Les cahiers du GERAD ISSN: 0711–2440

Stochastic Optimization by Simulation: Convergence Proofs and Experimental Results for the GI/G/1 Queue in Steady-State

P. L'Ecuyer, N. Giroux
P.W. Glynn

G–91–49

December 1991

Stochastic Optimization by Simulation: Convergence Proofs and Experimental Results for the GI/G/1 Queue in Steady-State

P. L'Ecuyer
Département d'I.R.O., Université de Montréal
C.P. 6128, Montréal, Canada, H3C 3J7

N. Giroux
Département d'informatique, Université Laval
Ste-Foy, Québec, Canada, G1K 7P4

P. W. Glynn
Operations Research Department
Stanford University, Stanford, CA94305, USA

December1991

Abstract

Approaches like infinitesimal perturbation analysis and the likelihood ratio method have drawn a great deal of attention recently, as ways of estimating the gradient of a performance measure with respect to continuous parameters in a dynamic stochastic system. In this paper, we experiment with the use of these estimators in stochastic approximation algorithms, to perform so-called "single-run optimizations" of steady-state systems. We also compare them to finite-difference estimators, with and without common random numbers. Under mild conditions, for an objective function that involves the mean system time in a GI/G/1 queue, we prove that many variants of these algorithms converge to the minimizer. In most cases, however, the simulation length must be increased from iteration to iteration; otherwise the algorithm may converge to the wrong value. One exception is a particular implementation of infinitesimal perturbation analysis, for which the single-run optimization converges to the optimum even with a fixed (and small) number of ends of service per iteration. We have performed extensive numerical experiments with a simple M/M/1 queue, which illustrate the basic convergence properties and possible pitfalls of the various techniques. Our convergence proofs are quite involved. As a by-product, we obtain new structural results for the GI/G/1 queue that could be of independent interest.

Résumé

Des techniques telles que l'analyse de perturbation infinitésimale et la méthode du rapport de vraisemblance ont attiré beaucoup l'attention récemment. Ces méthodes visent à estimer le gradient d'une mesure de performance par rapport à un vecteur de paramètres continus dans le contexte de systèmes dynamiques stochastiques à événements discrets. Nous avons expérimenté la combinaison de ces estimateurs avec des algorithmes d'approximation stochastique, en vue d'optimiser de tels systèmes sur horizon infini en une seule simulation. Dans ce contexte, nous les avons aussi comparés avec des estimateurs basés sur les différences finies, avec et sans valeurs aléatoires communes. Pour une fonction objectif impliquant la durée moyenne de séjour dans une file d'attente GI/G/1, nous montrons, sous des hypothèses relativement faibles, que plusieures variantes de ces algorithmes convergent vers l'optimum. Cependant, en général, la durée de la simulation par itération doit augmenter (vers l'infini) d'itération en itération; sinon l'algorithme peut converger vers la mauvaise valeur. Une implantation particulière de l'analyse de perturbation infinitésimale constitue une exception: dans ce cas, l'algorithme converge vers l'optimum même si chaque itération comporte le même nombre de clients (et ce nombre peut être très petit). Nous avons effectué des expériences numériques avec une simple file d'attente M/M/1, et les résultats illustrent bien les propriétés de convergence et les dangers à éviter avec les différentes méthodes.

I. Introduction

Simulation has traditionally been used to *evaluate* the performance of complex systems, especially when analytic formulas are not available. Using it to perform *optimization* is much more challenging. Consider a (stochastic) simulation model parametrized by a vector θ of continuous parameters, and suppose one seeks to minimize the expected value $\alpha(\theta)$ of some objective function. In principle, if $\alpha(\theta)$ is well behaved, one could estimate its derivative (or gradient) by simulation, and use adapted versions of classical nonlinear programming algorithms. Recently, the question of how to estimate the gradient of a performance measure (defined as a mathematical expectation), with respect to continuous parameters, by simulation, has attracted a great deal of attention [5, 13, 14, 15, 16, 17, 19, 20, 27, 35, 36, 37, 39, 41, 42]. For "steady-state" simulations, a "single-run" iterative optimization scheme based on stochastic approximation (SA) has been suggested [8, 21, 33, 35, 43, 44]. At each iteration, this scheme uses an estimate of the gradient of α to modify the current parameter value. Combined with appropriate variance reduction techniques, these methods could enlarge substantially the class of stochastic optimization problems that can be solved.

In this paper, we investigate the combination of SA with different derivative estimation techniques (DETs). The general theory of SA has been studied extensively (see [15, 21, 23, 24, 25, 34, 35, 37] and many reference cited there), but not very much their combination with various DETs, for discrete-event systems in the steady-state context, as we do here. Preliminary empirical experiments have been undertaken by Suri and Leung [43] for a M/M/1 queue. These authors look at two SA methods, which they presented as heuristics. One was based on infinitesimal perturbation analysis (IPA), while the other was an adaptation of the Kiefer-Wolfowitz (KW) algorithm, which uses finite differences (FD) to estimate the derivative. They observed empirically that for the problem considered, the approach based on IPA converged much faster than KW. Our results here imply that for the example they examined, their first method converges to the optimal solution, while their second might converge to the wrong value. We examine in this paper many other DETs, including some based on FD, with and without common random numbers (CRN), IPA [5, 17, 19, 20, 41, 42, 44, 46, 47], and variants of the likelihood ratio (LR) method [1, 13, 14, 16, 27, 36, 38, 39] (sometimes called the score function (SF) method). These techniques can be combined with SA in different ways. At each iteration, one can use for example a (deterministic) truncated horizon or perhaps a regenerative approach. The length of the horizon or the number of regenerative cycles for an iteration can remain fixed from one iteration to the next, or may vary (e.g., increase). We discuss convergence conditions for many such combinations and argue that some typically converge to the wrong value when using a fixed simulation length at each iteration. Increasing the simulation length between iterations typically reduces the bias. In many cases, the length of the simulation at each iteration should have little or no influence on the convergence rate in terms of the total simulation length, *provided* that the variance of the derivative estimator decreases linearly with the simulation length (we do not prove that, but for supporting arguments, see [23], Section 7.4). In fact, we do not study convergence rates in this paper. We suggest KW variants that converge to the optimal solution. One of these variants, which uses CRNs and increases the simulation length at each iteration, appears competitive with IPA, at least for our example. Fu and Ho [8] also obtained promising experimental results with different ("extended" or "improved") SA algorithms. (We do not consider these in this paper.)

In Section II, we consider a GI/G/1 queue for which the decision variable is a scale parameter of the service time distribution. The aim is to minimize a function of the average system time per customer. We take this example to explain what happens when SA is used to (try to) optimize infinite-horizon (steady-state) simulations. We feel that most of the important questions that would arise in more general models are well illustrated by this simple example. We recall the classical SA algorithm and give (simplified) sufficient conditions for its convergence to the optimum. More general conditions are given in Appendix I. Section III reviews different ways of estimating the derivative (DETs). For each of them, we prove convergence of SA to the optimum, under specific conditions. Later on, in Section V, we discuss how all this can be extended to more general systems. Our presentation is made in the steady-state setting, but it also applies (with lots of simplifications) to finite-horizon simulations or to the case of infinite-horizon total discounted cost, by taking $\alpha(\theta)$ as the total expected (possibly discounted) cost at parameter level θ. The initial simulation state is then fixed (or could be part of the parameter, or random with known distribution). For finite-horizon simulations, many things simplify since the initial bias problem disappears.

The proofs of most of our results of Sections II and III are relegated to Appendix II. These proofs are quite involved and one of the reason is that since the value of θ is constantly changing, some convergence properties of the derivative estimators (like, for instance, bounded variance and convergence in expectation to the steady-state derivative) must be shown to hold *uniformly* in θ. As a byproduct, we obtain original results concerning GI/G/1 queues that could be of independent interest. For instance, it follows from the renewal-reward theorem [45] that for a stable queue, the average sojourn time of the first t customers in the queue converges in expectation, as $t \to \infty$, to the infinite-horizon average sojourn time per customer. We prove, under appropriate conditions, that this convergence is uniform over θ and s, where s is the initial *state* (taken over some compact set), which corresponds to the waiting time of the first customer, and θ lies in a compact set in which the system is (uniformly) stable. We also derive a similar uniform convergence result for the derivative of the expected average sojourn time with respect to θ, and a few additional characterizations of this expectation.

In Section IV, we report an extensive numerical investigation, for an M/M/1 example similar to the one studied in [43]. The idea was to run the algorithm variants on a problem for which we could compute analytically the optimal solution. Our experiments deal with an example where the decision parameter vector θ has only one component. A multidimensional case would certainly involve more intensive computations. As always, since these experiments were done on a specific example, one should be careful in making any generalizations. The primary goal of this example is not really to compare performance, but to illustrate convergence properties and possible dangers. We also recall that in many cases, IPA and/or LR do not apply [27]. Other numerical results for other kinds of examples are given in the master's thesis of the second author [9], which has been the starting point of this paper. In the conclusion, we summarize our results and mention prospects for further research.

II. Example: a GI/G/1 queue

A. *The basic model*

Consider a GI/G/1 queue [2, 45] with interarrival and service-time distributions A and B_θ respectively, both with finite expectations and variances. The latter depends on a parameter $\theta \in \bar{\Theta} = [\ell_0, u_0] \subset \mathbb{R}$ and has a corresponding density function b_θ. To simplify analysis, we assume that θ is a scale parameter, i.e. $B_\theta(\zeta) = B(\zeta/\theta)$ for some distribution B. The more general case will be discussed later on. We assume that $\ell_0 < 1 < u_0$ and that for $\theta = u_0$, the system is stable. This implies that the system is also stable for smaller values of θ. Let $w(\theta)$ be the average sojourn time in the system per customer, in steady-state, at parameter level θ. The objective function is defined by

$$\alpha(\theta) = w(\theta) + C(\theta). \tag{1}$$

where $C : \bar{\Theta} \mapsto \mathbb{R}$ is continuously differentiable. We want to minimize $\alpha(\theta)$ over $\Theta = [\ell_1, u_1]$, where $\ell_0 < \ell_1 < u_1 < u_0$. Let θ^* be the optimum. The reason we define $\bar{\Theta}$ and Θ this way is to be able to do finite-difference derivative estimation at any point of Θ (see next Section). This is also useful for LR and IPA.

For many distributions, $\alpha(\theta)$ and its minimizer θ^* can be computed analytically or numerically. But let us ignore this momentarily and see how the problem can be solved using SA combined with different DETs. The solutions of some numerical examples can then be compared to the true optimal solutions for an empirical evaluation.

A GI/G/1 queue can be described in terms of a discrete-time Markov chain via Lindley's equation. For $i \geq 1$, let W_i, $\zeta_i = \theta Z_i$, and $W_i^* = W_i + \zeta_i$ be the *waiting* time, *service* time, and *sojourn* time for the i-th customer, and ν_i be the time between arrivals of the i-th and $(i+1)$-th customer. Here, Z_i follows the distribution B. For our purposes, W_i will be the state of the Markov chain at step i. The state space is $S = [0, \infty)$ and $W_1 = s$ is the initial state. $W_1 = 0$ corresponds to an initially empty system (zero wait for the first customer). For $i \geq 1$, one has

$$W_i^* := W_i + \zeta_i \qquad \text{and} \qquad W_{i+1} := (W_i^* - \nu_i)^+ \tag{2}$$

where x^+ means $\max(x, 0)$. Since $C(\theta)$ is deterministic, we will estimate only the derivative of $w(\theta)$ and then add $C'(\theta)$ separately to Y_n.

We can view the Markov chain $\{W_i, i = 1, 2, \ldots\}$ as being defined over a probability space $(\Omega, \Sigma, P_{\theta,s})$, where the sample point $\omega \in \Omega$ represents the "randomness" that drives the system and the associated probability measure $P_{\theta,s}$ depends (in general) on θ and s (where $W_1 = s \in S$ is deterministic). [As we will see later, there are different ways of defining that probability space and we sometimes define it in such a way that $P_{\theta,s}$ actually depends neither on θ nor on s, but that all the dependency on (θ, s) appears in the transformation from ω to the W_i's and W_i^*'s. For example, ω may represent the sequence of inter-arrival and service times, or may represent a sequence of i.i.d. $U(0,1)$ variates]. Let $E_{\theta,s}$ denote the corresponding mathematical expectation. When the quantities involved do not depend on s, we sometimes denote it by E_θ to simplify the notation.

For $t \geq 1$, let

$$h_t(\theta, s, \omega) = \sum_{i=1}^{t} W_i^*; \tag{3}$$

$$w_t(\theta, s) = \int_\Omega h_t(\theta, s, \omega) dP_{\theta,s}(\omega); \tag{4}$$

$$\alpha_t(\theta, s) = C(\theta) + w_t(\theta, s)/t. \tag{5}$$

Here, $h_t(\theta, s, \omega)$ represents the *total* sojourn time in the system for the first t customers, and $w_t(\theta, s)$ its expectation. Also, $C(\theta) + h_t(\theta, s, \omega)/t$ is the average cost for the first t customers, and $\alpha_t(\theta, s)$ its expectation. Let $\mathcal{F}_t$ be the sigma-field generated by $(\zeta_1, \nu_1, \ldots, \zeta_t, \nu_t)$. Then, $h_t(\theta, s, \omega)$ is $\mathcal{F}_t$-measurable. Also, if $s = 0$ and if τ denotes the number of customers in the first busy cycle (i.e., $\tau + 1$ is the smallest $i > 1$ such that $W_i = 0$), then $\tau + 1$ is a stopping time with respect to $\{\mathcal{F}_t,\ t \geq 1\}$.

Let $\bar{S} = [0, c]$, for some (perhaps large) constant c, which can be viewed as the set of admissible initial states. If $c = 0$, then all simulations are started from the empty state. It is well known from renewal theory that for each fixed $\theta \in \bar{\Theta}$ and $s \in \bar{S}$, $\lim_{t\to\infty} w_t(\theta, s)/t = w(\theta)$, and $\lim_{t\to\infty} \alpha_t(\theta, s) = \alpha(\theta)$.

B. *A stochastic approximation scheme*

We consider a stochastic approximation (SA) algorithm of the form

$$\theta_{n+1} := \pi_\Theta(\theta_n - \gamma_n Y_n), \tag{6}$$

for $n \geq 1$, where θ_n is the parameter value at the beginning of iteration n ($\theta_1 \in \Theta$ is fixed, or random with known distribution), Y_n is an estimate of the derivative $\alpha'(\theta_n)$ obtained at iteration n, $\{\gamma_n, n \geq 1\}$ is a (deterministic) positive sequence decreasing to 0 such that $\sum_{n=1}^{\infty} \gamma_n = \infty$, and π_Θ denotes the projection on the set Θ (i.e. $\pi_\Theta(\theta)$ is the point of Θ closest to θ). In what follows, except when stated otherwise, we will assume that $\gamma_n = \gamma_0 n^{-1}$ for some constant $\gamma_0 > 0$.

To obtain Y_n, we can compute directly the derivative of the deterministic term $C(\theta_n)$, and estimate only $w'(\theta_n)$. Each such estimation is obtained by simulating the system for one or more "subrun(s)" of finite duration. Each simulation subrun corresponds essentially to one copy of the Markov chain described above, with initial state $s \in \bar{S}$. Specific ways of obtaining Y_n are discussed in the next section. Some use a deterministic *truncated* horizon t_n at iteration n. Others exploit the regenerative structure, in which case the horizon t is the value taken by a random variable T (a stopping time). These estimators are usually biased. Some of the methods for computing Y_n require keeping information beyond the state of the chain $\{W_i,\ i \geq 1\}$. For IPA, for example, we usually maintain IPA "accumulators". To deal with that, we will extend the state space when necessary. But in all cases, that added part of the state has no influence on the future evolution of the chain in a given subrun and on Equations (3–9).

Let $s_n \in \bar{S}$ denote the *state* of the system at the beginning of iteration n. We assume that when θ_n and s_n are fixed, the distribution of (Y_n, s_{n+1}) is completely specified and independent of the past iterations (but may depend on n). In other words, $\{(Y_n, \theta_{n+1}, s_{n+1}), n \geq 0\}$ evolves as a (nonhomogeneous) Markov chain. (Here, Y_0 is a dummy value). Denote by $E_{n-1}(\cdot)$ the conditional expectation $E(\cdot \mid \theta_n, s_n)$, that is the expectation conditional on what is known at the beginning of iteration n. Assume that Y_n is integrable for all $n \geq 1$. Then, $E_{n-1}(Y_n)$ exists and we can write

$$Y_n = \alpha'(\theta_n) + \beta_n + \epsilon_n \tag{7}$$

where $\beta_n = E_{n-1}[Y_n] - \alpha'(\theta_n)$ represents the (conditional) bias on Y_n given (θ_n, s_n), while ϵ_n is a random sequence, with $E_{n-1}(\epsilon_n) = 0$ and $E_{n-1}(\epsilon_n^2) =$ var $(Y_n \mid \theta_n, s_n)$.

C. Convergence to the optimum

We now give (simplified) sufficient conditions for the convergence of (6) to an optimum. The following proposition is a special case of Theorem 2 in Appendix I. It treats the case where the (conditional) bias β_n goes to zero and the variance of Y_n does not increase too fast with n. When the DET uses the same simulation length at each iteration, β_n typically *does not* go to zero. But sometimes, $E_0(\beta_n) \to 0$ and the algorithm might still converge to the optimum. This is covered by Theorem 4 of Appendix I. This latter theorem insures only weak convergence, but this is good enough for practical applications. Chong and Ramadge [7] also analyze a situation where β_n does not converge to zero. They use a different approach than ours, estimate the derivative differently, and prove almost sure convergence to the optimum.

Proposition 1. *Suppose that α is differentiable and strictly unimodal over Θ. If $\lim_{n\to\infty} \beta_n = 0$ with probability one and $\sum_{n=1}^{\infty} E_0(\epsilon_n^2) n^{-2} < \infty$ with probability one, then $\lim_{n\to\infty} \theta_n = \theta^*$ with probability one.* ∎

For convenience in the following sections, we will decompose β_n as $\beta_n = \beta_n^F + \beta_n^R$, where β_n^F is the bias component due to the fact that we simulate over a finite (truncated) horizon and β_n^R represents the possibility that Y_n may itself be a biased estimator of the derivative of the finite-horizon expected cost. Typically, when we use finite differences, $\beta_n^R \neq 0$. If we use a deterministic (truncated) horizon t_n at iteration n, then $\beta_n^F = w'_{t_n}(\theta_n, s_n)/t_n - w'(\theta_n)$. Here and throughout the paper, the "prime" denotes the derivative with respect to θ. To make sure that the latter converges to zero with probability one, we will show, under appropriate conditions, that $w'_t(\theta, s)/t - w'(\theta)$ converges to zero *uniformly* in (θ, s) as t goes to infinity. This is discussed in the next subsection.

D. Continuous differentiability and uniform convergence

We want sufficient conditions under which α is strictly unimodal, w and each $w_t(\cdot, s)$ are differentiable, and the following uniform convergence results hold:

$$\lim_{t\to\infty} \sup_{\theta\in\bar{\Theta}, s\in\bar{S}} |w_t(\theta, s)/t - w(\theta)| = 0 \tag{8}$$

and

$$\lim_{t\to\infty} \sup_{\theta\in\bar{\Theta}, s\in\bar{S}} |w'_t(\theta, s)/t - w'(\theta)| = 0. \tag{9}$$

In Proposition 17 of Appendix II, we establish (8–9) under the following assumption 1. We also prove, under that Assumption, that $w_t(\theta, s)/t$ is convex and continuously differentiable in θ for each s and t, and that α is also strictly convex and continuously differentiable. Note that these continuous differentiability properties can be expected to hold only when appropriate regularity conditions are imposed on the distribution of Z. For example, if Z is deterministic, then each W_i^* and $w_t(\theta, s)$ are continuous, but only piecewise differentiable in θ. This does not means that SA will not work. The continuous differentiability that is exploited here is merely a sufficient condition for the validity of SA.

