1}^{i} \frac{dX_j}{d\theta}(\theta + \delta) \, | \ \ \le \ \ N_1(b) \sum_{i=1}^{N(b)} \xi_i \ . \tag{46}$$

Using the Cauchy-Schwartz inequality we obtain

$$E\left[N_1(b) \textstyle\sum_{i=1}^{N_1(b)} \xi_i\right] \ \le \ \left(EN_1^2(b)\right)^{1/2} \left(E\left[\textstyle\sum_{i=1}^{N_1(b)} \xi_i\right]^2\right)^{1/2} \ < \ \infty, \tag{47}$$

the last inequality following from Lemmas 8 and 9 of the Appendix.

Thus we have established the continuity of $D^+ET(\theta)$ on $[a, b)$. We also note that the above inequality gives a bound for the rhs of (40) for all $\theta \in [a, b]$. $D^+ET(\theta)$ can be defined at b by continuity.

(iii) Differentiability of $ET(\theta)$: Since $D^+ET(\theta)$ exists and is finite and continuous on $[a, b)$, at least two of the Dini derivatives, $\limsup_{h\downarrow 0} \frac{ET(\theta+h)-ET(\theta)}{h}$, and $\liminf_{h\downarrow 0} \frac{ET(\theta+h)-ET(\theta)}{h}$, are finite and both equal to $D^+ET(\theta)$. Hence they are continuous and bounded on $[a, b)$ and therefore Riemann-integrable. This allows us to use Lemma 14 to infer that $\frac{d}{d\theta}ET$ exists and is equal to D^+ET for all $\theta \in (a, b)$. □

When θ is a location parameter of the service time distribution, $\frac{dX}{d\theta} = 1$ and (40) becomes

$$\frac{dET}{d\theta} \ = \ \frac{E[N_1^2] \ + \ E[N_1]}{2E[N_1]} \ . \tag{48}$$

In the case of a scale parameter, $\frac{dX}{d\theta} = \frac{X}{\theta}$ and the corresponding expression is

$$\frac{dET}{d\theta} \ = \ \frac{E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} X_j]}{\theta E[N_1]} \ . \tag{49}$$

In view of examples 1 and 2, in both of the above cases, **C.1** is satisfied automatically. Also, assumption **A.2** is reduced to a simple moment condition: $E[X(b)^2] < \infty$ or $E[X(b)^3] < \infty$ for a location or scale parameter respectively. In applications, location and scale parameters arise naturally and the simplicity of (48) and (49) illustrate the ease of implementation of the corresponding perturbation analysis algorithms.

Theorem 3 *For a GI/GI/1 queue satisfying* **C.1, C.2**, *and assumptions* **M.1, A.2, A.3** *and* **A.4**, *the expected system time in steady state $ET(\theta)$ is twice continuously differentiable for $\theta \in [a,b]$ with second derivative given by*

$$\frac{d^2}{d\theta^2} ET(\theta) \quad = \quad \frac{E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} \frac{d^2 X_j}{d\theta^2}]}{E[N_1]} \quad + \quad E[\, \frac{g(Z_1)}{1-G(Z_1)} (\sum_{i=1}^{N_1} \frac{dX_i}{d\theta})^2] \ , \tag{50}$$

where Z_1 is the age of the arrival process at the end of the busy period.

Proof: From (23), and Lemmas 2, and 3 it follows that

$$ET(\theta+\Delta\theta) - ET(\theta) \ = \ \frac{E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} \Delta X_j]}{E[N_1]} + \frac{1}{2} E[\frac{g(Z_1)}{1-G(Z_1)} (\sum_{i=1}^{N_1} \frac{dX_i}{d\theta})^2] \Delta\theta^2 \ + \ \mathrm{o}(\Delta\theta^2) \ . \tag{51}$$

Since $\frac{d^2 X}{d\theta^2}(\theta)$ is assumed to be continuous in $[a,b]$ (Condition **C.2**), Taylor's theorem gives $\Delta X_j = \Delta\theta \frac{dX_j}{d\theta}(\theta) + \frac{1}{2}\Delta\theta^2 \frac{d^2 X_j}{d\theta^2}(\theta + \beta_j \Delta\theta)$ with $\beta_j \in [0,1]$ (depending in general on ω, θ, and $\Delta\theta$) and hence,

$$E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} (\frac{\Delta X_j}{\Delta\theta} - \frac{dX_j}{d\theta})] \quad = \quad \frac{\Delta\theta}{2} \ E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} \frac{d^2 X_j}{d\theta^2} (\theta + \beta_j \Delta\theta)] \ . \tag{52}$$

Arguing as in the proof of Theorem 2, the quantity inside the expectation on the rhs of (51) is dominated by $\sum_{i=1}^{N_1} \sum_{j=1}^{i} \psi_j$ (because of **A.4**) which in turn is dominated by $N_1 \sum_{i=1}^{N_1} |\psi_i|$. Hence

$$E\left[\sum_{i=1}^{N_1} \sum_{j=1}^{i} \psi_j\right] \le E[N_1 \sum_{i=1}^{N_1} |\psi_i| \,] \ \le \ (EN_1^2)^{1/2} \left(E\left[\sum_{i=1}^{N_1} |\psi_i|\right]^2 \right)^{1/2} \ < \infty \ ,$$

the second inequality following from Cauchy-Schwartz, while the last following from Lemmas 8 and 9 of the Appendix. From (52), the Dominated Convergence Theorem, and the continuity of $\frac{d^2X}{d\theta^2}(\theta)$ we obtain

$$\lim_{\Delta\theta\downarrow 0}\frac{1}{\Delta\theta}E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{\Delta X_j}{\Delta\theta} - \frac{dX_j}{d\theta}] = \lim_{\Delta\theta\downarrow 0}\frac{1}{2}E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{d^2X_j}{d\theta^2}(\theta+\beta_j\Delta\theta)] = \frac{1}{2}E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{d^2X_j}{d\theta^2}] . \quad (53)$$

This allows us to write (51) as

$$\begin{aligned} ET(\theta+\Delta\theta) - ET(\theta) &= \tfrac{1}{E[N_1]}E[\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{dX_j}{d\theta}]\Delta\theta \qquad (54)\\ &+ \tfrac{1}{2}\Big(E[\tfrac{g(Z_1)}{1-G(Z_1)}(\textstyle\sum_{i=1}^{N_1}\frac{dX_i}{d\theta})^2] + \frac{1}{E[N_1]}E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{d^2X_j}{d\theta^2}]\Big)\Delta\theta^2 + \mathrm{o}(\Delta\theta^2) . \end{aligned}$$

To simplify the notation let $h(\theta) = \frac{d}{d\theta}ET(\theta)$ and $q(\theta)$ be the quantity inside the parenthesis in the second term of the above expansion. Using a dominated convergence argument similar to that in the proof of Theorem 2, we can easily show that $q(\theta)$ is continuous. We also have

$$ET(\theta+2\Delta\theta) - ET(\theta) = h(\theta)2\Delta\theta + 2q(\theta)\Delta\theta^2 + o(\Delta\theta^2), \quad (55)$$

$$ET(\theta+2\Delta\theta) - ET(\theta+\Delta\theta) = h(\theta+\Delta\theta)\Delta\theta + \frac{1}{2}q(\theta)\Delta\theta^2 + o(\Delta\theta^2). \quad (56)$$

Subtracting (54) from (55) we obtain $\lim_{\Delta\theta\downarrow 0}\frac{ET(\theta+2\Delta\theta)-2ET(\theta+\Delta\theta)+ET(\theta)}{\Delta\theta^2} = q(\theta)$ while subtracting (54) from (56), we see that $\frac{1}{\Delta\theta}[g(\theta+\Delta\theta) - g(\theta)] = \frac{1}{\Delta\theta^2}[ET(\theta+2\Delta\theta) - 2ET(\theta+\Delta\theta) + ET(\theta)] + \frac{1}{2}[q(\theta+\Delta\theta) - q(\theta)] + o(1)$. Letting $\Delta\theta \downarrow 0$, in view of the previous equation and the continuity of $q(\theta)$ we see that the right derivative $D^+h(\theta)$ exists and is equal to $q(\theta)$. Using Lemma 14 and arguing along the lines of Theorem 2 we conclude that $ET(\theta)$ is twice continuously differentiable with second derivative

$$\frac{d^2ET}{d\theta^2} = \frac{1}{E[N_1]}E[\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{d^2X_j}{d\theta^2}] + E[\frac{g(Z_1)}{1-G(Z_1)}(\sum_{i=1}^{N_1}\frac{dX_i}{d\theta})^2] . \quad (57)$$

□

As before, location and scale parameters yield especially simple expressions: When θ is a location parameter of the service time distribution

$$\frac{d^2ET}{d\theta^2} = E[\frac{g(Z_1)}{1-G(Z_1)}N_1^2] , \tag{58}$$

and when θ is a scale parameter

$$\frac{d^2ET}{d\theta^2} = \frac{1}{\theta^2}E[\frac{g(Z_1)}{1-G(Z_1)}Y_1^2] , \tag{59}$$

where Y_1 is the length of the busy period. For location and scale parameters, and more generally, whenever $\frac{d^2X(\theta)}{d\theta^2} = 0$, (50), (58) and (59) reveal that the second derivative of ET is obtained simply by taking into account the "interactions" between adjacent busy periods. (In fact, Lemma 3 of §3 is a formalization of this idea.)

7 Vector parameters of the service time distribution: Gradient and Hessian Estimation

In this section we describe briefly an extension of the above results to vector parameters. While conceptually straightforward, this extension is important both in applications and in extending the results of the previous section to service time distributions satisfying **A.1** instead of the more restrictive monotonicity assumption **M.1**. We start by giving the vector parameter equivalents of the conditions in §2.

