in the space of MTA events $E$ and in the space of track states $X$. In this role $\beta$ is analogous to the gain parameter in a neural network or the inverse temperature in simulated annealing and thermodynamic systems.

The Hessian-based iterated schemes used here to solve (13) are similar to an iterated extended Kalman filter. The mean-field equations can also be approximately solved in a noniterated scheme analogous to a conventional extended Kalman filter. In this scheme, (13) is approximately solved for the $\eta$'s in a single iteration with $X_0$ replacing $X$ and $e^{1/2}$ replacing $\eta_0$ where $\eta_0$ is some nominal value. (Trials with either $\eta_0 = 0$ or $\eta_0$ obtained from the previous scan were both satisfactory.) The approximate $\eta$'s are then inserted in (13) and the expression for $\partial \text{tr}(X, \eta)/\partial x_t = 0$ with $X$ again replaced by $X_0$ in the exponents. The resulting approximation is linear in $X$ and can be solved directly. Preliminary results indicate that there is some degradation in estimation performance with this scheme but it may be useful when speed and deterministic response times are critical.

The statistical underpinnings of the MFEAMLE method are identical to those of the parametric PDA and JPDA filters [2] and the JPDA filter can be derived from (8) in an extended Kalman filter approximation. Mean-field equations (18) and (19) involve an average with weights $e^{1/2}/(1 + \sum e^{1/2})$ that closely resembles the construction of the combined innovation [2] in PDA. For this one-dimensional problem, PDA involves an average with weights of the same form based on $y_{r,t}$ PDA $= (-z_y - x) ^2 / 2R + x$, in contrast to $y_{r,t} = y_{r,t}$ PDA $\pm \eta$ for MFEAMLE. The significant difference from the PDA expression is that in MFEAMLE the factor $e^{1/2}$ modifies the association weights to account for the presence of other nearby targets.

ACKNOWLEDGMENT

The author would like to thank C. L. Lutes and D. G. Garrett for valuable discussions.

REFERENCES


III-Conditioned Performance Functions of Queuing Systems

Xi-Ren Cao, Wei-Bo Gong, and Yorai Wardi

Abstract—In this paper, we show that for queueing networks with deterministic or discrete service time distributions, the performance functions can be nondifferentiable at a dense subset of a given interval. We also show that when the service time densities are supported on small intervals, the performance function derivatives change rapidly. We prove these results for a two-server cyclic network and then point out a potential generality to other queueing networks. The results indicate that the nonsmooth analysis [3] may be useful in the area of stochastic discrete-event systems.

I. INTRODUCTION

Nonsmooth behavior for continuous variable dynamic systems has attracted considerable research attention in the system and control research. In this paper we want to show, by examples, that for a class of stochastic discrete-event systems the performance functions may be nondifferentiable or ill-conditioned at a dense set of a real interval. Specifically, we point out that for queueing networks consisting of servers with deterministic or discrete service time distributions the performance functions can be nondifferentiable with respect to a service parameter at a dense subset of a given interval. For networks with service time densities supported on small intervals, the first derivative may exist; however, the second derivative can be very

Manuscript received July 1, 1994; revised September 1, 1994. This work was supported in part by the Hong Kong UPQC under grant HKUST 9909/94E. X.-R. Cao is with Department of Electrical and Electronic Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong. W.-B. Gong is with Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA. Y. Wardi is with School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA, USA. IEEE Log Number 9406068.
large, implying that the numerical calculation of the first derivative is ill-conditioned.

We first prove the above statements for a two-server cyclic network and then give an explanation for more general networks. Specifically, we prove that in a cyclic network with two servers and two customers where the service times are discrete random variables, the steady-state throughput and response time as functions of the service time distribution parameters are differentiable at any rational number. We then show that if the service time density is supported by small intervals (we call such a service time distribution "almost discrete"), then a lower bound of the finite difference ratio of the derivative (i.e., the ratio of the derivative increment and the parameter variation) increases unboundedly when the interval lengths go to zero. This means that the first derivative changes too rapidly to be numerically useful. Finally, we point out that similar nondifferentiability and ill-conditioning may exist in common performance functions in more general queuing networks.

