Pergamon

PII: S0969-6016(97)00004-X

*Int. Trans. Opl Res.* Vol. 4, No. 3, pp. 223–233, 1997

0969-6016/97 $17.00 + 0.00

# A Modified Quasi-Newton Method for Optimization in Simulation

CHIANG KAO, WHEYMING TINA SONG[2] and SHIH-PIN CHEN[2]

[1]National Cheng Kung University, Republic of China and [2]National Tsing Hua University, Republic of China

Optimization in Simulation is an important problem often encountered in system behavior investigation; however, the existing methods such as response surface methodology and stochastic approximation method are inefficient. This paper presents a modification of a quasi-Newton method, in which the parameters are determined from some numerical experiments. To demonstrate the validity of the devised method, two examples resembling the $M/M/1$ queueing problem are solved. The closeness of the converged solutions to the optimal solutions and a comparison with two stochastic approximation methods indicate that the modified quasi-Newton method as devised in this paper is a robust and efficient method for solving optimization problems in simulation. © 1997 IFORS. Published by Elsevier Science Ltd.

*Key words*: Stochastic optimization, quasi-Newton method, simulation.

## INTRODUCTION

Computer simulation has been demonstrated to be a very powerful decision-making tool for evaluating complicated systems (Law and Kelton, 1991). In essence, simulation is a controlled statistical sampling technique for estimating the performance of complex stochastic systems with an attempt to find the best design of the systems. For example, in a queueing system, the system designer may wish to determine the optimal service rate and the optimal number of servers to minimize the queue waiting time. In fact, optimization in a stochastic environment has become an important branch of operations research (OR).

An important application of stochastic optimization is simulation optimization (Andradottir, 1995), in which the objective function can be evaluated only by computer simulation. In general, optimization problems in simulation can be formulated as follows:

$$\text{Opt. } f(X) = E[F(X)]$$

where $F(\cdot)$ is a stochastic response function and $f(\cdot)$ is a performance measure. Note that there is no analytical expression for $f(\cdot)$, consequently it is impossible to calculate the gradient of the objective function. In addition, $f(\cdot)$ is stochastic in nature, which presents a major problem in the estimation of gradients (Glynn, 1986; Ringuest, 1988; Fu *et al.*, 1995) and undermines the effectiveness of the gradient-based optimization methods.

Solution methods for optimization problems in simulation can be classified into four classes (Meketon, 1987; Jacobson and Schruben, 1989; Safizadeh, 1990; Azadivar, 1992; Fu, 1994a,b): heuristic, stochastic approximation, response surface methodology (RSM), and gradient-based nonlinear programming. Heuristic are derivative-free direct search methods, in that the modified Nelder–Mead simplex method (Barton and Ivey, 1991; Tomick *et al.*, 1995), the modified Box's complex method (Azadivar, 1992), simulated annealing (Haddock and Mittenthal, 1992), genetic algorithm (Yunker and Tew, 1994; Tompkins and Azadivar, 1995), and evolutionary algorithm (Nissen and Biethahn, 1995) are typical examples. Stochastic approximation methods are based on the root-finding algorithms for a function with random noise (Robbins and Monro, 1951; Kiefer and Wolfowitz, 1952). Some related studies include Azadivar and Talavage (1980), Glynn (1986), Meketon (1987), and Andradottir (1990, 1995, 1996). The RSM is a sequential procedure that fits metamodels between responses and experimental points and attempts to find search directions using the gradient of these metamodels (Montgomery, 1991; Law and Kelton, 1991). Finally, the gradient-based nonlinear programming methods estimate a usable direction and then perform a line

*Correspondence: Chiang Kao, Graduate School of Industrial Management, National Cheng Kung University, Tainan, Taiwan, Republic of China*

search along this direction. The steepest descent and Davidon's variable metric method (Reklaitis *et al.*, 1983) are of this kind.

Several comparative studies have examined the advantages and disadvantages of these solution methods. Azadivar and Talavage (1980) compare a type of stochastic approximation algorithm with RSM and find that the former outperforms the latter, especially for complex functions with sharp ridges and flat valleys (Azadivar, 1992). To speed up the convergence of the RSM, Safizadeh and Signorile (1994) use a quasi-Newton method in the neighborhood of the optimal point. The classical stochastic approximation algorithms such as those proposed by Robbins and Monro (1951) and Kiefer and Wolfowitz (1952) converge extremely slowly when the objective function is very flat and they even fail to converge when the objective function is steep (Azadivar, 1992; Andradottir, 1990; Andradottir, 1995). Meketon (1987) proposes a single-run technique to improve the efficiency of the classical stochastic approximation, and Suri and Leung (1989) verify its correctness by using an $M/M/1$ test example. Andradottir (1995) proposes a new stochastic approximation algorithm with varying bounds and presents empirical results to show that the proposed algorithm sometimes converges much faster than the classical stochastic approximation algorithm. Recently, Andradottir (1996) proposes a scaled stochastic approximation algorithm. Experimental results show that, when embedded with Kesten's acceleration, this method converges faster than the classical stochastic approximation algorithm.

Most of the existing solution methods use only the first derivatives because higher-order derivatives are very difficult to obtain in an efficient manner (Fu, 1994a,b). However, higher-order derivatives are more informative for solving the optimization problems in simulation. Therefore, in this study we borrow the idea of quasi-Newton methods which use first derivatives to approximate second derivatives. We propose a modified quasi-Newton method and provide several modifications and suggestions for parameter settings derived from empirical tests. To demonstrate the validity of the proposed method, a computer program coded in Fortran is implemented on a PC to solve two $M/M/1$ queueing problems.

## QUASI-NEWTON METHODS

Quasi-Newton methods are by far the most efficient methods for solving unconstrained nonlinear programming problems in deterministic environments (Bazaraa *et al.*, 1993). This type of method uses the first-order derivative to calculate a direction matrix $Z$ to approximate the inverse matrix of the Hessian matrix. Throughout the paper, suppose minimization is assumed. Starting with a given point $x^{(0)}$, successive points are derived according to the following iterative relationship:

$$x^{(k+1)} = x^{(k)} - \lambda^{(k)} Z^{(k)} g^{(k)}$$

where $g^{(k)}$ is the gradient of $f(\cdot)$ evaluated at $x^{(k)}$ and $\lambda^{(k)}$ is a step-length parameter. Based on this iterative formula, the original quasi-Newton method can be summarized as follows:

*Step* 1. Set $s^{(k)} = -Z^{(k)} g^{(k)}$ as the search direction.
*Step* 2. Perform a line search along $s^{(k)}$ to find the minimal point $x^{(k+1)}$, i.e.

$$f(x^{(k+1)}) = \min_{\lambda} f(x^{(k)} + \lambda s^{(k)})$$

*Step* 3. Update $Z^{(k)}$ to obtain $Z^{(k+1)}$ and return to Step 1.