Let $\theta = (\theta_1, \ldots, \theta_k)$, $\theta \in \mathbf{B} \stackrel{\text{def}}{=} [a_1, b_1] \times \cdots \times [a_k, b_k]$. As in §2 let U be a random variable uniformly distributed in $[0, 1]$, and $F^{-1}(u, \theta) = \inf\{v : F(v, \theta) > u\}$. Then $X(\theta) \stackrel{\text{def}}{=} F^{-1}(U, \theta)$ satisfies $P(X(\theta) \leq x) = F(x, \theta)$. We will assume that it satisfies the following conditions:

Condition VC.1 *The partial derivatives $D_l X(\theta)$, $l = 1, \ldots, k$, exist and are continuous functions of $\theta \in \mathbf{B}$ w.p.1.*

Condition VC.2 *The partial derivatives* $D_{lr}X(\theta)$, $l, r = 1, \ldots, k$, *exist and are continuous functions of* $\theta \in \mathbf{B}$ *w.p.1.*

We also introduce the vector parameter counterparts of assumptions **M.1, A.1, A.2**, and **A.4**. For economy of notation we will use the same symbols as in §3 whenever no confusion arises.

Assumption VM.1 $X(\theta_1, \ldots \theta_k)$ *is monotonic in each* θ_l, $l = 1, 2, \ldots k$, *w.p.1. Without loss of generality we will assume that for some* $k_1 \in \{1, 2, \ldots, k\}$, X *is nondecreasing in* $\theta_1, \ldots \theta_{k_1}$, *and nonincreasing in* $\theta_{k_1+1}, \ldots, \theta_k$. *Furthermore, if* A *is a random variable distributed according to the interarrival distribution,* $EX(b_1, \ldots, b_{k_1}, a_{k_1+1}, \ldots, a_k) < EA < \infty$.

As we shall see, the above monotonicity assumption can be replaced with the following, less restrictive,

Assumption VA.1 *Let* $\chi = \sup_{\theta \in \mathbf{B}} X(\theta)$. *Then* $E\chi < EA < \infty$ *where* A *is a random variable distributed according to the interarrival distribution* G.

Assumption VA.2 *Let* $\xi = \sup_{\theta \in \mathbf{B}} \sum_{l=1}^{k} |D_l X(\theta)|$. *Then* $E[\chi^2] < \infty$ *and* $E[\xi^3] < \infty$.

Assumption VA.4 *There exists* $\epsilon > 0$ *such that* $E[e^{\epsilon\chi}] < \infty$ *and* $E[e^{\epsilon\xi}] < \infty$. *Furthermore, if* $\psi = \sup_{\theta \in \mathbf{B}} \sum_{r,l=1}^{k} |D_{lr}X(\theta)|$, $E|\psi|^3 < \infty$.

The following theorem is obtained by an analysis entirely analogous to that of sections 4 and 5.

Theorem 4 *For a GI/GI/1 system with service time distribution* $F(x, \theta)$ *satisfying conditions* **VC.1** *and assumptions* **VM.1** *and* **VA.2**, *the expected system time in steady state is continuously differentiable with gradient*

$$\frac{\partial ET}{\partial \theta_l} = \frac{E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} D_l X_j]}{E[N_1]} \qquad l = 1, 2, \ldots, k.$$

If in addition **VC.2,** *and assumptions* **A.3,** *and* **VA.4** *hold, the expected system time in steady state is twice continuously differentiable with Hessian matrix given by*

$$\frac{\partial^2 ET}{\partial\theta_l\partial\theta_r} = \frac{E[\sum_{j=1}^{N_1}\sum_{j=1}^{i} D_{lr}X_j]}{E[N_1]} + E[\frac{g(Z_1)}{1-G(Z_1)}(\sum_{i=1}^{N_1} D_l X_i)(\sum_{i=1}^{N_1} D_r X_i)] \quad r,l = 1,2,\dots,k. \quad (60)$$

Proof: The analysis is carried out for each component of θ separately, following §4 and 5. If $X(\theta)$ is *nonincreasing* in θ_l, then we choose $\Delta\theta_l < 0$ and obtain a left derivative $D_l^- ET(\theta)$. Establishing the continuity of $D_l^+ ET(\theta)$ for $l \in \{1,2,\dots k_1\}$ and that of $D_l^- ET(\theta)$ for $l \in \{k_1+1,\dots k\}$ as in §4.1, we conclude that the partial derivatives $\frac{\partial ET}{\partial\theta_l}$ exist and are continuous and hence that $ET(\theta)$ is differentiable. □

8 Relaxing the monotonicity assumption

The main idea here is to convert a scalar parameter $\theta \in [a,b]$ which does not satisfy the monotonicity assumption **M.1** into a vector parameter (θ_1,θ_2) which satisfies **VM.1.**

Suppose we are given a stochastic service time function $X(\theta)$, $\theta \in [a,b]$ satisfying **C.1.** Let $\frac{dX}{d\theta} = (\frac{dX}{d\theta})^+ - (\frac{dX}{d\theta})^- = 1(\frac{dX}{d\theta} > 0)\frac{dX}{d\theta} + 1(\frac{dX}{d\theta} < 0)\frac{dX}{d\theta}$ where $(\frac{dX}{d\theta})^+$ and $(\frac{dX}{d\theta})^-$ denote the positive and negative part respectively of the derivative $\frac{dX}{d\theta}$ at θ. Define a stochastic service time function of two variables, by means of

$$\begin{aligned}\tilde{X}(\theta_1,\theta_2,\omega) &= X(a,\omega) + \textstyle\int_a^{\theta_1}(\frac{dX}{du}(u,\omega))^+du \;-\; \int_a^{\theta_2}(\frac{dX}{du}(u,\omega))^-du \qquad \text{for all } \omega\in\Omega \qquad (61)\\ &= X(a,\omega) + \textstyle\int_a^{\theta_1} 1(\frac{dX}{du}(u,\omega) > 0)\frac{dX}{du}(u,\omega)du \;+\; \int_a^{\theta_2} 1(\frac{dX}{du}(u,\omega) < 0)\frac{dX}{du}(u,\omega)du\end{aligned}$$

and $(\theta_1,\theta_2) \in \mathbf{D} \stackrel{\text{def}}{=} \{a \le \theta_1 \le b \;;\; a \le \theta_2 \le \theta_1\}$. Thus the rhs of (61) is nonnegative w.p.1. Furthermore $X(\theta) = \tilde{X}(\theta,\theta)$ w.p.1 and $\tilde{X}$ satisfies **VM.1.**

Example: Consider the family of uniform distributions with mean m and spread 2θ, $\theta \in [0,m]$:

$$F(x,\theta) = \begin{cases} 0 & \text{if } \; x < m-\theta \\ (x+\theta-m)/2\theta & \text{if } \; m-\theta \le x < m+\theta \\ 1 & \text{if } \; m+\theta \le x \;\; . \end{cases}$$

On the probability space $([0,1], \mathcal{B}_{[0,1]}, L_{[0,1]})$, the last two parts of the triplet designating the Borel sets on [0,1] and the Lebesgue measure on that same interval respectively, the corresponding stochastic service function would be $X(\theta,\omega) = m + \theta(2\omega - 1)$. Then $(\frac{dX}{d\theta}(\theta,\omega))^+ = (2\omega-1)^+$, $(\frac{dX}{d\theta}(\theta,\omega))^- = (2\omega-1)^-$, and for $0 \le \theta_i \le 2m$, $i = 1,2$, $\tilde{X}(\theta_1,\theta_2,\omega) = m + \theta_1(2\omega-1)^+ - \theta_2(2\omega-1)^-$, $\omega \in [0,1]$.

We can now state the following corollary to Theorem 4:

Corollary 1 *In Theorem 2, assumption* **M.1** *can be replaced by* **A.1**.

Proof: Given a queueing process with interarrival times $\{A_i(\omega)\}$ and service times $\{X_i(\theta,\omega)\}$, we construct for each ω and each $(\theta_1,\theta_2) \in \mathbf{D}$ a new queueing process with the same interarrival sequence $\{A_i(\omega)\}$ and service time sequence $\{\tilde{X}_i(\theta_1,\theta_2,\omega)\}$ with $\tilde{X}_i(\theta_1,\theta_2,\omega)$ given by (61). Let $\{\tilde{T}_i(\theta_1,\theta_2,\omega)\}$ be the corresponding sequence of system times. Notice that, for any $\theta \in [a,b]$, $\tilde{X}_i(\theta,\theta) = X(\theta)$ w.p. 1 and hence that $\tilde{T}_i(\theta,\theta) = T_i(\theta)$ w.p. 1, assuming that both systems start empty with the arrival of a customer at time $t = 0$. The new system with the vector parameter however satisfies the monotonicity assumption **VM.1**. In particular we have $D_1\tilde{X}_i = 1(\frac{dX_i}{d\theta} > 0)\frac{dX_i}{d\theta}$, $D_2\tilde{X}_i = 1(\frac{dX_i}{d\theta} < 0)\frac{dX_i}{d\theta}$, and hence

$$\begin{aligned}\frac{dET}{d\theta} &= D_1E\tilde{T}(\theta,\theta) + D_2E\tilde{T}(\theta,\theta) = \frac{1}{E[N_1]}E[\sum_{i=1}^{N_1}\sum_{j=1}^{i} D_1\tilde{X}_j] + \frac{1}{E[N_1]}E[\sum_{i=1}^{N_1}\sum_{j=1}^{i} D_2\tilde{X}_j] \\ &= \tfrac{1}{E[N_1]}E[\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{i} 1(\frac{dX_j}{d\theta} > 0)\frac{dX_j}{d\theta} + 1(\frac{dX_j}{d\theta} < 0)\frac{dX_j}{d\theta}] = \tfrac{1}{E[N_1]}E[\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{i}\frac{dX_j}{d\theta}(\theta)]\ .\end{aligned}$$

□

The same approach allows us to relax the monotonicity assumption in our second derivative results:

Corollary 2 *Theorem 3 also holds with assumption* **M.1** *replaced by* **A.1**.

