

**EVALUATING A QUEUEING APPROXIMATION
FOR THE MACHINE INTERFERENCE PROBLEM
WITH TWO TYPES OF STOPPAGES
VIA SIMULATION OPTIMIZATION**

Manuel D. Rossetti
Assistant Professor
Department of Systems Engineering
University of Virginia
Thornton Hall
Charlottesville, VA 22903

and

Gordon Clark
Professor Emeritus
Department of Industrial, Welding, and Systems Engineering
The Ohio State University
210 Baker Systems
1971 Neil Avenue
Columbus, OH 43210

**EVALUATING A QUEUEING APPROXIMATION
FOR THE MACHINE INTERFERENCE PROBLEM
WITH TWO TYPES OF STOPPAGES
VIA SIMULATION OPTIMIZATION ¹**

Abstract

An optimization framework for the evaluation of queueing approximations is proposed and illustrated on a queueing approximation for the two-type stoppage machine interference model. The optimization framework combines infinitesimal perturbation analysis and stochastic approximation in order to optimize the approximation's error function using simulation. The optimization procedure attempts to find the worst case error of the approximation and the associated parameter values. The optimization framework is compared against the standard factorial design approach for queueing approximation evaluation. The results indicate the potential of evaluating queueing approximations using an optimization framework. In addition, the performance of the machine interference approximation is shown to be worse than previous results indicated using factorial design approaches. This research illustrates to practitioners how queueing approximation evaluation is performed and should indicate to researchers new directions to take in queueing approximation evaluation methodology.

Keywords: Simulation, Queueing, Optimization, Machine Interference, Stochastic Approximation

¹Published in *Computers and Industrial Engineering* Vol. 34, No. 3, pp. 655-688

1 Introduction

The machine interference problem is a practical problem which occurs in a variety of systems. Machine interference is a common problem in manufacturing systems as well as communication systems. The classic example of machine interference occurs when operators or repairpersons tend to the operation of several production machines. The machines require the attention of the operators to load or unload material or to perform minor machine adjustments or repairs. When an operator is attending to the needs of a machine, the machine is considered idle. If another machine requires the operator for service, the latter machine must wait and is said to have interference time. Machine interference models are used to predict the performance of the machines in terms of operator efficiency, machine utilization, and interference time. The models are often used to determine the economic assignment of operators to machines and to decide the most appropriate maintenance policies. Stecke [1] presents a detailed survey of the machine interference problem with an extensive reference list and a discussion of the general approaches taken to model the machine interference system. Our problem involves the tending of operators to machines when there are two types of stoppages for service. The first type represents a stoppage for removal of the product, and the second type represents a stoppage for repair.

In analyzing the behavior of machine interference systems, analysis yielding closed form analytical solutions is often a difficult if not impossible task. When closed form solutions are unavailable, researchers and practitioners apply numerical techniques, simulation, or queueing approximations. Queueing approximations are simplified models which act as approximations for more “realistic” intractable mathematical queueing models. The accuracy of the approximation is an important issue when applying the approximation to analyze the behavior of the queueing system.

In this paper, we present the evaluation of an approximation for the complex two type stoppage machine interference model. The standard methodology for evaluating queueing approximations is the use of factorial designs over a representative sample of test cases. The factorial designs assess the accuracy of the approximation by comparing the results of the approximation to the results of a simulation for the various test cases. Researchers also assess the accuracy by comparing the results of competing approximations for the same queueing system. An example of the standard methodology can be found in reference [2] which evaluates the performance of the Queueing Network Analyzer(QNA). Our solution methodology deviates from the standard methodology by formulating the evaluation in terms of an opti-

mization problem. The optimization problem attempts to maximize the error between the approximation's output measure and the "true" output measure obtained via simulation. The motivation is to find the parameter settings of the approximation which cause the worse case error, and thus constitute the worse case parameter settings of the approximation. If the error under the worse case parameter settings is acceptable, then we might consider the approximation as acceptable. If the error under the worse case parameter settings is not acceptable, then we have identified the conditions under which the approximation should be used with caution, or modified for better results.

The purpose of this research is twofold. First, we want to evaluate the proposed approximation for the two type stoppage machine interference model. Secondly, we want to evaluate the usefulness of the proposed optimization methodology for evaluating queueing approximations. Our objective is to illustrate the proposed methodology and to give an indication of its potential for further research. The evaluation will be a detailed analysis of the approximation over an experimental set which although larger than previous studies is still comparable in scope. In the next section, we present background information on the machine interference problem, stochastic approximation, and infinitesimal perturbation analysis. In Section 3, we define the two type stoppage machine interference system, the proposed queueing approximation, and the optimization problem. In Section 4, we discuss briefly the simulation implementation, and then present a detailed discussion of the infinitesimal perturbation analysis and stochastic approximation algorithms. In Section 5, we present the experimental design and results of our investigations. Finally, in Section 7, we summarize the main conclusions of the research and present ideas for future research.

2 Background

In this research, we apply an optimization methodology to a single continuous parameter of the two type stoppage machine interference model. The other parameters of the model are held fixed under experimental conditions while performing the optimization. We apply a single dimensional stochastic approximation(SA) algorithm of the type defined by Robbins and Monroe, see reference [3], to solve the optimization problem. We implement infinitesimal perturbation analysis(IPA) on a queueing network representation of the system in order to obtain the gradient information used within the stochastic approximation algorithm. This section presents background information on stochastic approximation, and perturbation analysis. We give the motivation and concepts behind these topics as they

relate to this research while referring the reader to pertinent references for further details.

Glynn [4] offers a classification framework for stochastic optimization. According to his classification, the type of optimization problem encountered in this research is of the finite-dimensional continuous parameter type. Glynn also surveys the approaches to solving the finite-dimensional continuous parameter problem. We apply stochastic approximation in order to solve the optimization problem. Stochastic approximation is a sequential procedure designed to find the location of a root of a noisy function. Procedures of the stochastic approximation type can obtain the best possible monte carlo convergence rate, see for example Glynn [4]. We refer the reader to references [5], [6], [7], [8], and [9] for information on the properties of stochastic approximation and for further references on this topic. We present a detailed version of our stochastic approximation algorithm in Section 4.