The kind of systems discussed in this paper is in fact very common. For example, a queuing network consisting of servers with deterministic service times and feedback loops is such a system. Many practical systems, such as manufacturing systems and communication networks, can be considered as having discretely supported service times, where the processing time of a part or the transmission time of a packet usually takes values from a finite set. From an engineering design point of view, the results of this paper have the following implication: In optimization-based design of systems with discrete random service times, one should take into account that the first derivative of a given average performance measure may not exist, and if the directional derivatives exist they may be unequal, in this case the second derivative does not exist. If the system has "almost" discrete service times, then the computation of the first derivative may be ill-conditioned.

The research on this paper was stimulated by a conjecture in [8] that the nondifferentiables described above may hold for a class of open queuing networks with discrete random service times. Wardi et al. [8] suggested that, for a generic example, the steady-state function has a nondifferentiable additive component. Plambeck et al. [9] also noted such nondifferentiabilitys in production-system models. The above conjecture has been proved by Shapira and Wardi [10] for the case of convex sample performance functions.

The paper is organized as follows. In Section II, we analyze an example of a queuing network consisting of two servers with two customers circulating between them; each server has a discrete service time distribution. We prove that the expected sojourn time of a customer at a server is nondifferentiable with respect to a service time parameter at any rational value of the parameter. We also prove that other performance functions are even discontinuous at any rational value of the parameter. In Section III, we consider service times that are "almost" discrete, in the sense that they have densities that are supported on an interval of a short length, 2\epsilon. We show that when the service time distribution is almost discrete, the derivative of the performance function is ill-conditioned in the sense that, at any rational number, the finite difference ratio of the first derivative is bounded from below by a quantity in the order of 2^{-1}. In Section IV, we discuss the possible presence of such ill-conditioning for general queuing systems.

II. A CYCLIC QUEUING SYSTEM WITH DISCRETE SERVICE TIMES

Consider a two-server cyclic queuing network in which two customers circulate between the two servers, denoted as server 1 and server 2. The service time of a customer at server 1 takes integer values \( k = 1, 2, \ldots \) with probability \( q_k^{(1)} > 0 \). The service time of a customer at server 2 takes values of \( k/r, k = 1, 2, \ldots \), with probability \( q_k^{(2)} > 0 \), where \( r \) is a real number, is considered a parameter. For the convenience of the calculation we assume that the service processes of the two servers are independent. The mean service time of server 1 is \( w_1 = \sum_{k=1}^{\infty} k q_k^{(1)} \) and that of server 2 is \( w_2 x = \sum_{k=1}^{\infty} k q_k^{(2)} \). We assume that \( w_1 \) and \( w_2 \) are both finite.

Fig. 1 illustrates a typical sample path of the system. Since there are only two customers, their services at the two queues always start at the same time; this makes the analysis tractable. The service completion times are generally different. If \( x \) is a rational number, however, then the service completion times of the two servers may be the same. For example, if \( x = 2/3 \), then the customer with a service time of 2/r at server 2 will complete its service at the same time as the customer with a service time of 2/r at server 1. The state (the queue length) process is regenerative with the service starting times of the customers at the servers being the regeneration points.

Now we assume that \( x \) is a rational number, namely \( x = m/n \) with \( m \) and \( n \) being two coprime integers. Then a customer with a service time of \( kn/r \) at server 1 will complete its service at the same time as a customer with a service time of \( kn \) at server 1. With this in mind, we can classify the customers at server 2 into three classes, denoted as classes 0, 1, and 2, as follows: A customer is in class 0 if it arrives to server 2 after the other customer departed from it, it is in class 1 if it arrives to server 2 at the same time the other customer departs from it, and it is in class 2 if it arrives to server 2 before the other customer departs from it.