Clearly, the key point lies in updating $Z^{(k)}$. Let $\Delta Z^{(k)} = Z^{(k+1)} - Z^{(k)}$. Instead of updating $Z^{(k)}$ directly, one could also update $\Delta Z^{(k)}$ and calculate $Z^{(k+1)} = Z^{(k)} + \Delta Z^{(k)}$ subsequently. Two well-known formulae for updating $\Delta Z^{(k)}$ are the Davidon–Fletcher–Powell (DFP) formula and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula. Specifically, the DFP formula is:

$$\Delta Z^{(k)} = \frac{\Delta x^{(k)} (\Delta x^{(k)})^T}{(\Delta x^{(k)})^T \Delta g^{(k)}} - \frac{Z^{(k)} \Delta g^{(k)} (\Delta g^{(k)})^T (Z^{(k)})^T}{(\Delta g^{(k)})^T Z^{(k)} \Delta g^{(k)}} = A^{(k)} - B^{(k)}$$

where

$$\Delta x^{(k)} = x^{(k+1)} - x^{(k)} = \lambda^{(k)} s^{(k)}$$
$$\Delta g^{(k)} = g^{(k+1)} - g^{(k)}$$

$A^{(k)}$ and $B^{(k)}$ are rank-one symmetric matrices, and hence $\Delta Z^{(k)}$ is a rank-two symmetric matrix. The BFGS formula can be expressed as:

$$\Delta Z^{(k)} = [1 + \frac{(\Delta g^{(k)})^T Z^{(k)} \Delta g^{(k)}}{(\Delta x^{(k)})^T \Delta g^{(k)}}] \frac{\Delta x^{(k)} (\Delta x^{(k)})^T}{(\Delta x^{(k)})^T \Delta g^{(k)}} - \frac{\Delta x^{(k)} (\Delta g^{(k)})^T Z^{(k)} + Z^{(k)} \Delta g^{(k)} (\Delta x^{(k)})^T}{(\Delta x^{(k)})^T \Delta g^{(k)}}$$

Notably, these two methods are dual, or complementary, to each other. From the theoretical point of view, they have the same convergence property for solving unconstrained nonlinear programming problems. However, empirical results indicate that the BFGS method is more robust and efficient than the DFP method when an inexact line search is applied (Reklaitis *et al.*, 1983). Hence, in this study we develop a modified quasi-Newton method using the BFGS formula for optimization in simulation.

## A MODIFIED QUASI-NEWTON METHOD

Quasi-Newton methods have a sound theoretical basis and have yielded promising empirical results for solving unconstrained nonlinear programming problems in deterministic environments (Reklaitis *et al.*, 1983; Bazaraa *et al.*, 1993). For optimization problems in simulation the objective function is a stochastic function of deterministic decision variables. If we apply the sheer quasi-Newton methods to solve without modifications, many intractable problems may arise. Firstly, inadequate search directions may be derived due to the stochastic nature of the problem. Secondly, the line search procedure may terminate at non-minimal points. In other words, such an approach may take many simulation runs and much time only to produce a wrong answer. To solve the optimization problems in simulation efficiently and effectively, the stochastic behavior of the simulation system has to be considered. To cope with these difficulties, the traditional quasi-Newton methods have to be modified.

### *Search direction*

A major step in solving the optimization problem is to find a direction with descent property. The quality of the search direction seriously affects the rate of convergence. Since the responses of the simulation systems are stochastic, the chance of finding the correct response at an experimental point with one single replication is very low. Law and Kelton (1991) suggest that at least three replications should be conducted at any experimental point in simulation studies. In fact, replication is one basic principle of experimental design (Montgomery, 1991). It is expected that more replications would result in an average direction of better quality. Nevertheless, more replications implies longer execution time. Furthermore, inaccuracy in the subsequent operations such as line searches in the stochastic environment would make the effort devoted to accurate estimate of the search direction in vain. Therefore, a suitable number of replications has to be decided taking into account the quality of the search direction and the efficiency in deriving the direction. In this study, five choices for the number of replications $n_{repl}$ at every experimental point, namely, 1, 3, 5, 10, and 30, are considered.

### *Gradient approximation*

There are many gradient approximation methods for simulation systems (Azadivar, 1992), for instance, infinitesimal perturbation analysis (IPA), frequency domain analysis (FDA), and likelihood ratio estimators (LRE). However, these methods all have shortcomings in specific applications (Azadivar, 1992). More important, they are not applicable to all forms of simulation studies (Safizadeh and Singorile, 1994). In this study, we employ the simplest gradient approximation method, namely, the central finite difference (CFD) (Reklaitis *et al.*, 1983):

$$g(x_j) = \frac{f(x + \varepsilon_{CFD} e_j) - f(x - \varepsilon_{CFD} e_j)}{2\varepsilon_{CFD}}$$

where $e_j$ is the unit vector in the $j$th coordinate direction. For reasons similar to those stated in the preceding subsection, a proper choice of $\varepsilon_{CFD}$ will result in a gradient of better quality. To find a good value for $\varepsilon_{CFD}$, four values, i.e. $10^{-4}$, $10^{-3}$, $10^{-2}$, and $10^{-1}$ are examined.

*Line search*

Line search is a basic step in many algorithms for solving nonlinear programming problems. There are many line search methods, some well-known ones include golden section, Fibonacci series, cubic search, Davis–Swann–Campey (DSC) method, and Powell search (Bazaraa *et al.*, 1993), among which the DSC–Powell method is considered by many scholars to be the best (Himmelblau, 1972). In this study, we employ an inexact DSC–Powell search in one-dimensional optimization.