Finally, for the optimization of a noisy objective function, stochastic approximation searches for the root of the derivative of the objective function. The procedure requires the estimation of derivative information over the search region. We use infinitesimal perturbation analysis to estimate derivative information during the course of the simulation. Perturbation analysis is a technique which allows for sensitivity calculations on discrete event dynamic systems. Suri [10] presents a formal definition of a discrete event dynamic system along with information pertaining to the generation of perturbations. The system can be described by a tableau containing the event sequence over time which is called the nominal path. The perturbed path is obtained when the system is observed with a slight(infinitesimal) change in the parameters. Although, perturbation analysis offers a reduction in the number of simulation runs needed to estimate gradient information, the applicability of infinitesimal perturbation analysis to discrete event dynamic systems is problem dependent, see for example reference [11]. We refer the reader to reference [12] for the early developments of perturbation analysis and to references [13], [14], and [15] for details and further references as to perturbation analysis's applicability and consistency properties.

3 Problem Definition

We evaluate the same complex two type stoppage machine interference model evaluated in Guild and Hartnett [16]. The model is similar to, in terms of Kendall's notation, the M/M/R/K/K finite source repairperson model, but the machines have two types of stoppages where R is the number of repairpersons or operators and K is the number of machines. Type 1 stoppages occur after a constant amount of time, $RT_1 = 1/\lambda_1$, and have a constant

service time of $ST_1 = 1/\mu_1$. Type 2 stoppages occur after random intervals of time which are negative exponentially distributed with a mean of $RT_2 = 1/\lambda_2$. Service times for Type 2 stoppages are also negative exponentially distributed with a mean of $ST_2 = 1/\mu_2$. All of the operators have the same skill level, and the operators serve the machines on a first come first served basis with no priority given to either type of stoppage. After a stoppage occurs, the machine arrives for the operator and may join the queue. The service performed by one of the R operators depends on the type of stoppage. Each operator services only one machine at a time. The stoppage processes for the two types of stoppages occur independently of each other. The stoppage with the shortest time to the next stoppage will occur first. The time remaining on the other type of stoppage is machine running time and is kept for when the machine is next running.

The performance measure of interest in this research is the long run average machine utilization which is defined for the single type stoppage M/M/R/K/K system as the ratio of the average running time per machine per cycle to the average total cycle time per machine. A cycle for a machine corresponds to the time between when the machine enters production, has a stoppage, and then reenters production. Denoting the utilization as U , the average running time per machine per cycle as RT , and the average waiting time in the system (queue plus service) for a machine as W_s , we have that

$$U = \frac{RT}{RT + W_s} \quad (1)$$

Through the use of queueing conservation laws equation (1) can be rewritten as

$$U = 1 - \frac{L_s}{K} \quad (2)$$

where L_s is the average number of stopped machines waiting or in repair. The queueing approximation involves applying the single type stoppage M/M/R/K/K results to the two type stoppage case with the servicing factor, ρ , defined as

$$\rho = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} \quad (3)$$

For more information on the above approximation, we refer the reader to references [17] and [18].

3.1 Optimization Problem

We propose to evaluate the above approximation by simulation optimization applied to an error function representing the difference between the approximation and the output from

the simulation in terms of the utilization performance measure. We define the expected error as follows:

$$E[\text{error}] = U_a - E[U_s] \quad (4)$$

where U_a is the utilization obtained from the approximation and U_s is the utilization obtained from the simulation. Accordingly, U_a and U_s are functions of $(\lambda_1, \mu_1, \lambda_2, \mu_2, R, K)$. We propose to find the worse case value of $ST_2 = 1/\mu_2$, while considering the other parameters at fixed levels by finding the maximum of the absolute value of the error. The parameters $(\lambda_1, \mu_1, \lambda_2, R, K)$ will be fixed at the levels used in references [17] and [16]. We define the worse case error(WCE) as follows

$$\text{WCE} = \max_{ST_2} |E[\text{error}]| \quad (5)$$

The worse case error can be rewritten as the solution of two optimization problems as follows:

$$\text{WCE} = \max\{|g_1|, |g_2|\} \quad (6)$$

$$\text{where } g_1 = \max_{ST_2} |E[\text{error}]| \quad (7)$$

$$\text{and } g_2 = \min_{ST_2} |E[\text{error}]| \quad (8)$$

The solution to g_1 or g_2 involves finding where the derivative of $E[\text{error}]$ is equal to zero. Thus, the derivative of the utilization from the approximation, dU_a/dST_2 , and the derivative of the utilization from the simulation, dU_s/dST_2 , are required. The derivative of the utilization from the approximation is given in the appendix. The derivative of the utilization from the simulation is obtained from infinitesimal perturbation analysis (IPA). The IPA development is discussed in the following sections.

4 Solution Methodology

In this section, we present the details of the implementation of the infinitesimal perturbation analysis algorithm and the stochastic approximation algorithm as well as some of the simulation details. We developed the simulation program representing the complex two type stoppage machine interference model in Simscript II.5. Details of the simulation program are available from the authors upon request. We verified the simulation on a known analytical test case, and verified the IPA and SA algorithms on an optimization problem with a known solution. A typical simulation experiment involves the following: the setting of the parameters, simulating a required number of stoppages, deleting an initial transient number of stoppages, and preparing for the next experiment. We initialize each simulation experiment from the empty and idle state.

4.1 IPA Estimator and Algorithm

The implementation of the IPA algorithm involves the generation and propagation of perturbations in order to generate the perturbed sample path. To generate perturbations for estimating sensitivities due to the mean service time for a type two stoppage, $ST_2 = 1/\mu_2$, we followed the development given in Suri [19] for the negative exponential distribution. To propagate the perturbations, we view the system as a two station closed tandem queueing network with K servers at the first station and R servers at the second station and apply the propagation rules given in [20], see also reference [21]. For comparison purposes, we developed three estimators of dU/dST_2 based on the IPA algorithm. For this paper, we present only the estimator used within the optimization algorithm.

