Proceedings of the 1992 Winter Simulation Conference
ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson

A TUTORIAL ON SIMULATION OPTIMIZATION

Farhad Azadivar

Department of Industrial Engineering
Kansas State University
Manhattan, KS 66502

ABSTRACT

This tutorial discusses the issues and procedures for using simulation as a tool for optimization of stochastic complex systems that are modeled by computer simulation. It is intended to be a tutorial rather than an exhaustive literature search. Its emphasis is mostly on issues that are specific to simulation optimization instead of concentrating on the general optimization and mathematical programming techniques. Even though a lot of effort has been spent to provide a comprehensive overview of the field, still there are methods and techniques that have not been covered and valuable works that may not have been mentioned.

1 INTRODUCTION

Computer simulation has proved to be a very powerful tool in evaluating complex systems. These evaluations are usually in the from of responses to "what if" questions. In recent years the success of computer simulation has been extended to answering "how to" questions as well. "What if" questions demand answers on certain performance measures for a given set of values for the decision variables of the system. "How to" questions, on the other hand, seek optimum values for the decision variables of the system so that a given response or a vector of responses are maximized or minimized. In the past ten years a considerable amount of effort has been expended on simulation optimization procedures that deal with optimization of the quantitative decision variables of a simulated system. In addition, there seems to be an increasing need for procedures that address optimization of the structures of the complex systems. These are the problems where the performance of the system depends more on operation policies of the system than the values of the quantitative decision variables.

Comprehensive reviews of literature on simulation optimization has been provided by Glynn (1986), Meketon (1987), Jacobson and Schruben (1989) and Safizadeh (1990). In this tutorial these citations will not all be repeated. Instead, issues that make simulation optimization distinct from generic optimization procedures will be addressed, various classifications of these problems will be presented and solution procedures suggested in the literature and applied in practice will be explored.

2 ISSUES IN SIMULATION OPTIMIZATION

Using simulation as an optimization tool for complex systems presents several challenges. Some of these challenges are those involved in optimization of any complex and highly nonlinear function. Others are more specifically related to the special nature of the simulation modeling. Simply stated, a simulation optimization problem is an optimization problem where the objective function(objective functions in case of a multi-criteria problem), constraints, or both are responses that can only be evaluated by computer simulation. As such, these functions are only implicit functions of decision parameters of the system. In addition, these functions are often stochastic in nature as well. With these characteristics in mind, the major issues to address when comparing them to generic non-linear programming problems are as follows:

-There does not exist an analytical expression of the objective function or the constraints. This eliminates the possibility of differentiation or exact calculation of local gradients.

- The objective function(s) and constraints are stochastic functions of the deterministic decision variables. This presents a major problem in estimation of even approximate local derivatives. Furthermore, this works against even using complete enumeration because based on just one observation at each point the best decision point cannot be determined.

- Computer simulation programs are much more expensive to run than evaluating analytical functions. This makes the efficiency of the optimization algorithms more crucial.

- Most practitioners use some kind of simulation language for modeling their systems. Optimization, on the other hand, requires using some other kind of programming language which differs from one practitioner to the next. Interfacing simulation models with generic optimization routines is not always a simple task. This is especially true for newer higher level user friendly simulation languages.

We will address each of these issues in following sections.

3 GENERAL FORMULATION

The most common formulation for optimization of systems through simulation has been for maximization or minimization of the expected value of the objective function of the problem. This, however, does not have to be the case. Operation of a system might be considered optimal if the risk of exceeding a certain threshold is minimized. On other situations, one might be interested in minimizing the dispersion of the response rather than maximizing its expected value. In this tutorial we limit ourselves to optimization of the expected values.

Another pertinent issue in formulating simulation optimization problems is the treatment of stochastic constraints. These constraints, like the objective functions are functions of deterministic decision variables and are supposed to define a deterministic feasible region. To incorporate them into an optimization process they have to somehow be changed into deterministic functions. Again, the expected value has been used for this transformation by some. In practice, however, many decision makers prefer to deal with their constraints as the risk of violation of a particular constraint rather than being within the expected value of the feasible region.

Then two alternative ways of formulating the general simulation optimization problem are:

$$\text{Maximize(Minimize) } \mathbf{f}(\mathbf{X}) = E[\mathbf{z}(\mathbf{X})]$$

$$\text{Subject to:} \quad \mathbf{g}(\mathbf{X}) = E[\mathbf{r}(\mathbf{X})] \leq \mathbf{0} \quad (3.1)$$

$$\text{and} \quad \mathbf{h}(\mathbf{X}) \leq \mathbf{0}$$

where **z** and **r** are random vectors representing several responses of the simulation model for a given **X**, a p-dimensional vector of decision variables of the system. **f** and **g** are the unknown expected values of these vectors that can only be estimated by noisy observations on **z** and **r**. **h** is a vector of deterministic constraints on the decision variables.

