

## INFINITESIMAL PERTURBATION ANALYSIS AND THE REGENERATIVE STRUCTURE OF THE GI/G/1 QUEUE

Rajan Suri

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

Michael A. Zazanis

Dept. of Industrial Engineering  
 and Management Sciences  
 Northwestern University  
 Evanston, IL 60201

### ABSTRACT

The strong consistency of infinitesimal perturbation analysis estimates for the mean system time is shown using the regenerative structure of the GI/G/1 queue. The analysis throws some light on the conditions which are required for the consistency of infinitesimal perturbation analysis in more general systems with regenerative structure.

### 1. Introduction

Consider a sequence of r.v.'s  $T_{i,\theta}$ ,  $i=0, 1, \dots$ , which depend on a parameter  $\theta \in [a, b]$ . We will assume that, for each value of  $\theta$ , the sequence is regenerative w.r.t. the discrete time renewal process  $Q_{k,\theta}$ ,  $k=0, 1, \dots$ . As it is well known, for any given  $\theta$ , under very mild conditions, this process converges weakly to a steady state random variable  $T_\theta$  with expected value given by

$$E[T_\theta] = \frac{E\left[\sum_{i=0}^{Q_{1,\theta}-1} T_{i,\theta}\right]}{E[Q_{1,\theta}]} \quad (1)$$

We are interested in estimating the derivative  $\frac{d}{d\theta} E[T_\theta]$  (which we will assume to exist) without the use of finite difference estimators. We will assume that it is possible to consider this collection of sequences of r.v.'s indexed by  $\theta$  as a sequence of random functions of  $\theta$  such that the derivatives  $\frac{dT_{i,\theta}}{d\theta}$  exist w.p.1 and can

be computed from the sample path by means of the infinitesimal perturbation analysis (IPA) algorithm (e.g. see Glynn, 1987, Suri and Zazanis, 1987). It can be shown that, for given  $\theta$ , these derivatives converge weakly to a steady state r.v.  $\frac{dT_\theta}{d\theta}$  with expected value given by

$$E\left[\frac{dT_\theta}{d\theta}\right] = \frac{E\left[\sum_{i=0}^{Q_{1,\theta}-1} \frac{dT_{i,\theta}}{d\theta}\right]}{E[Q_{1,\theta}]} \quad (2)$$

Let us replace  $Q_{1,\theta}$  by  $Q_\theta$  for typographical convenience and let  $L_\theta = \sum_{i=0}^{Q_\theta-1} T_{i,\theta}$  and  $\frac{dL_\theta}{d\theta} = \sum_{i=0}^{Q_\theta-1} \frac{dT_{i,\theta}}{d\theta}$ . The condition for the consistency of the IPA estimate is then

$$\begin{aligned} \frac{d}{d\theta} E[L_\theta] &= E\left[\frac{dL_\theta}{d\theta}\right] \\ &+ \frac{E[L_\theta]}{E[Q_\theta]} \frac{d}{d\theta} E[Q_\theta] \end{aligned} \quad (3)$$

(see Heidelberger et al, 1987). In this paper we will prove that this condition is satisfied when  $T_{i,\theta}$  is the response time of the  $i^{\text{th}}$  customer in a GI/G/1 system, and we will give sufficient conditions for its validity for more general systems as well.



## 2. Analysis of the GI/G/1 queue

Consider a GI/G/1 system with interarrival times  $A_0, A_1, \dots$  distributed according to  $G$  (which we will assume absolutely continuous with density  $g$ ) and service requirements  $\theta X_0, \theta X_1, \dots$  where the  $X_i$  are distributed according to  $F$ . We will further assume that  $G$  has bounded hazard rate and that both  $F$  and  $G$  have finite second moments. Also let us denote by  $T_{i,\theta}$  the system time of the  $i^{\text{th}}$  customer and assume that  $T_{0,\theta} = \theta X_0$ . Without loss of generality let us assume that the system is stable when  $\theta \in [a, b]$ . Let  $Q_{k,\theta}$ ,  $k=0, 1, 2, \dots$ , be the index of the customer who initiates the  $(k+1)^{\text{th}}$  busy period. ( $Q_{0,\theta} = 0$ ). Also let us denote by  $L_\theta = \sum_{i=0}^{Q_\theta-1} T_{i,\theta}$  the area under the curve of the first busy period. Since  $T_{i,\theta} = \theta X_i + [T_{i-1,\theta} - A_{i-1}]^+$ , it is easy to show that the derivatives  $\frac{dT_{i,\theta}}{d\theta}$  exist w.p.1.

For reasons that will become apparent in the sequel, we will also consider a modified queueing system with interarrival sequence  $\hat{A}_1 = A_{Q_\theta}, \hat{A}_2 = A_{Q_\theta+1}, \dots$ , and service sequence  $\hat{X}_1 = R + (\theta + \Delta\theta)X_{Q_\theta}, \hat{X}_2 = (\theta + \Delta\theta)X_{Q_\theta+1}, \dots$ . Thus, when  $R = 0$  and  $\Delta\theta = 0$ , the sample paths of the modified system are replicas of the original (except for the fact that the first busy period of the original system is deleted). However, in general,  $\hat{X}_n$  can be any r.v. independent of the sequence  $\{(X_n, A_n)\}_{n \geq Q_\theta}$ .