The average machine utilization can be defined as the total running time per machine divided by the total simulation run length. If the simulation is run for a fixed number of stoppages, the simulation stopping time is a simple performance measure as defined in [10]. We develop the gradient estimator as follows:

Let T_s be the simulation stopping time

Let $M(t)$ be the number of running machines at time t , $0 \leq M(t) \leq K$

Let RT_{total} be the total running time per machine where

$$RT_{total} = \frac{\int_0^{T_s} M(t) dt}{K} \quad (9)$$

Let U be the average machine utilization where

$$U = \frac{RT_{total}}{T_s} \quad (10)$$

Differentiating equation (10) with respect to ST_2 and substituting equation (10) in order to remove RT_{total} , we have that

$$\frac{dU}{dST_2} = \frac{-U}{T_s} \frac{dT_s}{dST_2} \quad (11)$$

A perturbation propagation argument indicates that RT_{total} is not a function of ST_2 . We use IPA to estimate dT_s/dST_2 during the simulation while we estimate U and T_s in the normal manner during the simulation. Our gradient estimator of dU/dST_2 becomes

$$\frac{\hat{d}U}{dST_2} = \frac{-\hat{U}}{\hat{T}_s} \frac{\hat{d}T_s}{dST_2} \quad (12)$$

The IPA algorithm estimates dT_s/dST_2 for each operator and the estimate of dT_s/dST_2 is based on the average over the operators. The IPA algorithm is given in Figure 1. The key to the algorithm is the use of the variable *carry_perturbation*. Because the system is closed, perturbations to the service time of a machine will carry over to the next operator which serves the machine only if the next operator to serve the machine was idle immediately before being called for service. The variable *carry_perturbation* propagates the perturbations.

Figure 1: IPA Algorithm

1. initialize
 - for each server, let perturbation(server) = 0.0
2. during simulation
 - while number of stoppages \leq total number of stoppages
 - do
 - stoppage occurs*
 - record current server*
 - let $j =$ current server
 - generate service according to type*
 - if Type 2 stoppage needs service
 - generate service time $x_i \sim \text{expo}(ST_2)$
 - add x_i/ST_2 to perturbation(j)
 - endif
 - perform service*
 - capture perturbation for next server*
 - let carry_perturbation = perturbation(j)
 - stoppage finished*
 - generate time and type of next stoppage*
 - carry perturbation to next server*
 - if there is an idle server
 - let perturbation(next server) = carry_perturbation
 - endif
 - enddo
3. record simulation stopping time \hat{T}_s and average machine utilization \hat{U}
4. let $\frac{d\hat{T}_s}{dST_2} = \sum_{j=1}^K \text{perturbation}(j)/K$
5. let $\frac{d\hat{U}}{dST_2} = \frac{-\hat{U}}{\hat{T}_s} \frac{d\hat{T}_s}{dST_2}$

Because our problem has 2 types of stoppages it does not fall into the class of problems for which IPA consistency results are directly applicable; however, we checked the consistency properties of the estimator empirically by comparing to finite difference estimates using common random numbers(FDE with CRN). The results which are given in Table 1 indicate that there does not appear to be any serious consistency problems with the estimator, and we thus conclude that the estimator is adequate for our purposes. We also evaluated our IPA based estimators on a known analytical single type stoppage model. Although not the focus of this research, results given in Glasserman [13] and in Fu and Hu [22] may allow for further analytical analysis of the consistency properties of our estimators.

Table 1: Estimator Consistency Check

Test Case:

$$\lambda_1 = 0.02 \quad \mu_1 = 1.0 \quad \lambda_2 = 0.018 \quad \mu_2 = 0.1 \quad R = 3 \quad K = 8$$

Number of stoppages = 5500

Number of stoppages in transient period = 500

Number of replications = 31

Approximation of $\frac{dU}{dST_2} = -0.014159$

	FDE with CRN	IPA
$\frac{dU}{dST_2}$	-0.01376	-0.01479
s^2	2.667×10^{-3}	4.528×10^{-7}

4.2 The Stochastic Approximation Algorithm

The implementation of the stochastic approximation algorithm requires a reliable stopping criteria and a good initial start point. We investigated four different stopping rules and used the stopping criteria given in reference [23] in our procedure. We refer the reader to references [24], [25], and [23] for general details of stochastic approximation stopping rules. In order to find an initial starting point, we divide the search range for ST_2 into tenths and evaluate the error function at each of the ten levels of ST_2 . We select the value of ST_2 which corresponds to the worse error as the initial starting point. We limit the search range on ST_2 to a bounded interval using the same technique as in reference [23]. Whenever the updated value of the parameter obtained from the procedure is past the bounded range, a uniform random number between the current parameter setting and the bound was drawn, and used as the new parameter setting. The stochastic approximation algorithm is given in Figure 2.

Figure 2: The Stochastic Approximation Algorithm

1. Inputs
 - M - maximum number of iterations permitted
 - X_1 - initial starting parameter value
 - A - initial starting scale factor
 - α - exponential smoothing constant
 - ϵ - stopping constant
 - let $n = 1$; $a_1 = 1$
2. Evaluate the derivative of the error function
 - (a) run simulation to generate $d\hat{U}_s/dST_2$
 - (b) use approximation of M/M/R/K/K to get dU_a/dST_2
 - (c) evaluate the derivative of the error function, let $Z_n = dE[\text{error}]/dST_2$
3. Stopping Criteria
 - let $RSC_{n+1} = \alpha RSC_n + (1 - \alpha)Z_n$
 - if $|Aa_n RSC_{n+1}| < \epsilon$, let stop flag = true
4. Update a_n
 - if the sign of $Z_n =$ the sign of Z_{n-1}
 - let $a_{n+1} = a_n$
 - else
 - let $a_{n+1} = \frac{1}{1+1/a_n}$
5. Update X_n
 - let $X_{n+1} = X_n - Aa_n Z_n$
 - if $X_{n+1} \leq$ lower bound
 - let $X_{n+1} = \text{Uniform}(\text{lower bound}, X_n)$
 - elseif $X_{n+1} \geq$ upper bound
 - let $X_{n+1} = \text{Uniform}(X_n, \text{upper bound})$
 - let $n = n + 1$
6. if $n < M$ and stop flag = false, goto Step 2, else stop

