MATHEMATICAL PROGRAMMING-BASED PERTURBATION ANALYSIS FOR *GI*/*G*/1 QUEUES

He Zhang
Wai Kin (Victor) Chan

Department of Decision Sciences & Engineering Systems
CII 5015, 110 Eighth Street
Rensselaer Polytechnic Institute
Troy, NY 12180, U.S.A.

ABSTRACT

This paper addresses several issues of using the mathematical programming representations of discrete-event dynamic systems in perturbation analysis. In particular, linear programming techniques are used to perform Infinitesimal Perturbation Analysis (IPA) on GI/G/1 queues. A condition for unbiasedness is derived. For finite perturbation analysis (FPA), an upper bound is given for the error term of FPA.

1 INTRODUCTION

Perturbation analysis (PA) has been proved to be an effective method for sensitivity analysis of discrete-event dynamic systems (DEDSs). Most previous research focus on deriving conditions under which PA will give consistent estimates for different DEDSs.

One of the earliest important results can be found in Cao (1985), which gives conditions for convergence of different stochastic experiments. Other later research has shown that the IPA method is effective with some assumptions (see for example, Zazanis 1987; Heidelberger, et al. 1988; Suri and Zazanis 1988; Fu and Hu 1991). Considerable effort has been made to identify assumptions or conditions under which the dominated convergence theorem (Dudley 2002) can be applied to interchange the expectation and derivative—the key of IPA. Most of the existing results were established using renewal theory. More details can be found in (Glasserman and Ho 1991), (Ho and Cao 1991), and (Fu and Hu 1997).

Recently, mathematical programming representations (MPRs) for DSESs have been introduced (Schruben 2000, Chan and Schruben 2003, Chan 2005). These representations create a new way of analyzing DEDSs using mathematical programming techniques. Using the MPRs, Chan (2006) demonstrates how to compute IPA estimators from the dual variables of MPRs. This paper will show that for

GI/*G*/1 queues these estimators are consistent and establish bounds to support the validity of the IPA assumption.

One advantage of having MPRs is that perturbation analysis of DEDSs could be carried out by using sensitivity analysis of mathematical programming, in which a rich theory and tools have already existed. For example, Gal (1979), Gal and Greenberg (1997) provide an extensive discussion on sensitivity or postoptimal analysis of mathematical programming models. Renegar (1994) and Ward and Wendell (1994) review different approaches of sensitivity analysis in linear programming. Being the mathematical programming models for DEDSs, the MPRs could be considered as a bridge that allows these mathematical programming sensitivity analysis methods to be applied to perturbation analysis of DEDSs.

In this paper, we first introduce the MPR of *G*/*G*/1 queues and analyze it using sensitivity analysis. The first result is that the solutions of the MPR provide a consistent IPA estimator. In Section 3, we consider finite perturbation analysis and establish a convergent condition, which is found to be equivalent to the condition given by Cao (1985). We also provide an upper bound for the error of finite perturbation analysis.

2 GENERAL THEORY

2.1 Mathematical Programming Representation of G/G/1 Queue w.r.t Waiting Time

Different MPRs can be developed for the same DEDS; just like different simulation models can be built to model the same system. Chan (2005) gives several MPRs for a *G*/*G*/1 queue. These MPRs look distinct because they are based on different variables, such as finish times, start times, and waiting times. However, these MPRs are equivalent in the sense that they represent the same dynamics of the same system—a *G*/*G*/1 queue. The discussion in this paper will be carried out on the waiting-time MPR. However, the re-

sults of this paper can also be applied to the other two MPRs. The primal of the waiting-time MPR is:

GG1-LP(W):

$$\min \ \frac{1}{n}\sum_{i=1}^{n} W_i$$

$$s.t. \ \ W_i - W_{i-1} \ge s_{i-1} - a_i \quad i = 2,\cdots,n \quad (V_i)$$
$$W_i \ge 0, \qquad i = 1,\cdots,n \quad (U_i)$$
$$W_i \ \text{free} \ \forall i$$

where W_i, $i=1,\ldots,n$, is the waiting time of the ith customer in the sample path, U_i, $i=1,\ldots,n$ and V_i $i=2,\ldots n$ are the correspondent dual variables, s_i is the service time of the ith customer and a_i is the inter-arrival time between the i-1th and i^{th} customers. In this paper, we will assume that both the service time and the inter-arrival time follow a continuous distribution, and that the distribution of the service time has a parameter θ and cdf F. The dual model is:

GG1-LP-Dual(W):

$$\max \ \frac{1}{n}\sum_{i=2}^{n}(s_{i-1} - a_i)V_i$$

$$s.t. \ \ U_1 \qquad - V_2 = 1$$
$$U_i + V_i - V_{i+1} = 1, \ i = 2,\cdots,n-1$$
$$U_n \qquad - V_n = 1$$
$$U_i, V_i \ge 0, \ \forall i$$

Using the matrix notation, the dual model can be written as:

$$\max \ \mathbf{c}^T\mathbf{x}$$
$$s.t. \ \ \mathbf{Ax} = \mathbf{b}$$
$$x_i \ge 0, \text{ for } i = 1,2,\ldots 2n-1$$

where

$$\mathbf{c}_{1\times(2n-1)} = \frac{1}{n}\left[\underbrace{0,\ldots0}_{n}, \underbrace{s_1 - a_2,\ldots,s_{n-1} - a_n}_{n-1}\right]^T$$

$$\mathbf{A}_{n\times(2n-1)} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$\mathbf{b}_{1\times n} = [1,1,\ldots,1]^T$$

$$\mathbf{x}_{1\times 2n-1} = [x_1,\ldots,x_{2n-1}]^T = [U_1,\ldots,U_n,V_2\ldots,V_n]^T$$

Observe that $\mathbf{A}$ has full row rank.

2.2 Notations and Definitions

Define the set of homogeneous solutions of $\mathbf{Ax} = 0$ as $\text{HOM}(\mathbf{A}) = \{\mathbf{x} : \mathbf{Ax} = 0, \mathbf{x} \ge 0\}$. Define its polar as: $\text{HOM}(\mathbf{A})^* = \{\mathbf{y} \in R^{2n-1} : \mathbf{y}^T\mathbf{x} = 0 \text{ for all } \mathbf{x} \in \text{HOM}(\mathbf{A})\}$. Since $\text{HOM}(\mathbf{A})$ is a closed polyhedral cone, the polar $\text{HOM}(\mathbf{A})^*$ is another closed convex cone (see e.g., Theorem 14.1 in Rockafellar 1970). Denoted by Pc the possible value for $\mathbf{c}$, i.e., $Pc=\text{HOM}(\mathbf{A})^*$ (see e.g., Theorem 1 in Ward and Wendell 1990).

We define the critical regions (the subsets of Pc) in the following. Define a basic index set for each basic solution as $J = \{j_1, j_2, \ldots, j_n\}$, where $j_1 < \ldots < j_m$ are the indices of the variables in the solution and is a subset of $\{1,2,\ldots,2n-1\}$. For the same basic solution, let $\mathbf{B} = [a_{j_1}, a_{j_2}, \ldots, a_{j_n}]$ be a sub-matrix of $\mathbf{A}$, where a_{j_k} is the j_k^{th} column vector in matrix $\mathbf{A}$. Let $N = \{1,\ldots,2n-1\}\setminus J$ be the index of the non-basic variables and $\mathbf{A}_N$ be the matrix of non-basic columns. Let $\mathbf{c}_J = [c_{j_1},\ldots c_{j_n}]^T$ be the corresponding coefficients in the objective function and $\mathbf{c}_N$ be the coefficients for the non-basic variables in the objective function. For each particular set J_k and the corresponding $\mathbf{B}_k$, $\mathbf{c}_{J_k}$, $\mathbf{c}_{N_k}$ and $\mathbf{A}_{N_k}$, define Pc_k as: $Pc_k = \{\mathbf{c} \in \text{HOM}(\mathbf{A}) : \mathbf{c}_{N_k} - \mathbf{c}_{J_k}\mathbf{B}_k^{-1}\mathbf{A}_{N_k} \le 0\}$. Pc_k is called the critical region corresponding to the basis J_k. With this definition, one can perform sensitivity analysis without changing the basis. In other words, when $\mathbf{c} \in Pc_k$, the index set of the basis is always J_k.

2.3 Parameter Spaces for Perturbation Analysis

In this section, further properties of the parameter spaces defined in the previous section will be illustrated.