Let \( \epsilon = 0, 1, 2 \), be the event that the \( \epsilon \)th customer arriving at server 2 belongs to class 0, 1, or 2, respectively. Since the period between two service starting times at server 2 forms a regenerative period of the state process and there is only one arrival in such a regenerative period, events \( \epsilon, i = 1, 2, \ldots \) are independent and identically distributed. Let \( p_0, \epsilon_0, \) and \( \rho_0 \) be the probabilities that a customer belongs to classes 0, 1, and 2, respectively. These probabilities depend on \( x \) and sometimes are explicitly denoted as \( \rho_k(x), \epsilon = 0, 1, 2 \). Let \( s_0, s_1 = 1, 2, \) be the service times of server 1. Note that a customer is in class 0 if and only if its service time at server 1 is greater than the service time of the previous customer at server 2. Therefore, \( p_0(x) = \Pr \{ s_2 > s_1 \} \). Similarly, \( p_1(x) = \Pr \{ s_2 = s_1 \} \) and \( p_2(x) = \Pr \{ s_2 < s_1 \} \), where \( \Pr \{ \epsilon \} \) is the probability of event \( \epsilon \). Now we express explicitly these probabilities in terms of \( q_k^{(1)} \) and \( q_k^{(2)} \). Let

\[
u_j = \max \{ l : l \leq k \}, \quad j = 1, 2, \ldots ,
\]

and

\[
u_j = \max \{ l : l \leq j \}, \quad j = 1, 2, \ldots .
\]

Then

\[
p_0(x) = \prod_{j=1}^{\infty} \left\{ q_j^{(1)} \sum_{k=1}^{\nu_j} q_k^{(2)} \right\}, \tag{1}
\]

\[
p_0(x) + p_1(x) = \prod_{j=1}^{\infty} \left\{ q_j^{(1)} \sum_{k=1}^{\nu_j} q_k^{(2)} \right\}, \tag{2}
\]
and

$$p_2(x) = \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{\infty} q_i^{(2)}.$$  \hfill (3)

We choose the time that a customer stays in server 2 as the performance function and call it the system time. Since the events that a customer belongs to class 0, 1, or 2 are independent, we can write the sample system time for the ith customer as follows.

Let

$$\xi_i = \chi(s_2 = s_1) + 2\chi(s_2 > s_1)$$

where \(\chi\) is the indicator function. It is clear that \(\{\xi_i, i = 1, 2, \ldots\}\) is an i.i.d. random variable sequence and that if \(\xi_i = j\) then the ith customer at server 2 belongs to class \(j\).

Let \(f(\xi_i, k_i)\) be the system time of the ith customer at server 2, where \(k_i\) is a random integer with distribution \(q^{(2)}_{k_i}\), \(k_i = 1, 2, \ldots\). If the ith customer is class 0 or 1, \(f(\xi_i, k_i) = k_i x + r_i\), otherwise \(f(\xi_i, k_i) = k_i x + r_i\), where \(r_i\) is the customer’s waiting time, which equals the difference between the service times of the preceding customer in server 2 and the customer’s service time in server 1 who starts the service at the same time as the preceding customer (see Fig. 2). Note that the \(r_i\)’s are only defined for class 2 customers and are i.i.d. random variables. Therefore

$$f(\xi_i, k_i) = \begin{cases} k_i x & \text{if } \xi_i \leq 1 \\ k_i x + r_i & \text{if } \xi_i > 1. \end{cases}$$  \hfill (4)

From (4), the expected value of the system time is

$$f(x) = E[f(\xi_i, k_i)] = E[k_i x \mid s_2 < s_1]p_0 + E[k_i x \mid s_2 = s_1]p_1$$

+ \(E[k_i x + r_i \mid s_2 > s_1]p_2\)

$$= w_2 x + E[r_i]p_2.$$  \hfill (5)

where

$$E[r_i] = E[s_2 - s_1 \mid s_2 > s_1].$$

Note that

$$E[r_i]p_2 = \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{\infty} (i-x)q_i^{(2)}.$$  \hfill (6)

We now study the derivative of \(f(x)\).