The algorithm implements accelerated stochastic approximation, see reference [26], by only decreasing the step size, a_n , when a sign change occurs between consecutive observations. The optimal choice of the scale factor, A , is the reciprocal of the derivative of the function for which the root is sought evaluated at the optimal parameter value. For example, if $f(x)$ is the function for which the root is sought then the optimal value of A is equal to $1/f'(x^*)$ where x^* is the root. We refer the reader to references [5], [27], and [28] for more information on the scale factor. We estimate the value of A by fitting a regression line to the values of the derivative of the objective function on either side of the initial starting point. We then use the slope of the regression line as the estimate for A . Because an extremely large or small value of A can greatly hinder the procedure, we compare the slope of the regression line to the reciprocal of the derivative of the error function at the starting point and use the lesser of the two. The justification behind using the reciprocal of the value of the derivative of the objective function at the initial starting point is to obtain roughly a unit change in the parameter value during the initial steps of the procedure.

We verified that the stochastic approximation algorithm was working as intended by testing the procedure on an optimization problem with a known solution. We also compared the various stopping criteria, but our results were inconclusive as to which stopping rule was better. After the stochastic approximation procedure terminates, we form confidence intervals for the performance measures at the estimated worst case parameter setting by performing replication deletion.

5 Experimental Design and Results

In this section, we present the experimental designs which we used to evaluate the approximation and the results of the evaluation. We also present the experiments and the results of our assessment into the usefulness of the optimization methodology.

5.1 Comparison to Previous Results

Benson and Cox evaluated the approximation with a system having both the times between stoppages and the service times negative exponentially distributed. Guild and Hartnett evaluated the approximation via simulation for the system configuration considered in our research. For comparison purposes, we utilize the same design levels as in [17] and [16] for the parameters $(\lambda_1, \mu_1, \lambda_2, R, \text{ and } K)$ while optimizing on the parameter ST_2 . For comparison purposes, the results in reference [16] are tabulated in Table 2.

Table 2: Guild and Hartnett Results

$\lambda_1 = 0.02 \mu_1 = 1.0$
 $\Delta\% = 100\% \times (\hat{U}_s - U_a)/\hat{U}_s$
 s = standard deviation

				K	
				8	16
R	2	U_a	0.7800	0.5850	
		\hat{U}_s	0.7914	0.6056	
		$\Delta\%$	1.44(-1.14) [†]	3.4(-2.06)	
		s	0.0175	0.0257	
	3	U_a	0.8273	0.7502	
		\hat{U}_s	0.8268	0.7662	
		$\Delta\%$	0.06(-0.05)	2.09(-1.60)	
		s	0.0094	0.0176	
	4	U_a	0.8351	0.8116	
		\hat{U}_s	0.8322	0.8153	
		$\Delta\%$	0.35(-0.29)	0.45(-0.37)	
		s	0.0106	0.0121	

[†]() contains $100\% \times (U_a - \hat{U}_s)$

Table 3: Experimental Arrangement

$\lambda_1 = 0.02 \mu_1 = 1.0 \lambda_2 = 0.018$

		K	
		8	16
R	2	Exp. #1	Exp. #2
	3	Exp. #3	Exp. #4
	4	Exp. #5	Exp. #6

Table 3 represents the levels of the parameters examined which correspond to the design in [16]. The results of our experiments are given in Table 4. The objective of the experiments is to find the worse case value of ST_2 ; therefore, at each parameter value in the design we perform two optimization experiments. The first optimizing experiment finds the maximum positive error, and the second optimizing experiment finds the minimum negative error. The maximum absolute value of the two optimization experiments corresponds to the worse case error. For both optimization problems, we perform the optimization by minimizing an appropriate objective function.

Table 4: Experimental Results

Exp#	ST_2^*	U_a	\hat{U}_s	E[W $\hat{C}E$]%	E[W $\hat{C}E$]		$\Delta\%$ E[W $\hat{C}E$]	N
					LCL%	UCL%		
1	18.838	0.6083	0.5817	2.661	1.873	3.448	4.574	121
2	8.438	0.6811	0.6542	2.691	1.818	3.563	4.112	72
3	33.465	0.5310	0.5115	1.950	1.192	2.709	3.813	121
4	13.234	0.6690	0.6494	1.959	1.044	2.874	3.017	39
5	53.108	0.4528	0.4330	1.978	1.253	2.702	4.568	121
6	19.307	0.6279	0.6109	1.699	1.050	2.349	2.782	90

The steps in an optimization experiment are as follows:

1. set range for ST_2
2. pick initial starting point as discussed in Section 4.2
3. determine scale factor A as discussed in Section 4.2
4. perform SA algorithm
 - (a) set stopping constant, $\epsilon = 0.0001$
 - (b) set exponential smoothing constant, $\gamma = 0.75$
 - (c) set maximum number of iterations, $M = 121$
 - (d) set fixed parameters according to Table 3
 - (e) simulate 2,500 machine stoppages, with first 500 deleted
5. form confidence intervals for 31 replications with confidence coefficient, $\alpha = 0.01$

For all of the experiments, the worse case error was found during the maximum positive

error optimization problem. For each estimated ST_2^* , we construct a confidence interval for the derivative of the expected error objective function. All of the confidence intervals contained zero, so that we have some assurance that the procedure found a root of the derivative of the expected error objective function. For comparison purposes, we also compute the $\Delta\%$ error value reported in reference [16]. The maximum error of 2.691% for experiment#2 is the largest overall estimated error for all the experiments, as shown in Table 4. The value of ST_2^* for experiment#2 is 8.438. We can consider $ST_2^* = 8.438$ to be the worse case parameter value over the test cases examined. The results of Table 4 also indicate that the approximation is biased high because the utilization for the approximation is larger than the utilization for the simulation.