Usually many iterations are required along each search direction. To alleviate the computation burden, we limit the number of iterations in each search direction to 20. Many studies have proved that the BFGS method is able to converge to the minimal point even if the line searches are inexact (Fletcher, 1987). Hence, this modification is expected to result in a more efficient algorithm.

*Stopping rule*

In deterministic optimization, the general criteria for terminating the search are as follows (Reklaitis *et al.*, 1983):

$$(i)\ \left|\frac{f(x^{(k+1)}) - f(x^{(k)})}{f(x^{(k)})}\right| < \varepsilon_1$$

$$(ii)\ \left|\frac{x_i^{(k+1)} - x_i^{(k)}}{x_i^{(k)}}\right| < \varepsilon_2,\ i = 1,2,\ldots,n$$

$$(iii)\ \left|g_i(x^{(k+1)})\right| < \varepsilon_3,\ i = 1,2,\ldots,n$$

where $\varepsilon_1$, $\varepsilon_2$, and $\varepsilon_3$ are prespecified small values, and $n$ is the dimension of the problem. If $|f(x^{(k)})| < \varepsilon_1$, $(i)$ is replaced by $|f(x^{(k+1)}) - f(x^{(k)})| < \varepsilon_1$; if $|x_i^{(k)}| < \varepsilon_2$, $(ii)$ is replaced by $|x_i^{(k+1)} - x_i^{(k)}| < \varepsilon_2$, $i = 1,2,\ldots,n$. Safizadeh and Signorile (1994) employ $f(x^{(k)}) - f(x^{(k+1)}) \leqslant \varepsilon_1$ as the stopping rule in their study. However, since $f(\cdot)$ are stochastic in nature and the precise $g(\cdot)$ are difficult to obtain, they may terminate at points which are far from the optimal point. Evidently, these stopping rules are inadequate in the stochastic environment. To deal with the stochastic nature, in this study, we develop other stopping rules based on statistical theory.

Due to the random characteristic of simulation systems, variations in response in the optimization process should be considered. From a statistical point of view, several replications at each experimental point are required to test whether the difference in responses between two consecutive experimental points is significant. That is, we wish to test

$$H_0: \mu^{(k)} \leqslant \mu^{(k+1)}$$
$$H_1: \mu^{(k)} > \mu^{(k+1)}$$

where $\mu^{(k)}$ and $\mu^{(k+1)}$ are the population means of the responses at the $k$th and the $(k+1)$th experimental points, respectively. Assuming the population is normally distributed and the variances of the two responses are identical, then an appropriate test statistic to use for comparing two means in completely randomized design is (Montgomery, 1991)

$$t_0 = \frac{\overline{R^{(k)}} - \overline{R^{(k+1)}}}{S_P\sqrt{\frac{2}{n_{repl}}}}$$

$$S_p^2 = \frac{(n_{repl} - 1)S_k^2 + (n_{repl} - 1)S_{k+1}^2}{(n_{repl} - 1) + (n_{repl} - 1)} = \frac{S_k^2 + S_{k+1}^2}{2}$$

where $\overline{R^{(k)}} = \dfrac{\sum_{i=1}^{n_{repl}} f_i(x^{(k)})}{n_{repl}}$ and $\overline{R^{(k+1)}} = \dfrac{\sum_{i=1}^{n_{repl}} f_i(x^{(k+1)})}{n_{repl}}$ are the sample means, $S_p^2$ is an estimator of the common variances $\sigma_k^2 = \sigma_{k+1}^2 = \sigma^2$, and $S_k^2$ and $S_{k+1}^2$ are the two individual sample variances. Since $x^{(k)}$ and $x^{(k+1)}$ are usually very close when $x^{(0)}$ is not too far from the optimal point $x^*$, the assumption that $\sigma_k^2 = \sigma_{k+1}^2$ seems appropriate. To determine whether the null hypothesis $H_0$: $\mu^{(k)} \leqslant \mu^{(k+1)}$ should be rejected, we examine if $t_0 > t_{\alpha,2(n_{repl}-1)}$ holds, where $\alpha$ is a prespecified significance level and $2(n_{repl} - 1)$ is the degree of freedom. If $H_0$ is rejected, then $\mu^{(k+1)}$ is significantly smaller than $\mu^{(k)}$, and we should proceed to the next point $x^{(k+1)}$.

When $n_{repl} = 1$, the hypothesis testing cannot be performed. In this case the stopping rule is set to

$$|R^{(k)} - R^{(k+1)}| < \varepsilon_{stop}$$

A proper choice for $\varepsilon_{stop}$ is to be determined in this study. The values considered are 1, $10^{-1}$, $10^{-2}$, $10^{-3}$, $10^{-4}$, and $10^{-5}$.

In sum, the stopping rules to be adopted in this study can be summarized by the following two rules:

*Rule* 1. For $n_{repl} = 1$, if $|R^{(k)} - R^{(k+1)}| < \varepsilon_{stop}$, then terminate.
*Rule* 2. For $n_{repl} > 1$, if the null hypothesis $H_0$: $\mu^{(k)} \leqslant \mu^{(k+1)}$ is not rejected, then terminate. The significance level $\alpha$ is set to 5%.

With these modifications, the modified quasi-Newton method proposed in this study can be summarized as follows:

*Step 1.* Select an initial experimental point $x^{(0)}$, determine $n_{repl}$ and $\varepsilon_{CFD}$, and let $k = 0$. Set $Z^{(0)} = I$.
*Step 2.* Estimate the gradient $g(x^{(k)})$ using CFD from $n_{repl}$ replications.
*Step 3.* Set $s^{(k)} = -Z^{(k)}g^{(k)}$.
*Step 4.* Perform an inexact DSC–Powell line search along $s^{(k)}$ to find the next experimental point $x^{(k+1)}$.
*Step 5.* If two consecutive experimental points satisfy the stopping rules, then a presumably optimal solution to the simulation system is found; otherwise, continue on to Step 6.
*Step 6.* Update $Z^{(k)}$ according to the BFGS formula. Increment $k$ by one and return to Step 2.

The next problem is to determine the appropriate values for $n_{repl}$, $\varepsilon_{CFD}$, and $\varepsilon_{stop}$ which will result in an efficient and effective quasi-Newton method for optimization problems in simulation.