**Theorem 1:** \(f(x)\) is differentiable only at irrational points \(x \in (0, 1)\). If \(x = m/n\), with \(m\) and \(n\) being coprime, then

$$\frac{df(x)}{dx} = n w_2 + \left\{ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{\infty} i q_i^{(2)} \right\} + \sum_{k=1}^{\infty} k q_k^{(1)} q_k^{(2)}.$$  \hfill (7)

and

$$\frac{df(x)}{dx} = w_2 + \left\{ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{\infty} i q_i^{(2)} \right\} + n \sum_{k=1}^{\infty} k q_k^{(1)} q_k^{(2)}.$$  \hfill (8)

and if \(x\) is irrational, then

$$\frac{df(x)}{dx} = w_2 + \left\{ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{\infty} i q_i^{(2)} \right\}.$$  \hfill (9)

**Proof:** Consider first the case where \(x\) is rational. Let \(x = m/n\), where \(m\) and \(n\) are coprime. We first rewrite the right-hand side. Let \(\Delta x > 0\), and assume that \(x\) changes to \(x = \Delta x\). Noting that \(r_j\) is a function of \(x\) and \(r_j(x - \Delta x) \geq r_j(x)\), we have, by (6)

$$\Delta E[r_i]p_2$$

$$= \sum_{j=1}^{\infty} \left\{ q_j^{(1)} \sum_{i=r_j+1}^{\infty} [i(x - \Delta x) - i]q_i^{(2)} \right\}$$

$$- \sum_{j=1}^{\infty} \left\{ q_j^{(1)} \sum_{i=r_j+1}^{\infty} [i - i]q_i^{(2)} \right\}$$

$$= \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} [i(x - \Delta x) - i]q_i^{(2)}$$

$$- \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} [i - i]q_i^{(2)}$$

$$+ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} [i - i]q_i^{(2)}$$

$$- \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} [i - i]q_i^{(2)}$$

$$= \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} [i(x - \Delta x) - i]q_i^{(2)}$$

$$- \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} [i - i]q_i^{(2)}.$$  \hfill (10)

Note that \(r_j(x) < i \leq r_j(x - \Delta x)\) implies \(i > j\) and \(i(x - \Delta x) \leq j\).

Thus

$$0 \leq j - i(x - \Delta x) < i \Delta x$$  \hfill (11)

and

$$\Delta x \geq x - j = \frac{m - i}{n} \geq \frac{m - u - u}{n} = \frac{1}{n}$$

(since \(mi - uj = n(x - j) > 0\) or

$$i \geq \frac{1}{n}$$

Therefore, the first term in the right-hand side of (10) is positive, but also

$$\sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} [j - i(x - \Delta x)]q_i^{(2)}$$

$$\leq \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i=r_j+1}^{r_j+1} i q_i^{(2)}$$

$$\leq \sum_{j=1}^{\infty} q_j^{(1)} \left\{ \sum_{i=r_j+1}^{r_j+1} q_i^{(2)} \right\}$$

$$= \sum_{i=r_j+1}^{\infty} q_i^{(2)}.$$  \hfill (12)
Since \( w_i \) is finite, we have
\[
\lim_{\Delta x \to 0} \sum_{i \geq \lceil \frac{w_i}{h} \rceil} i q_i^{(2)} = 0.
\] (13)

From (10)-(13), we get
\[
\frac{dF(x_i) p_{x_i}}{dx_i} = \left\{ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i \leq j} i q_i^{(2)} \right\} \Delta x.
\] (14)

Thus, by (5) and (14), (7) holds.

Next, consider the right derivative. We assume that \( x \) changes to \( x + \Delta x, \Delta x > 0 \). We have \( r_j(x + \Delta x) \leq r_j(x) \) and, by a similar derivation to (10),
\[
\Delta(E[r_j] p_{x_j}) = \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i \leq j} [i(x + \Delta x) - j] q_i^{(2)} \Delta x
\]
\[
+ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i = j+1}^{\infty} i q_i^{(2)} \Delta x.
\] (15)

The first term in the right-hand side of (15) equals
\[
\sum_{j=1}^{\infty} q_j^{(1)} \sum_{i \leq j} [i(x + \Delta x) - j] q_i^{(2)} \Delta x
\]
\[
+ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i = j+1}^{\infty} i q_i^{(2)} \Delta x.
\] (16)