Returning to the original system, let us see how  $Q_\theta$  and  $L_\theta$  change when  $\theta$  is increased by  $\Delta\theta$ . Let  $Y_\theta$  and  $I_\theta$  be the length of the first busy period and busy cycle respectively and let  $\Delta T_{i,\theta} = T_{i,\theta+\Delta\theta} - T_{i,\theta}$ , and  $\Delta Y_\theta = Y_{\theta+\Delta\theta} - Y_\theta$ . Then

$$Q_{\theta+\Delta\theta} = Q_\theta + 1_{\{\Delta Y_\theta > I_\theta\}} \hat{Q} \quad (4)$$

$$L_{\theta+\Delta\theta} = L_\theta + \sum_{i=0}^{Q_\theta-1} \Delta T_{i,\theta} + 1_{\{\Delta Y_\theta > I_\theta\}} \hat{L} \quad (5)$$

where  $\hat{Q}$  is the number of customers and  $\hat{L}$  the area under the curve of the first busy period of the modified system with  $R = (\Delta Y_\theta - I_\theta)^+$ .

Let  $F = \sigma\{Q_\theta; X_0, X_1, \dots, X_{Q_\theta-1}; A_0, A_1, \dots, A_{Q_\theta-2}\}$ , when  $Q_\theta > 1$  and  $F = \sigma\{X_0\}$  when  $Q_\theta = 1$ . This  $\sigma$ -field describes the information available at the end of the first busy period. Taking conditional expectation w.r.t.  $F$ , and noting that  $\Delta T_{i,\theta} = \frac{dT_{i,\theta}}{d\theta} \Delta\theta$  for  $i=0, 1, \dots, Q_\theta-1$ , we obtain

$$E[Q_{\theta+\Delta\theta} | F] = Q_\theta + E[1_{\{\Delta Y_\theta > I_\theta\}} \hat{Q} | F] \quad (6)$$

$$E[L_{\theta+\Delta\theta} | F] = L_\theta + \sum_{i=0}^{Q_\theta-1} \frac{dT_{i,\theta}}{d\theta} \Delta\theta + E[1_{\{\Delta Y_\theta > I_\theta\}} \hat{L} | F] \quad (7)$$

Let  $\hat{Q}_+$  be the number of customers in the first busy period of the modified system when  $R = \Delta Y_\theta$  and  $\hat{Q}_-$  the number of customers when  $R = 0$ . Clearly, since  $0 \leq (\Delta Y_\theta - I_\theta)^+ \leq \Delta Y_\theta$  we have  $\hat{Q}_- \leq \hat{Q} \leq \hat{Q}_+$ . Notice that  $\hat{Q}_-$  is independent of  $1_{\{\Delta Y_\theta > I_\theta\}}$  (and of  $F$ ) and that, given  $F$ ,  $\hat{Q}_+$  and  $1_{\{\Delta Y_\theta > I_\theta\}}$  are conditionally independent. From the above and the fact that  $Q_\theta$  is  $F$ -measurable follows that

$$\begin{aligned} & \frac{1}{\Delta\theta} E[1_{\{\Delta Y_\theta > I_\theta\}} | F] E[\hat{Q}_-] \quad (8) \\ & \leq \frac{1}{\Delta\theta} E[Q_{\theta+\Delta\theta} - Q_\theta | F] \\ & \leq \frac{1}{\Delta\theta} E[1_{\{\Delta Y_\theta > I_\theta\}} | F] E[\hat{Q}_+ | F] \end{aligned}$$

Let  $z = \sum_{i=0}^{Q_\theta-1} \theta X_i - \sum_{i=0}^{Q_\theta-2} A_i$  (see fig.). Then clearly

$$E[1_{\{\Delta Y_\theta > I_\theta\}} | F] = \frac{G(z + \Delta Y_\theta) - G(z)}{1 - G(z)} \quad (9)$$



### 3. Implications for general regenerative systems

Here we will address the question of the validity of (3) for general regenerative systems. Throughout this section, we will restrict ourselves to the case where  $\{Q_{k,\theta+\Delta\theta}\}_{k=0,1,\dots} \subseteq \{Q_{k,\theta}\}_{k=0,1,\dots}$ , i.e. to the case where the increase by  $\Delta\theta$  in the value of the parameter has the effect that originally distinct regenerative cycles may collapse into a single one. In this case, (11) is shown to hold always (see Heidelberger et al., 1987) and under these circumstances IPA estimates are consistent iff (13) is also satisfied.