## PARAMETER DETERMINATION

To demonstrate the validity of the proposed method and to determine the parameter settings, a computer program is coded in Fortran to solve three standard test functions which are commonly used in unconstrained nonlinear program (Himmelblau, 1972) with a random term $\varepsilon$ being appended. The program is executed on a 586PC under DOS6.2 with double precision.
Function 1 (Rosenbrock function):

$$F(x) = h(x) + \varepsilon = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + \varepsilon$$

Initial point $x^{(0)} = (-1.2,1.0)$, $E[F(x^{(0)})] = 24.2$
Optimal point $x^* = (1.0,1.0)$, $E[F(x^*)] = 0$
Function 2 (Miele function):

$$F(x) = h(x) + \varepsilon = (e^{x_1} - x_2)^4 + 100(x_2 - x_3)^6 + \tan^4(x_3 - x_4) + x_1^8 + (x_4 - 1)^2 + \varepsilon$$

Initial point $x^{(0)} = (1.0,2.0,2.0,2.0)$, $E[F(x^{(0)})] = 2.2662$
Optimal point $x^* = (0,1.0,1.0,1 \pm n\pi)$, $E[F(x^*)] = 0$
Function 3 (Extended Rosenbrock function):

$$F(x) = h(x) + \varepsilon = \sum_{k=1}^{5} [100(x_{2k} - x_{2k-1}^2)^2 + (1 - x_{2k-1})^2] + \varepsilon$$

Table 1 Mean responses of the first test function from 1000 independent experiments for each parameter setting

| | $n_{repl} = 1$ | | | | $n_{repl} = 3$ | | | | $n_{repl} = 5$ | | | | $n_{repl} = 10$ | | | | $n_{repl} = 30$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\varepsilon_{CFD}$ | | | | $\varepsilon_{CFD}$ | | | | $\varepsilon_{CFD}$ | | | | $\varepsilon_{CFD}$ | | | | $\varepsilon_{CFD}$ | | | |
| $\sigma$ | 1.E-4 | 0.0001 | 0.01 | 0.1 | 1.E-4 | 0.001 | 0.01 | 0.1 | 1.E-4 | 0.001 | 0.01 | 0.1 | 1.E-4 | 0.001 | 0.01 | 0.1 | 1.E-4 | 0.001 | 0.01 | 0.1 |
| 0.01 | 0.882 | 0.198 | 0.194 | 0.166 | 0.616 | 0.292 | 0.193 | 0.190 | 0.437 | 0.269 | 0.194 | 0.189 | 0.321 | 0.269 | 0.194 | 0.189 | 0.231 | 0.213 | 0.194 | 0.184 |
| 0.05 | 0.760 | 0.435 | 0.251 | 0.235 | 1.131 | 0.306 | 0.267 | 0.206 | 1.047 | 0.255 | 0.320 | 0.198 | 1.104 | 0.233 | 0.196 | 0.193 | 0.977 | 0.234 | 0.194 | 0.191 |
| 0.10 | 0.736 | 1.014 | 0.309 | 0.273 | 1.128 | 0.651 | 0.249 | 0.226 | 1.014 | 0.417 | 0.396 | 0.245 | 1.050 | 0.298 | 0.290 | 0.208 | 1.162 | 0.223 | 0.218 | 0.195 |
| 0.15 | 0.803 | 1.079 | 0.318 | 0.293 | 1.078 | 0.870 | 0.297 | 0.238 | 1.020 | 0.789 | 0.357 | 0.233 | 1.068 | 0.451 | 0.320 | 0.217 | 1.027 | 0.261 | 0.324 | 0.205 |
| 0.20 | 0.853 | 1.140 | 0.382 | 0.343 | 1.065 | 1.193 | 0.325 | 0.256 | 1.310 | 0.923 | 0.311 | 0.241 | 1.191 | 0.699 | 0.359 | 0.227 | 1.010 | 0.409 | 0.311 | 0.214 |
| 0.25 | 0.953 | 1.151 | 0.447 | 0.401 | 1.402 | 1.264 | 0.330 | 0.290 | 1.067 | 1.123 | 0.284 | 0.242 | 1.097 | 0.825 | 0.393 | 0.241 | 1.091 | 0.477 | 0.338 | 0.219 |
| 0.30 | 0.956 | 1.323 | 0.442 | 0.411 | 1.324 | 1.611 | 0.297 | 0.286 | 1.246 | 1.158 | 0.290 | 0.254 | 1.154 | 1.063 | 0.262 | 0.237 | 1.112 | 0.565 | 0.278 | 0.230 |
| 0.35 | 1.116 | 1.459 | 0.487 | 0.476 | 1.315 | 1.474 | 0.309 | 0.318 | 1.266 | 1.308 | 0.351 | 0.240 | 1.038 | 1.027 | 0.330 | 0.267 | 1.071 | 0.709 | 0.332 | 0.230 |
| 0.40 | 1.262 | 1.305 | 0.577 | 0.565 | 1.334 | 1.587 | 0.345 | 0.299 | 1.212 | 1.404 | 0.405 | 0.340 | 0.840 | 1.277 | 0.297 | 0.295 | 1.038 | 0.760 | 0.350 | 0.240 |
| 0.45 | 1.187 | 1.353 | 0.664 | 0.563 | 1.298 | 1.566 | 0.350 | 0.318 | 1.240 | 1.423 | 0.401 | 0.311 | 1.199 | 1.247 | 0.297 | 0.261 | 1.098 | 0.896 | 0.300 | 0.240 |
| 0.50 | 1.335 | 1.525 | 0.826 | 0.536 | 1.571 | 1.427 | 0.341 | 0.434 | 1.158 | 1.384 | 0.303 | 0.309 | 1.123 | 1.189 | 0.320 | 0.296 | 0.991 | 1.141 | 0.268 | 0.249 |