Note that \( r_j(x + \Delta x) < i < r_j(x) \) implies \( i \leq j \) and \( i(x + \Delta x) > j \). By the same argument as that for (10)-(13), the contribution of the first term in (16) to the derivative is zero. In the second term of (16), \( i = r_j(x) > r_j(x + \Delta x) \) should hold for any \( \Delta x > 0 \). By the definition of \( r_j(x) \), we have \( r_j(x) = \frac{m}{k} \). Since \( r_j(x) \) is an integer and \( n \) and \( m \) are coprime, we have \( j = km, \ k = 1, 2, \ldots, \) and \( r_j(x) = km \). Therefore, the second term in (16) equals
\[
\sum_{k=1}^{\infty} \frac{1}{\nu} k q_k^{(2)} \Delta x = \sum_{k=1}^{\infty} k q_k^{(2)} p_k \Delta x.
\] (17)

Combining the results in (15)-(17), we get
\[
\frac{dF(x_i) p_{x_i}}{dx_i} = \left\{ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i \leq j} i q_i^{(2)} \right\} \Delta x.
\] (18)

By (5) and (18), (8) holds.

Now, let \( x \in (0, 1) \) be irrational. We shall show that \( f \) is differentiable at \( x \), and (9) holds. Consider first the left derivative. Let \( \Delta x < 0 \). Similar to (10), we have
\[
\Delta(E[r_j] p_{x_j}) = \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i \leq j} [i(x - \Delta x) - j] q_i^{(2)} \Delta x
\]
\[
+ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i = j+1}^{\infty} i q_i^{(2)} \Delta x.
\] (19)

The contribution of the first term in (19) to the derivative is zero. This can be seen as follows: Fix an \( \epsilon > 0 \). Let \( \delta \) be an integer satisfying
\[
\sum_{i \leq \delta} i q_i^{(2)} < \epsilon.
\] (20)

We now show that there exists a \( \delta > 0 \) such that, for every \( \Delta x < \delta \), \( \Delta x > 0 \), if \( r_j(x) < i \leq r_j(x - \Delta x) \), then \( i \geq j \). Note that \( i > r_j(x) \) implies \( i > j \), and \( i \leq r_j(x - \Delta x) \) implies \( i(x - \Delta x) \leq j \). If \( i(x - \Delta x) \leq j \), then \( \Delta x \geq x - j/i > 0 \). Since \( x \) is irrational, \( x - j/i \neq 0 \). Let \( \delta > 0 \) satisfy the following: For every pair of integers, \( i \) and \( j \), such that \( i < j, j \leq i \), and \( x - j/i > 0 \)
\[
\delta < x - j/i.
\]

Since \( x \) is irrational, such a \( \delta \) can be found. Now, if \( \Delta x < \delta, \Delta x > 0 \), and \( r_j(x) < i \leq r_j(x - \Delta x) \), then \( \Delta x \geq x - j/i > 0 \). Therefore, either \( i > I \), or \( i \geq I \). In either case, \( i > j \), since \( i > j \), hence, if \( j > I \), then \( i > I \). Therefore, by (11) and (20), with \( \Delta x < \delta \) the first term in (19) satisfies
\[
\sum_{j=1}^{\infty} q_j^{(1)} \sum_{i \leq j} [i(x - \Delta x) - j] q_i^{(2)} \Delta x \leq \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i = j+1}^{\infty} i q_i^{(2)} \Delta x \leq \epsilon \Delta x.
\] (21)

Thus, the contribution of the first term in (19) to the derivative is less than \( \epsilon \). Since \( \epsilon \) was arbitrary, the contribution of this term is zero. Therefore, by (5), we have
\[
\frac{df(x)}{dx} = w_2 + \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i = j+1}^{\infty} i q_i^{(2)} \Delta x.
\] (22)

Finally, consider the right derivative. Let \( \Delta x > 0 \). Similar to (15)-(16), we have
\[
\Delta(E[r_j] p_{x_j}) = \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i \leq j} [i(x + \Delta x) - j] q_i^{(2)} \Delta x
\]
\[
+ \sum_{j=1}^{\infty} q_j^{(1)} \sum_{i = j+1}^{\infty} i q_i^{(2)} \Delta x.
\] (23)

Similar to the proof of (21), the contribution of the first term in (23) to the derivative is zero. The second term is also equal to zero, since \( x \) is irrational, hence it is impossible that \( i = r_j(x) \). Thus, by (5), \( (df(x)/dx)_{+} \) is the same as the expression in (22). This establishes (9), completing the proof of the theorem.