Proof: Define $\tilde{X}(\theta_1,\theta_2)$, $\tilde{T}(\theta_1,\theta_2)$, as in Corollary 1. From (61) it is easy to see that w.p.1

$$\begin{aligned} D_{11}\tilde{X}(\theta,\theta) &= 1(\frac{dX}{d\theta} > 0)\frac{d^2X}{d\theta^2}, \\ D_{12}\tilde{X}(\theta,\theta) &= D_{21}\tilde{X}(\theta,\theta) = 0, \\ D_{22}\tilde{X}(\theta,\theta) &= 1(\frac{dX}{d\theta} < 0)\frac{d^2X}{d\theta^2}. \end{aligned} \tag{62}$$

Hence, from Theorems 2 and 3,

$$\begin{aligned}
\frac{d^2ET}{d\theta^2}(\theta) &= D_{11}E\tilde{T}(\theta,\theta) + D_{22}E\tilde{T}(\theta,\theta) + 2D_{12}E\tilde{T}(\theta,\theta) \\
&= \frac{E[\sum_{j=1}^{N_1}\sum_{j=1}^{i} D_{11}\tilde{X}_j(\theta,\theta)]}{E[N_1]} + E\left[\tfrac{g(Z_1)}{1-G(Z_1)}\left(\textstyle\sum_{i=1}^{N_1} D_1\tilde{X}_i(\theta,\theta)\right)^2\right] \\
&+ \frac{E[\sum_{j=1}^{N_1}\sum_{j=1}^{i} D_{22}\tilde{X}_j(\theta,\theta)]}{E[N_1]} + E\left[\tfrac{g(Z_1)}{1-G(Z_1)}\left(\textstyle\sum_{i=1}^{N_1} D_2\tilde{X}_i(\theta,\theta)\right)^2\right] \\
&+ 2\frac{E[\sum_{j=1}^{N_1}\sum_{j=1}^{i} D_{12}\tilde{X}_j(\theta,\theta)]}{E[N_1]} + 2E\left[\tfrac{g(Z_1)}{1-G(Z_1)}\left(\textstyle\sum_{i=1}^{N_1} D_1\tilde{X}_i(\theta,\theta)\right)\left(\textstyle\sum_{i=1}^{N_1} D_2\tilde{X}_i(\theta,\theta)\right)\right].
\end{aligned} \tag{63}$$

From (62) we get $D_{11}\tilde{X}_i + D_{22}\tilde{X}_i = \frac{d^2X_i}{d\theta^2}$. Also,

$$\left(\sum_{i=1}^{N_1} D_1\tilde{X}_i\right)^2 + \left(\sum_{i=1}^{N_1} D_2\tilde{X}_i\right)^2 + 2\left(\sum_{i=1}^{N_1} D_1\tilde{X}_i\right)\left(\sum_{i=1}^{N_1} D_2\tilde{X}_i\right) = \left(\sum_{i=1}^{N_1} D_1\tilde{X}_i + D_2\tilde{X}_i\right)^2 = \left(\sum_{i=1}^{N_1} \frac{dX}{d\theta}\right)^2.$$

From the above two equations and (63) the proof of the corollary follows. □

Extending the above ideas to vector parameters is straightforward. If $X(\theta_1,\ldots,\theta_k)$ satisfies **VA.1**, we can define using the same approach a stochastic service function depending on $2k$ parameters, say $X(\theta_1^+,\ldots,\theta_k^+;\theta_1^-,\ldots,\theta_k^-)$, which satisfies **VM.1**. We will state without proof

Corollary 3 *Theorem 4 also holds with assumption* **VM.1** *replaced by* **VA.1**. □

9 Derivatives with respect to parameters of the interarrival time distribution

We start by introducing the following assumptions:

Consider a parametric family of GI/GI/1 queueing processes with service and interarrival time distributions $F(x)$ and $G(x,\eta)$, $\eta \in [c,d]$, respectively. Define

$$A(\eta) = G^{-1}(U,\eta) \tag{64}$$

where U is uniformly distributed in $[0,1]$. We use the same notation as in §4. In order to avoid the proliferation of complicated notation we shall recycle χ,ξ,ψ, and α, used in assumptions **A.1** through **A.4** and use them with new meaning in the corresponding assumptions of this section. We will assume that the family of interarrival time distributions $G(x, \eta)$ is such that the corresponding random interarrival function $A(\eta)$ satisfies condition **C.1** and the assumptions that follow:

Assumption IA.1 *If* $\chi(\omega) = \inf_{\eta \in [c,d]} A(\eta, \omega)$ *for all* ω*, then* $EX < E\chi$*.*

If there exists $\eta^* \in [c, d]$ such that $A(\eta) \geq A(\eta^*)$ w.p.1 for all $\eta \in [c, d]$ then the above assumption is simply a stability condition for the corresponding family of queueing processes.

Assumption IA.2 *If* $\xi(\omega) = \inf_{\eta \in [c,d]} |\frac{dA}{d\eta}(\eta, \omega)|$ *then* $E[\xi^3] < \infty$*.*

To obtain second derivative estimates $A(\eta)$ must satisfy, in addition to the above, **C.2** and the following two assumptions:

Assumption IA.3 *For all* $\eta \in [c, d]$ *the interarrival distribution* $G(x, \eta)$ *is absolutely continuous with density* $g(x, \eta)$*, which without loss of generality we will assume right continuous, and has hazard function uniformly bounded in* $[c, d]$*:* $\sup_{\eta \in [c,d]} \frac{g(x,\eta)}{1-G(x,\eta)} \leq \alpha < \infty$*, for all* $x \in [0, \infty)$*.*

Assumption IA.4 *There exists* $\epsilon > 0$ *such that* $E[e^{\epsilon \chi}] < \infty$ *and* $E[e^{\epsilon \xi}] < \infty$*. Furthermore, if* $\psi(\omega) = \sup_{\eta \in [c,d]} |\frac{d^2 A}{d\eta^2}(\eta, \omega)|$*,* $E|\psi|^3 < \infty$*.*

We will again base the sample path analysis that follows on an additional monotonicity assumption which will be relaxed in the sequel:

Assumption MIA.1 $A(\eta + \Delta\eta) \leq A(\eta)$ *w.p.1 for* $\Delta\eta > 0$*.*

(Notice that to keep the analysis parallel to that of §3 we assume that $A(\eta)$ is *nonincreasing* in η w.p.1. As a result, a change from η to $\eta + \Delta\eta$ in the parameter may cause busy periods to coalesce, but not to break up.)

An important point in which the analysis for interarrival time parameters differs from that of §4 and 5 is the behavior of the associated system which is more complicated. In the following paragraphs we briefly describe the arguments highlighting the differences between the two cases and providing some details.

On the same probability space as in §4 we define

$$X_i \;=\; F^{-1}(U_{2i-1}) \quad \text{and} \quad A_i \;=\; G^{-1}(U_{2i},\eta), \quad i = 1, 2, \ldots\ .$$

We examine again two sample paths, the *nominal* with interarrival parameter value η and the *perturbed* with $\eta + \Delta\eta$, both starting with the arrival of the first customer, C_1, to an empty system at time $t = 0$. Let $T_i(\eta)$ be the system time of the ith customer and $\Delta T_i = T_i(\eta + \Delta\eta) - T_i(\eta)$. N_i is the number of customers in the ith busy period and $M_k = N_1 + N_2 + \cdots + N_k$, $M_0 = 0$, is the discrete time renewal process of the indices of customers whose departure *terminates* busy periods. These quantities depend of course on η but, unless the dependence is made explicit, they are determined from the nominal sample path. Hence when we refer to the index of the customer who terminates the kth busy period in the nominal sample path we may write either $M_k(\eta)$ or M_k, whereas for the index of the customer who terminates the kth busy period of the perturbed path we will always write $M_k(\eta + \Delta\eta)$.

Let $\Delta A_i = A_i(\eta + \Delta\eta) - A_i(\eta)$, and define

$$B_k(\eta) \;=\; \sum_{i=M_{k-1}+1}^{M_k} A_i(\eta), \qquad B_k(\eta + \Delta\eta) \;=\; \sum_{i=M_{k-1}+1}^{M_k} A_i(\eta + \Delta\eta),$$

$k = 1, 2, \ldots$, and

$$\Delta B_k \;\stackrel{\text{def}}{=}\; -\sum_{i=M_{k-1}+1}^{M_k} \Delta A_i\ , \quad k = 1, 2, \ldots\ . \tag{65}$$

Notice that for convenience we define ΔB_k to be positive. (In view of **MIA** $\Delta A_i \leq 0$ w.p.1.) $B_k(\eta)$ is the length of the kth busy cycle in the nominal sample path. However, this is not necessarily the case with $B_k(\eta + \Delta\eta)$ since both are defined with respect to M_k. Similarly let

$$S_k(\eta) \;=\; \sum_{i=M_{k-1}+1}^{M_k} T_i(\eta), \qquad S_k(\eta + \Delta\eta) \;=\; \sum_{i=M_{k-1}+1}^{M_k} T_i(\eta + \Delta\eta),$$

$k = 1, 2, \ldots$, and

$$\Delta S_k \stackrel{\text{def}}{=} \sum_{i=M_{k-1}+1}^{M_k} \Delta T_i \ , \quad k = 1, 2, \ldots \ , \tag{66}$$

The change in the system time of the ith customer is given by

$$\Delta T_i \ = \ - \sum_{j=M_{k-1}+1}^{i-1} \Delta A_j \ + \ V_{k-1} \qquad \text{for} \quad M_{k-1} + 1 \leq i \leq M_{k+1} \ , \tag{67}$$

where the above sum is taken to be 0 if empty and V_{k-1} is the waiting time of the kth customer in an *auxiliary* system given by

$$V_k \ = \ \max(0, \ \Delta B_k - I_k, \ \Delta B_k + \Delta B_{k-1} - I_k - I_{k-1}, \ \ldots, \\ \Delta B_k + \Delta B_{k-1} + \cdots + \Delta B_1 \ - \ I_k - I_{k-1} - \cdots - I_1) \ . \tag{68}$$

As in §4 the auxiliary process defined by (68) will be stable provided that $E\Delta B_1 < EI_1$ or equivalently $EX_1 < EA_1(\eta{+}\Delta\eta)$ (i.e. as long as the parameter change in the interarrival time distribution does not make the system unstable). We can also show that $E\Delta B_1^2 < \infty$ and hence, as $k \to \infty$, that V_k converges in distribution to a random variable V with $EV < \infty$.