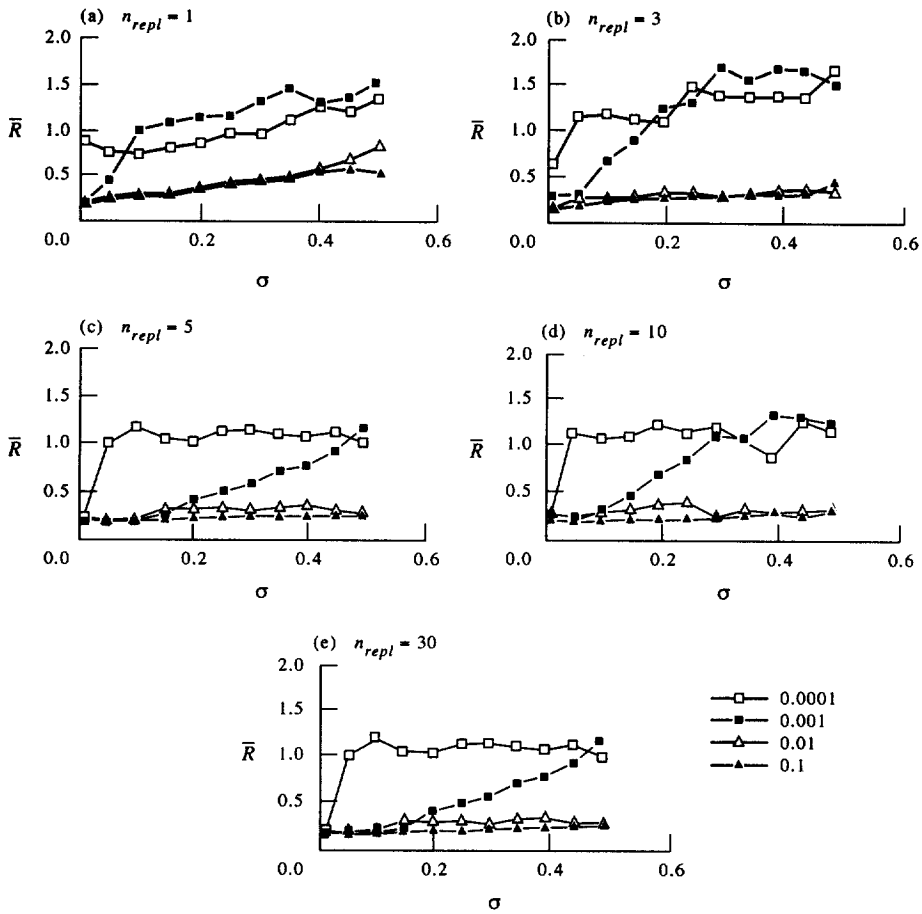


Fig. 1. Mean responses of the first test function for four choices of $\varepsilon_{CFD}$ when (a) $n_{repl} = 1$, (b) $n_{repl} = 3$, (c) $n_{repl} = 5$, (d) $n_{repl} = 10$, and (e) $n_{repl} = 30$.

Initial point $x^{(0)} = (-1.2, 1.0, 1.0, \ldots, 1.0)$, $E[F(x^{(0)})] = 24.2$
Optimal point $x^* = (1.0, 1.0, \ldots, 1.0)$, $E[F(x^*)] = 0$

In these three test functions, $\varepsilon$ is normally distributed with mean 0 and variance $\sigma^2$ in which $\sigma^2$ represents the variation of simulation systems. Intuitively, it is more difficult to converge to the minimum point when $\sigma$ is large. To examine the effect of different levels of the random noise, 11 values of $\sigma$ have been tested for each function. By applying the proposed method described in the preceding section, different points which are "conjectured" to be the minima are derived when different values of $n_{repl}$, $\varepsilon_{CFD}$, and $\sigma$ are used. Due to the random nature of the test functions, the procedure may not terminate at the same point even for the same choices of $n_{repl}$, $\varepsilon_{CFD}$, and $\sigma$. Therefore, for each function the whole process is repeated 1000 times, and the average is used to represent the presumably true result. The results of these three functions are very consistent in recommend parameter settings, hence we only list the results of Function 1. Table 1 shows the mean responses at the termination points for different values of $\sigma$, $n_{repl}$, and $\varepsilon_{CFD}$. Note that for $n_{repl} = 1$, the values are the minimal values for different values of $\varepsilon_{stop}$, which will be discussed later.

To determine which value of $\varepsilon_{CFD}$ is more appropriate for the modified quasi-Newton method, Fig. 1 depicts the mean responses of the Rosenbrock function for the four choices of $\varepsilon_{CFD}$. Visually, the responses for $\varepsilon_{CFD} = 0.01$ and $\varepsilon_{CFD} = 0.1$ are very stable as $\sigma$ increases regardless whether $n_{repl}$ is equal to 1, 3, 5, 10, or 30. A Duncan's multiple range test (Montgomery, 1991) also indicates that there is no difference between these two $\varepsilon_{CFD}$ in a statistical sense, while this group of $\varepsilon_{CFD}$ is significantly different from the others. Since in most cases the response for $\varepsilon_{CFD} = 0.1$ is better than that for $\varepsilon_{CFD} = 0.01$ (referring to Table 1), we recommend $\varepsilon_{CFD} = 0.1$.

To determine an appropriate value for $n_{repl}$, we focus on $\varepsilon_{CFD} = 0.1$. The Duncan's test reveals that there exists no significant difference among $n_{repl} = 5$, 10, and 30. However, when we expand $\sigma$ from

0.5 to 5, $n_{repl} = 5$ becomes significantly different from the group of $n_{repl} = 10$ and 30 as is evident from the results shown in Table 2. Since there is no difference between $n_{repl} = 10$ and 30, $n_{repl} = 10$ is obviously more efficient.

When $n_{repl} = 1$ the $t$-test cannot be performed, therefore this study uses $\varepsilon_{stop}$ to control the termination as discussed in the preceding section. Table 3 summarizes the results of six choices of $\varepsilon_{stop}$ under different values of $\sigma$. Across each row, the minimal value is denoted by "*". Evidently, $\varepsilon_{stop} = 1$ is the best choice.

In sum, from the numerical results of these three test functions, we recommend the following for parameter settings in the modified quasi-Newton method:

(i) Gradient approximation by central finite difference: $\varepsilon_{CFD} = 0.1$.
(ii) Number of replications at each experimental point: $n_{repl} = 10$.
(iii) Stopping rule for $n_{repl} = 1$: $|R^{(k)} - R^{(k+1)}| < \varepsilon_{stop} = 1$.

Note that identical parameter settings may produce different results in stochastic simulation. Next we shall apply the proposed method with this set of parameters to solve two examples.