Assumption 1. *(i) Suppose that θ is a scale parameter, i.e. that a service time can be written as $\zeta = \theta Z$ where $\theta > 0$ and Z is a random variable with distribution $B \stackrel{\text{def}}{=} B_1$ and density $b \stackrel{\text{def}}{=} b_1$. Equivalently, $b_\theta(\zeta) = b(\zeta/\theta)/\theta$, for $\theta > 0$. Assume that the set $\{z \geq 0 \mid b(z) > 0\}$, which is the support of b, is $[0, \infty)$.*

(ii) Let b be continuously differentiable and have a finite Laplace transform in a neighborhood of zero.

(iii) Suppose that for each $\theta_0 \in \bar{\Theta}$ and $K > 1$, there is an ϵ_0 such that $0 < \epsilon_0 < \theta_0$,

$$\sup_{|\theta-\theta_0|<\epsilon} \left(\frac{b_\theta(\zeta)}{b_{\theta_0+\epsilon_0}(\zeta)} \right) \leq K \quad \text{a.s.} \tag{10}$$

and

$$E_{\theta_0+\epsilon_0} \left[\sup_{|\theta-\theta_0|<\epsilon} \left(\frac{\partial}{\partial\theta} \ln b_\theta(\zeta) \right)^4 \right] < \infty. \tag{11}$$

(iv) Let $E_{\theta=1}[\zeta^8] < \infty$.

(v) Let C be strictly convex and continuously differentiable in $\bar{\Theta}$. ∎

III. Ways of estimating the derivative

One crucial ingredient for the SA algorithm considered here is an efficient derivative estimation technique (DET). In this section, we survey some possibilities, discussing their efficiencies and implementation difficulties.

A. Finite differences (FD)

This method is described, for instance, in [15, 23, 37, 46], without the projection operator. When used in conjunction with FD, the SA algorithm (6) is known as the Kiefer-Wolfowitz (KW) algorithm. Here, we describe and use *central* (or *symmetric*) FD. For other variants, like *forward* FD, see [15, 23, 37, 46]. When there are d parameters instead of just one, i.e. if θ is a d-dimensional vector, the latter uses only $d+1$ instead of $2d$ subruns per iteration. However, its asymptotic convergence rate is not as good [15].

Take a deterministic positive sequence $\{c_n, n \geq 1\}$ that converges to 0. At iteration n, simulate from some initial state $s_n^- \in \bar{S}$ at parameter value $\theta_n^- = \pi_\Theta(\theta_n - c_n)$ for t_n customers. Simulate also (independently) from state $s_n^+ \in \bar{S}$ at parameter value $\theta_n^+ = \pi_\Theta(\theta_n + c_n)$ for t_n customers. Ways of selecting s_n^- and s_n^+ will be discussed later. Let ω_n^- and ω_n^+ denote the respective sample points. The FD derivative estimator is

$$Y_n = C'(\theta_n) + \frac{h_{t_n}(\theta_n^+, s_n^+, \omega_n^+) - h_{t_n}(\theta_n^-, s_n^-, \omega_n^-)}{(\theta_n^+ - \theta_n^-)t_n}. \tag{12}$$

Here, the conditional bias $\beta_n^R = E_{n-1}[Y_n] - \alpha'_{t_n}(\theta_n, s_n)$ can itself be decomposed as

$$\beta_n^R = \beta_n^D + \beta_n^I \tag{13}$$

where

$$\beta_n^D = \frac{w_{t_n}(\theta_n^+, s_n) - w_{t_n}(\theta_n^-, s_n))}{(\theta_n^+ - \theta_n^-)t_n} - w'_{t_n}(\theta_n, s_n)/t_n \tag{14}$$

and

$$\beta_n^I = \frac{w_{t_n}(\theta_n^+, s_n^+) - w_{t_n}(\theta_n^+, s_n) + w_{t_n}(\theta_n^-, s_n) - w_{t_n}(\theta_n^-, s_n^-)}{(\theta_n^+ - \theta_n^-)t_n} \tag{15}$$

represent respectively the bias due to finite differences and the bias due to the possibly different initial states.

Proposition 2. *Let Asssumption 1 hold. If* $t_n \to \infty$, *then* $\lim_{n\to\infty} \beta_n^D = 0$.

PROOF. From Propositions 11 and 12 of Appendix II, $w(\cdot)$ and $w_t(\cdot, s)/t$ are continuously differentiable, for each $s \in \bar{S}$ and $t \geq 0$. Also, from Proposition 17, (9) holds. From Taylor's theorem, one has $\beta_n^D = (w'_{t_n}(\xi_n, s_n) - w'_{t_n}(\theta_n, s_n))/t_n$ for $\theta_n^- \leq \xi_n \leq \theta_n^+$. Note that as $n \to \infty$, one has $\theta_n^+ - \theta_n^- \to 0$ and therefore $\xi_n \to \theta_n$. Then, $\beta_n^D = [w'_{t_n}(\xi_n, s_n)/t_n - w'(\xi_n)] + [w'(\xi_n) - w'(\theta_n)] + [w'(\theta_n) - w'_{t_n}(\theta_n, s_n)/t_n]$ and each bracketed term converges to zero from (9) and from the continuity of w'. ∎

The term β_n^I can be eliminated by picking $s_n^- = s_n^+ = s_n$. Otherwise, if the distance between s_n^- and s_n^+ is in the order of 1, the numerator in (15) should decrease with t_n at rate in the order of $1/t_n$ (asymptotically). In that case, to get $\beta_n^I \to 0$, we could take $1/(t_n c_n) \to 0$. Even when $\beta_n^I = 0$, taking t_n constant may lead to problems, as will be illustrated later, because β_n^F is usually *not* zero. As c_n decreases to zero, when ω_n^- and ω_n^+ are distinct ("independent"), the variance on Y_n usually increases to infinity. However, we have the following proposition, whose proof is in Appendix II.

Proposition 3. *Let Assumption 1 hold, $t_n \to \infty$, $c_n \to 0$, and $\sum_{n=1}^{\infty} t_n^{-1}(nc_n)^{-2} < \infty$. Assume that $\beta_n^I \to 0$ a.s. as $n \to \infty$ (this can be achieved trivially by taking $s_n^- = s_n^+ = s_n$). Then, $\theta_n \to \theta^*$ with probability one.* ∎

Reasonable choices for the sequences might be for instance $t_n = t_a + t_b n$ and $c_n = c_0 n^{-1/6}$ for appropriate constants t_a, t_b, and c_0. This choice of c_n is motivated by what happens in the (different, but related) case where $\alpha(\theta)$ is an objective function over a fixed finite horizon and where t_n represents the number of i.i.d. replications. In that case, under reasonable assumptions, the best possible convergence rate for the derivative estimator is $t_n^{-1/3}$ and it is obtained when $c_n t_n^{1/6}$ converges to a constant. This choice of c_n also gives the best convergence rate to θ^* for the KW procedure. This rate is $t^{-1/3}$ in terms of the total simulation length $t = 2d\sum_{i=1}^{n} t_i$. (See [15, 23, 46]).

One simple way to choose the initial states of the subruns is as follows. Start the first subrun of iteration n from state $s_n \in \bar{S}$. We then take the terminal state of any given subrun as the initial state of the next one. (Project on $\bar{S}$ whenever necessary.) For s_{n+1}, take the terminal state of the last subrun of iteration n. Still, the two subruns of a given iteration can be ordered in two different ways. More generally, if θ has dimension d, one can permute the $2d$ subruns of a given iteration in any given way, and select the terminal state of any subrun for s_{n+1}. It is not clear what the best way of doing this is, if any. Another way is to take the same initial state for each subrun: $s_n^- = s_n^+ = s_n$. Again, there are different possibilities for the selection of s_{n+1}. One can also take (reset) $s_n = s_0$ for all n, for a fixed state s_0. But in any case, the KW method is usually plagued by a huge variance on Y_n, which makes it converge very slowly, at least when the subruns are performed with "independent" random numbers.

Glynn [13] describes an alternative KW approach based on regenerative analysis. It eliminates the bias β_n^F and β_n^I, but the variance is usually much higher. The package SAMOPT [3] is an implementation of KW with specially tuned parameters. It was designed for finite-horizon simulations. It also replaces Y_n by its sign.

B. Finite differences with common random numbers (FDC)

One way to reduce the variance in FD is to use common random numbers across the subruns at each iteration, start all the subruns from the same state: $s_n^- = s_n^+ = s_n$, and synchronize. More specifically, one views ω as representing a sequence of $U(0,1)$ variates, so that all the dependency on (θ, s) appears in $h_t(\theta, s, \cdot)$. Take $\omega_n^+ = \omega_n^- = \omega_n$. Since the subruns are aimed at comparing very similar systems, $h_{t_n}(\theta_n^+, s_n, \omega_n)$ and $h_{t_n}(\theta_n^-, s_n, \omega_n)$ should be highly correlated, especially when c_n is small, so that considerable variance reductions might be obtained. Conditions that *guarantee* variance reductions are given in [4, 37]. Proposition 2 still applies. For the related case of a fixed finite horizon with t_n i.i.d. replications, under additional assumptions, Glynn [15] shows that the best convergence rate for FDC is $t_n^{-2/5}$, which is attained if $c_n t_n^{1/5}$ converges to a constant. Under a different set of assumptions, which are essentially the same assumptions that one usually makes to show that IPA works, L'Ecuyer and Perron [28] show that FDC has the same convergence rate as IPA, namely $t_n^{-1/2}$ if $c_n t_n^{1/2}$ converges to a constant.

What could happen if t_n is kept constant ? Let us look at the simplest case, where $t_n = 1$ for each n.

Proposition 4. *Suppose that $C(\theta) + \theta$ has its minimum at $\tilde{\theta}^0$. Let $\theta^0 = \pi_\Theta(\tilde{\theta}^0)$. Then, with $t_n = 1$, SA with FDC converges to θ^0 almost surely.*

PROOF. When we estimate the average cost using $t_n = 1$, we actually look at the time spent in the system by *one* customer, i.e. the customer being served in that subrun. This time can be expressed as $h_1(\theta, s, \omega) = s + \theta B^{-1}(\omega)$, where s is the (waiting) time already spent in the system by that customer and ω is viewed as the $U(0,1)$ variate used to generate its service time. We then have

$$Y_n - C'(\theta) = \frac{h_1(\theta_n^+, s_n, \omega) - h_1(\theta_n^-, s_n, \omega)}{\theta_n^+ - \theta_n^-} = \frac{\theta_n^+ B^{-1}(\omega) - \theta_n^- B^{-1}(\omega)}{\theta_n^+ - \theta_n^-} = B^{-1}(\omega).$$

which has finite variance, from Assumption 1 (iv). Also, $E_n[Y_n] = 1 + C'(\theta)$. If we redefine for the moment $w(\theta) = \theta + C(\theta)$ and apply Proposition 1, the conclusion follows. ∎

As an illustration, take an M/M/1 queue with arrival rate $\lambda = 1$, mean service time $\theta \in \Theta = [\ell_1, u_1]$ for $u_1 < 1$, and $C(\theta) = 1/\theta$. Here, $C(\theta) + \theta$ has its minimum at $\tilde{\theta}^0 = 1$. Therefore, θ_n converges to u_1 with probability one. The problem here is that with a different θ, the time spent in the queue by the customers already there at the beginning of the iteration would have been different and the method does not take that into account. This flaw also exists for any fixed $t_n = t$. The difference $|\theta^0 - \theta^*|$ should decrease with t. In our numerical results of Section IV, for t as large as 100, the effect is still significant.

As initial states, one can take $s_n = s_0 \in \bar{S}$ for all n (for some fixed s_0), or s_{n+1} can be one of the two (or $2d$, in general) terminal states of iteration n. Whenever that state is outside $\bar{S}$, project back to $\bar{S}$. Since we are interested in steady-state behavior, taking a terminal state of the previous iteration appears intuitively better. A heuristic rule is to choose the state that was obtained from the subrun with the parameter value the closest to the new parameter value θ_{n+1}. Implementing this method for complex simulations is not without pain. Saving the simulation state means saving the states of the random number generators, the event list, all the objects in the model, etc. In practice, many objects in the model are pointers to data structures that can be created, modified or destroyed dynamically, and whose types have been defined by the programmer. When saving the state of the system, one cannot only save the pointer values, but must make an explicit "backup" copy of all these structures. When restoring the system to a given state, these must be recopied again. This is different than saving and restoring the state of the *program*, because some variables associated with the SA and FD algorithms (e.g., the index of the current subrun for FD) should *not* be saved and restored. Usually, the simulation package cannot do that and specific code must be written. In fact, it would be very difficult to implement "state saving" facilities in a general simulation package, because typically, the package has no way of knowing with certainty the structures of all the dynamic objects created by the user. All this implies overhead not only for the computer, but also for the programmer. Another source of programming overhead in FDC comes from the need to insure synchronization of the random numbers across the subruns.

When c_n is small, there is sometimes little change between the sample paths of the two subruns. One could then ask: is it possible to perform only *one* subrun and trace the few changes ? It is indeed sometimes possible, and this idea leads to what is called *finite perturbation analysis*. Taking that to the limit when c_n goes to zero, one obtains *infinitesimal perturbation analysis* (see below). These techniques permit "single-run" optimization algorithms, which can be applied not only to simulations, but also (on-line) to actual systems. KW is not really a "single-run" algorithm and its direct application to actual systems is more limited.

C. A likelihood ratio (LR) approach

The LR derivative estimation approach has drawn a lot of attention lately (e.g., [1, 13, 14, 16, 27, 36, 38, 39]). Here, for any $s \in S$, to differentiate the expectation (5) with respect to θ, take a probability measure G_s independent of θ that dominates the $P_{\theta,s}$'s for $\theta \in \bar{\Theta}$, and rewrite:

$$w_t(\theta, s) = \int_\Omega h_t(\theta, s, \omega) L_t(G_s, \theta, s, \omega) dG_s(\omega) \tag{16}$$

where $L_t(G_s, \theta, s, \omega) = (dP_{\theta,s}/dG_s)(\omega)$ is the *Radon-Nikodym* derivative of $P_{\theta,s}$ with respect to G_s. Under appropriate *regularity conditions* (see [27, 29, 36]), one can differentiate w_t by differentiating inside the integral:

$$w'_t(\theta, s) = \int_\Omega \psi_t(\theta, s, \omega) dG_s(\omega). \tag{17}$$

where

$$\psi_t(\theta, s, \omega) = L_t(G_s, \theta, s, \omega) h'_t(\theta, s, \omega) + h_t(\theta, s, \omega) L'_t(G_s, \theta, s, \omega). \tag{18}$$

When (17) holds, the LR estimator $\psi_t(\theta, s, \omega)$ can be used to estimate $w'_t(\theta, s)$. Only *one* simulation experiment (using G_s) is required to estimate the derivative.

In our case, we view ω as representing the sequence of inter-arrival and service times for the first t customers, that is $\omega = (\zeta_1, \nu_1, \ldots, \zeta_t, \nu_t)$, and take $G_s = P_{\theta_0,s}$ where θ_0 is the value of θ at which we want to estimate the derivative. [Note that here, G_s and $P_{\theta,s}$ do not depend on s.] The Radon-Nikodym derivative then becomes the *likelihood ratio*

$$L_t(P_{\theta_0,s}, \theta, s, \omega) = \prod_{i=1}^{t} \frac{b_\theta(\zeta_i)}{b_{\theta_0}(\zeta_i)}, \tag{19}$$

which is 1 at $\theta = \theta_0$. The derivative of (19) is $L_t(P_{\theta_0,s}, \theta, s, \omega) S_t(\theta, s, \omega)$, where

$$S_t(\theta, s, \omega) = \sum_{i=1}^{t} d_i, \tag{20}$$

and

$$d_i = \frac{\partial}{\partial \theta} \ln b_\theta(\zeta_i). \tag{21}$$

The expression (20) is called the *score function*. The LR estimator (18) then becomes

$$\psi_t(\theta, s, \omega) = h_t(\theta, s, \omega) L_t(P_{\theta_0,s}, \theta, s, \omega) S_t(\theta, s, \omega). \quad (22)$$

We show in Proposition 11 of Appendix II that under Assumption 1, (22) is an unbiased estimator of $w'_t(\theta, s)$ for θ in some neighborhood of θ_0. After adding the derivative of the deterministic part and letting $\theta_0 = \theta_n$, the LR derivative estimate at iteration n for SA becomes

$$Y_n = C'(\theta_n) + \psi_{t_n}(\theta_n, s_n, \omega_n)/t_n = C'(\theta_n) + h_{t_n}(\theta_n, s_n, \omega_n) S_{t_n}(\theta_n, s_n, \omega_n)/t_n. \quad (23)$$

Note that the variance of $S_t(\theta, s, \omega)$ (and of the derivative estimator (22) at $\theta = \theta_0$) increases with t. This is a significant drawback and must be taken into account when making the tradeoff between bias and variance. Here, $\beta_n^R = 0$ and β_n^F goes to zero as t_n goes to infinity. But the variance on Y_n then goes to infinity also. One remedy, as in FD, is to increase t_n with n, but not too fast. We show in Proposition 18 that under Assumption 1, the variance of Y_n does not increase faster than linearly in t_n. The conditions of Proposition 1 can then be verified with $\gamma_n = \gamma_0 n^{-1}$ and $t_n = t_a + t_b n^p$ for $0 < p < 1$. In the finite-horizon case, SA with LR converges at a rate of $t^{-1/2}$ [15] in terms of the total simulation length t. But when the variance increases with t_n and t_n increases with n, this is no longer true. L'Ecuyer [31] examines infinite-horizon derivative estimation with LR and shows that the best convergence rate for the *derivative estimator* as a function of the total computer budget is obtained when the simulation length is in the order of $n^{1/3}$ as a function of the number n of replications. This slow increase reflects the fact that the variance increases too rapidly when we try to get rid of the bias through longer simulations. In [30], a control variate scheme is proposed to eliminate that linear increase of the variance of $\psi_t(\theta, s, \omega)/t$. Instead, the variance remains in the order of 1 with respect to t. We will experiment with that scheme in Section IV.

Another way of reducing (less dramatically) the variance is to estimate the derivative of the expected sojourn time of each individual customer separately, and then add. This uses the idea that since the sojourn time of customer i is independent of the service times of the customers that follow him, the appropriate score function that should multiply W_i^* is the sum up to i instead of up to t. This gives the following *triangular* LR derivative estimator:

$$\tilde{\psi}_t(\theta, s, \omega) = \sum_{i=1}^{t} \left(W_i^* \sum_{j=1}^{i} d_j \right). \quad (24)$$

Some variants of the LR approach circumvent the bias/variance problem by using a regenerative approach [13, 14, 30, 36]. Let $s = 0$ and let τ be the number of customers in, say, the first regenerative cycle (busy period). From elementary renewal theory one has $w(\theta) = u(\theta)/\ell(\theta)$ where $u(\theta)$ and $\ell(\theta)$ are respectively the expected total system time (for all the customers) and the expected number of customers over a regenerative cycle:

$$\begin{aligned} u(\theta) &= E_{\theta,0}[\tau]; \\ \ell(\theta) &= E_{\theta,0}\left[\sum_{i=1}^{\tau} W_i^*\right]. \end{aligned}$$

If $w'(\theta)$ exists, then, from standard calculus, one has

$$w'(\theta) = \frac{u'(\theta)\ell(\theta) - \ell'(\theta)u(\theta)}{\ell^2(\theta)} = \frac{u'(\theta) - w(\theta)\ell'(\theta)}{\ell(\theta)}. \quad (25)$$

Combining estimators for each of the four quantities on the right-hand-side of (25) yields an estimator for $w'(\theta)$. One can estimate $\ell(\theta)$ and $w(\theta) = u(\theta)/\ell(\theta)$ as usual, using say r regenerative cycles, while $u'(\theta)$ and $\ell'(\theta)$ can be estimated through the LR method as follows.