5.2 Evaluation at Levels of RT_2

In order to assess any relationships among the parameters of the approximation, we implemented an experimental design on levels of $RT_2 = 1/\lambda_2$ for experiment#2 of Table 3. For each level of RT_2 in Table 5, we optimize on ST_2 in order to find the corresponding worse case value of ST_2 . The confidence intervals on the derivative of the objective function all contained zero, and the maximum error optimization problem yielded the worse case error for all experiments in Table 5. According to Table 5, the utilization from the approximation is biased high as compared to the utilization from the simulation. A relationship exists between RT_2 and ST_2^* . The value of ST_2^* decreases as the level of RT_2 increases. We found that a value of $\rho \approx 0.1639$ consistently gave the worst case error over the values of ST_2^* and RT_2 examined. The value of N in the tables containing the experimental results represents the number of iterations performed by the stochastic approximation algorithm.

Table 5: Results For Experiment#2 for Levels of RT_2

RT_2	ST_2^*	U_a	\hat{U}_s	E[\hat{WCE}]%	E[\hat{WCE}]		$s^2 \times 10^{-4}$	N
					LCL%	UCL%		
85.55	12.135	0.7089	0.6721	3.678	2.644	4.713	4.246	82
75.55	10.933	0.7009	0.6710	2.989	2.018	3.96	3.738	44
65.55	9.448	0.7025	0.6812	2.128	1.273	2.983	2.897	121
55.55	8.438	0.6811	0.6542	2.691	1.818	3.563	3.021	72
45.55	6.707	0.6811	0.6542	1.253	0.504	2.002	2.225	103
35.55	5.498	0.6736	0.6602	1.337	0.675	1.998	1.738	109
25.55	4.164	0.6511	0.6465	0.586	-0.199	1.371	2.445	117

6 Evaluation of Optimization Methodology

In this section, we present the evaluation of the optimization methodology as compared to the standard methodology. We select a representative test case and perform a single factor search on ST_2 . We also investigate the functional form of the error function for the first test case and investigate the effect of the initial starting point on the procedure.

6.1 Single Factor Test Case

For experiment#2 of Table 3, we performed a simplified search at different levels of ST_2 rather than performing the optimization. We divide the interval $(0, 40)$ for ST_2 into equally spaced points and then perform 31 replications at each point. The replications used the method of common random numbers. Table 6 presents the results of the experiments. The estimated error of the approximation is the highest at $ST_2 = 8.0$; however, because the confidence interval for $ST_2 = 12.0$ overlaps with the confidence interval for $ST_2 = 8.0$, a conclusion can only be made that there is a good chance that the worse case error falls in the interval, $8 \leq ST_2^* \leq 12$. We can not be sure that a larger error does not occur at a point not chosen as a design point. In order to improve the search for ST_2^* by this methodology, we could perform further experiments. We know that stochastic approximation examines “all” of the levels of ST_2 in the search range and has the most efficient monte carlo convergence rate. We found $ST_2^* = 8.438$ by our stochastic approximation methodology in a total of 254 simulation experiments while the single factor search needed 279 simulation experiments to achieve, $8 \leq ST_2^* \leq 12$, as a likely interval for containing ST_2^* . Of the 254 simulation experiments, the minimization optimization problem accounted for 152 experiments, and the maximization optimization problem found the worse case error in only 102 simulation experiments. Although in terms of the number of experiments performed the optimization methodology does not appear superior, we feel that the information gained is superior.

6.2 Functional Form Evaluation

In this section, we examine the function form of the error function for experiment#2 in order to assess the stochastic approximation algorithm’s performance and to check if there were any violations of the underlying assumptions. For experiment#2, we plot the estimated expected error and it’s derivative over the interval $(0.5, 42.5)$. At every point such that $ST_2 = 0.2n + 0.5$ for, $n = 0, 1, 2, \dots, 210$, five simulation runs were made and the average of the estimated expected error and it’s derivative are plotted in Figure 3 and Figure 4. By examining

Table 6: Single Factor Test Case Results

$$\lambda_1 = 0.02 \quad \mu_1 = 1.0 \quad \lambda_2 = 0.018 \quad R = 2 \quad K = 16$$

ST_2	E[error]%	LCL%	UCL%
4.0	0.844	0.4777	1.210
8.0	2.719	1.818	3.621
12.0	1.736	0.7909	2.682
16.0	0.9366	1.248	1.748
20.0	0.6200	-0.609	1.301
24.0	0.4602	-0.1364	1.057
28.0	0.3777	-0.1438	0.8991
32.0	0.3157	-0.1429	0.7744
36.0	0.2914	-0.1222	0.7051

Figure 3, we can see that the function has a large degree of variability (noisy), especially in areas which appear to be maxima or minima. In order to better realize any trends in the data, the data was smoothed using trewess (trimmed resistant weighted scatterplot smooth), see reference [29], with span set at 20% and trimming percentage set at 10% to produce the smoothed curves in the figures. Figure 3 shows that a peak(s) exist(s) in the range of $ST_2 = (8, 12)$, and that possible minima appear near $ST_2 = 1.0$ and as ST_2 increases. The derivative of the expected error seems more regular, in terms of noise, as can be seen in Figure 4. After the smoothing, it is clearly evident that a root exists near $ST_2 = 8.0$. The non-unimodality of the function and the fact that the derivative is clearly not monotonic over the entire search range is cause for concern since this violates one of the stochastic approximation assumptions.