Theorem 1 *Each critical region Pc_k is closed and the interior of Pc_k, defined as*

$$\text{int}(Pc_k) = \{\mathbf{c} \in \text{HOM}(\mathbf{A}) : \mathbf{c}_{N_k} - \mathbf{c}_{J_k}\mathbf{B}_k^{-1}\mathbf{A}_{N_k} < 0\}$$

is relatively open to Pc.

Proof: The first result is immediate because the expression $\mathbf{c}_{N_k} - \mathbf{c}_{J_k}\mathbf{B}_k^{-1}\mathbf{A}_{N_k} \le 0$ can be expressed as:

$$\left[I, -\mathbf{B}_k^{-1}\mathbf{A}_{N_k} \right] \begin{bmatrix} \mathbf{c}_{N_k} \\ \mathbf{c}_{J_k} \end{bmatrix} \le 0$$

which is a closed polyhedron. The second result follows from the fact that the interior of Pc_k is the interior of a closed polyhedron. Therefore, it is relative open to Pc.□

Theorem 2 *If both the service time and inter-arrival time follow a continuous distribution, the probability that the optimization problem GG1-LP(W) degenerates is zero.*

Proof: From Chan (2005), we know that the problem degenerates when a new customer enters a system that is just about to empty; that is, the instant at which the last customer is leaving the system. In this case, this new arriving customer can be considered either as a member of the previous busy period or as a founder who initializes a new busy period. Suppose this new customer is indexed by $n+1$ and the leaving customer, i.e., the nth customer, enters the system during the busy period started by the mth customer. The following equality holds:

$$\sum_{i=m+1}^{n+1} a_i = \sum_{i=m}^{n} s_i$$

Since both the inter-arrival time and service time have continuous distributions, the probability that this equation holds is zero. Therefore, the probability that the model degenerates is zero. □

A direct result of Theorem 2 is the following corollary.

Corollary 1 *In a specific simulation sample path with J_k being the index of the basis, the probability that vector $\mathbf{c}$ lies on the border of Pc_k is zero.*

Proof: Since matrix $\mathbf{A}$ has full row rank (see Section 2.1), i.e., the n constraints in the dual model are linear independent, the border of Pc_k can be represented as:

$$\text{int}\left(Pc_k\right) = \left\{ \mathbf{c} \in \text{HOM}\left(\mathbf{A}\right) : \mathbf{c}_{N_k} - \mathbf{c}_{J_k}\mathbf{B}_k^{-1}\mathbf{A}_{N_k} = 0 \right\}$$

By Theorem 2, this probability is zero. □

According to Corollary 1, we know that in a sample path, the vector $\mathbf{c}$ will stay in the interior of Pc_k with probability 1. By Theorem 1, we can conclude that there is a open neighborhood $B_\delta\left(\mathbf{c}\right)$, such that:

$$B_\delta\left(\mathbf{c}\right) \bigcap Pc \in Pc_k$$

In other words, in a finite length sample path, we can find a $\delta > 0$, such that $\mathbf{c} + \Delta\mathbf{c} \in Pc_k$ if $\left\| \Delta\mathbf{c} \right\|_1 < \delta$, where $\left\| . \right\|_1$ is the one-norm defined as: $\left\| \mathbf{c} \right\|_1 = \max_{1 \le j \le 2n-1}\left\{ \left| c_j \right| \right\}$, where vector $\mathbf{c} = \left[c_1, \dots, c_{2n-1} \right] \in R^{2n-1}$. This is equivalent to the usual IPA assumption: the perturbation is small enough such that the entire sample path will not deviate completely. For more details, see, e.g., Suri and Zazanis (1988).

2.4 Conditions for Convergence

The objective function of GG1-LP(W) is the mean waiting time. Since both the primal and dual models are feasible and bounded, according to the strong duality theorem, the primal and dual models will have the same optimal value. That is:

$$\frac{1}{n}\sum_{i=1}^{n} W_i^* = \frac{1}{n}\sum_{i=2}^{n}\left(s_{i-1} - a_i\right)V_i^*$$

where W_i^* and V_i^* $\forall i$ are the optimal primal and dual solutions, respectively.