An analysis similar to §4 yields

$$ET(\eta + \Delta\eta) - ET(\eta) \ = \ - \frac{1}{E[N_1]} E[\sum_{i=1}^{N_1-1} \sum_{j=1}^{i} \Delta A_j] \ + \ E[V] \ . \tag{69}$$

10 Analysis of the auxiliary system

This section consists of three lemmas, similar to their counterparts in §5. There are however enough differences both in the proofs and in the final results to warrant repetition.

Lemma 4 *For the auxiliary system defined above,* $\lim_{\Delta\eta \to 0} \frac{1}{\Delta\eta} E[V] \ = \ 0$.

The proof of Lemma 4 is similar to that of Lemma 1 and we will omit it.

Next we prove the counterpart of Lemma 2:

Lemma 5 *For any $r \in (0,1)$ and sufficiently small $\Delta\eta > 0$ there is a positive $L < \infty$ such that*

$$0 \;<\; E[V] - E[(\Delta B_1 - I_1)^+] \;\leq\; L\Delta\eta^{2+r}\ . \tag{70}$$

Notice that, unlike Lemma 2, here the exponent of $\Delta\eta$ can be arbitrarily close to 3, but not exactly.

Proof: As in the proof of Lemma 2 we will use Kingman's inequality (Kingman [16]). The only complication that arises here is the fact that ΔB_1 and I_1 are *not* conditionally independent given Z_1, the age of the interarrival process at the end of the busy period. Define $K^*(\gamma, \Delta\eta) = E[e^{\gamma(\frac{\Delta B_1}{\Delta\eta} - \frac{I_1}{\Delta\eta})}]$. Using Hölder's inequality we obtain

$$K^*(\gamma, \Delta\eta) \leq [Ee^{p\gamma\frac{\Delta B_1}{\Delta\eta}}]^{1/p}\,[Ee^{-q\gamma\frac{I_1}{\Delta\eta}}]^{1/q} \tag{71}$$

for any p and q such that $1/p + 1/q = 1$. As in the proof of Lemma 2, to compute the second term in the right hand side of (71) we first condition on the age of the arrival process at the end of the busy period, Z_1. This gives the inequality $E[\exp\{-q\gamma\frac{I_1}{\Delta\eta}\} \mid Z_1] \leq \frac{\Delta\eta\alpha}{q\gamma+\Delta\eta\alpha}$. Hence

$$K^*(\gamma, \Delta\eta) \;\leq\; [E\exp\{p\gamma\frac{\Delta B_1}{\Delta\eta}\}]^{1/p}[\frac{\Delta\eta\alpha}{q\gamma + \Delta\eta\alpha}]^{1/q}\ . \tag{72}$$

From Lemma 10 of the Appendix with obvious changes in the notation it follows that $E[\exp\{\gamma\frac{\Delta B_1}{\Delta\eta}\}] \leq \overline{K}$ for $\gamma < \epsilon$. Hence, letting $\tilde{K} = \overline{K}^{1/p}\left(\frac{\alpha}{q\gamma}\right)^{1/q}$, for $0 < \gamma < \epsilon$,

$$K^*(\gamma, \Delta\eta) \;\leq\; \tilde{K}\Delta\eta^{1/q}\ . \tag{73}$$

From this follows that $K^*(\gamma, \Delta\eta) < 1/2$ for γ and $\Delta\eta$ small enough (namely for $0 < \gamma < \epsilon/p$ and $\Delta\eta < (\tilde{K}/2)^q$). But then from Kingman's inequality (Kingman [16]),

$$0 \;\leq\; E[\sup(0\,,\ \frac{\Delta B_1}{\Delta\eta} - \frac{I_1}{\Delta\eta}, \ldots)] \;-\; E(\frac{\Delta B_1}{\Delta\eta} - \frac{I_1}{\Delta\eta})^+ \;\leq\; \frac{[K^*(\gamma, \Delta\eta)]^2}{2e\gamma[1 - K^*(\gamma, \Delta\eta)]}\ . \tag{74}$$

Fix $\gamma < \epsilon/p$ and let $\Delta\eta < (\tilde{K}/2)^q$. Then,

$$0 \;\leq\; \frac{1}{\Delta\eta}EV \;-\; E(\frac{\Delta B_1}{\Delta\eta} - \frac{I_1}{\Delta\eta})^+ \;<\; \frac{\tilde{K}^2}{e\gamma}\Delta\eta^{2/q} \tag{75}$$

From (75), for any $r \in (0,1)$, $q = 2/(1+r)$, $\Delta\eta < (\tilde{K}/2)^q$, and $L = (\tilde{K})^2/e\gamma$, we obtain (70). This concludes the proof. □

Finally, the limiting behavior of $E(\Delta B_1 - I_1)^+$ when $\Delta\eta \to 0$ is characterized by the following

Lemma 6 *Let Z_1 be the age of the arrival process at the end of the busy period. Then,*

$$\lim_{\Delta\eta\to 0} \tfrac{1}{\Delta\eta^2} E(\Delta B_1 - I_1)^+ \;=\; \tfrac{1}{2} E[\tfrac{g(Z_1)}{1-G(Z_1)} \; (\textstyle\sum_{i=1}^{N_1-1} \frac{dA_i}{d\eta} + \frac{dA_{N_1}}{d\eta}|_{A_{N_1}=Z_1})^2] \; . \tag{76}$$

where

$$\frac{dA_{N_1}}{d\eta}|_{A_{N_1}=Z_1} \;\; \stackrel{\text{def}}{=} \;\; \lim_{\delta\to 0} \frac{G^{-1}_{\eta+\delta}(G_\eta(Z_1)) \; - \; Z_1}{\delta} .$$

(In the above definition and in what follows we use $G_\eta(x)$ as a more compact notation for $G(x,\eta)$.)

The presence of the term $\frac{dA_{N_1}}{d\eta}|_{A_{N_1}=Z_1}$ in (76) does not come as a surprise since $(\Delta B_1 - I_1)^+$ is positive only on $\{I_1 < \Delta B_1\}$ and as $\Delta\eta \to 0$, $\Delta B_1 \to 0$ w.p.1 and hence $A_{N_1} \to Z_1$ w.p.1.

Proof: Let Z_1 be the time that has elapsed since the last arrival, at the instant when BP_1 ends. To compute the limit in (76) we condition on the information available at the end of the busy period and more specifically on $\sum_{i=1}^{N_1-1} \Delta A_i$ and Z_1. To simplify the notation define $\Delta R_1 = -\sum_{i=1}^{N_1-1} \Delta A_i$. Then

$$E(\Delta B_1 - I_1)^+ \;\; = \;\; E[\, E[(\Delta R_1 + \Delta A_{N_1} - I_1)^+ | \Delta R_1, Z_1]\,] \; . \tag{77}$$

Turning our attention to the conditional expectation in (77) and taking into account that $A_{N_1} = Z_1 + I_1$ and that

$$\Delta A_{N_1} = G^{-1}_{\eta+\Delta\eta}(G_\eta(A_{N_1})) - A_{N_1} = G^{-1}_{\eta+\Delta\eta}(G_\eta(I_1 + Z_1)) - Z_1 - I_1 \; ,$$

we can rewrite (77) as

$$\int_0^\infty [\Delta R_1 + G^{-1}_{\eta+\Delta\eta}(G_\eta(Z_1 + x)) - Z_1 - x]^+ \, \tfrac{g_\eta(Z_1+x)}{1-G_\eta(Z_1)} dx \; . \tag{78}$$

Let $DR_1(x) \stackrel{\text{def}}{=} -\sum_{i=1}^{N_1-1} \frac{dA_i}{d\eta}(x)$ denote the value of the derivative $\frac{d}{d\eta}R_1$ evaluated at $\eta = x$. Using Taylor's theorem, we get $\Delta R_1 = DR_1(\eta + \beta\Delta\eta)$, where $\beta \in [0,1]$, depending in general on η, $\Delta\eta$, and ω. This together with the change of variable $y = x/\Delta\eta$ in (78) gives

$$\int_0^\infty [DR_1(\eta+\beta\Delta\eta)\Delta\eta + G^{-1}_{\eta+\Delta\eta}(G_\eta(Z_1+y\Delta\eta)) - Z_1 - y\Delta\eta]^+ \, \tfrac{g_\eta(Z_1+y\Delta\eta)}{1-G_\eta(Z_1)} \Delta\eta dy \; . \qquad (79)$$

We thus have

$$\lim_{\Delta\eta\downarrow 0} \tfrac{1}{\Delta\eta^2} \; E\left[E[(\Delta R_1 + \Delta A_{N_1} - I_1)^+ | \Delta R_1, Z_1]\right] \;=$$
$$= \lim_{\Delta\eta\downarrow 0} \; E[\textstyle\int_0^\infty [DR_1(\eta+\beta\Delta\eta) + \frac{1}{\Delta\eta}(G^{-1}_{\eta+\Delta\eta}(G_\eta(Z_1+y\Delta\eta)) - Z_1) \; - y]^+ \, \frac{g_\eta(Z_1+y\Delta\eta)}{1-G_\eta(Z_1)} dy \;] \; . \quad (80)$$