6.3 Starting Point Convergence Evaluation

Because the assumptions of the stochastic approximation procedure may be violated, we examined the convergence behavior of the procedure for a variety of starting points for experiment#2. The results of the experiments are given in Table 7. Because the scale factor A in our procedure depends on the starting point, we performed all experiments with the scale factor A equal to 2000. The experiments for $ST_2 = (8, 12, 16)$ all passed the derivative confidence interval test, and that the experiment on $ST_2 = 4$ appears to be converging on $ST_2^* = 8.438$. The experiments with $ST_2 = (20, 24)$ did not converge towards the earlier

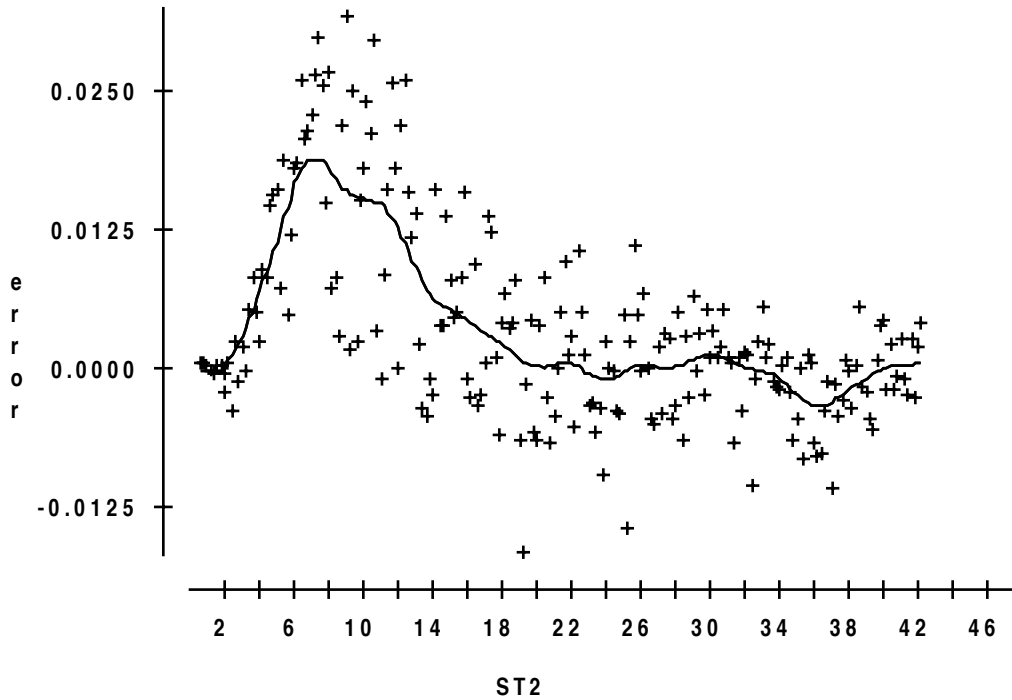


Figure 3: $E[\text{erfor}]$ vs ST_2

reported maximum. Examination of Figure 4 indicates that the stochastic approximation procedure is stalling in flat areas of the function which have values close to zero. The results of the experiments indicate that the procedure is sensitive to the initial starting point.

7 Summary and Conclusions

The goal of this research was to investigate an optimization framework for evaluating queueing approximations using infinitesimal perturbation analysis and stochastic approximation. The investigation was performed on a common and useful queueing approximation, namely, the complex two type machine interference model. This investigation has yielded valuable insights into queueing approximation evaluation under an optimization framework and opens many issues for future research.

A problem that often arises in evaluating queueing approximations by the factorial design approach is whether or not the test cases chosen are representative. By evaluating an approximation by the optimization methodology, we try in essence to examine “all” of the test cases for a continuous parameter. When using the standard methodology, researchers

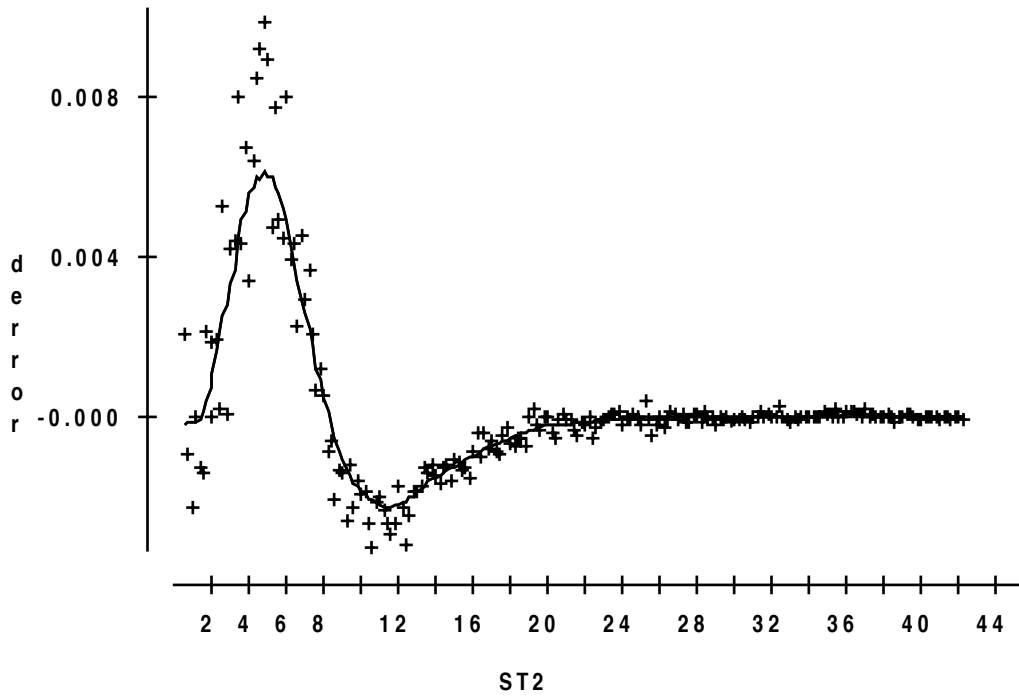


Figure 4: $\frac{dE[\text{error}]}{dST_2}$ vs ST_2

Table 7: Starting Point Evaluation Results

initial ST_2	final ST_2	$E[\widehat{WCE}]%$	LCL%	UCL%	N	derivative CI test
4.0	7.843	2.676	1.821	3.531	121	no
8.0	8.682	2.618	1.685	3.551	121	yes
12.0	8.244	2.711	1.837	3.584	121	yes
16.0	8.220	2.707	1.807	3.608	121	yes
20.0	23.464	0.465	0.142	1.072	58	no
24.0	28.492	0.363	0.150	0.877	121	yes