Table 2 Mean responses of the first test function from 1000 independent experiments for $\varepsilon_{CFD} = 0.1$ and some larger values of $\sigma$

| $\sigma$ | $n_{repl} = 1$ | $n_{repl} = 3$ | $n_{repl} = 5$ | $n_{repl} = 10$ | $n_{repl} = 30$ |
|---|---|---|---|---|---|
| 0.01 | 0.166 | 0.190 | 0.189 | 0.189 | 0.184 |
| 0.05 | 0.235 | 0.206 | 0.198 | 0.193 | 0.191 |
| 0.10 | 0.273 | 0.226 | 0.245 | 0.208 | 0.195 |
| 0.15 | 0.293 | 0.238 | 0.233 | 0.217 | 0.205 |
| 0.20 | 0.343 | 0.256 | 0.241 | 0.227 | 0.214 |
| 0.25 | 0.401 | 0.290 | 0.242 | 0.241 | 0.219 |
| 0.30 | 0.411 | 0.286 | 0.254 | 0.237 | 0.230 |
| 0.35 | 0.475 | 0.318 | 0.240 | 0.267 | 0.230 |
| 0.40 | 0.565 | 0.299 | 0.340 | 0.295 | 0.240 |
| 0.45 | 0.563 | 0.318 | 0.311 | 0.261 | 0.240 |
| 0.50 | 0.536 | 0.434 | 0.309 | 0.296 | 0.249 |
| 0.60 | 0.596 | 0.367 | 0.333 | 0.265 | 0.260 |
| 0.70 | 0.651 | 0.400 | 0.389 | 0.281 | 0.252 |
| 0.80 | 0.742 | 0.404 | 0.321 | 0.321 | 0.281 |
| 0.90 | 0.810 | 0.336 | 0.364 | 0.336 | 0.273 |
| 1.00 | 0.816 | 0.354 | 0.387 | 0.365 | 0.307 |
| 2.00 | 1.652 | 0.433 | 0.406 | 0.403 | 0.441 |
| 3.00 | 2.493 | 0.527 | 0.546 | 0.277 | 0.417 |
| 4.00 | 3.598 | 0.862 | 0.598 | 0.437 | 0.364 |
| 5.00 | 4.617 | 1.158 | 0.815 | 0.511 | 0.380 |

Table 3 Mean responses of the first test function for $n_{repl} = 1$ and $\varepsilon_{CFD} = 0.1$ from 1000 independent experiments

| $\sigma \backslash \varepsilon_{stop}$ | 1 | 0.1 | 0.01 | 0.001 | 0.0001 | 0.00001 |
|---|---|---|---|---|---|---|
| 0.01 | 0.195 | 0.191 | 0.178 | 0.170 | 0.166* | 0.166 |
| 0.05 | 0.244 | 0.239 | 0.239 | 0.235* | 0.238 | 0.238 |
| 0.10 | 0.273* | 0.276 | 0.282 | 0.275 | 0.285 | 0.285 |
| 0.15 | 0.293* | 0.307 | 0.395 | 0.383 | 0.384 | 0.384 |
| 0.20 | 0.379 | 0.343* | 0.410 | 0.371 | 0.373 | 0.373 |
| 0.25 | 0.401* | 0.434 | 0.431 | 0.404 | 0.402 | 0.401 |
| 0.30 | 0.411* | 0.485 | 0.602 | 0.463 | 0.463 | 0.463 |
| 0.35 | 0.540 | 0.568 | 0.518 | 0.476* | 0.480 | 0.480 |
| 0.40 | 0.565* | 0.565 | 0.602 | 0.605 | 0.568 | 0.566 |
| 0.45 | 0.563* | 0.621 | 0.644 | 0.608 | 0.605 | 0.605 |
| 0.50 | 0.574 | 0.536* | 0.591 | 0.578 | 0.577 | 0.577 |
| 0.60 | 0.596* | 0.733 | 0.768 | 0.770 | 0.757 | 0.757 |
| 0.70 | 0.651* | 0.768 | 0.844 | 0.791 | 0.785 | 0.785 |
| 0.80 | 0.742* | 1.707 | 1.726 | 1.725 | 1.747 | 1.719 |
| 0.90 | 0.810* | 0.897 | 0.928 | 0.947 | 0.958 | 0.958 |
| 1.00 | 0.816* | 1.071 | 1.066 | 1.040 | 1.058 | 1.058 |
| 2.00 | 1.652* | 1.826 | 1.923 | 1.865 | 1.861 | 1.861 |
| 3.00 | 2.493* | 2.774 | 2.775 | 2.882 | 2.924 | 2.924 |
| 4.00 | 3.598* | 15.783 | 3.925 | 4.060 | 4.079 | 4.079 |
| 5.00 | 4.617* | 4.861 | 5.099 | 5.112 | 5.399 | 5.399 |

* Minimum value across the row.

## EXAMPLES

Finding a real world complex problem to test the proposed algorithm is rather difficult. The problem lies on that analytical solutions for justification do not exist for real world problems. Therefore, following Andradottir (1996), this study selects two $M/M/1$ problems quite often used in stochastic optimization for illustration. These two examples are similar to that studied by Suri and Leung (1989), L'Ecuyer *et al.* (1994), and Andradottir (1995).

*Example 1*

$$\min_{\theta\in(0,1)} R(\theta) = \frac{\alpha}{\theta} + \beta L(\theta)$$

where $L(\theta)$ denotes the mean number of customers in an $M/M/1$ queueing system with arrival rate 1, $\theta$ is the mean service time which is to be determined, and $\alpha$ and $\beta$ are prespecified costs. The analytical solution of this problem is readily obtainable from queueing theory (Gross and Harris, 1985):

$$\theta^* = \frac{\sqrt{\alpha}}{\sqrt{\alpha} + \sqrt{\beta}}$$

In Andradottir (1996), four sets of $(a,\beta)$ values, i.e. (1,1), (1,10), (10,1), and (10,10) are considered. In this paper, $(\alpha,\beta) = (10,1)$ is selected. The corresponding theoretical optimal mean service time is $\theta^* = 0.760$ and the optimal response is $R(\theta^*) = 16.325$.