From assumption **IA.2** we have

$$DR_1(\eta+\beta\Delta\eta) \;\le\; \sum_{i=1}^{N_1-1} \xi_i \quad \text{w.p.1},$$

and from **IA.2** and the triangular inequality,

$$\begin{aligned} |G^{-1}_{\eta+\Delta\eta}(G_\eta(Z_1+y\Delta\eta)) - Z_1| \;&\le\; |G^{-1}_{\eta+\Delta\eta}(G_\eta(Z_1+y\Delta\eta)) - Z_1 - y\Delta\eta| \;+\; y\Delta\eta \\ &\le\; \xi_N\Delta\eta + y\Delta\eta \qquad \text{w.p.1.} \end{aligned}$$

Hence, arguing as in Lemma 3 we can show that the quantity inside the integral on the rhs of (80), for $\Delta\eta \le \frac{1}{2}$ is dominated by $2\alpha(\sum_{i=1}^{N_1}\xi_i - y)^+$. Since $E\int_0^\infty 2\alpha(\sum_{i=1}^{N_1}\xi_i - y)^+ dy \;=\; \alpha E(\sum_{i=1}^{N_1}\xi_i)^2 < \infty$ (by Lemma 8 of the appendix), an appeal to the Dominated Convergence Theorem shows that we can pass the limit inside the expectation and the integral in (80) to obtain

$$E[\int_0^\infty (DR_1(\eta) + \lim_{\Delta\eta\downarrow 0}\frac{1}{\Delta\eta}[G^{-1}_{\eta+\Delta\eta}(G_\eta(Z_1) - Z_1) - y]^+ \frac{g_\eta(Z_1)}{1-G_\eta(Z_1)} dy \;\; ,$$

which is equal to the rhs of (76). □

Theorem 5 *Suppose that* $A(\eta)$ *satisfies* **C.1** *and assumptions* **IA.1** *and* **IA.2**. *Then the expected system time in steady state,* $ET(\eta)$*, is continuously differentiable with respect to* $\eta \in [c,d]$ *with*

$$\frac{dET}{d\eta} \quad = \quad - \frac{E[\sum_{i=1}^{N_1-1}\sum_{j=1}^{i}\frac{dA_j}{d\eta}]}{E[N_1]} \quad . \tag{81}$$

Furthermore, if in addition to the above $A(\eta)$ *satisfies* **IA.3** *and* **IA.4** *then* $ET(\eta)$ *is twice continuously differentiable on* $[c,d]$ *with*

$$\frac{d^2ET}{d\eta^2} \quad = \quad - \frac{E[\sum_{i=1}^{N_1-1}\sum_{j=1}^{i}\frac{d^2A_j}{d\eta^2}]}{E[N_1]} \quad + \quad E[\frac{g(Z_1,\eta)}{1-G(Z_1,\eta)}(\sum_{i=1}^{N_1-1}\frac{dA_i}{d\eta}+\frac{dA_{N_1}}{d\eta}|_{A_{N_1}=Z_1})^2\ ] \ . \tag{82}$$

The proof follows from (69) and Lemmas 4 through 6, proceeding along the same steps as in the proof of Theorems 2 and 3. After obtaining a right derivative $D^+ET(\eta)$, a continuity argument for it is used to show differentiability as in §6. To extend the result beyond the realm of the monotonicity assumption **MIA**, we can use an argument similar to that presented in §8.

11 Queues with parametric families of interarrival and service time distributions

Consider a family of GI/GI/1 queueing processes with interarrival time distribution $G(x,\eta)$ and service time distribution $F(x,\theta)$, $(\eta,\theta) \in [c,d] \times [a,b]$. For the purpose of derivative estimation we can of course consider one of the two parameters fixed and vary the other, thereby transforming the problem to a case already addressed. However for the purpose of estimating the partial derivative $\frac{\partial^2}{\partial\eta\partial\theta}ET$ we have no recourse but to repeat the steps described in §4 and §9 with both parameters changed at the same time. We briefly sketch the approach: With ΔY_k and ΔB_k defined in (8) and (65) let

$$EV(\Delta\eta,\Delta\theta) \quad \stackrel{\text{def}}{=} \quad E\sup(0, \Delta B_1+\Delta Y_1-I_1, \Delta B_1+\Delta B_2+\Delta Y_1+\Delta Y_2-I_1-I_2,\ldots) \tag{83}$$

If we set $\Delta\eta = 0$ in (83) we obtain the expectation of the rhs of (13). $\Delta\theta = 0$ would give the last term of (69))

Let $\Delta = \sqrt{\Delta\eta^2 + \Delta\theta^2}$. Following the same steps as before we obtain

$$ET(\eta + \Delta\eta, \theta + \Delta\theta) - ET(\eta, \theta) = \frac{1}{EN_1} E[\sum_{i=1}^{N_1} \sum_{j=1}^{i} \Delta X_j] - \frac{1}{EN_1} E[\sum_{i=1}^{N_1-1} \sum_{j=1}^{i} \Delta A_j] + EV(\Delta\theta, \Delta\eta) . \quad (84)$$

To establish the existence of $\frac{\partial^2}{\partial\theta\partial\eta} ET$ it is enough to show that the following iterated limit exists and is a continuous function of $(\eta, \theta) \in [c, d] \times [a, b]$.

$$\lim_{\Delta\theta\downarrow 0} \lim_{\Delta\eta\downarrow 0} \frac{1}{\Delta\theta\Delta\eta} [\, ET(\eta + \Delta\eta, \theta + \Delta\theta) - ET(\eta, \theta + \Delta\theta)] - ET(\eta + \Delta\eta, \theta) + ET(\eta, \theta)] . \quad (85)$$

Taking into account (84), (20), and (69) we can show that the quantity inside the iterated limits in (85) is equal to

$$EV(\Delta\eta, \Delta\theta) - EV(0, \Delta\theta)] - EV(\Delta\eta, 0) = E(\Delta B_1 + \Delta Y_1 - I_1)^+ - E(\Delta Y_1 - I_1)^+ - E(\Delta B_1 - I_1)^+ + o(\Delta^2) , \quad (86)$$

where in the last equality we have used Lemma 2, Lemma 4, and its obvious counterpart for $EV(\Delta\eta, \Delta\theta)$. It remains to compute

$$\lim_{\Delta\theta\downarrow 0} \lim_{\Delta\eta\downarrow 0} \frac{1}{\Delta\theta\Delta\eta} \left[E(\Delta B_1 + \Delta Y_1 - I_1)^+ - E(\Delta Y_1 - I_1)^+ - E(\Delta B_1 - I_1)^+\right] .$$

The analysis is similar to that of Lemma 5. We summarize the results in the following

Theorem 6 *If the system satisfies* **C.1, C.2,** *assumptions* **A.1 - A.4**, *and* **IA.1 -IA.4** *then* $ET(\theta, \eta)$ *is twice continuously differentiable with*

$$\frac{\partial^2 ET}{\partial\eta\partial\theta} = E\left[\frac{g(Z_1, \eta)}{1 - G(Z_1, \eta)} \left(\sum_{j=1}^{N_1} \frac{dX_i}{d\theta}\right) \left(\sum_{i=1}^{N_1-1} \frac{dA_i}{d\eta} + \frac{dA_{N_1}}{d\eta}|_{A_{N_1}=Z_1}\right)\right] . \quad (87)$$

The continuity of the rhs of (87) is established by means of arguments similar to those of §4. It also guarantees the equality $\frac{\partial^2}{\partial\theta\partial\eta} ET = \frac{\partial^2}{\partial\eta\partial\theta} ET$.

12 Perturbation Analysis Algorithms

Here we present perturbation analysis algorithms that provide strongly consistent estimates of the first and second derivative of ET. Algorithm 1 describes a standard regenerative procedure for estimating $\frac{dET}{d\theta}$ and $\frac{d^2ET}{d\theta^2}$ from observations on a single sample path consisting of m busy periods.

Algorithm 1

Start with arrival to idle system
INITIALIZE:
  Set the following variables to zero: DX, DDX, SDX, SDDX, BP, n, nbp
  Set m= number of busy periods to be observed

NEXT-SERVICE:
  Observe next service time X
  $n = n + 1$
  $DX = DX + \frac{dx}{d\theta}$
  $DDX = DDX + \frac{d^2X}{d\theta^2}$
  $SDX = SDX + DX$
  $SDDX = SDDX + DDX$

NEXT-ARRIVAL:
    If arrival finds system idle then
    Z = Length of last interarrival time – Idle period
    $BP = BP + \frac{g(Z)}{1-G(Z)}(DX)^2$
    $DX = 0$
    $DDX = 0$
    $nbp = nbp + 1$
  If arrival finds system busy do nothing

STOPPING CRITERION
  If $nbp < m$ then go to NEXT-SERVICE
FINAL VALUE
  $T1(m) = \frac{SDX}{n}$
  $T2(m) = \frac{SDDX}{n} + \frac{BP}{m}$

Print $T1(m)$, $T2(m)$.
STOP

The outputs $T1(m)$ and $T2(m)$ printed by Algorithm 1 are respectively, the PA estimates for the first and second derivatives of ET with respect to θ based on observing m busy periods. Note that $T1$ is the same estimate stated in Suri and Zazanis [29].

The following is a corollary to Theorems 2 and 3.