Now we study another performance function, \( g(x) \), which is the probability that a customer belongs to the second class, i.e., \( g(x) = p_2(x) \). This performance measure has a practical meaning: in a communication network a class 2 customer indicates a collision between two packets; we certainly would like to minimize the chance of such collisions. Following the same argument as that for \( f(x) \), we can prove that if \( x \) is an irrational number, then
\[
\lim_{\Delta x \to 0} g(x + \Delta x) - g(x) = 0
\] (24)

and if \( x = (m/n) \) is a rational number, then
\[
\lim_{\Delta x \to 0} g(x + \Delta x) - g(x) = 0
\] (25)

but
\[
\lim_{\Delta x \to 0} g(x + \Delta x) - g(x) = \sum_{k=1}^{\infty} q_k^{(1)} q_k \Delta x.
\] (26)

Equations (24)-(26) show that the performance function \( g(x) \) is continuous at any irrational number but is discontinuous at any rational number.
III. THE “ALMOST DISCRETE” SERVICE TIME CASE

It can be argued that a network with discrete random service time is not realistic because in practical, systems service times (e.g., the service times of different parts on the same machine) cannot be exactly the same, and therefore the nondifferentiability result obtained above does not apply to real systems. Pushing this argument further, one may mistakenly feel that it is not an issue to use first derivatives in optimization of real systems.

This section is to address the above plausible argument. To mathematically formulate a small deviation from a discrete random variable, we assume that the service time distribution density functions are supported on a set of small intervals. More precisely, we add to the discrete service time a random noise, which is uniformly distributed in the interval $[\lambda - \delta, \lambda + \delta]$. We call this kind of service time “almost discrete.” We shall see that for this network the steady-state performance function is indeed differentiable everywhere. If the support of the random noise density is small, however, then the first derivative changes very rapidly. The computation of the first derivative may be ill-conditioned, and the value of the first derivative at any particular point does not have any significant meaning.

To this end, we consider the same network analyzed in the previous section except that each service time is added by a random variable uniformly distributed in the interval $[-\delta, \delta]$. We show that as $\delta$ goes zero, the finite difference ratio (the ratio of the first derivative increment and the parameter variation) approaches infinity. More precisely, we prove that the finite difference ratio of the first derivative is bounded from below by a quantity in the order of $1/\delta$.

We first introduce the following notation for the system time of the $i$th arrival to server 2. Suppose that the $i$th arrival to server 2 has service time $k_{i-1}$ at server 2 and $j_i$ at server 1. Suppose also that the previous arrival to server 2 has service time $k_{i-1}$ at server 2. We denote the system time for the $i$th arrival to server 2 as $S_i(x, k_{i-1}, j_i)$ where $x$ is still the decision parameter. We have

$$S_i(x, k_{i-1}, j_i) = k_{i-1} x + [k_{i-1} x + j_i]$$

where $[x]$ is the greatest integer less than or equal to $x$.

Now replace $x$ by $x = x + w$ where $w$ is the random noise with uniform distribution over $[-\delta, \delta]$. We assume $w$ is independent of everything else. The service time of server 2 is $k_{i-1} x$ with probability density $\frac{1}{\delta}$.

$S_i(x, k_{i-1}, j_i)$ we now calculate the mean system time for the $i$th arrival to server 2.