report the relative error and the absolute relative error for the approximation in order to measure the usefulness of the approximation over the test cases. The relative error(RE) and absolute relative error(ARE) are defined as follows:

$$\text{RE} = \frac{100\% \theta_a - \theta_s}{\theta_s} \quad (13)$$

$$\text{ARE} = \frac{\sum_{i=1}^N |\text{RE}_i|}{N} \quad (14)$$

where i ranges over the number of test cases observed and θ represents the performance measure of interest. Based on the results of Guild and Hartnett, the ARE for the average machine utilization equaled 1.298%. Over the same test cases by our optimization methodology, the ARE equaled 3.811%. The experiments indicate that the approximation is biased high at the worse case error for the test cases examined. We discovered that the ratio of RT_2 to ST_2^* is approximately constant at the apparent worse case levels, so that $\rho \approx 0.1639$ appears to be a worse case value of ρ over the test cases examined. As the individual values of RT_2 and ST_2^* increases, the expected worse case error increases. This is because as RT_2 and ST_2^* increase, there are less of the random stoppages for which the approximation applies. Although the worse case error results are higher than the Guild and Hartnett results, we conclude that the approximation is acceptable for use over the test cases examined.

In evaluating the optimization methodology, we found that much more effort is needed in order to implement the methodology as compared to the factorial design approach. The optimization methodology was difficult to implement even for the relatively simple model and approximation which we investigated. The major difficulty was the development of the IPA estimators. This problem could be alleviated through the use of finite difference estimators (with the corresponding penalty of an increase in the number of simulation experiments) or through the use of other gradient estimation techniques. We feel that the sum gain in information can out weigh the information gained from a strictly factorial design approach. Our approach involved solving the optimization problem by two subproblems. We note that because the approximation was biased high, the maximization subproblem was the optimization problem that found the worse case error. If we had known this property in advance, we could have halved the number of simulation runs performed. An investigation of a different error function which does not involve two subproblems is warranted in future research. For example, a competing error function could simply be the squared difference between the approximation and the true value. Finally, the non-unimodality of the error function places the convergence properties of stochastic approximation algorithms in question. We were able

to reduce the severity of this problem by using good initial starting points. Although the optimization methodology can be challenging to implement, we were able to overcome the major difficulties encountered. We feel that the methodology allowed us to learn more about the approximation than previously published.

Further research areas include extending the search to the multidimensional case which would include a combined continuous-discrete optimization problem. The stochastic approximation procedure can be improved by finding better ways to estimate the scale factor A and by finding ways to eliminate the initial starting point dependency. Research is also needed in finding and evaluating stochastic approximation stopping rules. The area of estimating gradients in simulation experiments should also continue to be an area of intense research interest. Finally, research is needed into applying the optimization methodology to other queueing models and their approximations.

Appendix

In this section, we present the derivations for dU_a/dST_2 . The variables used within this section are:

Let K be the number of machines assigned to the operator(s)

Let R be the number of operators

Let ST be the average service or repair time per machine

Let μ be the average service rate = $1/ST$

Let RT be the average running time per machine

Let λ be the average stoppage rate = $1/RT$

Let $\rho = \lambda/\mu$ be the servicing factor

Let U be the average machine utilization

Let n represent the current number of stopped machines

Let P_n be the steady state probability of n machines stopped

Let L_s be the long run expected number of stopped machines

The standard machine interference results for the M/M/R/K/K model are as follows:

$$\lambda_n = \begin{cases} (K - n)\lambda & 0 \leq n \leq K \\ 0 & n \geq K \end{cases} \quad (15)$$

$$\mu_n = \begin{cases} n\mu & 0 \leq n \leq R \\ R\mu & R \leq n \leq K \\ 0 & n \geq K \end{cases} \quad (16)$$

$$P_n = \begin{cases} \binom{K}{n} \rho^n P_0 & 0 \leq n \leq R \\ \binom{K}{n} \frac{n! \rho^n}{R! R^{n-R}} P_0 & R \leq n \leq K \end{cases} \quad (17)$$

$$P_0 = \left\{ \sum_{n=0}^R \binom{K}{n} \rho^n + \sum_{n=R+1}^K \binom{K}{n} \frac{n! \rho^n}{R! R^{n-R}} \right\}^{-1} \quad (18)$$

$$L_s = \sum_{n=1}^K n P_n \quad (19)$$

$$U = 1 - \frac{L_s}{K} \quad (20)$$

We want to compute

$$\frac{dU}{dST} = \frac{dU}{d\rho} \frac{d\rho}{dST} = \frac{-\lambda}{K} \frac{dL_s}{d\rho} \quad (21)$$

where

$$\frac{dL_s}{d\rho} = \rho^{-1} \left[\sum_{n=1}^K n^2 P_n - L_s^2 \right] \quad (22)$$

Finally, therefore we have,

$$\frac{dU}{dST} = \frac{-\lambda}{K\rho} \left[\sum_{n=1}^K n^2 P_n - L_s^2 \right] \quad (23)$$

Now, because for the approximation ρ is defined as follows

$$\rho = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} = \lambda_1 ST_1 + \lambda_2 ST_2 \quad (24)$$

we have that the derivative with respect to ST_2 is equal to

$$\frac{d\rho}{dST_2} = \lambda_2 \quad (25)$$

Finally, we have,

$$\frac{dU_a}{dST_2} = \frac{-\lambda_2}{K\rho} \left[\sum_{n=1}^K n^2 P_n - L_s^2 \right] \quad (26)$$