Corollary 4 *For a system satisfying* **C.1, A.1,** *and* **A.2***, the output $T1(m)$ of Algorithm 1 is a strongly consistent estimate of $\frac{dET}{d\theta}$. If in addition* **C.2, A.3,** *and* **A.4** *are satisfied, $T2(m)$ is a strongly consistent estimate of $\frac{d^2ET}{d\theta^2}$.*

Proof: With the notation of §3, the total number of customers in m busy periods is equal to $\sum_{k=1}^{m} N_k$ and the output of Algorithm 1 is given by

$$T1(m) \;=\; \frac{1}{\sum_{k=1}^{m} N_k} \sum_{k=1}^{m} \left[\sum_{i=M_{k-1}+1}^{M_k} \sum_{j=M_{k-1}+1}^{i} \frac{dX_j}{d\theta}\right] \ , \tag{88}$$

and

$$\begin{aligned} T2(m) \;&=\; \frac{1}{\sum_{k=1}^{m} N_k} \sum_{k=1}^{m} \left[\sum_{i=M_{k-1}+1}^{M_k} \sum_{j=M_{k-1}+1}^{i} \frac{d^2X_j}{d\theta^2}\right] \\ &+\; \frac{1}{m} \sum_{k=1}^{m} \left[\frac{g(Z_k)}{1-G(Z_k)} \left(\sum_{i=M_{k-1}+1}^{M_k} \frac{dX_i}{d\theta}\right)^2\right] \ . \end{aligned} \tag{89}$$

Each one of the three quantities indexed by k inside the square brackets in (88) and (89) form an i.i.d. sequence of random variables, $k = 1, 2, \ldots, m$, since they are obtained from consecutive busy periods. Consider the following inequalities:

$$|\sum_{i=M_{k-1}+1}^{M_k} \sum_{j=M_{k-1}+1}^{i} \frac{dX_j}{d\theta}| \;\leq\; N_k \sum_{i=M_{k-1}+1}^{M_k} \xi_i = N_k \Phi_k \ ,$$

$$|\sum_{i=M_{k-1}+1}^{M_k} \sum_{j=M_{k-1}+1}^{i} \frac{d^2X_j}{d\theta^2}| \;\leq\; N_k \sum_{i=M_{k-1}+1}^{M_k} \psi_i = N_k \Psi_k \ ,$$

and

$$\frac{g(Z_k)}{1-G(Z_k)}\left(\sum_{i=M_{k-1}+1}^{M_k}\frac{dX_i}{d\theta}\right)^2 \;\le\; \alpha\left(\sum_{i=M_{k-1}+1}^{M_k}\xi_i\right)^2 = \alpha\Phi_k^2.$$

Since by Lemma 7 of the Appendix $EN_k^2 < \infty$ and by Lemma 8 $E\Phi_k^2 < \infty$ and $E\Psi_k^2 < \infty$, it follows from the Cauchy-Schwartz inequality that the lhs of the above inequalities have finite expectations. The corollary follows then from the Strong Law of Large Numbers and Theorems 2 and 3. □

Algorithm 2 shows that estimating Hessian matrices simply involves using additional variables accumulating quantities computed from each busy period but does not complicate in any other way our task. This fact, which is clear from (44),(82), and (87) results in considerable computational savings. (For simplicity only two entries of the Hessian are estimated in Algorithm 2).

Algorithm 2

Start with arrival to idle system

INITIALIZE:

Set the following variables to zero:

$DX1$, $DX2$, DA, SDA, DDX, $SDDX$, BPA, BPX, nbp, n

Set m =number of busy periods to be observed

NEXT-SERVICE:

Observe next service time $X(\theta_1, \theta_2)$

$n = n + 1$

$DX1 = DX1 + \frac{\partial X}{\partial \theta_1}$

$DX2 = DX2 + \frac{\partial X}{\partial \theta_2}$

$DDX = DDX + \frac{\partial^2 X}{\partial \theta_1 \partial \theta_2}$

$SDDX = SDDX + DDX$

NEXT-ARRIVAL:

Observe next interarrival time $A(\eta)$.

If arrival finds system busy then

$DA = DA + \frac{dA}{d\eta}$

$SDA = SDA + DA$

$DDA = DDA + \frac{d^2 A}{d\eta^2}$

$SDDA = SDDA + DDA$

If arrival finds system idle then

Z =Length of last interarrival time - Length of idle period

$BPX12 = BPX12 + \frac{g(Z)}{1-G(Z)}(DX1)(DX2)$

$BPA = BPA + \frac{g(Z)}{1-G(Z)}(DA)^2$

$DX1 = 0$, $DX2 = 0$, $DDX = 0$, $DA = 0$, $DDA = 0$

$nbp = nbp + 1$

STOPPING CRITERION

If $nbp < m$ then go to NEXT-SERVICE

FINAL VALUE

$T12(m) = \frac{SDDX}{n} + \frac{BPX}{M}$ (An estimate for $\frac{\partial^2 ET}{\partial\theta_1\partial\theta_2}$).

$TA(m) = \frac{SDDA}{n} + \frac{BPA}{m}$ (An estimate for $\frac{\partial^2 ET}{\partial\eta^2}$).

Print $T12(m)$, $TA(m)$.

STOP

Corollary 5 *If* **VC.2, VA.2, A3,** *and* **VA.4** *hold, Algorithm 2 gives a strongly consistent esitmate of the Hessian matrix.*

The proof is similar to that of Corollary 3 we will omit it.

13 Experimental Results

The results presented here were obtained by simulating a GI/GI/1 queue using regenerative techniques along with the bias reduction method of Meketon and Heidelberger [19]. The length of the sample paths was approximately 100000 customers and the confidence intervals were at a 95% level.

In Table 1 we present the output of the PA algorithm for an M/G/1 queue with arrival rate a^{-1} and service time distribution uniform in $[\theta - \delta, \theta + \delta]$ and we compare the PA derivative estimates with the true values obtained from the P-K formula (Kleinrock [17]). Experiments were done for (a, θ, δ) equal to (100, 20, 16), (100, 50, 40), and (100, 80, 64), corresponding to traffic intensities $\rho = 0.2, 0.5$, and 0.8. With three parameters, a, θ, and δ, the Hessian matrix of ET has six different entries and if we were to use finite difference estimates we would need 19 experiments for each value of ρ instead of a single one required by our approach. It is interesting to notice the accuracy of the results and the surprisingly tight confidence intervals.

In Table 2 we present results for an analytically intractable system with interarrival times uniformly distributed in $[a-\delta, a+\delta]$ and service times with the following triangular *density* function

$$f(x,\theta) \;=\; \begin{cases} x/\theta^2 & \text{if } 0 \le x < \theta \\ 2/\theta - x/\theta^2 & \text{if } \theta \le x < 2\theta \\ 0 & \text{otherwise} \end{cases} .$$

The nominal experiment is observed for parameter values $a = 100, \delta = 80, \theta = 70$. Since no analytic results are available, we compare the PA estimates with finite difference estimates. The first row of the table shows the PA estimates, while the second shows the finite difference estimates obtained using increments $\Delta\theta = 3, \Delta a = 5, \Delta\delta = 10$. Thus the second row required 19 experiments while the first row was obtained from a single one.

In addition to the fact that the PA estimates required 19 times fewer experiments, it is interesting to notice the extent to which PA estimates outperform the conventional symmetric difference estimate T''_{SD}. In fact the confidence intervals of the former are an order of magnitude smaller than that of the latter. This is not accidental. As a matter of fact, in Zazanis and Suri [31], it is shown that the Mean Square Error (MSE) of T''_{PA} goes to zero as $O(\frac{1}{m})$ when the number of busy periods in the sample path m goes to infinity. On the other hand, the MSE of T''_{SD} goes to zero at best as $O(\frac{1}{M^{1/3}})$. This demonstrates the asymptotic superiority of the PA estimate from a numerical point of view as well.

These experimental results also suggest that algorithms 1 and 2 may provide unbiased estimates for a wider class of distributions than those restricted by **A.1 - A.4** (since the hazard rate for the arrival density is not bounded in this example).

Table 1: Results for an M/G/1 Queue

Traffic Intensity $\rho =$		0.2		0.5		0.8	
$\frac{\partial^2 T}{\partial \theta^2}$	PA	1.97 ± 0.02	$\times 10^{-2}$	8.23 ± 0.26	$\times 10^{-2}$	1.44 ± 0.18	
	P-K Formula	1.97		8.43		1.42	
$\frac{\partial^2 T}{\partial \delta^2}$	PA	4.13 ± 0.04	$\times 10^{-3}$	6.59 ± 0.14	$\times 10^{-3}$	1.68 ± 0.11	$\times 10^{-2}$
	P-K Formula	4.17		6.70		1.67	
$\frac{\partial^2 T}{\partial \alpha^2}$	PA	9.45 ± 0.13	$\times 10^{-4}$	2.38 ± 0.08	$\times 10^{-2}$	0.98 ± 0.13	
	P-K Formula	9.48		2.43		0.97	
$\frac{\partial^2 T}{\partial \theta \partial \delta}$	PA	8.20 ± 0.21	$\times 10^{-4}$	5.23 ± 0.27	$\times 10^{-3}$	5.42 ± 0.90	$\times 10^{-2}$
	P-K Formula	8.33		5.33		5.33	
$\frac{\partial^2 T}{\partial \theta \partial \alpha}$	PA	-4.07 ± 0.05	$\times 10^{-3}$	-4.34 ± 0.14	$\times 10^{-2}$	-1.18 ± 0.15	
	P-K Formula	-4.07		-4.43		-1.17	
$\frac{\partial^2 T}{\partial \delta \partial \alpha}$	PA	-7.92 ± 0.85	$\times 10^{-4}$	-5.18 ± 0.46	$\times 10^{-3}$	-5.44 ± 1.10	$\times 10^{-2}$
	P-K Formula	-8.33		-5.33		-5.33	

Table 2: Results for a GI/GI/1 Queue		
Partial derivatives	Symmetric Difference (SD) estimate	Perturbation Analysis (PA) estimate
$\frac{\partial^2 T}{\partial \theta^2}$	$1.73 \pm 2.48 \times 10^{-1}$	$1.32 \pm 0.09 \times 10^{-1}$
$\frac{\partial^2 T}{\partial \alpha^2}$	$1.22 \pm 0.78 \times 10^{-1}$	$1.01 \pm 0.07 \times 10^{-1}$
$\frac{\partial^2 T}{\partial \delta^2}$	$8.9 \pm 19.0 \times 10^{-3}$	$10.5 \pm 0.51 \times 10^{-3}$
$\frac{\partial^2 T}{\partial \theta \partial \alpha}$	$-1.27 \pm 0.31 \times 10^{-1}$	$-1.17 \pm 0.08 \times 10^{-1}$
$\frac{\partial^2 T}{\partial \theta \partial \delta}$	$2.62 \pm 1.44 \times 10^{-2}$	$-3.05 \pm 0.20 \times 10^{-2}$
$\frac{\partial^2 T}{\partial \delta \partial \alpha}$	$-3.0 \pm 0.9 \times 10^{-2}$	$-2.98 \pm 0.18 \times 10^{-2}$

14 APPENDIX

We begin by establishing (7) which we repeat here for ease of reference.