In section 2.3, we have proved that in a finite-length sample path we can change the random variables a little bit without changing the basis. Since the n constraints are linearly independent, if the basis does not change, we can compute the values of the n basic variables by solving n linearly independent equations. This solution is also unique and does not change. By the assumption made in Section 2.1 regarding the service time random variable, i.e., s is continuous w.r.t θ, we can calculate the derivative of random variable s as: (see also Suri and Zazanis 1988, Freimer and Schruben 2001)

$$\frac{ds_i}{d\theta} = -\left.\frac{\partial F/\partial\theta}{\partial F/\partial s}\right|_{(s_i,\theta)}$$

In a non-degenerate sample path, since the basic does not change, the values of the dual variables $U_1, \dots, U_n$, $V_2, \dots, V_n$ will remain unchanged. Taking the derivative of the objective function w.r.t θ yields

$$\frac{d}{d\theta}\left(\frac{1}{n}\sum_{i=1}^{n} W_i \right) = \frac{d}{d\theta}\left(\frac{1}{n}\sum_{i=2}^{n}\left(s_{i-1} - a_i\right)V_i \right)$$
$$= \frac{1}{n}\sum_{i=2}^{n}\frac{d}{d\theta}\left[\left(s_{i-1} - a_i\right)V_i \right]$$
$$= \frac{1}{n}\sum_{i=2}^{n} V_i \frac{ds_{i-1}}{d\theta}$$

To facilitate our discussion, let us define the IPA estimator, Z, obtained from the dual variables as LP-IPA, which is:

$$Z = \frac{1}{n}\sum_{i=2}^{n} V_i \frac{ds_{i-1}}{d\theta}$$

We note that we have interchanged the sum and derivative in the second equality of above derivation because n is finite. However, when $n \to \infty$, interchanging the infinite sum and the derivative might not be valid. To ensure such interchangeability, we will need the conditions discussed below. For a *GI/G/1* queue, the busy period is a renewal process (Crane, 1975). Let B represent the length of the busy period. If the queue is stable, $E(B) < \infty$. From the Strong Law of Large Numbers, we have:

$$\frac{1}{n}\sum_{i=1}^{n}(s_{i-1}-a_i)V_i \to E\left(\sum_{i=1}^{B}(s_{i-1}-a_i)V_i\right), \text{ as } n \to \infty$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{d}{d\theta}\left[(s_{i-1}-a_i)V_i\right] \to E\left(\frac{d}{d\theta}\sum_{i=1}^{B}(s_{i-1}-a_i)V_i\right), \text{ as } n \to \infty$$

Therefore, the key issue in perturbation analysis is to show that:

$$E\left(\frac{d}{d\theta}\sum_{i=1}^{B}(s_{i-1}-a_i)V_i\right) = \frac{d}{d\theta}E\left(\sum_{i=1}^{B}(s_{i-1}-a_i)V_i\right)$$

Now, suppose that we change the parameter by $\Delta\theta$ without changing the basis. Using the Taylor expansion, we have:

$$\frac{1}{n}\sum_{i=2}^{n}(s_{i-1}(\theta+\Delta\theta)-a_i)V_i - \frac{1}{n}\sum_{i=2}^{n}(s_{i-1}(\theta)-a_i)V_i$$

$$= \frac{1}{n}\sum_{i=2}^{n}(s_{i-1}(\theta+\Delta\theta)-s_{i-1}(\theta))V_i$$

$$= \frac{1}{n}\sum_{i=2}^{n}\left(\frac{\partial s_{i-1}}{\partial\theta}\Delta\theta + o(\Delta\theta)_i\right)V_i$$

$$= \frac{1}{n}\sum_{i=2}^{n}\left(\frac{\partial s_{i-1}}{\partial\theta}\Delta\theta\right)V_i + \frac{1}{n}\sum_{i=2}^{n}V_i \cdot o(\Delta\theta)_i$$

where $o(\Delta\theta)_i$ is a residual term under the first order Taylor expansion for s_i. That is:

$$\left[\frac{1}{n}\sum_{i=1}^{n}W_i(\theta+\Delta\theta) - \frac{1}{n}\sum_{i=1}^{n}W_i(\theta)\right]\Big/\Delta\theta$$

$$= \left[\frac{1}{n}\sum_{i=2}^{n}(s_{i-1}(\theta+\Delta\theta)-a_i)V_i - \frac{1}{n}\sum_{i=2}^{n}(s_{i-1}(\theta)-a_i)V_i\right]\Big/\Delta\theta$$

$$= \left[\frac{1}{n}\sum_{i=2}^{n}(s_{i-1}(\theta+\Delta\theta)-s_{i-1}(\theta))V_i\right]\Big/\Delta\theta$$

$$= \frac{1}{n}\sum_{i=2}^{n}\frac{\partial s_{i-1}}{\partial\theta}V_i + \frac{1}{n}\sum_{i=2}^{n}V_i \cdot \frac{o(\Delta\theta)_i}{\Delta\theta}$$