*Example 2*

$$\min_{0\leq\theta_1,\theta_2\leq 1} R(\theta_1,\theta_2) = \frac{\alpha_1}{\theta_1} + \beta_1 L(\theta_1) + \frac{\alpha_2}{\theta_2} + \beta_2 L(\theta_2) + \frac{\gamma}{\theta_1\theta_2}$$

where $L(\cdot)$ has the same meaning as that defined in Example 1, $\theta_1$ and $\theta_2$ are the mean service times which are to be determined, and $\alpha_1$, $\alpha_2$, $\beta_1$, $\beta_2$, and $\gamma$ are prespecified costs. There are several choices for $(\alpha_1, \alpha_2,\beta_1,\beta_2,\gamma)$, and in this paper they are set to (1,1,1,1,10). Accordingly, the theoretical optimal mean service times are $\theta_1^* = \theta_2^* = 0.787$ and the optimal response is $R(\theta_1^*,\theta_2^*) = 26.083$.

To eliminate the bias caused by the initial conditions, we delete the first 200 observations in conducting the optimization procedure. At each experimental point, the simulation terminates when a number of customers $n_{cus}$ have completed services. Apparently, as $n_{cus}$ increases, the variance of the estimated mean number of customers in the system decreases. In other words, $n_{cus}$ serves as an indicator for the variance. In this study, 11 values for $n_{cus}$ have been selected somewhat arbitrarily. After the whole process terminates, the total number of service completions $n_{compl}$ can be accumulated. For each value of $n_{cus}$, the experiment is repeated 10 times to construct confidence intervals.

Tables 4 and 5 list the numerical results from this study and that of Andradottir (1996), including $n_{compl}$, 90% confidence interval for $\theta^*$, and the standard error for the estimates of $\theta^*$. As expected, larger $n_{compl}$ converges to a better point and has narrower confidence interval and smaller standard error. Termination of the optimization process is controlled by mean response values. In Example 1, we find that for $n_{compl}$ = 3,000,000 and 4,000,000, the averages of the estimated optimal service times are 0.769 and 0.763, respectively, which are very close to the true optimal service time 0.760. Moreover, the associated standard errors are very small and the 90% confidence intervals really cover the true value. In Example 2, for $n_{compl}$ = 9,696,000 and 12,870,000, the estimated service times are almost the same as the real optimal service time. The 90% confidence intervals are also very narrow.

Compared with the scaled stochastic approximation of Andradottir (1996), the proposed method converges much faster to the true optimal design point. In Example 1, the 90% confidence interval constructed for 10,000 customers is $0.738 \pm 0.0323$, which is even better than that constructed by the scaled stochastic approximation algorithm for 10,000,000 customers, namely $0.784 \pm 0.0441$. In

Example 2, for $n_{compl}$ equal to 4,490,000, the optimal solution estimated by the proposed method is already better than that estimated by the scaled stochastic approximation method with $n_{compl}$ as large as 10,000,000. In Tables 4 and 5 the estimates, which are less bias and have smaller variance than those obtained from the scaled stochastic approximation algorithm, are denoted by "*".

These two examples indicate that the modified quasi-Newton method proposed in this paper is able to solve optimization problems in simulation more efficiently and effectively.

## CONCLUSION

Quasi-Newton methods have a sound theoretical basis and have demonstrated to be promising for solving deterministic unconstrained nonlinear programming problems. In this paper, a stochastic version of a quasi-Newton method is devised for solving optimization problems in simulation. To adapt to the randomness of stochastic functions, we have made several modifications. On the basis of

Table 4 Numerical results of the first $M/M/1$ example from 10 experiments

| Method | $n_{compl}$ | 90% confidence interval | Standard error |
|---|---|---|---|
| Modified quasi-Newton | 2000 | $0.733 \pm 0.0536$ | 0.1570 |
| | 10,000 | $0.738 \pm 0.0323$* | 0.0974 |
| | 20,000 | $0.739 \pm 0.0318$* | 0.0928 |
| | 100,000 | $0.741 \pm 0.0147$* | 0.0512* |
| | 200,000 | $0.742 \pm 0.0174$* | 0.0691* |
| | 400,000 | $0.744 \pm 0.0146$* | 0.0526* |
| | 1,000,000 | $0.744 \pm 0.0144$* | 0.0457* |
| | 2,000,000 | $0.746 \pm 0.0095$* | 0.0465* |
| | 2,400,000 | $0.749 \pm 0.0096$* | 0.0396* |
| | 3,000,000 | $0.769 \pm 0.0094$* | 0.0480* |
| | 4,000,000 | $0.763 \pm 0.0086$* | 0.0412* |
| Scaled stochastic approximation | 10,000,000 | $0.784 \pm 0.0441$ | 0.0763 |
| Classical stochastic approximation | 10,000,000 | $0.945 \pm 0.0556$ | 0.2060 |

* Better than scaled stochastic approximation.

Table 5 Numerical results of the second $M/M/1$ example from 10 experiments

| Method | $n_{compl}$ | Dimension | 90% confidence interval |
|---|---|---|---|
| Modified quasi-Newton | 12,680 | 1 | $0.816 \pm 0.0258$ |
| | | 2 | $0.773 \pm 0.0255$ |
| | 58,400 | 1 | $0.762 \pm 0.0351$ |
| | | 2 | $0.796 \pm 0.0273$ |
| | 99,600 | 1 | $0.780 \pm 0.0149$ |
| | | 2 | $0.803 \pm 0.0181$ |
| | 206,400 | 1 | $0.776 \pm 0.0107$ |
| | | 2 | $0.779 \pm 0.0061$ |
| | 477,000 | 1 | $0.782 \pm 0.0067$ |
| | | 2 | $0.775 \pm 0.0091$ |
| | 904,000 | 1 | $0.788 \pm 0.0219$ |
| | | 2 | $0.776 \pm 0.0149$ |
| | 1,780,000 | 1 | $0.780 \pm 0.0276$ |
| | | 2 | $0.790 \pm 0.0254$ |
| | 4,490,000 | 1 | $0.790 \pm 0.0018$* |
| | | 2 | $0.792 \pm 0.0016$* |
| | 8,620,000 | 1 | $0.783 \pm 0.0060$* |
| | | 2 | $0.787 \pm 0.0053$* |
| | 9,696,000 | 1 | $0.790 \pm 0.0050$* |
| | | 2 | $0.788 \pm 0.0057$* |
| | 12,870,000 | 1 | $0.789 \pm 0.0080$* |
| | | 2 | $0.789 \pm 0.0071$* |
| Scaled stochastic approximation | 10,000,000 | 1 | $0.780 \pm 0.0129$ |
| | | 2 | $0.787 \pm 0.00282$ |
| Classical stochastic approximation | 10,000,000 | 1 | $0.990 \pm 0$ |
| | | 2 | $0.696 \pm 0.274$ |