Let $G_0 = P_{\theta_0,0}$. The likelihood ratio and score function associated with the first regenerative cycle are respectively

$$L_\tau(P_{\theta_0,0}, \theta, 0, \omega) = \prod_{i=1}^{\tau} \frac{b_\theta(\zeta_i)}{b_{\theta_0}(\zeta_i)} \tag{26}$$

and

$$S_\tau(\theta, 0, \omega) = \sum_{i=1}^{\tau} d_i. \tag{27}$$

The LR estimators of $u'(\theta)$ and $\ell'(\theta)$ based on that cycle are respectively

$$\psi_u(\theta, \omega) = \left(\sum_{i=1}^{\tau} W_i^*\right) L_\tau(P_{\theta_0,0}, \theta, 0, \omega) S_\tau(\theta, 0, \omega) \tag{28}$$

and

$$\psi_\ell(\theta, \omega) = \tau L_\tau(P_{\theta_0,0}, \theta, 0, \omega) S_\tau(\theta, 0, \omega). \tag{29}$$

Proposition 12 in Appendix II states that these estimators are unbiased under Assumption 1. One can then repeat this and take the average over r regenerative cycles. More precisely, let τ_j be the number of departures during the j-th regenerative cycle, h_j the total system time for those τ_j customers who left during that cycle, and S_j the score function associated with that cycle. Then, an estimator of $w'(\theta_0)$ is given by

$$\frac{\sum_{j=1}^{r} \tau_j \sum_{j=1}^{r} h_j S_j - \sum_{j=1}^{r} h_j \sum_{j=1}^{r} \tau_j S_j}{\left(\sum_{j=1}^{r} \tau_j\right)^2}. \tag{30}$$

This estimator is biased, because it estimates the expectation of a ratio using correlated estimators for the numerator and denominator, and the expectation of products using correlated estimators for the factors. But the variance does not increase linearly with the simulation length as for the truncated horizon case. Algorithm B in [13] proposes a different regenerative estimator, which is unbiased. However, according to our experience, its variance is usually rather high.

In Proposition 19, we show that under Assumption 1 and if

$$\sup_{\theta \in \bar{\Theta}} E_\theta \left[d_i^8\right] < \infty, \tag{31}$$

then, as $r \to \infty$, (30) has bounded variance and converges in expectation to $w'(\theta_0)$, uniformly with respect to θ_0. This yields the following regenerative LR estimator of $\alpha'(\theta_n)$ for iteration n:

$$Y_n = C'(\theta_n) + \frac{\sum_{j=1}^{t_n} \tau_j \sum_{j=1}^{t_n} h_j S_j - \sum_{j=1}^{t_n} h_j \sum_{j=1}^{t_n} \tau_j S_j}{\left(\sum_{j=1}^{t_n} \tau_j\right)^2}. \tag{32}$$

The estimators (23) and (32) can be integrated into the SA algorithm. The following proposition, proved in Appendix II, tells us about the convergence of such a scheme.

Proposition 5. *Let Assumption 1 and Equation (31) hold.*

(a) *Suppose one uses SA with the LR estimator (23). If $s_n \in \bar{S}$ for all n, $t_n \to \infty$, and $\sum_{n=1}^{\infty} t_n n^{-2} < \infty$, then $\theta_n \to \theta^*$ with probability one.*

(b) *Suppose one uses SA with the regenerative LR estimator (32). If $t_n \to \infty$, then $\theta_n \to \theta^*$ with probability one.* ∎

D. Infinitesimal Perturbation analysis (IPA)

A straightforward approach to attack the variance problem in LR above is to define the sample space in such a manner that $P_{\theta,s}$ is independent of θ. For instance, one can view ω as a sequence of independent $U(0,1)$ variates that drive the simulation. Then, $L(P_{\theta_0,s}, \theta, s, \omega) = 1$ for all θ and the last term in (17) vanishes, yielding (under the appropriate regularity conditions):

$$\psi_t(\theta, s, \omega) = h_t'(\theta, s, \omega). \tag{33}$$

This is in fact the infinitesimal perturbation analysis (IPA) derivative estimator for $w_t'(\theta, s)$ [5, 12, 17, 19, 20, 41, 42, 44].

The basic idea of IPA is to generate a sample point ω, viewed as a sequence of $U(0,1)$ variates, and, for ω fixed, observe the effect of an infinitesimal perturbation on θ (around θ_0) by propagating it over the sample path, assuming that the sequence of events does not change, and that the events can only "slide smoothly" in time. The derivative estimation is taken as the derivative of the objective function for that fixed value of ω. Again, the derivative and expectation can be interchanged only under the appropriate regularity conditions [10, 19, 27]. Regenerative versions can be defined as for LR; see [19]. When IPA does not work for the original system, various devices can sometimes be used to "smooth out" or transform the original problem into a problem for which IPA will work correctly (see [11, 17, 28] for instance). These devices are usually problem-dependent. In principle, when (33) is unbiased, IPA can be viewed as a limiting version of FDC as c_n becomes infinitesimal. Note however that one must be careful about implementation "details", which can sometimes make a big difference between IPA and "infinitesimal" FDC. This is illustrated by Propositions 4 and 7.

Here, an infinitesimal perturbation on ζ_j affects the system time of customer j and of all the customers (if any) following him in the same busy period. Therefore,

$$h_t'(\theta, s, \omega) = \sum_{i=1}^{t} \sum_{j=v_i}^{i} \frac{\partial B_\theta^{-1}(U_j)}{\partial \theta} = \sum_{i=1}^{t} \sum_{j=v_i}^{i} Z_j, \tag{34}$$

where v_i is the first customer, with index ≥ 1, in the busy period to which customer i belongs. That is, $v_i = i$ if $W_i = 0$, and $v_i = \min\{j \geq 1 \mid W_j = 0 \text{ and } W_k > 0 \text{ for } j < k \leq i\}$ if $W_i > 0$. Then,

$$Y_n = C'(\theta_n) + h_t'(\theta_n, s_n, \omega)/t_n. \tag{35}$$

This can be computed easily during the simulation. The inside sum in (34) is called the *IPA accumulator.* Observe that imposing $v_i \geq 1$ means that we consider only the service time perturbations of the customers who left during the current iteration. In other words, we assume that the IPA accumulator is reset to zero between iterations. The initial state s_n of iteration n can be either 0 for all n (always restart from an empty system), or be the value of $(W_t^* - \nu_t)^+$ from the previous iteration (for $n > 1$).

We can consider another variant of IPA in which the IPA accumulator is *not* reset to zero between iterations. In that case, both s_n and the initial value a_n of the IPA accumulator are taken form the previous iteration. The value of a_n is the value of $\sum_{j=\nu_t}^{t} Z_j$ from the previous iteration if $s_n > 0$, and is 0 otherwise. The value of a_n must then be considered as part of the "state". Let

$$k_t^* = \min(t,\ \min\{i \geq 0 \mid W_{i+1} = 0\}). \tag{36}$$

When $W_1 = s > 0$, k_t^* represents the number of customers in the current iteration who are in the same busy period as the last customer of the previous iteration. For this IPA variant, (34) must be modified to:

$$h_t'(\theta, s, a, \omega) = ak_t^* + \sum_{i=1}^{t} \sum_{j=v_i}^{i} Z_j, \tag{37}$$

where a is the initial value of the IPA accumulator.

For the regenerative version, let $s = 0$ and τ be the number of customers in, say, the first regenerative cycle. The value of (34) for that cycle becomes

$$h_\tau'(\theta, 0, \omega) = \sum_{i=1}^{\tau} \sum_{j=1}^{i} Z_j. \tag{38}$$

With r regenerative cycles, let τ_j and h_j' denote the respective values of τ and $h_\tau'(\theta, 0, \omega)$ for the j-th regenerative cycle. The regenerative IPA estimator is then

$$\frac{\sum_{j=1}^{r} h_j'}{\sum_{j=1}^{r} \tau_j}. \tag{39}$$

At iteration n, one takes $r = t_n$ regenerative cycles and t is the value taken by a random variable $T_n = \sum_{j=1}^{t_n} \tau_j$. This yields

$$Y_n = C'(\theta_n) + \frac{\sum_{j=1}^{t_n} h_j'}{\sum_{j=1}^{t_n} \tau_j}. \tag{40}$$

In [19], the authors argue that the derivative estimator (37) is a consistent estimator of $w'(\theta)$ for a rather general class of GI/G/1 queues, and give a proof for the M/G/1 case. A proof for the GI/G/1 case is given in [47] under assumptions slightly different than ours. To prove convergence of the stochastic optimization algorithm using Proposition 1, what we need is not convergence of (34) divided by t to $w'(\theta)$ with probability one (as $t \to \infty$), but convergence in expectation, uniformly over θ. In fact, *both* kinds of convergence, as well as variance boundedness, follow from Propositions 13, 17, and 20 in Appendix II. Proposition 13 also shows that the IPA estimator (34) is unbiased under Assumption 1. This leads to the following Proposition, proven in Appendix II.

Proposition 6. *Let Assumption 1 hold.*

(a) *Suppose one uses SA with the IPA estimator (35). If $s_1 = 0$, $s_n \in \bar{S}$ and $a_n = 0$ for all n, and $t_n \to \infty$, then $\theta_n \to \theta^*$ with probability one.*

(b) *Suppose one uses SA with the regenerative IPA estimator (40), with t_n regenerative cycles at iteration n. If $t_n \to \infty$, then $\theta_n \to \theta^*$ with probability one.* ∎

If the IPA accumulator a_n is not reset to 0 between iterations, proving Proposition 6 (a) appears slightly more difficult. But we believe that the result still holds.

For this GI/G/1 example, IPA has the stronger property that even when using a truncated horizon t_n that is constant with n, if the IPA accumulator is kept between iterations and under mild additional assumptions, SA converges to the optimizer. But on the other hand, if the IPA accumulator is reset to zero at the beginning of each iteration, then we have the same problem as with FDC. When we keep the value of the accumulator across iterations, the estimator takes into account the service time perturbations due to all preceding customers, including those who left during previous iterations. It is true that the structure of the busy periods, and (in general) the individual terms of the sum (34), could depend on θ, which changes between iterations. But as θ_n converges to some value, that change becomes negligible under appropriate continuity assumptions. (In the present GI/G/1 context, the Z_j's are in fact *independent* of θ, but not the v_i's) With this intuitive reasoning, we should expect that SA with IPA converges to θ^* even with fixed t_n. Proposition 7, whose proof is in Appendix II, states that this is effectively true. Here, we cannot use Proposition 1 because we do not have $\beta_n \to 0$. Instead, we will give a *weak convergence* proof, based on Kushner and Shwartz [25] (see Theorem 4 of that paper).

Proposition 7. *Consider the SA algorithm with IPA, under Assumption 1, with $\{\gamma_n, n \geq 0\}$ satisfying* W4 *of Appendix I, and constant truncated horizon $t_n = t$. Let the interarrival time distribution have a bounded density. Suppose that the IPA accumulator is not reset to 0 between iterations. Then, θ_n converges in probability to the optimum θ^*.* ∎

With the regenerative IPA estimator (40), SA does not converge to the optimum in general if t_n does not converge to infinity. On the other hand, Chong and Ramadge [6] propose a somewhat different regenerative approach which converges to the optimum for constant t_n, in the case of an M/G/1 queue. The basic idea is to replace the factor $1/t$ in (34) by $1/E_\theta[\tau] = 1 - \lambda E_\theta[\zeta]$, for $t_n = 1$. The average service time $E_\theta[\zeta]$ is assumed to be known. When λ is unknown, for a given regenerative cycle, one can use an estimate $\hat{\lambda}$ obtained from the previous cycles. Alternatively, one can use at each iteration one or more cycle(s) to estimate λ and another cycle to compute the sum in (34). Further, in [7], they generalize their approach to the GI/G/1 queue.

IV. Numerical experiments with the M/M/1 queue

This example is inspired by Suri and Leung [43]. Consider an M/M/1 queue with arrival rate $\lambda = 1$ and mean service time $\theta \in \bar{\Theta}$. One has $B_\theta(\zeta) = 1 - e^{-\zeta/\theta}$, $B_\theta^{-1}(u) = -\theta \ln(1-u)$, $b_\theta(\zeta) = (1/\theta)e^{-\zeta/\theta}$, and $\frac{\partial}{\partial\theta} \ln b_\theta(\zeta) = (\zeta - \theta)/\theta^2$. Let $C(\theta) = C_1/\theta$ for some constant $C_1 > 0$. The optimal value θ^* can be computed easily in this case. Indeed, $w(\theta) = \theta/(1-\theta)$ and $\theta^* = \sqrt{C_1}/(1+\sqrt{C_1})$ (if this value is not in Θ, the optimum is at the nearest boundary). We will compare our empirical results to this theoretical value. Assumption 1 is easily verified. Indeed, (i) holds trivially, $b(\zeta) = e^{-\zeta}$ is continuously differentiable, and its Laplace transform $\int_0^\infty e^{s\zeta} e^{-\zeta} d\zeta$ is finite for $|s| < 1$. For (iii), take $\epsilon_0 = \theta_0(K-1)/(K+1)$. Then,

$$\begin{aligned}
\sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{b_\theta(\zeta)}{b_{\theta_0+\epsilon_0}(\zeta)} \right) &\le \frac{\theta_0+\epsilon_0}{\theta} \exp\left[-\zeta \left(\frac{1}{\theta} - \frac{1}{\theta_0+\epsilon_0} \right) \right] \\
&\le \frac{\theta_0+\epsilon_0}{\theta} \le \frac{\theta_0+\epsilon_0}{\theta_0-\epsilon_0} = K
\end{aligned}$$

and

$$\begin{aligned}
E_{\theta_0+\epsilon_0}\left[\sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{\partial}{\partial\theta} \ln b_\theta(\zeta) \right)^4 \right] &= E_{\theta_0+\epsilon_0}\left[\sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{\zeta-\theta}{\theta^2} \right)^4 \right] \\
&= \frac{E_{\theta_0+\epsilon_0}\left[(\theta_0+\epsilon_0-\zeta)^4 + (\theta_0-\epsilon_0-\zeta)^4 \right]}{(\theta_0-\epsilon_0)^8} < \infty,
\end{aligned}$$

since the exponential distribution has finite moments of all orders. Finally, $\int_0^\infty z^8 b(z) dz = \int_0^\infty z^8 e^{-z} dz < \infty$, which is (iv), and $C(x) = C_1/x$ obeys (v), which completes the verification of Assumption 1.

In these experiments, we have tried many SA-GET combinations, or variants. Henceforth, we refer to them as *algorithms*. The final state of each simulation subrun was taken as the initial state for the next one, except when stated otherwise. For FDC, the initial state s_{n+1} was the final state of the subrun at iteration n with parameter value the closest to θ_{n+1}. When the queue was not empty at the end of an iteration, we were careful to generate the new service time only at the beginning of the next iteration, i.e. *after* modifying the parameter. Kesten [22] has proposed a rule under which instead of diminishing γ_n at each iteration, one diminishes it only when the sign of the gradient estimate (for one parameter) is different from the one of the previous iteration (i.e. when the change on the parameter changes direction). The heuristic idea is that if the parameter keeps moving in the same direction, it should be because we are still far away from the optimum and so, let's give it a chance to move faster. That heuristic might help in situations where we start really far away from the optimum, and where the change on the parameter at each iteration tends to be very small. We have implemented this rule for some of our experiments. This is indicated in the results. Besides the GETs described in the previous section, we also implemented the regenerative algorithms described in [13] (with and without the arctan transformation), SAMOPT [3], and other variants, for which we do not give the details here.

A. *The experimental setup*

We actually performed the following experiment. For each algorithm, we made N simulation runs, each yielding an estimation of θ^*. The N initial parameter values were randomly chosen, uniformly in Θ, and the initial state was $s = 0$ (an empty system). Across the algorithms, we used common random numbers and the same set of initial parameter values. This means that the different entries of Table 1 are strongly correlated. Each run was stopped after a (fixed) total of $\bar{T}$ ends of service. Hence, all algorithms were subjected to approximately the same sequence of random numbers and, if we neglect the differences in overhead for the GETs, used about the same CPU time. (The overhead was quite low in general, except for very small values of t_n, like $t_n = 1$.) The programs were written using SIMOD [26], a simulation package based on the language Modula-2. We insisted on using exactly the same simulation program for all algorithms. In fact, the simulation model and the algorithms were implemented in two different modules, the latter being totally model independent.

For each algorithm variant, we computed the empirical mean $\bar{\theta}$, standard deviation s_d and standard error s_e of the N retained parameter values. If y_i denotes the retained parameter value for run i (i.e. the value of θ_n after the last iteration, for that run), the above quantities are defined by

$$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N} y_i; \qquad s_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{\theta})^2; \qquad s_e^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \theta^*)^2. \tag{41}$$

We also computed a confidence interval I_θ on the expectation of $\bar{\theta}$, assuming that $\sqrt{N}(\bar{\theta} - E(\bar{\theta}))/s_d$ follows a Student distribution with $N - 1$ degrees of freedom.

B. *Numerical results*

The third column of Table 1 gives some of the results of an experiment we made with $\bar{T} = 10^6$, $N = 10$, $\Theta = [0.01, 0.95]$ and $C_1 = 1$. The optimal solution is $\theta^* = 0.5$. We computed the 95% confidence intervals I_θ as described above, and the entries for which I_θ does not contain θ^* are indicated in table 1. FD, FDC, IPA and LR are as described in the previous section. LRR refers to the regenerative version of LR given in (12), while IPAR refers to the regenerative version of IPA. The symbol -K following the name of the algorithm signifies that Kesten's rule was used. The symbol -0 means that the state was reset to $s = 0$ at the beginning of each iteration. The symbol -Z following IPA means that the IPA accumulator was reset to 0 between iterations. The symbol -S following FD means that instead of always simulating first at $\theta_n - c_n$ and then at $\theta_n + c_n$, we adopted the following heuristic rule: if the parameter has just decreased, simulate first on the right (at $\theta_n + c_n$), otherwise simulate first on the left. The rationale is that if the parameter has just decreased, the current state has been reached by simulating at a parameter value larger than θ_n, and should thus be a statistically "better" initial state for a simulation at $\theta_n + c_n$ than at $\theta_n - c_n$ (and symmetrically if the parameter has just increased). LR-D means the "triangular" version of LR given by (24). LR-C [LR-DC] means LR [LR-D] in which $h_t(\theta, s, \omega)$ was replaced by $h_t(\theta, s, \omega) - 1$. This does not change the expectation of $\psi_t(\theta, s, \omega)$, but reduces its variance from $O(t)$ to $O(1)$ at $\theta = \theta^*$, because $w(\theta^*) = 1$ (see [30]). In all cases, we had $\gamma_n = 1/n$. We took $c_n = 0.1n^{-1/6}$ for FD and $c_n = 0.1n^{-1/5}$ for FDC. For FDC, we also tried $c_n = 0.001n^{-2}$, which is denoted by FDC-NN.