Lemma 7 *Under assumptions* **A.1** *and* **A.2**,

$$ET(\theta + \Delta\theta) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} T_i(\theta + \Delta\theta) = \lim_{m\to\infty} \frac{1}{M_m(\theta)} \sum_{i=1}^{M_m(\theta)} T_i(\theta + \Delta\theta) \text{ w.p. } 1. \qquad (7)$$

Proof: Define the counting process corresponding to the discrete time renewal process $\{M_k(\theta); k = 0, 1, 2 \ldots\}$ via $R_n(\theta) = \sum_{k=0}^{\infty} 1\{M_k(\theta) < n\}$, for $n = 1, 2, \ldots$. (Thus $R_n(\theta) = l$ iff at parameter value θ the n'th customer, C_n, belongs to the lth busy period). For typographical convenience the number of customers in the busy period which contains customer C_n will be denoted by $N_{R_n}(\theta)$ (instead of $N_{R_n(\theta)}(\theta)$). Let us also denote by $J_n(\theta)$ the index of the customer who initiates the busy period in which C_n belongs. With the above convention, we write as $J_n(\theta) = M_{R_n-1}(\theta) + 1$. We also denote by $Q_k(\theta)$ the area under the kth busy period as a function of θ (not to be confused with $S_k(\theta)$ defined in §4). Then $Q_{R_n}(\theta+\Delta\theta) \stackrel{\text{def}}{=} \sum_{i=M_{R_n-1}(\theta+\Delta\theta)+1}^{M_{R_n}(\theta+\Delta\theta)} T_i(\theta+\Delta\theta)$ is the area under the busy period in which C_n belongs at parameter value $\theta+\Delta\theta$.

To establish Lemma 7 it suffices to show that the second limit in (70) exists and equals the first. Indeed the first limit can be written as

$$\lim_{n\to\infty} \frac{J_n(\theta)-1}{n} \left(\frac{1}{J_n(\theta)-1} \sum_{i=1}^{J_n(\theta)-1} T_i(\theta+\Delta\theta) + \frac{1}{J_n(\theta)-1} \sum_{i=J_n(\theta)}^{n} T_i(\theta+\Delta\theta) \right) . \tag{90}$$

As a consequence of the above definitions $\lim_{n\to\infty} \frac{1}{J_n(\theta)-1} \sum_{i=1}^{J_n(\theta)-1} T_i(\theta+\Delta\theta)$ exists iff the limit on the rhs of (70) exists and in that case they have the same value. We first establish that

$$\lim_{n\to\infty} \frac{J_n(\theta)}{n} = 1 \quad \text{w.p. } 1. \tag{91}$$

In view of the inequality $J_n \le n < J_n + N_{R_n}$ it suffices to show that $\lim_{n\to\infty} \frac{N_{R_n}}{n} = 0$ w.p. 1. This in turn follows readily from the fact that, with probability one, $R_n \le n$, $\lim_{n\to\infty} R_n = \infty$, and $\lim_{k\to\infty} N_k/k = 0$. (Since the r.v.'s $\{N_k\}$ have the same marginal distribution and $EN_1 < \infty$, the last claim is a direct consequence of the Borel-Cantelli lemma).

Similarly, for $\Delta\theta > 0$, a monotonicity argument shows that $J_n(\theta+\Delta\theta) \le J_n(\theta)$ w.p. 1 and hence the inequality $\sum_{i=J_n(\theta)}^{n} T_i(\theta+\Delta\theta) \le \sum_{i=J_n(\theta+\Delta\theta)}^{J_n(\theta+\Delta\theta)+N_n(\theta+\Delta\theta)} T_i(\theta+\Delta\theta) = Q_{R_n}(\theta+\Delta\theta)$ holds w.p. 1. The second term inside the parenthesis of (90) is dominated by $\frac{R_n(\theta+\Delta\theta)}{J_n(\theta)-1} \frac{1}{R_n(\theta+\Delta\theta)} Q_{R_n}(\theta+\Delta\theta)$ and is nonnegative. Since $\{Q_k(\theta+\Delta\theta)\}$ is an i.i.d. sequence and $\lim_{n\to\infty} R_n(\theta+\Delta\theta) = \infty$ w.p. 1, the argument of the preceding paragraph applies again and $\frac{1}{J_n(\theta)-1} \sum_{i=J_n(\theta)}^{n} T_i(\theta+\Delta\theta)$ is thus seen to vanish w.p. 1 as $n \to \infty$. In view of (91) and the remarks that precede it, the proof of the lemma is complete. □

Lemma 8 *Under assumption* **A.2***,* $EN_1(\theta)^2 < \infty$ *for all* $\theta \in [a, b]$.

Proof: **A.2** guarantees that $EX_1(\theta)^2 < \infty$ which implies that $EN_1(\theta)^2 < \infty$ on $[a, b]$ (see Wolff [30]). □

Lemma 9 *Let* $\Phi_1 = \sum_{i=1}^{N_1} \xi_i$, $\Psi_1 = \sum_{i=1}^{N_1} \psi_i$. *Assumption* **A.2** *implies that* $E\Phi_1^2 < \infty$, *and assumption* **A.4** *that* $E\Psi_1^2 < \infty$.

Proof: $EN_1^2 < \infty$ (Lemma 5). From this and the fact that $E\xi_1^3 \leq \infty$ it follows that $E\Phi_1^2 \leq \infty$. (See Gut [11, p.22].) The proof of the second statement is identical. □

We continue with an elementary inequality for N_1, the number of customers in the busy period of a GI/GI/1 queue:

Lemma 10 *Under assumptions* **A.1, A.2***, and* **A.4***,*

$$P(N_1(\theta) > k) \;\leq\; l^k, \qquad \text{with } \; 0 < l < 1 \;, \tag{92}$$

for $\theta \in [a, b]$.

Proof: The difference between interarrival and service times can be bounded for all $\theta \in [a, b]$ as follows: $X_i(\theta) - A_i \leq \chi_i - A_i \stackrel{\text{def}}{=} \Gamma_i$. Then

$$\begin{aligned} P(N_1(\theta) > k) \;&=\; P(\sum_{j=1}^{i} X_j(\theta) - A_j \;>\; 0, i = 1, 2, \ldots, k) \;\leq\; P(\sum_{j=1}^{k} X_j(\theta) - A_j > 0) \\ &\leq\; P(\sum_{j=1}^{k} \Gamma_j > 0) \;\; . \end{aligned}$$

From Chernoff's theorem [6] we have

$$P(\sum_{j=1}^{k} \Gamma_j \;>\; 0\,) \;\;\leq\;\; l^k \;, \tag{93}$$

with $l = \inf_{t \in \mathbf{R}_+} E[e^{t\Gamma_1}] < 1$. (The existence of such l is guaranteed by **A.4**. Because of the definition of Γ_i, Inequality (92) holds for all $\theta \in [a, b]$). □

Now we are ready to state and prove

Lemma 11 *For a stable GI/GI/1 queue with service time satisfying assumptions* **A.2***, and* **A.4***, there exists* $\gamma > 0$ *such that*

$$E[e^{\gamma \sum_{i=1}^{N_1} \xi_i}] \leq \overline{K} < \infty \; . \tag{94}$$

Proof: It is enough to show that, for sufficiently large u, there exist positive c and A such that $P(\sum_{i=1}^{N_1} \xi_i > u) \leq Ae^{-cu}$. Let $a > E[\xi_i]$ and k be an integer such that

$$ka \;\leq\; u \;<\; (k+1)a \; .$$

Then,

$$P(\sum_{i=1}^{N_1} \xi_i \;>\; u) \;\leq\; P(\sum_{i=1}^{k} \xi_i \;>\; ka) \;+\; P(N_1 > k) \; .$$

Also, let $m(a) = \inf_{t \in \mathbf{R}_+} E[e^{t[\xi_1 - a]}]$. Clearly since $a > E[\xi_1]$, $m(a) < 1$. Using Chernoff's theorem once more we get

$$P(\sum_{i=1}^{k} \xi_i \;>\; ka) \;\leq\; [m(a)]^k \;\leq\; \frac{1}{m(a)}\, e^{-ua^{-1}|\log m(a)|} \;=\; A_1 e^{-uc_1} \;\; ,$$

with $c_1 = a^{-1}|\log m(a)|$. On the other hand, from Lemma 10 it readily follows that

$$P(N_1 \;>\; k) \;\leq\; A_2 e^{-uc_2} \; .$$

From the above follows that

$$P(\sum_{i=1}^{N_1} \xi_i \;>\; u\,) \;<\; Ae^{-cu} \; . \tag{95}$$

This establishes the existence of $\overline{K} < \infty$ such that (94) holds. □

The next lemma provides the martingale stability argument which establishes (21).