References

- [1] Stecke, K. E., “Machine Interference: Assignment of Machines to Operators”, In *The Handbook of Industrial Engineering*, edited by Gavriel Salvendy (New York: John Wiley & Sons), 2nd Edition, pp. 1460–1494, (1992).
- [2] Whitt, W., “Performance of the Queueing Network Analyzer”, *The Bell System Technical Journal*, vol. **62**, no. **9**, (1983), pp. 2779–2815.
- [3] Robbins, H., and Monroe, S., “On a Stochastic Approximation Technique”, *Annals of Mathematical Statistics*, **22**, pp. 400–407, (1951) .
- [4] Glynn, P. W., “Optimization of Stochastic Systems”, In *Proceedings of the 1986 Winter Simulation Conference*, J. Wilson, J. Henriksen, S. Roberts (Eds.), Washington, D.C., pp. 52–58, (1986).
- [5] Meketon, M. S., “Optimization in Simulation: A Survey of Recent Results”, In *Proceedings of the 1987 Winter Simulation Conference*. A. Thesen, H. Grant, and W. D. Kelton (Eds.), Atlanta, GA, pp. 58–67, (1987).
- [6] Fu, M. C. “Optimization via simulation: A review”, *Annals of Operations Research*, **53**, pp. 199–247, (1994).
- [7] Kushner, H. J. and Clark, D. S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems* (New York NY:Springer-Verlag), (1978).
- [8] Kushner, H. J. and Yang, J. “Stochastic Approximation With Averaging and Feedback: Rapidly Convergent “on-line” Algorithms”, *IEEE Transactions on Automatic Control*, **40**, pp. 24–34, (1995).
- [9] Andradóttir, S. “A Stochastic Approximation Algorithm With Varying Bounds”, *Operations Research*, **43**, pp. 1037–48 (1995). ,
- [10] Suri, R., “Infinitesimal Perturbation Analysis for General Discrete Event Systems”, *Journal of the Association for Computing Machinery(ACM)*, vol. **37**, no. **3**, pp. 686–717, (1987).
- [11] Heildelberger, P., Cao, X. R., Zanazani, S. and Suri, R., “Convergence Properties of Infinitesimal Perturbation Analysis Estimates”, *Management Science*, **34**, pp. 1281–1302, (1987) .
- [12] Ho, Y. C., and Cao, X., “Optimization and Perturbation Analysis of Queueing Networks” *Journal of Optimization Theory and Applications*, **40**, **4**, pp. 559–582, (1983).
- [13] Glasserman, P., “Gradient Estimation Via Perturbation Analysis”, Kluwer Academic Publishers, (1991).
- [14] Suri, R., and Zanzanis, M., 1988, “Perturbation Analysis Gives Strongly Consistent Sensitivity Estimates for the M/G/1 Queue”, *Management Science*, Vol. **34**, pp. 39–64, (1988).

- [15] Gong, W. L., and Ho, Y. C., “Smoothed(Conditional) Perturbation Analysis of Discrete Event Dynamical Systems”, *IEEE Transactions on Automatic Control*, vol. **AC-32**, no. **10**, pp. 858–866, (1987).
- [16] Guild, R. D., and Hartnett, R. J., “A Simulation Study of a Machine Interference Problem with M Machines Under the Care of N Operators Having Two Types of Stoppages”, *International Journal of Production Research*, vol. **20**, no. **1**, pp. 47–55, (1982).
- [17] Benson, F., and Cox, D. R., “The Productivity of Machines Requiring Attention at Random Intervals”, *Journal of the Royal Statistical Society*, **B,13**, pp. 65, (1951).
- [18] Chandra, J., and Palesano, J., “A Machine Interference Problem with Multiple Types of Failures”, *International Journal of Production Research*, **24**, no. **3**, pp. 567–582, (1986).
- [19] Suri, R., “Technical Note: Implementation of Sensitivity Calculations on a Monte-Carlo Experiment” *Journal of Optimization Theory and Applications*, vol. **40**, no. **4**, pp. 625–620, (1983).
- [20] Ho, Y. C., and Cao, X., *Discrete Event Dynamic Systems and Perturbation Analysis*, Kluwer Academic, (1991).
- [21] Rubenstein, R. Y., *Monte-Carlo Optimization, Simulation, and Sensitivity of Queuing Networks*(New York: John Wiley & Sons, Inc), (1986).
- [22] Fu, M. C. and Hu, J., “Bias Properties of Infinitesimal Perturbation Analysis for Multi-Server Queues”, In *Proceedings of the 1990 Winter Simulation Conference*, Osman Balci, Randall P. Sadowski, Richard E. Nance (Eds), New Orleans, LA, pp. 377–381 (1990).
- [23] Suri, R., and Leung, Y. T., “Single Run Optimization of a Siman Model for Closed Loop Flexible Assembly Systems”, In *Proceedings of the 1987 Winter Simulation Conference*, A. Thesen, H. Grant, and W. D. Kelton (Eds.), Atlanta, GA, pp. 738–748, (1987).
- [24] Sielken, R. L., “Stopping Rules for Stochastic Approximation Procedures”, *Zeitschrift Wahrscheinlichkeitstheorie*, **39**, pp. 135–150, (1973).
- [25] Stroup, D. F., and Braun, M. I., “On a New Stopping Rule for Stochastic Approximation”, *Zeitschrift Wahrscheinlichkeitstheorie*, **60**, pp. 535–554, (1982).
- [26] Kesten, H., “Accelerated Stochastic Approximation”, *Annals of Mathematical Statistics*, **29**, pp. 41–59, (1958).
- [27] Abdelhamid, S. N., “Transformations of Observations in Stochastic Approximation”, *The Annals of Statistics*, **1**, pp. 1158–1174, (1973).
- [28] Anbar, D., “On Optimal Estimation Methods Using Stochastic Approximation Procedures”, *The Annals of Statistics*, **1**, pp. 1175–1184, (1973).
- [29] Hoaglin, D. C. and Velleman, P. E. *Applications, Basics and Computing of Exploratory Data Analysis*, Belmont, CA, Wadsworth, (1981).