We see that IPA performs well, even when t_n is fixed at a small constant. IPA-Z, IPAR, FDC, FDC-K and FDC-NN, with a linearly increasing t_n, are approximately as good. When t_n is fixed to a small constant, convergence is also quick with FDC, IPA-Z, or IPAR (small s_d), but the standard error s_e is very large, which indicates that convergence is not to the right value. Even for $t_n = 100$, the bias is still quite apparent for FDC. The problem with IPAR is that with the regenerative

		$C_1 = 1$ ($\theta^* = 1/2$)		$C_1 = 1/25$ ($\theta^* = 1/6$)		$C_1 = 25$ ($\theta^* = 5/6$)	
	T_n	s_d	s_e	s_d	s_e	s_d	s_e
FD	n	.00979	.00967				
FD	$100+n$	.01075	.01044				
FD-S	n	.00761	.00732				
FDC	5	.00149	.15343 ◁				
FDC	100	.00340	.00721 ◁				
FDC	n	.00193	.00184	.00030	.00030	.02354	.02234
FDC	$100+n$	.00204	.00198	.00027	.00029	.02875	.02824
FDC-K	n	.00193	.00184	.00030	.00030	.02354	.02234
FDC-0	n	.00243	.00231	.00039	.00037	.03019	.02867
FDC-NN	n	.00203	.00196	.00031	.00031	.02270	.02177
FDC	$n^{1/2}$	.00181	.00684 ◁				
IPA	1	.00227	.00217				
IPA	10	.00227	.00216	.00053	.00051	.02402	.02575
IPA	100	.00229	.00219				
IPA	n	.00195	.00185	.00046	.00044	.03208	.03416
IPA	$100+n$	.00203	.00193	.00046	.00043	.02685	.02849
IPA-K	n	.00195	.00185	.00046	.00044	.03208	.03416
IPA-Z	10	.00169	.07365 ◁				
IPA-Z	n	.00192	.00189	.00046	.00044	.02449	.02597
IPA-0	n	.00246	.00233	.00042	.00040	.01721	.01956
IPAR	5	.00228	.06175 ◁				
IPAR	n	.00200	.00197	.00046	.00044	.02981	.03110
LR	$n^{1/3}$	.01221	.02062 ◁				
LR	$n^{1/2}$	.03012	.02876	.02454	.02355	.04473	.05214
LR	$n^{2/3}$	.07494	.07115				
LR-C	$n^{1/2}$	.00772	.00749	.00221	.00291 ◁	.03433	.04864 ◁
LR-C0	$n^{1/2}$	.00709	.00725				
LR-D	$n^{1/2}$	.01502	.01658				
LR-CD	$n^{1/2}$	.00533	.00615	.00175	.00176	.03000	.05141 ◁
LR-CD	$n^{2/3}$	.00706	.00688	.00264	.00255	.04893	.04857
LRR	n	.00447	.00453	.00124	.00118	.07608	.07446
LRR	$n^{1/2}$	.00443	.01775 ◁				

Table 1: Some experimental results for $\bar{T} = 10^6$, $N = 10$ and $C_1 = 1, 1/25$, and 25. For the values marked with ◁, the 95% confidence interval does not contain θ^*.

approach, the number of ends of service during the t_n regenerative cycles is now random, and we get a bias due to the fact that we estimate a ratio with that number on the denominator. Of course, this bias goes to zero as t_n goes to infinity, and this is why IPAR with $t_n = n$ works fine. FD has approximately the same behavior as FDC, but with larger variance. FD-S is slightly better than FD, but not competitive with FDC or IPA.

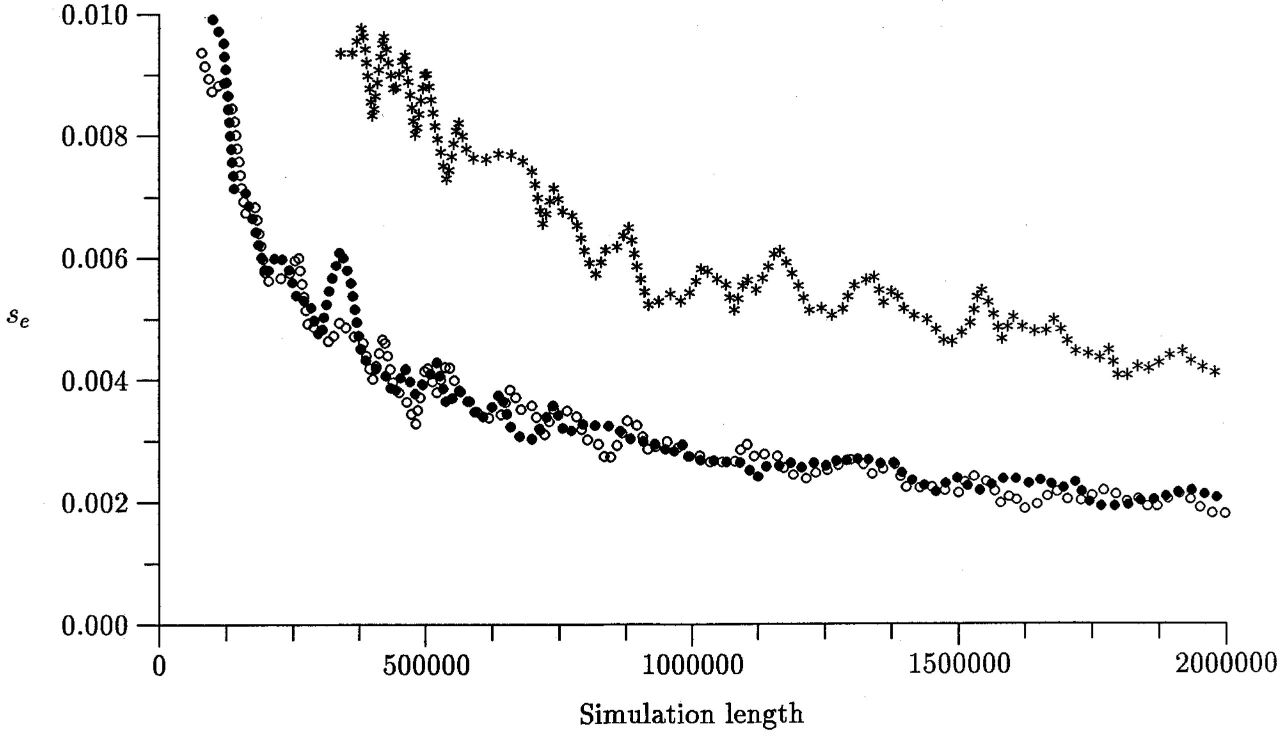

Figure 1: Evolution of s_e (N=100) for FDC (black), PA (white) and LRR (star).

The LR methods in general have trouble due to their large associated variance. That variance stays low when t_n grows slowly, but then, the bias becomes more of a problem. LR with $t_n = n^p$ has large variance for large p, and for small p, the bias goes down much too slowly compared to the variance. As a result, the confidence interval I_θ, based on the N final values of θ_n, is very likely not to cover θ^*. This is what happens, for instance, with $p = 1/3$. Among the truncated-horizon variants, LR-C and LR-CD provide significant improvements over LR. The LR variant that gives the best results here is LRR (regenerative) with t_n increasing linearly. With $t_n = n^{1/2}$, both LRR and FDC have the same bias problem as described above: the bias goes down too slowly and I_θ does not contain θ^*. Nevertheless, they converge (slowly) to the right answer (we verified it empirically with longer simulation runs). Kesten's rule does not appear to help for any of the methods in this example. SAMOPT [3] and the algorithms described in [13] gave rather bad results (huge variances) and they do not appear in the figure. They are obviously not competitive, at least for this example. The problem with SAMOPT is that near the optimum, the gradient is very small in absolute value, and replacing it by its sign is really not a good idea. We also obtained bad results with other variants, like for instance IPA with $t_n = 100 + n$ but $\gamma_n = n^{-1/2}$ instead of n^{-1}. Independent sets of experiments were also performed with $\bar{T} = 10^5$ and the results were quite similar to the ones given here [9].

We have also observed that for IPA, the evolution of the parameter θ_n with the total simulation length depends very little on t_n. We made experiments with different (constant) values of t_n between 1 and 1000, and observed that the evolution of θ was practically identical in all cases, independently of t_n. These simulations were performed with common random numbers. Even if the starting values are different, the evolution is almost identical if common random numbers are used.

In Figure 1, we see the evolution of the standard error s_e with the simulation length, for FDC, IPA, and LRR, all with $t_n = n$. This is the result of a new experiment, with longer and much more numerous runs: we took $\bar{T} = 2 \times 10^6$ and $N = 100$. We see that FDC and IPA appear roughly comparable, and that LRR has about twice their standard error (four times more variance). These graphs agree with the expected convergence rates of $O(\bar{T}^{-1/2})$: the standard error gets approximately cut in half when the total simulation length $\bar{T}$ is multiplied by four.

How does the speed of convergence of θ_n to θ^* compares to the speed of convergence of the cost estimator to the true average cost when θ is fixed ? We note that simply comparing the widths of the confidence intervals at the end doesn't make sense, since the parameter and the cost are not necessarily measured on the same scale. Dividing by the means to obtain relative values doesn't make sense either, there might be cases where θ^* or the average cost is zero or near zero. In any case, it is well known that the average cost estimator converges at rate $\bar{T}^{-1/2}$, and we have observed the same convergence rate for θ_n. This means that we can estimate the optimum as fast (in terms of orders of convergence rates) as we can estimate the cost at a given point! This was already observed in [44].

We made other sets of experiments with $C_1 = 1/25$ (for which $\theta^* = 1/6$) and $C_1 = 25$ (for which $\theta^* = 5/6$). The results appear in the last two columns of Table 1. For $C_1 = 1/25$, the traffic intensity for θ near θ^* is low, and we get a much lower variance than for $C_1 = 1$. The opposite is true for $C_1 = 25$. The relative "rankings" of the algorithms are about the same. For $C_1 = 1/25$, FDC appears better than IPA. But note that these standard error estimates are based on only 10 observations, which means that they themselves have non-negligible variance. The entries of Table 1 are also correlated, because of the common random numbers. For $C_1 = 25$, LR-CD, and $t_n = n^{1/2}$, the variance for θ_n goes down quickly and the bias does not go to zero fast enough to cope with that. The result is that for this experiment, the confidence interval I_θ does not contain θ^*. A possible remedy is to increase t_n faster: for $t_n = n^{2/3}$, the problem disappears. But in any case, this shows that one must be *very careful* about confidence intervals in these kinds of experiments, even if they are asymptotically valid. For $C_1 = 25$, LR is now competitive with LRR (θ is larger and the regenerative cycles are much longer in this case).

V. GENERALIZATIONS

A. *More general service time parameters*

For simplicity, the development of the previous sections and of Appendix II was made under the assumption that θ is a scale parameter of the service time distribution. This assumption can of course be relaxed. We now give more general sufficient conditions under which our results hold. Let B_θ be the service time distribution, with density b_θ, and suppose that the service time of customer i is generated by inversion, i.e., that $\zeta_i = B_\theta^{-1}(U_i)$, where U_i is a $U(0,1)$ random variate. Assume that $B_\theta^{-1}(u)$ is differentiable in θ, for each $0 < u < 1$, and let $Z_i = \frac{\partial}{\partial\theta} B_\theta^{-1}(U_i)$.

Assumption 2.

(i) *The set* $\{\zeta \geq 0 \mid b_\theta(\zeta) > 0\}$, *which is the support of* b_θ, *is independent of* θ.

(ii) *Everywhere in* $\bar{\Theta}$, $b_\theta(\zeta)$ *is continuously differentiable with respect to* θ, *for each* $\zeta \geq 0$.

(iii) *There is a distribution* $\tilde{B}$ *such that* $\sup_{\theta\in\bar{\Theta}} B_\theta^{-1}(u) \leq \tilde{B}^{-1}(u)$ *for each* u. *The queue remains stable when the service times are generated according to* $\tilde{B}$. *Also,* $\int_0^1 (\tilde{B}^{-1}(u))^8 du < \infty$. *In other words,* $\tilde{E}[\zeta^8] < \infty$, *where* $\tilde{E}$ *is the expectation that corresponds to* $\tilde{B}$.

(iv) *There exists a measurable function* $\Gamma : (0,1) \mapsto \mathbb{R}$ *such that* $\int_0^1 (\Gamma(u))^4 du < \infty$ *and* $\sup_{\theta\in\bar{\Theta}} |\frac{\partial}{\partial\theta} B_\theta^{-1}(u)| \leq \Gamma(u)$ *for each* u.

(v) *For each* $\theta_0 \in \bar{\Theta}$ *and* $K > 1$, *there is an open interval* Υ *containing* θ_0 *such that* b_θ *is well defined for* $\theta \in \Upsilon$, *and a* $\bar{\theta} \in \bar{\Theta}$ *such that*

$$\sup_{\theta\in\Upsilon} \left(\frac{b_\theta(\zeta)}{b_{\bar{\theta}}(\zeta)} \right) \leq K, \tag{42}$$

and

$$E_{\bar{\theta}} \left[\zeta^8 + \left(\sup_{\theta\in\Upsilon} \frac{\partial}{\partial\theta} \ln b_\theta(\zeta) \right)^4 \right] < \infty. \tag{43}$$

Also, the moment generating function associated with $\tilde{B}$ *is finite in some neighborhood of zero.*

(vi) $Z_i \geq 0$ *for all* i *with probability one.*

(vii) *For each* $0 < u < 1$, $B_\theta^{-1}(u)$ *is non-decreasing and convex in* θ. ∎

Other variants of these conditions also work. For example, in (vii), non-decreasing can be replaced by non-increasing, provided that the same substitution is also made in the statement and proof of Proposition 15. In (vi), ≥ 0 can be replaced by ≤ 0. In the latter case, in Proposition 14, $w_t'(\theta, 0)$ becomes non-increasing (instead of non-decreasing) in t. Condition (vi) can also be replaced by the more general one: $w_t'(\theta,0)/t$ is non-decreasing in t for each θ, or non-increasing in t for each θ. A generalization of condition (v) is (v') For each $\theta_0 \in \bar{\Theta}$ and $K > 1$, there is an open interval Υ containing θ_0 such that b_θ is well defined for $\theta \in \Upsilon$, and a density q whose support contains the support of b_θ, such that

$$E_q\left[\zeta^8 + \sup_{\theta\in\Upsilon}\left(\frac{\frac{\partial}{\partial\theta}b_\theta(\zeta)}{q(\zeta)}\right)^4 + \sup_{\theta\in\Upsilon}\left(\frac{b_\theta(\zeta)}{q(\zeta)}\right)^2 + \prod_{i=1}^{\tau}\sup_{\theta\in\Upsilon}\left(\frac{b_\theta(\zeta)}{q(\zeta)}\right)^8\right] < \infty. \tag{44}$$

where τ is the number of customers in a regenerative cycle.

Proposition 8. *The results of Propositions 2–3, 5–7, and 10–21 still hold with Asumption 1 replaced by Assumption 2.* ∎

Proposition 8 can be verified by going through all the proofs with the new assumptions. This is discussed in Appendix II. Finally, the following result, proved in Appendix II, shows that the uniform convergence (8) holds under much weaker conditions than those of Assumption 1 or 2. This result could be of independent interest.

Proposition 9. *Suppose that there is a distribution $\tilde{B}$ such that $\sup_{\theta\in\Theta} B_\theta^{-1}(u) \le \tilde{B}^{-1}(u)$ for each u, that $\int_0^1(\tilde{B}^{-1}(u))du < \infty$, and that the queue remains stable when the service times are generated according to $\tilde{B}$. Then, (8) holds.* ∎

B. General Markov chain models

The convergence results of Section III can be extended to more general models than the GI/G/1 queue. Consider for example a general discrete-time Markov chain model parametrized by θ. Let $\alpha_t(\theta, s) = C(\theta) + w_t(\theta, s)/t$ be the expected average cost per step for the first t steps, if the initial state is s. Suppose that (8–9) hold (which implies that the derivative exists), that $\alpha(\theta) = C(\theta) + w(\theta)$ is strictly unimodal, and that an unbiased LR or IPA derivative estimator for $w_t'(\theta, s)$ is available. If the variance of the LR estimator is in $O(t)$, then Proposition 5 (a) applies, while if the variance of the IPA estimator is in $O(1/t)$, then Proposition 6 (a) applies. Further, if the system is regenerative, and if unbiased LR estimators are available for the derivative of the expected (regenerative) cycle length and the derivative of the expected cost per cycle, then one can construct an estimator for $w'(\theta)$ as in (30). If that estimator has bounded variance and converges in expectation to $w'(\theta)$ uniformly in θ, as $r \to \infty$, then Proposition 5 (b) applies. If a FD or FDC estimator is used and if $w(\cdot)$ and $w_t(\cdot, s)$ are continuously differentiable (for each s), then Proposition 3 applies.

One can also consider continuous-time models, in which costs are incurred continuously, models with discounting, etc. All this generalizes to the setup of Appendix I, where θ is a vector of parameters. Derivatives are then replaced by gradients.

VI. CONCLUSION

Through a simple example, we have seen how a derivative estimation technique, such as FD, IPA, or LR, can be incorporated into a stochastic approximation algorithm to get a provably convergent stochastic optimization method. We also pointed out some dangers associated with different kinds of bias. For the example considered, IPA gave the best results, but this may not be true in general.

The performance of these algorithms when there are many parameters to optimize, the incorporation of proper variance reduction techniques, and the study of convergence rates, are other interesting subjects for further investigation. In principle, IPA and LR can be used to estimate higher order derivatives, but the variance is typically quite high. Is it too high to permit the implementation of good second order algorithms based on these estimates ? Again, further investigation is needed.

Appendix I. Sufficient convergence conditions

In this appendix, we give sufficient conditions for the convergence of the SA algorithm (6) to an optimum. The first set of conditions implies almost sure convergence, whereas the second set implies weak convergence. These conditions are adaptated from [23, 25].

In what follows, we assume that Θ is a compact and convex set of the form $\Theta = \{\theta \in \mathbb{R}^d \mid q(\theta) \le 0\}$, equal to the closure of its interior, where q is a κ_q-dimensional vector of continuously differentiable functions (constraints), as in [32, chap.10], and at any point on the boundary of Θ, the gradients of the active constraints are linearly independent. We want to minimize the function $\alpha : \Theta \to \mathbb{R}$, over Θ. We assume that there is a companion stochastic process $\{s_n,\ n \ge 1\}$, defined over some Borel space S, and a family of probability measures $\{P_{\theta,s},\ (\theta, s) \in \Theta \times S\}$, such that $P[(Y_{n+1}, s_{n+1}) \in \cdot \mid (\theta_n, s_n) = (\theta, s)] = P_{\theta,s}[\cdot]$ for all $n \ge 1$. Let E_{n-1}, β_n, and ϵ_n be defined as in Section 2.

Define $z_n = \sum_{i=1}^{n} \gamma_i$ and $m(z) = \max\{n \mid z_n \le z\}$ for $z \ge 0$. Let $x^0 : [0, \infty) \mapsto \mathbb{R}^d$ be the piecewise linear interpolation of the set of points $\{(z_n, \theta_n), n \ge 1\}$, and x^n the left shift of x^0 defined by $x^n(z) = x^0(z + z_n)$, for $z \ge 0$. Hence, $x^n(0) = x_0(z_n) = \theta_n$, and if x^n converges to a limit x, the asymptotic properties of $x(z)$ as $z \to \infty$ can provide information on the asymptotic behavior of θ_n as $n \to \infty$. For any function $v : \Theta \to \mathbb{R}^d$, define (when the limit exists):

$$\bar{\pi}(v(\theta)) = \lim_{\delta \to 0+} \left(\frac{\pi_\Theta(\theta + \delta \cdot v(\theta)) - \theta}{\delta} \right). \tag{45}$$

Consider the differential equation

$$x'(z) = \bar{\pi}(-v(x(z))). \tag{46}$$

The theorems below give conditions under which as $n \to \infty$, x^n converges in some sense to a solution of (46) for a proper v. This convergence property permits one to analyze the behavior of $\{\theta_n, n \ge 1\}$. We give a list of assumptions that will be used selectively in the next two theorems.