If the LP-IPA estimator is unbiased, the residual term $\frac{1}{n}\sum_{i=2}^{n}V_i \cdot \frac{o(\Delta\theta)_i}{\Delta\theta}$ should converge to zero as $n \to \infty$. On the other hand, if the queue is stable, the expected length of the busy period is finite. Therefore, $E(V_i)$ is finite $\forall 2 \le i \le n$, and there exist an $M > 0$, such that $|E(V_i)| < M$. With this stability condition, we can now state the following theorem.

Theorem 3 Let $s_1,\ldots,s_n \in R_+, \theta \in [a,b]$. If $\frac{o(\Delta\theta)}{\Delta\theta}$ uniformly converges to 0 for $\forall\theta \in [a,b], s \in R_+$, then the LP-IPA estimator, $\frac{1}{n}\sum_{i=2}^{n}\frac{\partial s_{i-1}}{\partial\theta}V_i$, will be unbiased.

Proof: Since $\frac{o(\Delta\theta)}{\Delta\theta}$ uniformly converges to 0 for $\forall\theta \in [a,b], s \in R_+$, $\forall\varepsilon > 0$, $\exists\delta_1 > 0$, such that $\left|\frac{o(\Delta\theta)}{\Delta\theta}\right| < \varepsilon$, as $|\Delta\theta| < \delta_1, s \in R_+$. Choose $\delta_2 > 0$, such that the basis will not change as $|\Delta\theta| < \delta_2$. Let $\delta = \min\{\delta_1, \delta_2\}$. We have

$$\left|\frac{1}{n}\sum_{i=2}^{n}V_i \cdot \frac{o(\Delta\theta)_i}{\Delta\theta}\right| \le \frac{1}{n}\sum_{i=2}^{n}M\cdot\varepsilon = M\cdot\varepsilon, \text{ as } |\Delta\theta| < \delta.$$

Since ε is arbitrary, the error term will converge to zero as $n \to \infty, \Delta\theta \to 0$. □

Therefore, the condition imposed on the service time distribution ensures the unbiasedness of the IPA estimator. Observe that this condition, while derived using a linear programming-based argument, is equivalent to the condition of uniformly differentiability given in Cao (1985).

There are, of course, some other convergent conditions that will also lead to the unbiasedness result of IPA estima-

tors. The general approach, however, is similar; that is, in order to apply the dominated convergence theorem to interchange the expectation and derivative, one needs to develop certain conditions to give bounds for the estimators. For more details, see, for example, Heidelberger et al. (1988), Suri and Zazanis (1988), Fu and Hu (1991).

3 FINITE PERTURBATION ANALYSIS

In this section, we discuss the use of the dual variables for finite perturbation analysis. Sometimes, it is not easy to calculate the LP-IPA estimator because the distribution of the service time does not have an analytical form. In this case, we need to estimate the perturbation by using finite perturbation methods.

There are two methods given in Cao (1985). The first one is to do two independent simulations with parameters θ and $\theta + \Delta\theta$, calculate the difference, and then divide it by $\Delta\theta$. In these two independent simulations, the random numbers are different. Therefore, this method will introduce more randomness and require more conditions for convergence. Another method is to use the same random number with different parameters. In this case, the perturbation estimator is calculated using the difference $\dfrac{\Delta W}{\Delta\theta}$

given in Section 2.4 as:

$$\frac{\Delta W}{\Delta\theta} = \left[\frac{1}{n}\sum_{i=1}^{n}W_i(\theta+\Delta\theta) - \frac{1}{n}\sum_{i=1}^{n}W_i(\theta)\right]\Big/\Delta\theta$$

From the analysis performed in Section 2.4, we know that the error term is give by: $\dfrac{1}{n}\sum_{i=2}^{n}V_i\cdot\dfrac{o(\Delta\theta)_i}{\Delta\theta}$. In order to satisfy the condition of unbiasedness we still assume the uniformly convergence condition for $\dfrac{o(\Delta\theta)}{\Delta\theta}$ as described in Theorem 3.