Denoting the expectation of $S_i(x, k_{i-1}, j_i)$ over $w$ by $\overline{S}_i(x, k_{i-1}, j_i)$, we have for $x \in \left[\frac{k_{i-1}}{\delta} - \delta, \frac{k_{i-1}}{\delta} + \delta\right]$

$$\overline{S}_i(x, k_{i-1}, j_i) = k_{i-1} x + \frac{1}{2\delta} \int_{-\delta}^{\delta} \left[ k_{i-1} x + k_{i-1} w - j_i \right] dw$$

In general we have

$$\overline{S}_i(x, k_{i-1}, j_i) = \begin{cases} k_{i-1} x, & \text{if } x < \frac{k_{i-1}}{\delta} - \delta \\ k_{i-1} x + \frac{k_{i-1}}{\delta} \left( x - \frac{k_{i-1}}{\delta}, \frac{k_{i-1}}{\delta} + \delta \right), & \text{if } x \in \left[\frac{k_{i-1}}{\delta} - \delta, \frac{k_{i-1}}{\delta} + \delta\right] \\ k_{i-1} x + k_{i-1} \left( x - \frac{k_{i-1}}{\delta} \right), & \text{if } x \geq \frac{k_{i-1}}{\delta} + \delta \end{cases}$$

It can be seen that $\overline{S}_i(x, k_{i-1}, j_i)$ is continuously differentiable and the first derivative is

$$\frac{d}{dx} \overline{S}_i(x, k_{i-1}, j_i) = \begin{cases} k_{i-1}, & \text{if } x < \frac{k_{i-1}}{\delta} - \delta \\ k_{i-1} x + k_{i-1} \left( x - \frac{k_{i-1}}{\delta}, \frac{k_{i-1}}{\delta} + \delta \right), & \text{if } x \in \left[\frac{k_{i-1}}{\delta} - \delta, \frac{k_{i-1}}{\delta} + \delta\right] \\ k_{i-1} x + k_{i-1} \left( x - \frac{k_{i-1}}{\delta} \right), & \text{if } x \geq \frac{k_{i-1}}{\delta} + \delta \end{cases}$$

We now consider the second-order behavior of the mean system time $\overline{ES}_i(x, k_{i-1}, j_i)$. To avoid complication we do not calculate the second-order derivative (which does not always exist) in this case. Rather, we show that the finite difference ratio for the first derivative $(d/dx) \overline{ES}_i(x, k_{i-1}, j_i)$ increases unboundedly when $\delta$ decreases.

Since $(d/dx) \overline{ES}_i(x, k_{i-1}, j_i) \leq k_{i-1} + 1$ and $\overline{ES}_i(k_{i-1}, j_i) < \infty$, the following interchange of the expectation and the differentiation is guaranteed by the Lebesgue dominant convergence theorem

$$\frac{d}{dx} \overline{ES}_i(x, k_{i-1}, j_i) = E(\frac{d}{dx} \overline{S}_i(x, k_{i-1}, j_i))$$

From (29) we know that $(d/dx) \overline{ES}_i(x, k_{i-1}, j_i)$ is a nonnegative function of $x$. Therefore we have

$$\overline{ES}_i(x + \Delta x, k_{i-1}, j_i) - \overline{ES}_i(x, k_{i-1}, j_i) \geq 0, \forall x, k_{i-1}, j_i, \Delta x$$

For $|\Delta x| < \delta$ and $x = (j_i/k_{i-1})$ we have, again, from (29)

$$\frac{d}{dx} \overline{ES}_i(x + \Delta x, k_{i-1}, j_i) - \frac{d}{dx} \overline{ES}_i(x, k_{i-1}, j_i) = \frac{k_{i-1}}{2\delta}$$

Thus for $|\Delta x| < \delta$ and $x = j_i/k_{i-1}$ we have, $j_i = 1, 2, \ldots, k_{i-1}$, we have

$$\frac{d}{dx} \overline{ES}_i(x + \Delta x, k_{i-1}, j_i) - \frac{d}{dx} \overline{ES}_i(x, k_{i-1}, j_i) \geq \frac{k_{i-1}}{2\delta} \cdot \text{Pr}(j_i = j, k_{i-1} = k)$$

Since $\text{Pr}(j_i = j, k_{i-1} = k) = \frac{k_{i-1}}{2\delta} \cdot \text{Pr}(j_i = j, k_{i-1} = k)$ we see that the finite difference ratio for the first derivative increases unboundedly when $\delta$ decreases.