S1. For all $n \ge 1$, $\gamma_n \ge \gamma_{n+1} > 0$, and $\sum_{n=1}^{\infty} \gamma_n = \infty$.

S2. $\lim_{n\to\infty} \beta_n = 0$ almost surely.

S3. For each $T > 0$ and $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\sup_{j \ge n,\ z \le T} \left\| \sum_{i=m(jT)}^{m(jT+z)-1} \gamma_i \epsilon_i \right\| \ge \epsilon \right) = 0. \tag{47}$$

S4. There is a positive sequence $\{\delta_n,\ n \ge 1\}$ such that $E_0[\epsilon_n' \epsilon_n] \le 1/\delta_n^2$ and $\sum_{n=1}^{\infty} (\gamma_n/\delta_n)^2 < \infty$.

S5. There is a $\theta^* \in \Theta$, an asymptotically stable point of $x'(z) = \bar{\pi}(-\alpha'(x(z)))$, with domain of attraction $D_A(\theta^*)$ (in the sense of Liapounov), and almost surely, infinitely many θ_n belong to $D_A(\theta^*)$.

Theorem 1. *(Kushner and Clark). Assume S1 to S3. Then, almost surely, x^0 is uniformly continuous on $[0, \infty)$ and any limit x of a convergent subsequence of $\{x^n, n \ge 1\}$ satisfies (46) with $v = \alpha'$. If θ^* also satisfies S5, then $\lim_{n\to\infty} \theta_n = \theta^*$ almost surely.* ∎

Theorem 1 is proved in [23, Theorem 5.3.1]. Condition S3 is quite general, but has low intuitive appeal. The theorem below uses a more restrictive but more "familiar" condition. It is a variant of Proposition F in section II of [34].

Theorem 2. *Under conditions S1, S2, S4 and S5, SA converges almost surely to θ^*.*

PROOF. It suffices to show that S3 holds, and the result will follow from part (b) of the previous Theorem. Note that under S4, for each n, the sequence $\{\sum_{i=n}^{j} \gamma_i \epsilon_i,\ j \geq 1\}$ is a martingale. For each $\epsilon > 0$, from Doob's inequality and from S4, we have

$$P\left(\sup_{j\geq n}\left\|\sum_{i=n}^{j}\gamma_i\epsilon_i\right\| \geq \epsilon\right) \leq \frac{K}{\epsilon^2}\sum_{i=n}^{\infty}\gamma_i^2 E_0[\epsilon_i'\epsilon_i] \leq \frac{K}{\epsilon^2}\sum_{i=n}^{\infty}(\gamma_i/\delta_i)^2. \tag{48}$$

for some constant K. This upper bound goes to zero as $n \to \infty$. Hence, we obtain condition $A2.2.4''$ of [23], which implies S4. ∎

Often, S2 is not satisfied. But if $E_0[\beta_n] \to 0$ as $n \to \infty$, the algorithm should converge as well to the optimum. This is addressed by the following (weaker) results. Theorem 3 follows easily from the results of [25], while Theorem 4 is an adaptation of the second part of Theorem 4.2.1 in [23], and can be proved in the same way (note that in the last paragraph of the proof of Theorem 4.2.1 in [23], the max should be replaced by a min). We give a new list of assumptions.

W1. Denote $\xi_n = (Y_n, s_{n+1}) \in \mathbb{R} \times S$. Assume that $P_{\theta,s}$ is weakly continuous in (θ, s) and that $E[Y_{n+c} \mid (\theta_n, s_n) = (\theta, s)]$ is continuous in (θ, s), for some integer $c \geq 0$. Assume that for each fixed $\theta \in \Theta$, i.e. if $\gamma_n = 0$ for all n, $\{\xi_n, n \geq 1\}$ is a Markov process with unique invariant measure P^θ and corresponding mathematical expectation E^θ. Denote $v(\theta) = E^\theta(Y_n)$. Let $\{P^\theta, \theta \in \Theta\}$ and $\{\xi_n, n \geq 1\}$ be tight (the latter uniformly over θ and s; see [25]).

W2. For each compact $C \subset \mathbb{R} \times S$, there is an integer $n_C < \infty$ such that for each $T > 0$, the set of probability measures $\{P[(\theta_{n+j}, \xi_{n+j-1}) \in \cdot \mid \theta_n = \theta, \xi_n = \xi],\ \theta \in \Theta,\ \xi \in C,\ n \geq 1,\ j \geq n_C,\ \sum_{i=n+1}^{n+j} \gamma_i \leq T,\ C \text{ compact subset of } S\}$ is tight.

W3. For some constant $\kappa > 0$, $\sup_{n\geq 1} E_0[|Y_n|^{1+\kappa}] < \infty$.

W4. $\gamma_n > 0$ for all n, $\lim_{n\to\infty} \gamma_n = 0$, $\sum_{n=1}^{\infty} \gamma_n = \infty$ and $\sum_{n=1}^{\infty} |\gamma_{n+1} - \gamma_n| < \infty$.

W5. There is a $\theta^* \in \Theta$, an asymptotically stable point of $x'(z) = \bar{\pi}(-v(x(z)))$, with domain of attraction $D_A(\theta^*) = \Theta$.

Theorem 3. *Under Assumptions W1—W4, $\{x^n, n \geq 1\}$ is tight and any weak limit of one of its subsequences satisfies the projected differential equation (46) almost everywhere with probability one. Also, $v(\theta)$ is continuous in θ.*

PROOF. This follows from Theorems 1 and 4 and the remarks that follow Theorem 4 in [25]. ∎

Theorem 4. *Under Assumptions W1—W5, θ_n converges to θ^* in probability, i.e. for each $\epsilon > 0$,* $\lim_{n\to\infty} P(\|\theta_n - \theta^*\| \geq \epsilon) = 0$. ∎

Appendix II. Some convergence proofs

In this Appendix, we prove that under Assumption 1, LR and IPA provide unbiased estimators for $w_t'(\theta, s)$, the derivative of the expected total sojourn time of the first t customers. We obtain variance bounds for these derivative estimators and for their regenerative counterparts, which are asymptotically unbiased and converge in quadratic mean, uniformly in θ. Recall that uniform convergence in quadratic mean implies uniform convergence in expectaton. We also show that $w_t(\cdot, s)$ and $w(\cdot)$ are continuously differentiable, and that (8–9) hold. We then prove Propositions 3, and 5 to 9. Along the road, we obtain a few additional characterizations of w_t, w_t', w, and w'.

Proposition 10. *Under Assumption 1, one has*

$$\sup_{k\leq 8,\, \theta\in\bar{\Theta}} E_\theta[\zeta^k + \tau^k] < \infty.$$

PROOF. From Assumption 1 (i) and (iv), $E_\theta[\zeta^8] = \int_0^\infty (\theta z)^8 b(z) dz \leq (u_0)^8 \int_0^\infty z^8 b(z) dz < \infty$. In particular, $E_{u_0}[\zeta^8] < \infty$. Then, from Theorem II.3.1 (i) in [18], one has $E_{u_0}[\tau^8] < \infty$. This implies that $E_{u_0}[\zeta^k + \tau^k] < \infty$ for each $k \leq 8$. Now, we will argue that ζ and τ are stochastically non-decreasing in θ. From that and from basic stochastic ordering principles [45], the result will follow automatically. Let ω represent the underlying sequence of $U(0,1)$ variates. For ω fixed, each ζ_i is (linearly) increasing in θ. But increasing any service time cannot increase the number of customers in the first busy cycle. Therefore, ζ and τ are stochastically non-decreasing in θ. ∎

Proposition 11. *Consider the truncated horizon LR derivative estimator (22), under Assumption 1. For each $\theta_0 \in \bar{\Theta}$, there is a neighborhood Υ of θ_0 such that for all $(\theta, s) \in \Upsilon \times S$, (22) is an unbiased estimator of $w_t'(\theta, s)$. Further, for (θ, s) in $\bar{\Theta} \times S$, $w_t(\cdot, s)$ is differentiable and $w_t'(\theta, s)$ is continuous (jointly) in (θ, s).*

PROOF. For fixed ω, $h_t(\theta, s, \omega)$ does not depend on θ. Since each $b_\theta(\zeta_i) = b(\zeta_i/\theta)/\theta$ is assumed differentiable in θ, Assumption A2 (a) in [29] is satisfied. Observe that for each i, $W_i \leq s + \sum_{j=1}^{i-1} \zeta_j$, so that $h_t(\theta, s, \omega) = \sum_{i=1}^t (W_i + \zeta_i) \leq t\left(s + \sum_{i=1}^t \zeta_i\right)$. Then,

$$(h_t(\theta, s, \omega))^2 \leq t^2 \left(s + \sum_{i=1}^t \zeta_i\right)^2 \leq 2t^2 \left(s^2 + \left(\sum_{i=1}^t \zeta_i\right)^2\right) \leq 2t^2 s^2 + 2t^3 \sum_{i=1}^t \zeta_i^2.$$

Let $K > 1$ and $\epsilon_0 > 0$ satisfy Assumption 1 (iii). Let $\bar{\theta} = \theta_0 + \epsilon_0$. Then,

$$\begin{aligned}
& E_{\bar{\theta}}\left[\sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{\frac{\partial}{\partial\theta} b_\theta(\zeta)}{b_{\bar{\theta}}(\zeta)}\right)^2 + \sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{b_\theta(\zeta)}{b_{\bar{\theta}}(\zeta)}\right)^2 + (h_t(\theta, s, \omega))^2\right] \\
\leq\ & E_{\bar{\theta}}\left[\sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{b_\theta(\zeta)}{b_{\bar{\theta}}(\zeta)} \frac{\frac{\partial}{\partial\theta} b_\theta(\zeta)}{b_\theta(\zeta)}\right)^2\right] + K^2 + 2t^2 s^2 + 2t^3 \sum_{i=1}^t E_{\bar{\theta}}(\zeta_i^2) \\
\leq\ & E_{\bar{\theta}}\left[K^2 \sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{\partial}{\partial\theta} \ln b_\theta(\zeta)\right)^2\right] + K^2 + 2t^2 s^2 + 2t^4 E_{\bar{\theta}}(\zeta^2) \\
<\ & \infty.
\end{aligned}$$

This implies A2 (b) in [29] with $q = b_{\bar\theta}$. Therefore, from Proposition 2 of [29], the conclusion follows, except for the joint continuity of $w_t'(\theta, s)$. For fixed ω, it is easily seen, by induction on i, that each W_i^* is continuous in s. Indeed, this is clearly true for $i = 1$, since $W_1^* = s + \zeta_1$. If W_i^* is continuous in s, then $W_{i+1}^* = (W_i^* - \nu_i)^+ + \zeta_i$ is also continuous in s, since ν_i does not depend on s. This implies that for fixed ω, $h_t(\theta, s, \omega)$ is continuous in s (and constant in θ). Also, from Assumption 1 (i) and (ii), $S_t(\theta, s, \omega)$ and $L_t(P_{\theta_0,s}, \theta, s, \omega)$ are continuous in θ and constant in s. Therefore, $\psi_t(\theta, s, \omega)$ defined in (22) is continuous in (θ, s). (Recall that the product of continuous functions is continuous; see Theorem 4.10 (b) of [40]). Let $s_0 \geq 0$ and assume that (θ, s) remains in $(\theta_0 - \epsilon_0, \theta_0 + \epsilon_0) \times S$ when taking the limit. The first part of the proof and Proposition 2 of [29] also implies that $\psi_t(\theta, s, \omega) \leq \Gamma(\omega)$ in some neighborhood of (θ_0, s_0), where $\Gamma(\omega)$ is integrable with respect to $P_{\bar\theta}$. As a consequence, we can apply Lebesgues's dominated convergence theorem,

$$\begin{aligned}
\lim_{(\theta,s)\to(\theta_0,s_0)} w_t'(\theta, s) &= \lim_{(\theta,s)\to(\theta_0,s_0)} E_{\bar\theta}[\psi_t(\theta, s, \omega)] \\
&= E_{\bar\theta}\left[\lim_{(\theta,s)\to(\theta_0,s_0)} \psi_t(\theta, s, \omega)\right] \\
&= E_{\bar\theta}[\psi_t(\theta_0, s_0, \omega)] = w_t'(\theta_0, s_0).
\end{aligned}$$

If $s_0 = 0$, the convergence of s to s_0 should be interpreted as convergence from the right. This can be done for each $(\theta_0, s_0) \in \bar\Theta \times S$, which completes the proof. ∎

Proposition 12. *Under Assumption 1, $u(\theta)$, $\ell(\theta)$, and $w(\theta)$ are finite and continuously differentiable in θ, for $\theta \in \bar\Theta$. Also, $\psi_u(\theta, \omega)$ and $\psi_\ell(\theta, \omega)$, defined in (28) and (29), are unbiased estimators of $u'(\theta)$ and $\ell'(\theta)$, respectively, for θ in a small enough neighborhood of θ_0.*

PROOF. We first prove the second part of the proposition and for that, we will use Proposition 3 of [29]. For fixed ω, τ and $\sum_{i=1}^{\tau} W_i^*$ do not depend on θ. Therefore, τ, $\sum_{i=1}^{\tau} W_i^*$, and each $b_\theta(\zeta_i)$ are differentiable in θ everywhere in $\bar\Theta$. This implies Assumption A2 (a) in [29] with t replaced by τ.

From Assumption 1 (ii), there is an $\tilde s > 0$ such that for all s that satisfy $su_0 \leq \tilde s$, $E_{u_0}\left[e^{s\zeta}\right] = E_1\left[e^{su_0\zeta}\right] < \infty$. Then, from Theorem III.3.2 in [18], page 81, there is an $\epsilon_1 > 0$ such that $E_{u_0}[e^{\epsilon_1 \tau}] < \infty$. In words, the service time distribution and the number of customers in a busy period have finite moment generating functions in a neighborhood of zero. Let $0 < K \leq e^{\epsilon_1/8}$, ϵ_0 satisfy Assumption 1 (iii), and $\bar\theta = \theta_0 + \epsilon_0$. One has

$$\prod_{i=1}^{\tau} \sup_{|\theta-\theta_0|<\epsilon_0} \left(\frac{b_\theta(\zeta)}{b_{\bar\theta}(\zeta)}\right)^8 \leq K^{8\tau} \leq e^{\epsilon_1 \tau}$$

and, since τ is stochastically non-decreasing in θ,

$$E_{\bar\theta}\left[K^{8\tau}\right] \leq E_{\bar\theta}[e^{\epsilon_1\tau}] \leq E_{u_0}[e^{\epsilon_1\tau}] < \infty.$$

From Proposition 10 (used in the last inequality) and Theorem I.5.2 in [18, p,22] (used in the next to last inequality), there is a constant $K_1 < \infty$ such that

$$E_{\bar\theta}\left[(h_\tau(\bar\theta, 0, \omega))^4\right] \leq E_{\bar\theta}\left[\left(\sum_{i=1}^{\tau} W_i^*\right)^4\right] \leq E_{\bar\theta}\left[\left(\tau \sum_{i=1}^{\tau} \zeta_i\right)^4\right]$$

$$\leq E_{\bar\theta}[\tau^8] + E_{\bar\theta}\left[\left(\sum_{i=1}^{\tau}\zeta_i\right)^8\right]$$
$$\leq E_{\bar\theta}[\tau^8] + K_1(E_{\bar\theta}[\tau^8]E_{\bar\theta}[\zeta^8])^{1/2} < \infty.$$

From Wald's equation and Assumption 1 (iii),

$$E_{\bar\theta}\left[\sum_{j=1}^{\tau}\sup_{|\theta-\theta_0|<\epsilon_0}\left(\frac{\frac{\partial}{\partial\theta}b_\theta(\zeta_j)}{b_{\bar\theta}(\zeta_j)}\right)^4\right] = E_{\bar\theta}[\tau]E_{\bar\theta}\left[\sup_{|\theta-\theta_0|<\epsilon_0}\left(\frac{\frac{\partial}{\partial\theta}b_\theta(\zeta)}{b_{\bar\theta}(\zeta)}\right)^4\right]$$
$$\leq E_{\bar\theta}[\tau]E_{\bar\theta}\left[K^4\sup_{|\theta-\theta_0|<\epsilon_0}\left(\frac{\partial}{\partial\theta}\ln b_\theta(\zeta)\right)^4\right] < \infty.$$

Then, all the requirements of Assumption A3 in [29] are satisfied, with $h(\theta,\omega)$ there replaced by either τ or $\sum_{i=1}^{\tau}W_i^*$ (which here do not depend on θ for ω fixed), and $\Gamma_{1i}(\zeta) = K^8$. This holds in a neighborhood of θ_0 for each $\theta_0 \in \bar\Theta$. This implies the result, except for the continuous differentiability of w. But the latter follows from (25) and the continuous differentiability of u and ℓ. ∎

Proposition 13. *For each $(\theta,s) \in \bar\Theta \times \bar S$, the IPA estimator (34) is an unbiased estimator of $w_t'(\theta,s)$.*

PROOF. For fixed ω (viewed as equivalent to a sequence of independent $U(0,1)$ variates), $h_t(\theta,s,\omega)$ is continuous in each ζ_j, and therefore continuous in θ from Assumption 1 (i). It is also differentiable in θ everywhere except when two events (arrival or departure) occur simultaneously, which happens at most for a finite number of values of θ. One has

$$|h_t'(\theta,s,\omega)| \leq \sum_{i=1}^{t}\sum_{j=1}^{t}\left|\frac{\partial B_\theta^{-1}(U_j)}{\partial\theta}\right| \leq t\sum_{j=1}^{t}Z_j.$$

For fixed t, from Proposition 10, $E_\theta\left[t\sum_{j=1}^{t}Z_j\right] = t\sum_{j=1}^{t}E_1[\zeta_j] < \infty$. Then, from Theorem 1 in [27], $h_t'(\theta,s,\omega)$ is an unbiased estimator of $w_t'(\theta,s)$. ∎

Proposition 14. *Under Assumption 1, $w_t(\theta,0)/t$ and $w_t'(\theta,0)/t$ are non-decreasing in t.*

PROOF. We use stochastic order arguments [45, Section 11.4]. For $s = 0$, and fixed ω (representing the underlying sequence of uniform variates), consider the cost estimator (3) and the IPA derivative estimator (34). For $i \geq 1$, from Proposition 13, W_i^* and $\sum_{j=v_i}^{i}Z_j$ are unbiased estimators of the expected sojourn time of customer i and of its derivative, respectively. To obtain estimators of the corresponding quantities for customer $i+1$, the straightforward way is simply to replace i by $i+1$ in the above expressions. But what we will do, rather, is to augment the sample path by adding a customer at the beginning. This will permit us to exploit stochastic ordering. Add a new customer before customer 1 and call it customer 0. Generate a new service time $\zeta_0 = \theta Z_0$ and a new interarrival time ν_0, according to the distributions A and B_θ, respectively. For $i \geq 1$, let $v_i = 0$ if and only if customer i (which is now the $(i+1)$-th customer) is in the same busy period as customer 0 in this new (augmented) sample path. Otherwise, v_i remains the same as in (34). In the latter case, W_i^* and $\sum_{j=v_i}^{i} Z_j$ (the inside sum in (34)) also remain the same. On the other hand, if v_i is now 0, W_i^* and $\sum_{j=v_i}^{i} Z_j$ are increased by $\zeta_0 - \nu_0 \geq 0$ and $Z_0 \geq 0$, respectively. But the values of W_i^* and $\sum_{j=v_i}^{i} Z_j$ for the *new* sample path have in fact the same distribution as W_{i+1}^* and $\sum_{j=v_{i+1}}^{i+1} Z_j$ for the *original* sample path. This means that these expressions are stochastically non-decreasing in i, which implies that their expectations, namely $E_{\theta,0}[W_i^*]$ and $E_{\theta,0}\left[\sum_{j=v_i}^{i} Z_j\right] = \frac{\partial}{\partial\theta} E_{\theta,0}[W_i^*]$ (from Proposition 13), are non-decreasing in i. Then, the averages $w_t(\theta,0)/t = \sum_{i=1}^{t} E_{\theta,0}[W_i^*]/t$ and $w_t'(\theta,0)/t = \sum_{i=1}^{t} \frac{\partial}{\partial\theta} E_{\theta,0}[W_i^*]/t$ are non-decreasing in t. ∎

Observe that the result of the previous proposition is not true in general, for $s \neq 0$.