Since $\dfrac{o(\Delta\theta)}{\Delta\theta}$ converges to 0 uniformly, for a given $\lambda>0$, $\exists\,\delta>0$ such that the basis does not change and $\left|\dfrac{o(\Delta\theta)}{\Delta\theta}\right|<\lambda$ as $|\Delta\theta|<\delta$. Now suppose that there are totally m busy periods in the whole sample path. Let the length of these busy periods be $k_1, k_2, \ldots, k_m$. Define $k_0 = 0$ and $V_1 = 0$. Then this error term is bounded as shown in the following.

$$\left|\frac{1}{n}\sum_{i=2}^{n}V_i\cdot\frac{o(\Delta\theta)_i}{\Delta\theta}\right| \le \frac{\lambda}{n}\sum_{i=2}^{n}V_i$$

$$= \frac{\lambda}{n}\left(\sum_{l=1}^{m}\sum_{r=\sum_{j=0}^{l-1}k_j+1}^{\sum_{j=0}^{l}k_j}V_i\right)$$

$$= \frac{\lambda}{n}\sum_{l=1}^{m}\sum_{r=\sum_{j=0}^{l-1}k_j+1}^{\sum_{j=0}^{l}k_j}\left[k_l - \left(r - \sum_{j=0}^{l-1}k_j\right)\right]$$

$$= \frac{\lambda}{n}\sum_{l=1}^{m}\frac{(k_l-1)k_l}{2}$$

$$= \frac{\lambda}{2n}\left[\sum_{l=1}^{m}k_l^2 - \sum_{l=1}^{m}k_l\right]$$

When $n\to\infty$, this bound can be computed as

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=2}^{n}V_i\cdot\frac{o(\Delta\theta)_i}{\Delta\theta}\right| \le \lim_{n\to\infty}\frac{\lambda}{2n}\left[\sum_{l=1}^{m}k_l^2 - \sum_{l=1}^{m}k_l\right]$$

$$= \frac{\lambda}{2}\lim_{n\to\infty}\frac{1}{n/m}\cdot\frac{1}{m}\left[\sum_{l=1}^{m}k_l^2 - \sum_{l=1}^{m}k_l\right]$$

Let random variable B represent the length of the busy period. If the queue is stable, $E(B)<\infty$. Another fact is that $m\to\infty$ as $n\to\infty$. Since $n/m\to E(B)$ w.p.1, as $n\to\infty$, we have

$$\frac{\lambda}{2}\lim_{n\to\infty}\frac{1}{n/m}\cdot\frac{1}{m}\left[\sum_{l=1}^{m}k_l^2 - \sum_{l=1}^{m}k_l\right]$$

$$= \frac{\lambda}{2E(B)}\left[\text{var}(B) + E(B)^2 - E(B)\right]$$

$$= \frac{\lambda}{2}\left[\frac{\text{var}(B)}{E(B)} + E(B) - 1\right]$$

yielding,

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=2}^{n}V_i\cdot\frac{o(\Delta\theta)_i}{\Delta\theta}\right| \le \frac{\lambda}{2}\left[\frac{\text{var}(B)}{E(B)} + E(B) - 1\right]$$

where $\text{var}(B)$ and $E(B)$ are, respectively, the variance and the expectation of the distribution of the busy period.

Therefore, we obtain a bound for the error term when $n\to\infty$. In other words, when n is big enough, we can choose a $\Delta\theta$ small enough so that $\left|\dfrac{o(\Delta\theta)}{\Delta\theta}\right|<\lambda$. The error

term is then bounded by the expectation and the variance of the distribution of the busy period in the way described above.

4 CONCLUSION AND FUTURE WORK

In this paper, we show that the mathematical programming representation (MPR) for $G/G/1$ queues can be used to perform perturbation analysis using sensitivity analysis techniques in linear programming. Corollary 1 can be considered as another description of the usual IPA assumption. We derive a condition, based on the dual variables of the MPR, for the convergence and unbiasedness of IPA estimators. Finally, we discuss the use of the dual variables for finite perturbation and give a bound for the error term.

We note that the MPRs provide another interpretation for the perturbation on the parameter: that is, when changing the parameter, the vector of random variables changes in the space Pc. Perturbation is introduced by altering the random vector within the same critical region. This is the reason why we can use the second FPA method discussed in Section 3 to estimate the perturbation information.