IV. DISCUSSION AND EXTENSIONS

Before discussing more general cases, we shall give an intuitive explanation of the nondifferentiability in the above queuing network example. The explanation reveals the fundamental mechanism behind the mathematics and helps to clarify the nondifferentiability problem for more complex systems.
Let $x$ be a rational number. Then with a positive probability the two servers in the example may complete their services at the same time. In Fig. 3, a shows the scenario of such a service completion at server 2. Suppose that $x$ changes by $\Delta x > 0$, and b shows that the first customer's service time increases by $\Delta x$ but that of the second customer by $2\Delta x$. Now, if $x$ changes by $\Delta x < 0$, as shown in Fig. 3-c, both customers' system times decrease by $\Delta x$. Thus, we see a difference between the changes of the system times corresponding to $\Delta x > 0$ and $\Delta x < 0$. This scenario occurs no matter how small $\Delta x$ is. Because the probability of this scenario is positive, the contribution of this difference to the derivatives is not negligible. Thus, the two one-sided derivatives of the average system time are different at any rational number. On the other hand, if $x$ is an irrational number, then the two servers will not complete their service simultaneously. The above scenario will not occur. If the service distribution function is continuous, then the probability that the two servers will simultaneously complete their service is zero, the above scenario will occur with probability zero, and the two one-sided derivatives of the average system time are the same.

The above explanation indicates that the average system time may be nondifferentiable at $x$ if the following scenario occurs with a positive probability: for a parameter taking value $\varepsilon$, a customer arrives to a server at the same time as another customer, who is the only one remaining at the server, leaves the server. As was pointed out in Section II for closed queuing networks and in [8] for open networks, this scenario often arises. Verifying that such a probability is positive may not be easy, and further proving that the corresponding steady state function is nondifferentiable can be rather difficult. Therefore, the issue in its generality remains open. We do suggest, however, that given a queuing network with discrete service time distributions, one cannot blindly assume that a steady-state function of interest is differentiable.

The analysis in Section II can be extended. For example, if we let $q_{11}^{(1)} = q_{12}^{(2)} = (1 - q)q^{k - 1}$, $q < 1$, $k = 1, 2, \ldots$, in the queuing network considered in Section II, then each server is equivalent to a server with a service time 1 or $x$ and a feedback probability $q$ (a customer, upon the completion of its service at a server, has a probability $q$ of feeding back to the same server) and the network is equivalent to a cyclic network of two such servers. This system can serve as a model of communication systems where a fixed length packet can be successfully transmitted with probability $q$ and has to be retransmitted with probability $1 - q$. Furthermore, extensions of the analysis to more general closed networks with such servers appears to be possible. For systems consisting of some (not all) such servers, whether the nondifferentiability holds depends on the specific configuration of the network. In short, the nondifferentiable property of a performance function may exist for networks consisting of servers with deterministic or discrete random service times.

The set at which the performance function is nondifferentiable may be a dense set other than the rational set. For example, if the service time of server 1 takes values $k\alpha$, $k = 1, 2, \ldots$, with $\alpha$ being an irrational number, then the nondifferentiable set is $r\alpha$ with $r$ being any rational number. The nondifferentiability can also occur at a nondense set but still cause problem for gradient-based optimization. Such an example can be found by setting the service time of server 1 in our example to be deterministic and letting that of server 2 remain the same.

Our observation could have some implications in optimization. After all, most convergence analyses of algorithms require continuous differentiability of the performance function in question, see, e.g., [7] for deterministic problems and [5] for stochastic problems. If the gradient does not exist or it exists but its computation is ill-conditioned on a dense set in the parameter-space, it is not clear how an algorithm would perform. In particular, stochastic variants of second-order methods, like quasi-Newton-Raphson, probably would not perform well. Regarding the theoretical question of algorithms’ asymptotic convergence, most of the results that pertain to queuing networks [4], [2], [6] are based on differentiability of the functions involved. Thus, the issue of nondifferentiability of the performance functions may have to be addressed from the standpoints of both theory and implementation.

In summary, we have shown that for a class of networks consisting of servers with deterministic or discrete random service times, the performance functions may be nondifferentiable or discontinuous at a dense set on an interval and that for networks with almost discrete service times the first derivative changes rapidly. This phenomena has to be taken into consideration in gradient-based optimization and the nonsmooth analysis of Clarke [3] may be useful.

REFERENCES