Proposition 15. *Under Assumption 1, $w(\theta)$ and $w_t(\theta, s)$, for each $t \geq 1$ and $s \in \bar{S}$, are non-decreasing and convex in θ.*

PROOF. Since $\zeta_i = \theta Z_i$ and from (2), it is easily seen that (for fixed Z_i's) each W_i and W_i^* is non-decreasing and convex in θ. Therefore, for each (s,t), $w_t(\theta, s)$ is non-decreasing and convex in θ. This implies that $w(\theta) = \lim_{t\to\infty} w_t(\theta, s)/t$ is also non-decreasing and convex in θ. From Assumption 1 (v), it follows that $\alpha(\theta)$ is strictly convex. ∎

Proposition 16. *For a given regenerative cycle, let τ be the number of customers in that cycle and $h_\tau'(\theta, 0, \omega)$ as in (38). Under Assumption 1, one has*

$$w'(\theta) = \frac{E_\theta\left[h_\tau'(\theta, 0, \omega)\right]}{E_\theta[\tau]}. \tag{49}$$

PROOF. From Proposition 13 (for the first equality), and the expected value version of the renewal-reward Theorem [45] (for the second one), one has

$$\lim_{t\to\infty} \frac{w_t'(\theta,0)}{t} = \lim_{t\to\infty} \frac{E_\theta\left[h_t'(\theta,0,\omega)\right]}{t} = \frac{E_\theta\left[h_\tau'(\theta,0,\omega)\right]}{E_\theta[\tau]}. \tag{50}$$

It now remains to show that the latter ratio is equal to $w'(\theta)$. For that, we will show that the ratio is continuous in θ. Then, since $w_t'(\theta,0)$ is continuous in θ (Propositions 11), and since $w_t'(\theta,0)/t$ is non-decreasing in t (Proposition 14), it follows from the Theorem of Dini [40, Theorem 7.13] that the convergence in (50) is uniform:

$$\lim_{t\to\infty} \sup_{\theta\in\bar{\Theta}} \left| \frac{w_t'(\theta,0)}{t} - \frac{E_\theta\left[h_\tau'(\theta,0,\omega)\right]}{E_\theta[\tau]} \right| = 0. \tag{51}$$

From Theorem 7.17 in [40] and since $w_t(\theta,0)/t$ converges to $w(\theta)$, this implies that the limit in (50) must be equal to $w'(\theta)$.

We still have to show that the ratio of expectations is continuous in θ. We have already shown in Proposition 12 the continuity of $E_\theta[\tau] = \ell(\tau)$. We can prove the continuity of the other expectation in a similar way, as follows. From Proposition 10, Assumption 1 (iv), and Theorem I.5.2 in [18], there is a constant $1 \le K < \infty$, independent of θ, such that $E_\theta[\tau^4] \le K$ and $E_\theta\left[\left(\sum_{j=1}^{\tau} Z_j\right)^4\right] \le K$. Then,

$$E_\theta[(h'_\tau(\theta,0,\omega))^2] \;\le\; E_\theta\left[\tau^2\left(\sum_{j=1}^{\tau} Z_j\right)^2\right] \le \left(E_\theta[\tau^4] E_\theta\left[\left(\sum_{j=1}^{\tau} Z_j\right)^4\right]\right)^{1/2} \le K. \qquad (52)$$

(This bound will also be re-used in a later proof.) Now, by simulating for one regenerative cycle at $\theta = \theta_0$, one can estimate $E_\theta[h'_\tau(\theta,0,\omega)]$ by using a likelihood ratio:

$$E_\theta[h'_\tau(\theta,0,\omega)] \;=\; E_{\theta_0}\left[\left(\prod_{i=1}^{\tau} \frac{b_\theta(\zeta_i)}{b_{\theta_0}(\zeta_i)}\right) \sum_{i=1}^{\tau}\sum_{j=1}^{i} Z_j\right].$$

Since this estimator (inside the brackets) is continuous in θ, from (52) and Lebesgue's dominated convergence Theorem, it follows that the expectation is also continuous in θ. ∎

Proposition 17. *Under Assumption 1, (8–9) hold.*

PROOF. First recall that from renewal theory [2, 45], for fixed $(\theta, s) \in \bar\Theta \times \bar S$, one has $\lim_{t\to\infty} w_t(\theta,s)/t = w(\theta)$, which is pointwise convergence. But we want uniform convergence. We will first prove (8), and then (9), under the assumption that $\bar S = \{0\}$. Then, we will generalize.

Let $\bar S = \{0\}$. From the pointwise convergence, since $w_t(\theta, 0)$ and $w(\theta)$ are continuous in θ (Propositions 11 and 12), and since $w_t(\theta,0)/t$ is non-decreasing in t (Proposition 14), it follows from the Theorem of Dini [40, Theorem 7.13] that

$$\lim_{t\to\infty} \sup_{\theta\in\bar\Theta} |w_t(\theta,0)/t - w(\theta)| = 0.$$

We have already shown the analogue for the derivative in (51).

Now, let $\bar S = [0,c]$ for $c > 0$. Recall that (3) and (34) give unbiased estimators for $w_t(\theta,s)$ and $w'_t(\theta,s)$. Since these estimators are non-decreasing in s (trivial to verify; in (34), each v_i is non-increasing in s), $w_t(\theta,s)$ and $w'_t(\theta,s)$ are both non-decreasing in s. For the moment, assume that there is an infinite stream of customers (not just t), and let τ^* denote the index of the last customer in the first busy cycle, i.e., $\tau^* = \min\{i \ge 1 \mid W_{i+1} = 0\}$. Note that τ^* is a stopping time with respect to $\{\mathcal{F}_t,\ t \ge 1\}$. From Proposition 10, $E_{u_0}[\zeta^2] < \infty$, so that from Theorem III.3.1 (i) in [18], $E_{u_0,c}[(\tau^*)^2] < \infty$. Therefore, $E_{\theta,c}[(\tau^*)^2]$ is bounded uniformly in θ, because τ^* is stochastically non-decreasing in θ. We want to bound $|w_t(\theta,s) - w_t(\theta,0)|$ by a constant K_3 that does not depend on t and s. For any given $t \ge 1$, let $\tau_t^* = \min(\tau^*, t)$. For $s = c$, one has

$$\begin{aligned}
\sum_{i=1}^{\tau_t^*} W_i^* &= \sum_{i=1}^{\tau_t^*}\left(c + \sum_{j=1}^{i}\zeta_j\right) \\
&\le c\tau^* + \tau^* \sum_{j=1}^{\tau^*}\zeta_j \\
&\le c\tau^* + (\tau^*)^2 + \left(\sum_{j=1}^{\tau^*}\zeta_j\right)^2
\end{aligned}$$

and, using Theorem I.5.2 in [18],

$$E_{\theta,c}\left[\sum_{i=1}^{\tau_t^*} W_i^*\right] \le E_{\theta,c}\left[c\tau^* + (\tau^*)^2\right] + K_2 E_{\theta,c}[(\tau^*)^2] E_{\theta,c}[\zeta_j^2] \le K_3$$

for some finite constants K_2 and K_3. Also,

$$E_{\theta,c}\left[\sum_{i=\tau^*+1}^{t} W_i^* \middle| \tau^*\right] = \begin{cases} w_{t-\tau^*}(\theta,0) \le w_t(\theta,0) & \text{if } \tau^* < t; \\ 0 & \text{otherwise.} \end{cases}$$

(Here, the sum is defined to be 0 whenever $\tau^* + 1 > t$.) Therefore,

$$\begin{aligned} w_t(\theta,0) &\le w_t(\theta,s) \le w_t(\theta,c) \\ &= E_{\theta,c}\left[\sum_{i=1}^{\tau_t^*} W_i^* + \sum_{i=\tau_t^*+1}^{t} W_i^*\right] \\ &= E_{\theta,c}\left[\sum_{i=1}^{\tau_t^*} W_i^* + E_{\theta,c}\left[\sum_{i=\tau_t^*+1}^{t} W_i^* \middle| \tau^*\right]\right] \\ &\le K_3 + w_t(\theta,0). \end{aligned}$$

Then,

$$\begin{aligned} \sup_{\theta\in\bar{\Theta}, s\in\bar{S}} \left|\frac{w_t(\theta,s)}{t} - w(\theta)\right| &\le \sup_{\theta\in\bar{\Theta}, s\in\bar{S}} \left(\left|\frac{w_t(\theta,s) - w_t(\theta,0)}{t}\right| + \left|\frac{w_t(\theta,0)}{t} - w(\theta)\right|\right) \\ &\le \frac{K_3}{t} + \sup_{\theta\in\bar{\Theta}, s\in\bar{S}} \left|\frac{w_t(\theta,0)}{t} - w(\theta)\right| \end{aligned}$$

converges to zero as $t \to \infty$. This proves (8).

The proof of (9) is similar. For $s = c$, one has

$$\sum_{i=1}^{\tau_t^*}\sum_{j=1}^{i} Z_j \le \tau^* \sum_{j=1}^{\tau^*} Z_j \le (\tau^*)^2 + \left(\sum_{j=1}^{\tau^*} Z_j\right)^2.$$

From Theorem I.5.2 in [18], there are then finite constants K_4 and K_5 such that

$$E_{\theta,c}\left[\sum_{i=1}^{\tau_t^*}\sum_{j=1}^{i} Z_j\right] \le E_{\theta,c}\left[(\tau^*)^2\right] + K_4 E_{\theta,c}[(\tau^*)^2] E_{\theta,c}[Z_j^2] \le K_5.$$

Also, since $w_t'(\theta,0)$ is non-decreasing in t,

$$E_{\theta,c}\left[\sum_{i=\tau^*+1}^{t}\sum_{j=1}^{i} Z_j \middle| \tau^*\right] = w_{t-\tau^*}'(\theta,0) \le w_t'(\theta,0)$$

if $\tau^* < t$, and is zero otherwise. Therefore,

$$\begin{aligned} w'_t(\theta,0) &\le w'_t(\theta,s) \le w'_t(\theta,c) \\ &= E_{\theta,c}\left[\sum_{i=1}^{\tau_t^*}\sum_{j=1}^{i} Z_j + \sum_{i=\tau_t^*+1}^{t}\sum_{j=1}^{i} Z_j\right] \\ &= E_{\theta,c}\left[\sum_{i=1}^{\tau_t^*}\sum_{j=1}^{i} Z_j + E_{\theta,c}\left[\sum_{i=\tau_t^*+1}^{t}\sum_{j=1}^{i} Z_j \Big| \tau^*\right]\right] \\ &\le K_5 + w'_t(\theta,0). \end{aligned}$$

This, with the uniform convergence for $s = 0$, implies (9). ∎

Proposition 18. *Consider the LR estimator (22), under Assumption 1. Then,*

$$\sup_{\theta\in\bar{\Theta},\, s\le c,\, t\ge 1} E_{\theta,s}\left[\frac{\psi_t^2(\theta,s,\omega)}{t^3}\right] < \infty.$$

PROOF. We will first show that

$$\sup_{\theta\in\bar{\Theta}} E_\theta[d_i^4] < \infty. \tag{53}$$

Let $K > 1$. From Assumption 1 (iii), for each $\theta_0 \in \bar{\Theta}$, there is an open interval $\Upsilon(\theta_0) = (\theta_0 - \epsilon_0, \theta_0 + \epsilon_0)$ and a constant $\tilde{K}(\theta_0) < \infty$ such that

$$E_{\bar{\theta}}\left[\sup_{\theta\in\Upsilon(\theta_0)} \left(\frac{\partial}{\partial\theta}\ln b_\theta(\zeta)\right)^4\right] \le \tilde{K}(\theta_0),$$

where $\bar{\theta} = \theta_0 + \epsilon_0$. It follows that

$$\begin{aligned} \sup_{\theta\in\Upsilon(\theta_0)} E_\theta[d_i^4] &= \sup_{\theta\in\Upsilon(\theta_0)} E_{\bar{\theta}}\left[\left(\frac{\partial}{\partial\theta}\ln b_\theta(\zeta)\right)^4 \frac{b_\theta(\zeta)}{b_{\bar{\theta}}(\zeta)}\right] \\ &\le K E_{\bar{\theta}}\left[\sup_{\theta\in\Upsilon(\theta_0)} \left(\frac{\partial}{\partial\theta}\ln b_\theta(\zeta)\right)^4\right] \\ &\le K\tilde{K}(\theta_0). \end{aligned}$$

Now, $\{\Upsilon(\theta_0),\ \theta_0 \in \bar{\Theta}\}$ is a family of open sets that covers $\bar{\Theta}$. Since $\bar{\Theta}$ is compact, there is a finite subset of that family, say $\{\Upsilon(\theta^{(1)}), \ldots, \Upsilon(\theta^{(N)})\}$, that covers $\bar{\Theta}$, and one has

$$\sup_{\theta\in\bar{\Theta}} E_\theta[d_i^4] \le \max_{1\le i\le N} K\tilde{K}(\theta^{(i)}) < \infty.$$

Since $E_\theta[\zeta_i^4] < \infty$ (Proposition 10), from section VIII.2 of [2], since $E_{\theta,s}[(W_i^*)^4] \le E_{\theta,c}[(W_i^*)^4]$, and from (53), there exists a constant $\delta < \infty$ such that

$$\sup_{\theta\in\bar{\Theta},\, s\le c,\, i\ge 1} E_{\theta,s}[(W_i^*)^4 + d_i^4] \le \delta.$$

Recall that $E[d_j] = 0$ and that the d_j's are independent. Then, using Cauchy-Schwartz inequality, one gets

$$\begin{aligned} E_{\theta,s}[\psi_t^2(\theta,s,\omega)] &= E_{\theta,s}\left[\left(\sum_{i=1}^t W_i^*\right)^2\left(\sum_{j=1}^t d_j\right)^2\right] \\ &\leq \left[E_{\theta,s}\left(\sum_{i=1}^t W_i^*\right)^4 E_{\theta,s}\left(\sum_{j=1}^t d_j\right)^4\right]^{1/2} \\ &\leq \left[t^4 \sup_{1\leq i\leq t} E_{\theta,s}((W_i^*)^4)\sum_{i=1}^t\sum_{j=1}^t E_{\theta,s}(d_i^2 d_j^2)\right]^{1/2} \\ &\leq [t^4\delta t^2\delta]^{1/2} = t^3\delta. \blacksquare \end{aligned}$$

Proposition 19. *Consider the regenerative LR estimator (30), with r regenerative cycles. Suppose that Assumption 1 holds and that*

$$\sup_{\theta\in\bar{\Theta}} E_\theta\left[d_i^8\right] < \infty. \tag{54}$$

Then, as $r \to \infty$, that estimator has bounded variance and converges in quadratic mean to $w'(\theta)$, uniformly with respect to θ in $\bar{\Theta}$.

PROOF. Here, almost sure convergence follows easily from the law of large numbers applied to each sum divided by r. But uniform convergence in expectation is more difficult to prove. What we will show is that the estimator (30), minus $w'(\theta)$, converges to zero in quadratic mean, uniformly in θ. This implies our results.

For a given regenerative cycle, let τ be the number of customers in that cycle. One has $S_\tau(\theta,0,\omega) = \sum_{j=1}^{\tau} d_j$ and $h_\tau(\theta,0,\omega) = \sum_{j=1}^{\tau} W_j^* \leq \tau\sum_{j=1}^{\tau}\zeta_j$. From Proposition 10 and Theorem I.5.2 in [18], there is a constant K independent of θ, $1 \leq K < \infty$ such that $E_\theta[\tau^8] \leq K$, $E_\theta\left[\left(\sum_{j=1}^{\tau}\zeta_j\right)^8\right] \leq K$, and $E_\theta\left[\left(\sum_{j=1}^{\tau} d_j\right)^8\right] \leq K$. Then,

$$E_\theta[(h_\tau(\theta,0,\omega))^4] \leq E_\theta\left[\tau^4\left(\sum_{j=1}^{\tau}\zeta_j\right)^4\right] \leq \left(E_\theta[\tau^8]E_\theta\left[\left(\sum_{j=1}^{\tau}\zeta_j\right)^8\right]\right)^{1/2} \leq K.$$

Define $A_{1j} = h_jS_j - w(\theta)\tau_jS_j - w'(\theta)\tau_j$ and $A_{2j} = h_j - w(\theta)\tau_j$. Note that $E_\theta[A_{1j}] = E_\theta[A_{2j}] = 0$, since $w(\theta) = E_\theta[h_j]/E_\theta[\tau_j]$ and $w'(\theta) = (E_\theta[h_jS_j] - w(\theta)E_\theta[\tau_jS_j])/E_\theta[\tau_j]$ (from Proposition 12). Also, since $E_\theta[\tau_j] \geq 1$ (used in the first two lines) and $E_\theta[A_{2j}] = 0$ (used in the last line), we have

$$\begin{aligned} w(\theta) &\leq E_\theta[h_j] \leq (E_\theta[h_j^4])^{1/4} \leq K^{1/4}, \\ w'(\theta) &\leq E_\theta[h_jS_j] - w(\theta)E_\theta[\tau_jS_j] \\ &\leq |E_\theta[h_jS_j]| + w(\theta)|E_\theta[\tau_jS_j]| \\ &\leq (E_\theta[h_j^2S_j^2])^{1/2} + K^{1/4}(E_\theta[\tau_j^2S_j^2])^{1/2} \\ &\leq (E_\theta[h_j^4])^{1/4}(E_\theta[S_j^8])^{1/8} + K^{1/4}(E_\theta[\tau_j^8]E_\theta[S_j^8])^{1/8} \end{aligned}$$

$$\begin{aligned}
&\leq K^{3/8}+K^{1/2} \leq 2K^{1/2}, \\
E_\theta[A_{1j}^2] &\leq 2E_\theta[h_j^2 S_j^2]+4w^2(\theta)E_\theta[\tau_j^2 S_j^2]+4(w'(\theta))^2 E_\theta[\tau_j^2] \\
&\leq 2(E_\theta[h_j^4])^{1/2}(E_\theta[S_j^8])^{1/4}+4K^{1/2}(E_\theta[\tau_j^8]E_\theta[S_j^8])^{1/4}+16K(E_\theta[\tau_j^8])^{1/4} \\
&\leq 2K^{3/4}+4K+16K^{5/4} \leq 22K^{5/4}, \\
E_\theta[A_{2j}^4] &= E_\theta[(h_j-w(\theta)\tau_j)^4] \\
&\leq 8E_\theta[h_j^4]+8(w(\theta))^4 E_\theta[\tau_j^4] \\
&\leq 8K+8K\cdot K^{1/2} \leq 16K^{3/2}, \\
E_\theta\left[\left(\frac{1}{r}\sum_{i=1}^r A_{2i}\right)^4\right] &= E_\theta\left[\frac{1}{r^4}\sum_{i=1}^r\sum_{j=1}^r A_{2i}^2A_{2j}^2\right) \leq \frac{1}{r^2}E_\theta[A_{2j}^4] \leq 16K^{3/2}/r^2.
\end{aligned}$$