Lemma 12 *Under assumptions* **A.1** *and* **A.2** *(21) holds, i.e.*

$$\frac{1}{m-1}\sum_{k=1}^{m-1} V_k(N_{k+1}-EN_1) \;\to\; 0 \quad \text{w.p.1} .$$

Proof: From Lemma 8 $EN_1^2 < \infty$, hence $EN_1^{3/2} < \infty$. However, $E\Phi_1^3 \leq 2c_3\, E[\xi_1]^3\, EN_1^{3/2}$, where c_3 is a numerical constant (Gut [11, p.20]). Using **A.2** we conclude that $E\Phi_1^3 < \infty$. From this follows that the second moment of the steady state distribution of the auxiliary system is finite, i.e. $EV^2 < \infty$.

Now let $\{\mathcal{F}_k; k=1,2,...\}$ be the history of the process until the end of the k' th *busy cycle.* From (11) we check that V_k is $\mathcal{F}_k$- measurable and a moment's reflection shows that $V_k(N_{k+1}-EN_1)$ is a $\mathcal{F}_k$- martingale difference sequence. Now use the independence of N_{k+1} and V_k and the fact that $EV_k^2 < EV^2 < \infty$ which follows from the stochastic monotonicity of V_k and the finiteness of $E\Phi_1^3$ established in the previous paragraph:

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{1}{k^2}E[V_k^2(N_{k+1}-EN_1)]^2 \quad &= \quad \mathrm{Var}(N_1)\sum_{k=1}^{\infty}\frac{EV_k^2}{k^2} \qquad (96)\\ &< \quad \mathrm{Var}(N_1)\; EV^2\sum_{k=1}^{\infty}\frac{1}{k^2} \;<\; \infty .\end{aligned}$$

From (96) and a standard martingale stability argument (Stout [25, p.5]) we conclude that (21) holds. □

The final two lemmas presented here are needed in the proof of Theorems 1 and 2. The first states an inequality necessary in establishing the continuity of $ET(\theta)$ while the second is a standard result from real analysis presented without proof.

Lemma 13 *Under* **M.1** *the following inequality holds for $\theta - a \geq \Delta\theta \geq 0$:*

$$0 \geq ET(\theta-\Delta\theta) - ET(\theta) \geq \frac{E[\sum_{i=1}^{N_1}\sum_{j=1}^{i} X_j(\theta-\Delta\theta) - X_j(\theta)]}{E[N_1]} . \qquad (97)$$

Proof: Our analysis here parallels that of sections 4 and 5 and examines the effect of negative perturbations introduced to the sample path of the system. It is not carried out at the same level of detail since our aim here is simply to establish inequality (97). With the same notation, we have the following counterpart of (12)

$$0 \geq T_i(\theta - \Delta\theta) - T_i(\theta) \geq \sum_{j=M_{k-1}+1}^{i} X_j(\theta - \Delta\theta) - X_j(\theta) \,, \qquad \text{for} \quad M_{k-1} < i \leq M_k \,. \tag{98}$$

Inequality (98) expresses the fact that, when perturbations are negative, busy periods cannot coalesce. Hence, the change in the system time of customer C_i is negative with absolute value equal to $\sum_{j=M_{k-1}+1}^{i} |X_j(\theta - \Delta\theta) - X_j(\theta)|$, if BP_k, the busy period in which C_i belongs in the unperturbed sample path does not break up, or smaller if it does. The counterpart of (13) becomes

$$\begin{aligned} 0 \geq \Delta S_k = S_k(\theta - \Delta\theta) - S_k(\theta) &= \sum_{i=M_{k-1}+1}^{M_k} T_i(\theta - \Delta\theta) - T_i(\theta) \\ &\geq \sum_{i=M_{k-1}+1}^{M_k} \sum_{j=M_{k-1}+1}^{i} X_j(\theta - \Delta\theta) - X_j(\theta) \,. \end{aligned} \tag{99}$$

Summing the terms of the above inequality for $k = 1, \ldots, m$, dividing both sides by $\sum_{k=1}^{m} N_k$ and letting $m \to \infty$ yields (97). The details of the argument are similar to those in section 4. □

Lemma 14 *[22, p.95] Let $f : [a, b] \to \mathbf{R}$ be continuous. Let one of the Dini derivatives of f be Riemann integrable. Then so are the others, and all Dini derivatives are equal a.e. on $[a, b]$. If $\tilde{D}f$ denotes any one of the four Dini derivatives of f, then*

$$\int_x^y \tilde{D}f(t)dt = f(y) - f(x) \qquad (x, y \in [a, b]) \,.$$

□

References

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

[2] Baccelli, F., and P. Brémaud, (1986). *Palm Probabilities*, Lecture Notes in Statistics, Springer Verlag.

[3] Brémaud and Lasgouttes (1992). Stationary IPA estimates for non-smooth functions of the GI/GI/1/∞ workload. *Rapp. de Rech. INRIA No. 1677.*

[4] Konstantopoulos, P. and M. Zazanis (1992). "Sensitivity Analysis for Stationary and Ergodic Queues", *J. Appl. Prob.* **24**, 738-750.

[5] Konstantopoulos, P. and M. Zazanis (1993). A note on the sensitivity analysis for stationary and ergodic queues. To appear in *Adv. Appl. Prob.*

[6] Chernoff, H. (1952). "A measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations", *Ann. Math. Stat.,* **23**, 493-507.

[7] Glasserman, P., (1990). *Gradient Estimation via Perturbation Analysis,* Kluwer Academic.

[8] Glasserman, P., (1992). "Structural Conditions for Perturbation Analysis Derivative Estimation: Finite Time Performance Indices", *Oper. Res.,* **39,** 5, 724-738.

[9] Glasserman, P., J.Q. Hu, and S.G. Strickland (1991). "Strongly Consistent Steady State Derivative Estimates", *Probability in the Informational and Engineering Sciences,* **5,** 4, 391-414.

[10] Glynn, P.W. (1987). "Likelihood Ratio Gradient Estimation: An overview", *Proceedings of the 1987 Winter Simulation Conference,* pp. 366-375.

[11] Gut, A., (1988). *Stopped Random Walks,* Springer.

[12] Heidelberger, P., X.R. Cao, M.A. Zazanis, and R. Suri (1988). "Convergence Properties of Infinitesimal Perturbation Analysis", *Mgmt. Sci.*, **34**, 11, 1281-1302.

[13] Ho, Y.C., and X.R. Cao (1983). "Perturbation Analysis and Optimization of Queueing Networks", *J. Optim. Theory Applic.*, **40**, 4, 559-582.

[14] Ho, Y.C., and X.R. Cao (1991). *Perturbation Analysis of Discrete Event Dynamic Systems,* Kluwer Academic.

[15] Hu, J.Q. (1992). "Convexity of Sample Path Performance and Strong Consistency of Infinitesimal Perturbation Analysis Estimates," *IEEE Trans. Aut. Contr.,* **37,** 2, 258-262.

[16] Kingman, J.F.C., (1962). "Some Inequalities for the Queue GI/G/1", *Biometrika*, **49,** 3 and 4, 315-324.

[17] Kleinrock, L. (1975). *Queueing Systems I,* John Wiley.

[18] L'Ecuyer, P. N. Giroux, and P.W. Glynn (1989). "Stochastic Optimization by Simulation: Some Experiments with a Simple Steady State Queue", Working Paper, Universite Laval.

[19] Meketon, M.S. and P. Heidelberger (1982). "A Renewal Theoretic Approach to Bias Reduction in Regenerative Simulations", *Mgmt. Sci.*, **28**, No.2, 173-181.

[20] Reiman M.I. and A. Weiss, (1989). "Sensitivity Analysis of Simulations via Likelihood Ratios", *Oper. Res.*, **37**, 5, 830-844.

[21] Reiman M.I. and B. Simon, (1989). "Open Queueing Systems in Light Traffic" , *Math. Oper. Res.*, **14**, 1, 26-59.

[22] Rooij, A.C.M. van and W.H. Schikhof (1982). *A Second Course on Real Functions,* Cambridge University Press.

[23] Rubinstein, R.Y., (1989). "Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models", *Oper. Res.*, **37**, 1, 72-81.

[24] Shiryayev, A.N., (1984). *Probability,* Springer.

[25] Stout, W.F., (1974). *Almost Sure Convergence,* Academic Press.

[26] Suri, R. (1987). "Infinitesimal Perturbation Analysis for General Discrete Event Systems," *J. ACM*, **34**, 3, 686-717.

[27] Suri, R. (1989). "Perturbation Analysis: The State of the Art and Research Issues Explained Via the GI/G/1 Queue", *Proceedings of the IEEE*, **77**, 1, 114-137.

[28] Suri, R. and Y.T. Leung (1989). "Single Run Optimization of Discrete Event Systems: Experimental Investigation for the M/M/1 Queue", *IIE Trans.*, **21**, 1, 35-49.

[29] Suri, R. and Zazanis, M.A. (1988). "Perturbation Analysis Gives Strongly Consistent Sensitivity Estimates for the M/G/1 Queue", *Mgmt. Sci.*, **34**, 1, 39-64.

[30] Wolff, R.W. (1989). *Stochastic Modeling and the Theory of Queues,* Prentice Hall.

[31] Zazanis, M.A. and R. Suri (1993). "Convergence Rates of Finite-Difference Sensitivity Estimates for Stochastic Systems", *Oper. Res.,* **41,** 694-703.

[32] Zazanis, M.A. (1987). "Weak Convergence of Sample Path Derivatives for GI/G/1 Queues", *Proceedings, 25th Allerton Conference on Communication, Control, and Computing,* 297-304.