As it is shown in Heidelberger et al. (1987), (13) is in general replaced by

$$\begin{aligned} \frac{d}{d\theta} E[L_\theta] &= E\left[\frac{dL_\theta}{d\theta}\right] \\ &+ p_\theta \lim_{\Delta\theta \rightarrow 0} E[L_{\theta-\Delta\theta} - L_\theta \mid Q_{\theta+\Delta\theta} > Q_\theta]. \end{aligned} \tag{14}$$

Intuitively, the meaning of (14) is clear. Assuming for simplicity all changes to be positive,  $E\left[\frac{dL_\theta}{d\theta}\right]$  represents the rate of increase of  $E[L_\theta]$  with  $\theta$  if merging of regenerative cycles is not taken into account,  $p_\theta$  can be seen to be the rate with which merging occurs, and  $\lim_{\Delta\theta \rightarrow 0} E[L_{\theta+\Delta\theta} - L_\theta \mid Q_{\theta+\Delta\theta} > Q_\theta]$  is the effect of the merging given that it occurs. Thus, in this framework, the condition for consistency of IPA estimates becomes

$$\begin{aligned} \lim_{\Delta\theta \rightarrow 0} E[L_{\theta+\Delta\theta} - L_\theta \mid Q_{\theta+\Delta\theta} > Q_\theta] \\ = E[L_\theta]. \end{aligned} \tag{15}$$

The above condition is not always satisfied. As a counterexample consider the case where  $T_{i,\theta}$  is the "backward busy period" of the  $i^{\text{th}}$  customer in a GI/G/1 queue. (This is the time that has elapsed from the beginning of the busy period in which the  $i^{\text{th}}$  customer belongs till his arrival).

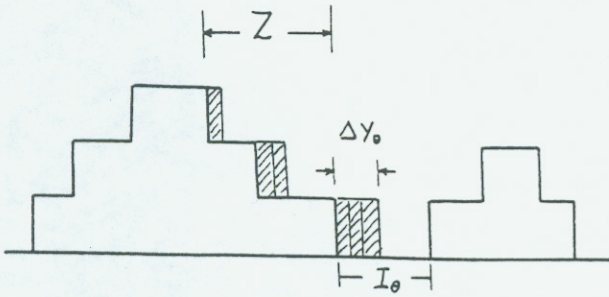


Figure 1.

and it can be shown that

$$\begin{aligned} \lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta} E[1_{\{\Delta Y_\theta > I_\theta\}} \mid F] \\ = \frac{g(z)}{1 - G(z)} \frac{dY_\theta}{d\theta}. \end{aligned} \tag{10}$$

From (8), (9), and (10) it can be shown using the dominated convergence theorem that

$$\frac{d}{d\theta} E[Q_\theta] = p_\theta E[Q_\theta] \tag{11}$$

where

$$p_\theta = E\left[\frac{g(z)}{1 - G(z)} \frac{dY_\theta}{d\theta}\right]. \tag{12}$$

The counterpart of (11) for  $L_\theta$  can be derived in a similar way and is given by

$$\begin{aligned} \frac{d}{d\theta} E[L_\theta] &= E\left[\frac{d}{d\theta} L_\theta\right] \\ &+ p_\theta E[L_\theta]. \end{aligned} \tag{13}$$

By eliminating  $p_\theta$  from (11) and (13) we can see that (3) is satisfied and this establishes the strong consistency of IPA estimates for the GI/G/1 queue. This result was obtained in



A sufficient condition for (15) to hold is given in Zazanis (1987). There it is proved that if (in addition to the conditions given in Heidelberg et al., 1987) for some  $\theta_0 \in (a, b)$  there exists a sequence of r.v.'s  $K_i$ ,  $i=0, 1, \dots$ , such that for all  $\Delta\theta$  with  $\theta_0 + \Delta\theta \in (a, b)$ ,  $|\frac{\Delta T_{i, \theta_0}}{\Delta\theta}| \leq K_i$ ,  $i=0, 1, \dots, Q_b - 1$ , and  $E[\sum_{i=0}^{Q_b - 1} K_i] < \infty$ , then (15) holds and the IPA estimate of  $\frac{d}{d\theta} E[T_\theta]$  at  $\theta = \theta_0$  is strongly consistent.

## REFERENCES

- Glynn, P.W. (1987). "Construction of Process-Differentiable Representations for Parametric Families of Distributions", Technical Report, Mathematics Research Center, University of Wisconsin-Madison.
- Heidelberg, P., X.R. Cao, M.A. Zazanis and R. Suri (1987). "Convergence Properties of Infinitesimal Perturbation Analysis", Technical Report RC 11891. IBM Thomas J. Watson Research Center, Yorktown Heights, New York.
- Suri, R. and Zazanis, M.A. (1987). "Perturbation Analysis Gives Strongly Consistent Sensitivity Estimates for the M/G/1 Queue", to appear in *Management Science*.
- Zazanis, M.A. and R. Suri (1986b). "Estimating First and Second Derivatives of Response Time for GI/G/1 Queues from a Single Sample Path", Working paper # 86-121, Dept. of Industrial Engineering, University of Wisconsin-Madison.
- Zazanis, M.A. (1987). "Infinitesimal Perturbation Analysis for Regenerative Systems Revisited", Technical Report 87-08, IE/MS Dept., Northwestern University.