Keeping in mind that $\tau_j \geq 1$ and $E_\theta[A_{1j}]=0$ for each j, one then has

$$\begin{aligned}
&E_\theta\left[\frac{\sum_{j=1}^r h_jS_j}{\sum_{j=1}^r \tau_j}-\frac{\sum_{j=1}^r h_j\sum_{j=1}^r \tau_jS_j}{\left(\sum_{j=1}^r \tau_j\right)^2}-w'(\theta)\right]^2 \\
&\leq E_\theta\left[\frac{1}{r}\left(\sum_{j=1}^r h_jS_j-w(\theta)\sum_{j=1}^r\tau_jS_j-w'(\theta)\sum_{j=1}^r\tau_j+\left(\sum_{j=1}^r\tau_j\right)^{-1}\left(w(\theta)\sum_{j=1}^r\tau_j-\sum_{j=1}^r h_j\right)\sum_{j=1}^r\tau_jS_j\right)\right]^2 \\
&\leq 2E_\theta\left[\left(\frac{1}{r}\sum_{j=1}^r A_{1j}\right)^2\right]+2E_\theta\left[\left(\frac{1}{r}\sum_{j=1}^r A_{2j}\right)^2\left(\frac{1}{r}\sum_{j=1}^r \tau_jS_j\right)^2\right] \\
&\leq \frac{2}{r}E_\theta[A_{1j}^2]+2\left(E_\theta\left[\left(\frac{1}{r}\sum_{j=1}^r A_{2j}\right)^4\right]E_\theta\left[\left(\frac{1}{r}\sum_{j=1}^r \tau_jS_j\right)^4\right]\right)^{1/2} \\
&\leq \frac{2}{r}E_\theta[A_{1j}^2]+2\left(\frac{16K^{3/2}}{r^2}E_\theta[\tau_j^4S_j^4]\right)^{1/2} \\
&\leq \frac{2}{r}(22K^{5/4}+16K^{3/4}K^{1/2}) \leq 76K^{5/4}/r.
\end{aligned}$$

As r goes to infinity, this clearly converges to zero uniformly in θ. These inequalities also provide a uniform upper bound of $76K^{5/4}/r$ on the variance of (30). ∎

Proposition 20. *Suppose that Assumption 1 holds, that the system was originally started from the empty state $s=0$, and that the service time of the j-th customer overall has distribution B_{θ_j} with $\theta_j \in \bar{\Theta}$. The θ_j's can be different and might even be random, but we assume in that case that θ_j is $\mathcal{F}_{j-1}$-measurable for each j and that $\theta_j \in \bar{\Theta}$ with probability one. Then, B_{θ_j} must be viewed as the service time distribution of customer j conditional on θ_j, i.e. conditional on the past. Let v_i*

be defined as in (34). Then, there is a constant $K < \infty$ such that

$$\sup_{k\geq 0,\, t\geq 1} E\left[\left(\frac{1}{t}\sum_{i=k+1}^{k+t}\sum_{j=v_i}^{i} Z_j\right)^2\right] \leq K. \tag{55}$$

Here, E denotes the expectation associated with the above sequence of θ_j's and we assume that it is well defined.

PROOF. Suppose first that we use the service time distribution B_{u_0} for all the customers, namely $\zeta_j = u_0 Z_j$. Then, the queue is a stable regenerative system [2, Chap. VIII]. Let τ be the number of customers in a regenerative cycle (say, the first one). From Assumption 1 and Proposition 10,

$$E_{u_0}[\tau^4 + Z^4] < \infty. \tag{56}$$

For each customer i, define

$$\xi_i = \sum_{j=v_i}^{i} Z_j$$

and let us view for the moment ξ_i^2 as a "cost" associated to customer i. The expected "cost" per regenerative cycle is then, using Theorem I.5.2 in [18].

$$\begin{aligned} E_{u_0}\left[\sum_{i=1}^{\tau}\xi_i^2\right] &\leq E_{u_0}\left[\tau\left(\sum_{i=1}^{\tau} Z_j\right)^2\right] \\ &\leq \left(E_{u_0}[\tau^2]E_{u_0}\left[\left(\sum_{i=1}^{\tau} Z_j\right)^4\right]\right)^{1/2} \\ &\leq \left(E_{u_0}[\tau^2]K_1 E_{u_0}[\tau^4]E_{u_0}[Z^4]\right)^{1/2} \\ &\leq K \end{aligned}$$

for some finite constants K_1 and K. From the renewal-reward theorem [45], one then has

$$\lim_{t\to\infty}\frac{1}{t}\sum_{i=1}^{t} E_{u_0}[\xi_i^2] = E_{u_0}\left[\sum_{i=1}^{\tau}\xi_i^2\right]/E_{u_0}[\tau] \stackrel{\text{def}}{=} \tilde{K} \leq K. \tag{57}$$

It has been shown in the proof of Proposition 14 that ξ_i is stochastically non-decreasing in i. This implies that $E_{u_0}[\xi_i^2]$ is non-decreasing in i. Using (57), it follows that $E_{u_0}[\xi_i^2] \leq \lim_{t\to\infty} E_{u_0}[\xi_t^2] = \tilde{K} \leq K$.

Now, we will complete the proof using stochastic ordering arguments similar to those used in the proof of Proposition 10. For a given sequence of underlying uniform variates, increasing θ_j increases the service time of customer j and does not affect the other service and interarrival times. Clearly, increasing a service time can never split a busy period, i.e. can never increase any v_i. Therefore, τ and each ξ_i are stochastically non-decreasing in each θ_j. Since $\theta_j \leq u_0$ for each j, ξ_i generated under the assumptions of the proposition is stochastically dominated by ξ_i generated under the assumption that $\theta_j = u_0$ for all j. This implies that $E[\xi_i^2] \leq E_{u_0}[\xi_i^2] \leq K$, where E is the same as in (55). The expectation in (55), which is the second moment of the average of $\xi_{k+1}, \ldots, \xi_{k+t}$, is then bounded by K. ∎

Proposition 21. *As $r \to \infty$, (39) has bounded variance and converges in quadratic mean to $w'(\theta)$, uniformly with respect to θ.*

PROOF. Let K be as in the proof of Proposition 16. Then, from (52) and since $\tau \geq 1$,

$$w'(\theta) \leq E_\theta\left[|h'_\tau(\theta, 0, \omega)|\right] \leq K^{1/2}. \tag{58}$$

For $j = 1, \ldots, r$, define $A_j = h'_j - w'(\theta)\tau_j$. These A_j's are i.i.d. and, from Proposition 16,

$$E_\theta[A_j] = E_\theta[h'_j] - w'(\theta)E_\theta[\tau_j] = 0.$$

Also,

$$E_\theta[A_j^2] \leq 2(E_\theta[(h'_j)^2] + (w'(\theta))^2 E_\theta[\tau_j^2]) \leq 2(K + K^{3/2})$$

and, for $j \neq i$,

$$E_\theta[A_j A_i] = E_\theta[A_j]E_\theta[A_i] = 0.$$

Then,

$$\begin{aligned} E_\theta\left[\frac{\sum_{j=1}^{r} h'_j}{\sum_{j=1}^{r} \tau_j} - w'(\theta)\right]^2 &= E_\theta\left[\frac{\sum_{j=1}^{r}(h'_j - w'(\theta)\tau_j)}{\sum_{j=1}^{r}\tau_j}\right]^2 \\ &\leq E_\theta\left[\left(\frac{1}{r}\sum_{j=1}^{r} A_j\right)^2\right] \\ &\leq \frac{1}{r^2}E_\theta\left[\sum_{j=1}^{r} A_j^2\right] = \frac{1}{r}E_\theta\left[A_j^2\right] \leq 4K^{3/2}/r. \end{aligned}$$

As r goes to infinity, this converges to zero uniformly in θ. This also provides a uniform upper bound of $4K^{3/2}$ on the variance of (39). ∎

PROOF of Proposition 3. We will show that the mean-square error of $h_t(\theta, s, \omega)/t$ is in the order of $1/t$, uniformly in (θ, s). From that, $E_{n-1}[\epsilon_n^2]$, which is the variance of Y_n, is in the order of $t_n^{-1}c_n^{-2}$ and $\sum_{n=1}^{\infty} E_{n-1}[\epsilon_n^2]n^{-2} < \infty$ with probability one. Then, from Proposition 17, $\lim_{n\to\infty} \beta_n^F = 0$ with probability one and the result follows from Proposition 1.

We now bound the mean-square error. We convene that the j-th busy cycle ends when the system empties out for the j-th time and that the first "busy cycle" starts with customer 1. So, when $s \neq 0$, the first "busy cycle" does not obey the same probability law than the other ones, but all these busy cycles are nevertheless independent. For $j \geq 1$, let τ_j be the number of customers in the j-th busy cycle, h_j the total sojourn time of those τ_j customers, and $A_j = h_j - w(\theta)\tau_j$. For $j \geq 2$, one has $E_\theta[A_j] = 0$ and, from the proof of Proposition 19, there are finite constants K_1 and K_2 independent of θ such that $w(\theta) \leq K_1$ and $E_\theta[A_j^2] \leq E_\theta[h_j^2 + K_1^2\tau_j^2] \leq K_2$. For $j = 1$, by the same kind of argument as in the proof of Proposition 17, the are finite constants K_3 and K_4 independent of θ such that

$$\begin{aligned} E_{\theta,s}[\tau_1^4] &\leq K_3; \\ E_{\theta,s}[h_1^2] &\leq E_{\theta,c}\left[\left(c\tau_1 + \tau_1^2 + \left(\sum_{j=1}^{\tau_1}\zeta_j\right)^2\right)^2\right] \\ &\leq 3E_{\theta,c}\left[c\tau_1^2 + \tau_1^4 + \left(\sum_{j=1}^{\tau_1}\zeta_j\right)^4\right] \leq K_4. \end{aligned}$$

Then, there is a finite constant K_5 such that

$$E_{\theta,s}[A_1^2 + h_1^2] \leq E_{\theta,s}[2h_1^2 + (w(\theta))^2\tau_1^2] \leq K_5.$$

Let $M(t) = \sup\{i \geq 0 \mid \sum_{j=1}^{i}\tau_j \leq t\}$ be the number of busy cycles completed when the t-th customer leaves and $\Lambda(t) = \sum_{j=1}^{M(t)+1}\tau_j$. Applying Wald's equation and observing that $M(t) \leq t$ (for the last two inequalities), one obtains:

$$\begin{aligned} & E_{\theta,s}\left[\left(\frac{h_t(\theta,s,\omega)}{t} - w(\theta)\right)^2\right] \\ = \;& E_{\theta,s}\left[\left(\frac{1}{t}\sum_{i=1}^{t}(W_i^* - w(\theta))\right)^2\right] \\ \leq \;& \frac{1}{t^2}E_{\theta,s}\left[\left(\left(\sum_{j=1}^{M(t)+1}A_j\right) - \sum_{i=t+1}^{\Lambda(t)}(W_i^* - w(\theta))\right)^2\right] \\ \leq \;& \frac{2}{t^2}E_{\theta,s}\left[\left(\sum_{j=1}^{M(t)+1}A_j\right)^2 + \left(h_{M(t)+1}\right)^2 + \left(w(\theta)\tau_{M(t)+1}\right)^2\right] \\ \leq \;& \frac{2}{t^2}E_{\theta,s}\left[\sum_{i,j=1}^{M(t)+1}A_iA_j + \sum_{j=1}^{M(t)+1}(h_j^2 + w^2(\theta)\tau_j^2)\right] \\ \leq \;& \frac{2}{t^2}E_{\theta,s}\left[A_1^2 + h_1^2 + 2A_1\sum_{j=2}^{M(t)+1}A_j + \sum_{i,j=2}^{M(t)+1}A_iA_j + \sum_{j=1}^{M(t)+1}(h_j^2 + w^2(\theta)\tau_j^2)\right] \\ \leq \;& \frac{2}{t^2}\Big[E_{\theta,s}\left[A_1^2 + h_1^2\right] + 2E_{\theta,s}[A_1]E_{\theta,0}[M(t)]E_{\theta,0}[A_2] \\ & \quad + K_5 + E_{\theta,0}[M(t)]E_{\theta,0}[A_2^2 + h_2^2 + w^2(\theta)\tau_2^2]\Big] \\ \leq \;& \frac{2}{t^2}[2K_5 + tK_2] \in O(1/t). \;\blacksquare \end{aligned}$$

PROOF of Proposition 5. From Proposition 11, $\psi_{t_n}(\theta_n, s_n, \omega_n)$ is an unbiased estimator of $w'_{t_n}(\theta_n, s_n)$, so that $\beta_n^R = 0$. From Proposition 17, we know that $\beta_n = \beta_n^F \to 0$ when $t_n \to 0$. From Proposition 18, there exists a constant K such that $E_{n-1}[\epsilon_n^2] \leq Kt_n$ for all n. Therefore, $\sum_{n=1}^{\infty}E_{n-1}[\epsilon_n^2]n^{-2} \leq \sum_{n=1}^{\infty}Kt_nn^{-2} < \infty$. The first result then follows from Proposition 1.

For the regenerative case, it is shown in Proposition 19 that as the number of regenerative cycles $t_n \to \infty$, the expression (30) has bounded variance and converges in quadratic mean to $w'(\theta)$, uniformly in θ. This implies uniform convergence in expectation. Then, $\lim_{n\to\infty} \beta_n = 0$, the variance of Y_n is uniformly bounded, and Proposition 1 applies. ∎

PROOF of Proposition 6. From Proposition 13, $h'_{t_n}(\theta_n, s_n, \omega_n)$ is an unbiased estimator of $w'_{t_n}(\theta_n, s_n)$, so that $\beta_n^R = 0$. From Proposition 17, we know that $\beta_n = \beta_n^F \to 0$ when $t_n \to 0$. From Proposition 20, the variance of $h'_{t_n}(\theta_n, s_n, \omega_n)$ (conditional on $s_1 = 0$) is bounded uniformly in θ_n and t_n (even when θ varies between customers and iterations). The first result then follows from Proposition 1.

For the regenerative case, it is shown in Proposition 21 that as the number of regenerative cycles $t_n \to \infty$, the expression (39) has bounded variance and converges in quadratic mean to $w'(\theta)$, uniformly in θ. This implies uniform convergence in expectation. Then, $\lim_{n\to\infty} \beta_n = 0$, the variance of Y_n is uniformly bounded, and Proposition 1 applies. ∎

PROOF of Proposition 7. For the proof of this proposition, we will redefine differently the state of the Markov chain. Remove the restriction $s_n \le c$ and redefine the system state at iteration n as $s_n = (x_n, a_n)$, where x_n is the *sojourn* time of the last customer of iteration $n-1$ ($x_1 = 0$), and a_n is the value of the IPA accumulator at the beginning of iteration n. Here, we assume that the arrival time of the first customer of an iteration is "unknown" (not part of the state) at the beginning of the iteration. We do that in order to facilitate the verification of the continuity conditions required in W1 of Appendix I. Let $s = (x, a)$ be the system state at the beginning of an iteration, k_t^* be defined as in (36),

$$\psi^* = ak_t^* + \sum_{i=1}^{t} \sum_{j=v_i}^{i} Z_j \tag{59}$$

and

$$\xi = \left(\psi^*,\ W_t^*,\ I(k_t^* = t)a + \sum_{j=v_t}^{t} Z_j \right). \tag{60}$$

Here, ψ^* is the value of the IPA estimator (37), while the other two components of ξ give the initial state for the next iteration. If the current iteration is iteration n, then $(\theta, x, a) = (\theta_n, x_n, a_n)$ and $\xi = \xi_n = (\psi_n^*, x_{n+1}, a_{n+1})$. We need to verify assumptions W1 to W5 of Appendix I and the result will follow from Theorem 4. Since t_n is fixed at t, $P_{\theta,x,a}(\xi_n \in \cdot)$ does not depend on n.

To prove the weak continuity, let $g : \mathbb{R}^3 \to \mathbb{R}$ be continuous and bounded in absolute value by a constant K_g. We need to show that $E_{\theta,x,a}[g(\xi)]$ is continuous in (θ, x, a). Let $\theta_0 \in \Theta$, $K > 1$, and ϵ_0 and Υ as in Assumption 1 (iii). Let $\bar{\theta} = \theta_0 + \epsilon_0$, $x_0 \ge 0$, and $a_0 \ge 0$. Now, for $|\theta - \theta_0| < \epsilon_0$, $x \ge 0$ and $a \ge 0$, define

$$\begin{aligned} A(\theta, x, a, \omega) &= g\left(ak_t^* + \sum_{i=1}^{t} \sum_{j=v_i}^{i} Z_j,\ W_t^*,\ I(k_t^* = t)a + \sum_{j=v_t}^{t} Z_j \right) \prod_{i=1}^{t} \frac{b_\theta(\zeta_i)}{b_{\bar{\theta}}(\zeta_i)} \\ &\quad - g\left(a_0 k_{t,0}^* + \sum_{i=1}^{t} \sum_{j=v_{i,0}}^{i} Z_j,\ W_{t,0}^*,\ I(k_{t,0}^* = t)a_0 + \sum_{j=v_{t,0}}^{t} Z_j \right) \prod_{i=1}^{t} \frac{b_{\theta_0}(\zeta_i)}{b_{\bar{\theta}}(\zeta_i)}, \end{aligned}$$

where $k^*_{t,0}$, $v_{i,0}$, and $W^*_{t,0}$ are the respective values of k^*_t, v_i, and W^*_t when x is replaced by x_0 and $\omega = (\nu_0, \zeta_1, \ldots, \nu_{t-1}, \zeta_t)$ remains the same. We assume that ω is generated using $\bar{\theta}$, i.e. that $\zeta_i = \bar{\theta} Z_i$ for each i. Let $I(x, x_0) = 1$ if $v_{i,0} \neq v_i$ for at least one i, and $I(x, x_0) = 0$ otherwise. Note that $I(x, x_0) = 0$ implies that $k^*_{t,0} = k^*_t$. Also,

$$\begin{aligned}
P_{\bar{\theta}}(I(x, x_0) = 1) &\leq \sum_{i=1}^{t} P_{\bar{\theta}} \left[\left| \nu_0 - x_0 + \sum_{j=1}^{i-1} (\nu_j - \zeta_j) \right| < |x - x_0| \right] \\
&= \sum_{i=1}^{t} E_{\bar{\theta}} \left[P_{\bar{\theta}} \left[|\nu_0 - x_0 + z| < |x - x_0| \,\middle|\, \sum_{j=1}^{i-1} (\nu_j - \zeta_j) = z \right] \right] \\
&\leq 2tK_\nu |x - x_0|,
\end{aligned}$$

where $E_{\bar{\theta}}$ integrates over the values of z, and K_ν is a bound on the density of the interarrival time ν_0. Conditional on $I(x, x_0) = 0$, $A(\theta, x, a, \omega)$ is continuous in (θ, x, a), because g is continuous, W^*_t is continuous in x and does not depend on (θ, a), $b_\theta(\zeta)$ is continuous in θ for each ζ, $k^*_{t,0} = k^*_t$, and $v_{i,0} = v_i$ for each i. Further, $|A(\theta, x, a, \omega)|$ is bounded by $2K_g K^t$ and is zero when $(\theta, x, a) = (\theta_0, x_0, a_0)$. Therefore,

$$\begin{aligned}
& \lim_{(\theta,x,a) \to (\theta_0, x_0, a_0)} |E_{\theta,x,a}[g(\xi)] - E_{\theta_0,x_0,a_0}[g(\xi)]| \\
= & \lim_{(\theta,x,a) \to (\theta_0, x_0, a_0)} |E_{\bar{\theta}}[A(\theta, x, a, \omega)]| \\
\leq & \lim_{(\theta,x,a) \to (\theta_0, x_0, a_0)} \left| E_{\bar{\theta}}[A(\theta, x, a, \omega)(1 - I(x, x_0))] + E_{\bar{\theta}} \left[2K_g K^t I(x, x_0) \right] \right| \\
\leq & E_{\bar{\theta}} \left[\lim_{(\theta,x,a) \to (\theta_0, x_0, a_0)} \left| A(\theta, x, a, \omega)(1 - I(x, x_0)) \right| \right] + \lim_{(\theta,x,a) \to (\theta_0, x_0, a_0)} 2K_g K^t 2tK_\nu |x - x_0| \\
= & 0,
\end{aligned}$$

where Lebesgue's dominated convergence theorem has been used to pass the limit inside the expectation to get the last inequality. This proves the required weak continuity. This also implies (as a special case) that $E_{\theta,x,a}[\psi^*]$ is continuous in (θ, x, a), which verifies the second requirement of W1, with $c = 0$.