When using the first method of FPA, the difference may still give us a good approximation if other conditions are satisfied (see Cao 1985). However, since these two sample paths could belong to different critical regions, further analysis is needed to address the properties of the estimator under this case. Another research topic is how much we can change the parameters. Some preliminary results given in matrix form are provided in Chan 2006. One on-going research is to find an efficient way to estimate the range of the parameters.

ACKNOWLEDGMENTS

REFERENCES

Cao, X. R. 1985. Convergence of parameter sensitivity estimates in a stochastic experiment. *IEEE Transactions on Automatic Control,* 30(9):845–853.

Chan, W. K. V. and L. W. Schruben. 2003. Properties of discrete event system from their mathematical programming representations. In *Proceeding of the 2003 Winter Simulation Conference*, ed. S. E. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, 496–502. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc. .

Chan, W. K. V. 2005. *Mathematical representation of discrete-event system dynamics*. Ph.D dissertation, University of California, Berkeley.

Dudley, R. M. 2002. *Real analysis and probability*. 2nd ed. Cambridge, England: Cambridge University Express.

Freimer, M. and L. W. Schruben. 2001. Graphical representation of IPA estimation. In *Proceeding of the 2001 Winter Simulation Conference*, ed. Peter, B. A., Smith, J. S., Medeiros, D. J. and Rohrer, M. W. 422–427. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc. .

Fu, M. C. and J. Q. Hu. 1991. Consistency of infinitesimal perturbation analysis for the GI/G/m queue. *European Journal of Operations Research,* 34(1):39–64.

Fu, M. and J. Q. Hu. 1997. *Conditional monte carlo: gradient estimation and optimization applications*. Boston, Massachusetts: Kluwer Academic Publishers.

Gal, T. 1979. *Postoptimal analysis, parametric programming, and related topics*. 1st Ed. New York, USA: McGraw Hill.

Gal, T. and H. J. Greenberg. 1997. Advances in sensitivity analysis and parametric programming. Boston, Massachusetts: Kluwer Academic Publishers.

Glasserman, P. and Y. C. Ho. 1991. *Gradient estimation via perturbation analysis*. Boston, Massachusetts: Kluwer Academic Publishers.

Heidelberger, P., X. R. Cao, M. A. Zazanis and R. Suri 1988. Convergence properties of infinitesimal perturbation analysis estimates. *Management Science,* 34(11):1281–1302.

Ho, Y. C. and X. R. Cao. 1991. *Perturbation analysis of discrete event dynamic systems*. Boston, Massachusetts: Kluwer Academic Publishers.

Renegar, J. 1994. Some perturbation theory of linear programming. *Mathematical Programming,* 65(1-3):73–91.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton, New Jersey: Princeton University Press.

Schruben, L. W. 2000. Mathematical Programming Models of Discrete Event System Dynamics. Proceedings of the 2000 Winter Simulation Conference Proceedings (Cat. No.00CH37165). IEEE. Part vol.1, 2000, pp.381-385. Piscataway, NJ, USA.

Suri, R. and M. A. Zazanis. 1988. Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue. *Management Science,* 34(1):39–64.

Ward, J. E. 1990. Approaches to sensitivity analysis in linear programming. *Annals of Operations Research ,* 27(1):3–38.

Zazanis, M. A. 1987. Weak convergence of sample path derivatives for the waiting time in a single-server queue. *In Proceedings of the 25th Allerton Conference on Communication, Computing and Control* :297–304.

AUTHOR BIOGRAPHIES

He Zhang is an Ph.D. student in the Department of Decision Sciences & Engineering systems at Rensselaer Polytechnic Institute. His research interest includes sensitivity analysis of discrete-event simulation and mathematical programming. His e-mail is <zhangh2@rpi.edu>. His web page is <www.rpi.edu/~zhangh2>.

WAI KIN (VICTOR) CHAN is an Assistant Professor of the Department of Decision Sciences & Engineering Systems at Rensselaer Polytechnic Institute. He received a bachelor's degree and a master's degree in electrical engineering from, respectively, Shanghai Jiao Tong University, China in 1997, and Tsinghua University, China in 2000, and a M.S. and Ph.D. degrees in industrial engineering and operations research from University of California, Berkeley in 2001 and 2005, respectively. His research interests include discrete-event simulation, agent-based simulation, mathematical programming and sensitivity analysis, and semi-conductor manufacturing. His e-mail address is <chanw@rpi.edu>.