* Better than scaled stochastic approximation.

some numerical experiments, we recommend that 10 replications be conducted at each experimental point. In gradient approximation, the perturbation for the central finite difference is set to 0.1. Under the stochastic environment there is not much meaning to conduct exact line searches, therefore, an inexact DSC–Powell method is adopted in this study. Most important, the criterion for terminating the search is based on the result of statistical hypothesis testing rather than on simply examining the mean responses.

Numerical results from two $M/M/1$ queueing problems indicate that the modified quasi-Newton method devised in this study is able to solve the optimization problems in simulation successfully. Furthermore, the proposed method is more efficient than two versions of a stochastic approximation method in solving these two queueing problems. The modified quasi-Newton method is a robust and efficient method which deserves further investigation.

*Acknowledgement*—The authors are grateful to the valuable suggestions of Professor B. Schmidt and an anonymous referee which makes this paper more readable.

## REFERENCES

Andradottir, S. (1990) A new algorithm for stochastic optimization. *Proceedings of the 1990 Winter Simulation Conference*, pp. 364–366.

Andradottir, S. (1995) A stochastic approximation algorithm with varying bounds. *Operations Research* **43**(6), 1037–1048.

Andradottir, S. (1996) A scaled stochastic approximation algorithm. *Management Science* **42**(4), 475–498.

Azadivar, F. (1992) A tutorial on simulation optimization. *Proceedings of the 1992 Winter Simulation Conference*, pp. 198–204.

Azadivar, F. and Talavage, J. (1980) Optimization of stochastic simulation models. *Mathematics and Computers in Simulation* **22**, 231–241.

Barton, R.R. and Ivey, Jr, J.S. (1991) Modifications of the Nelder–Mead simplex method for stochastic simulation optimization. *Proceedings of the 1991 Winter Simulation Conference*, pp. 945–953.

Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (1993) *Nonlinear Programming: Theory and Algorithms*, 2nd Edition. Wiley, New York.

Fletcher, R. (1987) *Practical Methods of Optimization*, 2nd Edition. Wiley, Chichester.

Fu, M.C. (1994a) Optimization via simulation: a review. *Annals of Operations Research* **53**, 199–247.

Fu, M.C. (1994b) A tutorial review of techniques for simulation optimization. *Proceedings of the 1994 Winter Simulation Conference*, pp. 149–156.

Fu, M.C., Hu, J.-Q. and Nagi, R. (1995) Comparison of gradient estimation techniques for queues with non-identical servers. *Computers and Operations Research* **22**(7), 715–729.

Glynn, P.W. (1986) Optimization of stochastic systems. *Proceedings of the 1986 Winter Simulation Conference*, pp. 52–59.

Gross, D. and Harris, C.M. (1985) *Fundamentals of Queueing Theory*, 2nd Edition. Wiley, New York.

Haddock, J. and Mittenthal, J. (1992) Simulation optimization using simulated Annealing. *Computers and Industrial Engineering* **22**(4), 387–395.

Himmelblau, D.M. (1972) *Applied Nonlinear Programming*. McGraw-Hill, New York.

Jacobson, S.H. and Schruben, L.W. (1989) Techniques for simulation response optimization. *Operations Research Letters* **8**(1), 1–9.

Kiefer, J. and Wolfowitz, J. (1952) Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics* **23**, 462–466.

Law, A.M. and Kelton, E.D. (1991) *Simulation Modeling and Analysis*, 2nd Edition. McGraw-Hill, New York.

L'Ecuyer, P., Giroux, N. and Glynn, P.W. (1994) Stochastic optimization by simulation: numerical experiments with the $M/M/1$ queue in steady-state. *Management Science* **40**, 1245–1261.

Meketon, M.S. (1987) Optimization in simulation: a survey of recent results. *Proceedings of the 1987 Winter Simulation Conference*, pp. 58–67.

Montgomery, D.C. (1991) *Design and Analysis of Experiments*, 2nd Edition. Wiley, New York.

Nissen, V. and Biethahn, J. (1995) Determining a good inventory policy with a genetic algorithm. In *Evolutionary Algorithms in Management Applications*. Eds J. Biethahn and V. Nissen, pp. 240–249. Springer, New York.

Reklaitis, G.V., Ravindran, A. and Ragsdell, K.M. (1983) *Engineering Optimization: Methods and Applications*. Wiley, New York.

Ringuest, J.L. (1988) Substitute derivatives in unconstrained optimization: a comparison of finite difference and response surface approximation. *Computers and Operations Research* **15**, 314–352.

Robbins, H. and Monro, S. (1951) A stochastic approximation method. *Annals of Mathematical Statistics* **22**, 400–407.

Safizadeh, M.H. (1990) Optimization in simulation: current issues and the future outlook. *Naval Research Logistics* **37**, 807–825.

Safizadeh, M.H. and Signorile, R. (1994) Optimization of simulation via quasi-Newton methods. *ORSA Journal on Computing* **6**(4), 398–408.

Suri, R. and Leung, Y.T. (1989) Single run optimization of discrete event simulations—an empirical study using the $M/M/1$ queue. *IIE Transactions* **21**(1), 35–49.

Tomick, J.J., Arnold, S.F. and Barton, R.R. (1995) Sample size selection for improved Nelder–Mead performance. *Proceedings of the 1995 Winter Simulation Conference*, pp. 341–345.

Tompkins, G. and Azadivar, F. (1995) Genetic algorithms in optimizing simulated systems. *Proceedings of the 1995 Winter Simulation Conference*, pp. 757–762.

Yunker, J.M. and Tew, J.D. (1994) Simulation optimization by genetic search. *Mathematics and Computers in Simulation* **37**, 17–28.