For fixed $\theta \in \Theta$, since the system is stable, $\{\xi_n, n \geq 1\}$ is regenerative and is a Markov chain with some steady-state distribution P^θ (see [2], chapter VIII). Regeneration occurs whenever an iteration starts with an empty system. From the proof of Proposition 20, there exists $K_1 < \infty$ such that $\sup_{n \geq 1} E_0[(\psi^*_n / t_n)^2] \leq K_1$ and $\sup_{n \geq 1} E_0[a_n^2] \leq K_1$. This yields W3. By similar arguments, one can show that $\sup_{n \geq 1} E_0[x_n^2] \leq K_2$ for some constant $K_2 < \infty$. Take $K = \max(K_1, K_2)$. For any $\epsilon > 0$, one has $K \geq E_0[(\psi^*_n / t_n)^2] \geq (3K/\epsilon) P[(\psi^*_n / t_n)^2 > 3K/\epsilon]$, so that $\sup_{n \geq 1} P[(\psi^*_n / t_n)^2 > 3K/\epsilon] \leq \epsilon/3$. Similarly, $\sup_{n \geq 1} P[x_n^2 > 3K/\epsilon] \leq \epsilon/3$ and $\sup_{n \geq 1} P[a_n^2 > 3K/\epsilon] \leq \epsilon/3$. Then, $\sup_{n \geq 1} P[\max((\psi^*_n / t_n)^2, x_{n+1}^2, a_{n+1}^2) \leq 3K/\epsilon] \geq 1 - \epsilon$. This reasoning also holds for θ varying in any manner inside $\bar{\Theta}$. This implies the tightness properties required in W1.

For W2, let C be a compact subset of $\mathbb{R} \times S$, $c < \infty$ such that $C \subseteq [0, c]^3$, and let $\xi_n \in C$. Let i denote the i-th customer overall and $nt + 1 + \tau^*_n$ be the index of the first non-waiting customer from the beginning of iteration $n + 1$. One has $\tau^*_n = 0$ if iteration $n + 1$ starts with a new busy cycle and otherwise, τ^*_n is the number of customers, from the beginning of iteration $n + 1$, who are in the same busy cycle as the last customer of iteration n. From the same argument as in the proof of Proposition 17, there exists $K_\tau(c) < \infty$ such that $E_{u_0,c}[(\tau_1^*)^2] \le K_\tau(c)$. Then, from straightforward stochastic ordering, $E[(\tau_n^*)^2 \mid \xi_n] \le E_{u_0,c}[(\tau_1^*)^2] \le K_\tau(c)$. This implies that for all $\epsilon > 0$, $P[\tau_n^* \ge K_\tau(c)/\epsilon \mid \xi_n] \le \epsilon$. Let $\epsilon > 0$, $n^*(c) = \lceil K_\tau(c)/\epsilon \rceil$, $\tilde{c} = (3K/\epsilon)^{1/2}$, and $\tilde{C} = [0, \tilde{c}]^3$. Let $n_c = 1 + \lceil n^*(c)/t \rceil$ and $i \ge n_c$. For each $0 \le j \le n_c$, from the same argument as we used above to prove W1, one has $P[\xi_{n+i} \in \tilde{C} \mid \tau_n^* = j] \ge 1 - \epsilon$. Then,

$$\begin{aligned} P[\xi_{n+i} \in \tilde{C} \mid \xi_n] &\ge \sum_{j=0}^{n^*(c)} P[\xi_{n+i} \in \tilde{C},\ \tau_n^* = j \mid \xi_n] \\ &= \sum_{j=0}^{n^*(c)} P[\tau_n^* = j \mid \xi_n] P[\xi_{n+i} \in \tilde{C} \mid \tau_n^* = j] \\ &\ge (1-\epsilon) P[\tau_n^* \le n^*(c) \mid \xi_n] \\ &\ge (1-\epsilon)^2. \end{aligned}$$

Here, P denotes the probability law associated with the Markov chain $\{\xi_n, n \ge 1\}$ when θ varies according to the algorithm and n_c can be viewed as a time that we give to the system to stabilize. Roughly, if c is larger, the initial state could be larger (e.g. large initial queue size), and we will take a larger n_c. This implies W2.

When θ is fixed, according to the proof of Proposition 4, $\sum_{j=v_i}^{i} Z_j$ is an unbiased estimator of the gradient of the expected system time of the i-th customer (overall). Then, ψ_n^* is unbiased for the gradient of the expected total system time of customers $nt, \ldots, (n+1)t - 1$. When $n \to \infty$, from (9), the expectation of $\psi_n^*/t_n + C'(\theta)$ thus converges to $\alpha'(\theta)$. Therefore, $v(\theta) = \alpha'(\theta)$ and W5 follows. ∎

PROOF of Proposition 8. We just have to check that the arguments in the proofs still hold, sometimes with slight adaptations. We will quickly discuss these adaptations. The other proofs basically remain the same.

In all the proofs, replace B_{u_0} by $\tilde{B}$, and E_{u_0} by $\tilde{E}$ (the expectation that corresponds to $\tilde{B}$). In Proposition 17, if $Z_i \ge 0$ in (vi) is replaced by $Z_i \le 0$, then $w_t'(\theta, s)$ becomes non-increasing in t and s. In the proof, after "The proof of (9) is similar", replace Z_j by $-Z_j$ in the first two equations, then replace the last four "$\le$" (in the last two equations) by "$\ge$". For the proof of Proposition 20, observe that $|Z_j| \le \Gamma(U_j)$. Define $\tilde{\xi}_i = \sum_{j=v_i}^{i} \Gamma(U_j)$. Then,

$$\begin{aligned} \tilde{E}\left[\sum_{i=1}^{\tau} \tilde{\xi}_i^2\right] &\le \tilde{E}\left[\tau \left(\sum_{i=1}^{\tau} \Gamma(U_j)\right)^2\right] \\ &\le \left(\tilde{E}[\tau^2] \tilde{E}\left[\left(\sum_{i=1}^{\tau} \Gamma(U_j)\right)^4\right]\right)^{1/2} \\ &\le \left(\tilde{E}[\tau^2] K_1 \tilde{E}[\tau^4] \tilde{E}[(\Gamma(U_j))^4]\right)^{1/2} \\ &\le K \end{aligned}$$

for some finite constants K_1 and K. Using an argument similar to that in the proof of Proposition 14, it is easily seen that $\tilde{\xi}_i$ is stochastically non-decreasing in i. Therefore, $\tilde{E}[\tilde{\xi}_i^2]$ is stochastically non-decreasing in i and bounded by K, and it follows that $E[\xi_i^2] \le E[\tilde{\xi}_i^2] \le \tilde{E}[\tilde{\xi}_i^2] \le K$. Then, the last sentence of the proof is still valid. Note that the boundedness of the fourth moment in Assumption 2 (iv) is used in the proof of the analogue of (52). ∎

PROOF of Proposition 9. The following notation is inspired from [2]. Define $D_i = \zeta_i - \nu_i$, $S_0 = 0$, $S_i = D_1 + \cdots + D_i$, $M_t = \max_{1 \le i \le t} S_i$, $M = \lim_{t\to\infty} M_t$, and $i_*(t) = \max\{i \mid S_i = M_t\}$. Recall that $B_\theta^{-1}(u) \le \tilde{B}^{-1}(u)$ for each $u \in (0,1)$. Let $(\zeta_i^B, D_i^B, S_i^B, M_t^B, i_*^B(t))$ be the values corresponding to $(\zeta_i, D_i, S_i, M_t, i_*(t))$ when $\tilde{B}$ is used instead of B_θ, with the same uniform variates, and let E denote the associated mathematical expectation (which does not depends on (θ, s)). Note that $\zeta_i^B \ge \zeta_i$, $D_i^B \ge D_i$, $S_i^B \ge S_i$, $M_t^B \ge M_t$, and $i_*^B(t) \ge i_*(t)$. The latter holds because $i_*^B(t) = i$ if and only if $D_{i+1}^B + \cdots + D_j^B < 0$ for all j, $i < j \le t$, which implies $D_{i+1} + \cdots + D_j < 0$ for all j, $i < j \le t$, because $D_j \le D_j^B$.

Since W_t has the same distribution as $\max(s + S_t, M_{t-1})$ (see [2, p.80, Corollary 7.4]), we have

$$E_{\theta,s}[M_{t-1} - M] \le E_{\theta,s}[W_t - M] \le E_{\theta,s}[(s + S_t)^+] + E_{\theta,s}[M_{t-1} - M].$$

But $E_{\theta,s}[(s + S_t)^+] \le E[(c + S_t^B)^+] \to 0$ as $t \to \infty$ since $S_t^B \to -\infty$, $(c + S_t^B)^+ \le c + M^B$, and $E[c + M^B] < \infty$ (and from Lebesgue's dominated convergence theorem). This convergence does not depend on (θ, s). Also, using the definition of $i_*(t)$ for the second inequality,

$$0 \le M - M_t = \sum_{i \ge i_*(t)} D_i \le \sum_{i \ge i_*^B(t)} D_i \le \sum_{i \ge i_*^B(t)} D_i^B = M^B - M_t^B.$$

But $E[M^B - M_t^B] \to 0$ as $t \to \infty$ (from the monotone convergence theorem) and this convergence does not depend on (θ, s). Therefore, $\lim_{t\to\infty} |E_{\theta,s}[W_t - M]| = 0$ uniformly in (θ, s), which implies that

$$\lim_{t\to\infty} |w_t(\theta, s) - w(\theta)| = \lim_{t\to\infty} \frac{1}{t} \left| \sum_{i=1}^{t} E_{\theta,s}[W_t - M] \right| = 0$$

uniformly in (θ, s). ∎

Acknowledgements

The work of the first author was supported by NSERC-Canada grant no. OGP0110050 and FCAR-Québec grant no. EQ2831. The third author's work was supported by the IBM Corporation under SUR-SST Contract 12480042 and by the U.S. Army Research Office under Contract DAAL-03-91-G-0101. Part of the work was done while the first author was visiting the Operations Research Department at Stanford University. Abderrahim Loukili and Philippe Locong performed some preliminary numerical experiments which were helpful. We also wish to thank S. Asmussen, W. Hengartner, Y. T. Leung, G. Perron, and R. Suri for helpful suggestions.

References

[1] Aleksandrov, V. M., V. I. Sysoyev and V. V. Shemeneva, "Stochastic Optimization", *Engineering Cybernetics*, **5** (1968), 11–16.

[2] Asmussen, S., *Applied Probability and Queues*, Wiley, 1987.

[3] Azadivar, F. and Talavage, J. "Optimization of Stochastic Simulation Models", *Math. and Computers in Simulation*, **XXII** (1980), 231–241.

[4] Bratley, P., B. L. Fox, and L. E. Schrage, *A Guide to Simulation*, Springer-Verlag, New York, Second Edition, 1987.

[5] Cao, X. R., "Sensitivity Estimates Based on One Realization of a Stochastic System", *J. Stat. Comput. Simul.*, **27** (1987), 211–232.

[6] Chong, E. K. P. and Ramadge, P. J., "Convergence of Recursive Optimization Algorithms using Infinitesimal Perturbation Analysis Estimates", Draft paper, Dept. of Electrical Engineering, Princeton University, 1990.

[7] Chong, E. K. P. and Ramadge, P. J., "Optimization of Queues using an IPA based Stochastic Algorithm with General Update Times", Draft paper, Dept. of Electrical Engineering, Princeton University, 1991.

[8] Fu, M. C. and Ho, Y.-C., "Using Perturbation Analysis for Gradient Estimation, Averaging and Updating in a Stochastic Approximation Algorithm", *Proceedings of the Winter Simulation Conference 1988*, IEEE Press (1988), 509–517.

[9] Giroux, N. "Optimisation Stochastique de Type Monte Carlo", Mémoire de maîtrise, dépt. d'informatique, Univ. Laval, jan. 1989.

[10] Glasserman, P., "Performance Continuity and Differentiability in Monte Carlo Optimization", *Proceedings of the Winter Simulation Conference 1988*, IEEE Press (1988), 518–524.

[11] Glasserman, P., "Derivative Estimates From Simulation of Continuous Time Markov Chains", Technical report, AT&T Bell Laboratories, Holmdel, New Jersey (1989).

[12] Glasserman, P., *Gradient Estimation via Perturbation Analysis*, Kluwer Academic, 1991.

[13] Glynn, P. W. "Stochastic Approximation for Monte-Carlo Optimization", *Proceedings of the Winter Simulation Conference 1986*, IEEE Press (1986), 356–364.

[14] Glynn, P. W. "Likelihood Ratio Gradient Estimation: an Overview", *Proceedings of the Winter Simulation Conference 1987*, IEEE Press (1987), 366–375.

[15] Glynn, P. W. "Optimization of Stochastic Systems Via Simulation", *Proceedings of the Winter Simulation Conference 1989*, IEEE Press (1989), 90–105.

[16] Glynn, P. W. "Likelihood Ratio Gradient Estimation for Stochastic Systems", *Communications of the ACM*, **33**, 10 (1990), 75–84.

[17] Gong, W. B. and Ho, Y. C. "Smoothed (Conditional) Perturbation Analysis of Discrete Event Dynamical Systems", *IEEE Trans. on Automatic Control*, **AC-32**, 10 (1987), 858–866.

[18] Gut, A. *Stopped Random Walks*, Springer-Verlag, 1988.

[19] Heidelberger, P., Cao, X.-R., Zazanis, M. A. and Suri, R., "Convergence Properties of Infinitesimal Perturbation Analysis Estimates", *Management Science*, **34**, 11 (1989), 1281–1302.

[20] Ho, Y.-C., "Performance Evaluation and Perturbation Analysis of Discrete Event Dynamic Systems", *IEEE Transactions of Automatic Control*, **AC-32**, 7 (1987), 563–572.

[21] Jacobson, S. H. and Schruben, L. W., "Techniques for Simulation Response Optimization", *Operations Research Letters*, **8** (1989), 1–9.

[22] Kesten, H., "Accelerated Stochastic Approximation", *Ann. Math. Stat.*, **23** (1952), 463–467.

[23] Kushner, H. J. and Clark, D. S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, Applied Math. Sciences, vol. 26, 1978.

[24] Kushner, H. J., *Approximation and Weak Convergence Methods for Random Processes*, The MIT Press, Cambridge, Mass., 1984.

[25] Kushner, H. J. and Shwartz, A., "An Invariant Measure Approach to the Convergence of Stochastic Approximations with State Dependent Noise", *SIAM J. on Control and Optim.*, **22**, 1 (1984), 13–24.

[26] L'Ecuyer, P., "SIMOD: Définition fonctionnelle et guide d'utilisation (version 2.0)", Report no. DIUL-RT-8804, Département d'informatique, Univ. Laval (1988).

[27] L'Ecuyer, P., "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques", *Management Science*, **36**, 11 (1990), 1364–1383.

[28] L'Ecuyer, P. and Perron, G., "On the Convergence Rates of IPA and FDC Derivative Estimators for Finite-Horizon Stochastic Simulations", Submitted for publication (1990).

[29] L'Ecuyer, P., "On the Interchange of Derivative and Expectation for Likelihood Ratio Derivative Estimators", Submitted for publication (1991).

[30] L'Ecuyer, P. and Glynn, P. W., "A Control Variate Scheme for Likelihood Ratio Gradient Estimation", In preparation (1991).

[31] L'Ecuyer, P., "Convergence Rates for Steady-State Derivative Estimators", To appear in *Annals of Operations Research* (1991).

[32] Luenberger, D. G., *Linear and Nonlinear Programming*, Addison-Wesley, second ed., 1984.

[33] Meketon, M. S., "Optimization in Simulation: a Survey of Recent Results", *Proceedings of the Winter Simulation Conference 1987*, IEEE Press (1987), 58–67.

[34] Métivier, M. and Priouret, P., "Application of a Kushner and Clark Lemma to General Classes of Stochastic Algorithms", *IEEE Trans. on Information Theory*, **IT-30**, 2 (1984), 140–151.

[35] Pflug, G. Ch., "On-line Optimization of Simulated Markovian Processes", *Math. of Operations Research*, **15**, 3 (1990), 381–395.

[36] Reiman, M. I. and Weiss, A., "Sensitivity Analysis for Simulation via Likelihood Ratios", *Operations Research*, **37**, 5 (1989), 830–844.

[37] Rubinstein, R. Y., *Monte-Carlo Optimization, Simulation and Sensitivity of Queueing Networks*, Wiley, 1986.

[38] Rubinstein, R. Y., "The Score Function Approach for Sensitivity Analysis of Computer Simulation Models", *Math. and Computers in Simulation*, **28** (1986), 351–379.

[39] Rubinstein, R. Y., "Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models", *Operations Research*, **37**, 1 (1989), 72–81.

[40] Rudin, W., *Principles of Mathematical Analysis*, Third Edition, McGraw-Hill, 1976.

[41] Suri, R., "Infinitesimal Perturbation Analysis of General Discrete Event Dynamic Systems", *J. of the ACM*, **34**, 3 (1987), 686–717.

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

[43] Suri, R. and Leung, Y. T., "Single Run Optimization of Discrete Event Simulations—An Empirical Study Using the M/M/1 Queue", *IIE Transactions*, **21**, 1 (1989), 35–49.

[44] Suri, R. and Zazanis, M., "Perturbation Analysis Gives Strongly Consistent Estimates for the M/G/1 Queue", *Management Science*, **34**, 1 (1988), 39–64.

[45] Wolff, R., *Stochastic Modeling and the Theory of Queues*, Prentice-Hall, 1989.

[46] Zazanis, M. A. and Suri, R., "Comparison of Perturbation Analysis with Conventional Sensitivity Estimates for Stochastic Systems", Submitted for publication (1988).

[47] Zazanis, M. A. and Suri, R., "Perturbation Analysis of the GI/G/1 Queue", Submitted for publication (1989).