The alternative formulation is:

$$\text{Maximize(Minimize) } \mathbf{f}(\mathbf{X}) = E[\mathbf{z}(\mathbf{X})]$$

$$\text{Subject to:} \quad \mathbf{P}\{\mathbf{g}(\mathbf{X}) \leq \mathbf{0}\} > \mathbf{1} - \boldsymbol{\alpha} \quad (3.2)$$

$$\text{and} \quad \mathbf{h}(\mathbf{X}) \leq \mathbf{0}$$

where $\boldsymbol{\alpha}$ is the vector of risks of violation of constraints the decision maker is prepared to accept. This formulation yields itself well to simulation analysis because the constraints can easily be transformed into a manageable form as follows:

$$UCL_{1-\alpha j} g_j(X) \leq 0 \quad (3.3)$$

where $UCL_{1-\alpha j}$ indicates the upper confidence limit calculated for the response g_j at $1-\alpha_j$ level. This form of constraint can be easily used to check whether a decision point is feasible, because one can use available means of estimating confidence intervals for a given **X**.

4 CLASSES OF SIMULATION OPTIMIZATION PROBLEMS

There are several ways simulation optimization problems can be classified. Each class can be considered as a special case of the above general formulation. If **f**(X) is a one-dimensional vector, the problem is reduced to a single objective optimization while in its general form it is a multiple objective problem. If elements of **X** are continuous variables the problem is often easier to solve by available stochastic search methods. If they are discrete but still quantitative, the problem will be closer to those addressed by integer programming techniques. If **X** represents a vector of qualitative decision policies, optimization becomes more difficult because of the lack of available analytical tools to treat this type of problems. In addition, for such problems there will be a need for automatic generation of simulation models according to a systematic process. In this tutorial, we refer to those problems as non-parametric optimization problems.

In following sections we will cover available solution procedures for various classes of these problems. Most of the efforts will be spent on exploring procedures applied to single objective problems with continuous or discrete quantitative decision variables subject to deterministic or stochastic constraints. Several approaches to solving multiple objective problems will be discussed next. Finally, a short discussion on non-parametric optimization problems will be presented.

5 SINGLE OBJECTIVE PROBLEMS

There have basically been four major approaches to solving these problems. These are:

- Gradient based search methods
- Stochastic approximation methods
- Response surface methods
- Heuristic search methods

5.1 Gradient Based Search Methods

These methods attempt to take advantage of the vast amount of literature available on search methods developed for non-linear programming problems. The major contribution of practitioners in simulation optimization to this field has been the various methods of efficient estimation of gradients. Two major factors in determining the success of these methods are the reliability and the efficiency. Reliability is important because simulation responses are stochastic and a large error in gradient estimation may result in a movement in an entirely wrong direction. The efficiency is a major factor because simulation experiments are expensive and it is desirable to estimate gradients with minimum number of function evaluations. The gradient estimation methods often employed in simulation optimization are as follows:

5.1.1 Finite Difference Estimation

This is the crudest method of estimating the gradient. Partial derivatives of f(**X**) in this case are estimated by:

$$\delta f/\delta X_i = [f(X_1,..,X_i+\Delta X_i,..,X_p)-f(X_1,..,X_p)]/\Delta X_i \qquad (5.1.1.1)$$

As a result, to estimate the gradient at each point at least p+1 evaluations of the simulation model will be required. Furthermore, to obtain a more reliable estimate of the derivatives there may be a need for multiple observations for each derivative. An example of applying this method in conjunction with the Hooke and Jeeves pattern search technique is presented by Pegden and Gately (1977).

5.1.2 Infinitesimal Perturbation Analysis(IPA)

Perturbation analysis, when applied properly and to models that satisfy certain conditions estimates all gradients of the objective function from a single simulation experiment. In a relatively short time since its introduction to simulation field a significant volume of work on this topic is reported in the literature. A sample of these works can be found in Ho (1984), Ho et al (1983), Ho et al (1984), and Suri (1983). A complete discussion of all issues in IPA has been published in a recent book by Ho and Cao (1991).

The main principle behind perturbation analysis is that if a decision parameter of a system is perturbed by an infinitesimal amount, the sensitivity of the response of the system to that parameter can be estimated by tracing its pattern of propagation through the system. This will be a function of the fraction of the propagations that die before having a significant effect on the response of interest. The fact that all derivatives can be derived from the same simulation run, represents a significant advantage to IPA in terms of the efficiency. However, some restrictive conditions have to be satisfied for IPA to be applicable. For instance if as a result of perturbation of a given parameter, the sequence of events that govern the behavior of the system changes, the results obtained by perturbation analysis may not be reliable. Considering the complex nature of most simulation models this condition may not be satisfied most of the time. Heidelburger (1986) presents a study of deficiencies of IPA in estimating the gradients. There are also reports that additional work done in this area in recent years may alleviate some of the problems in its application to simulation optimization.

One difficulty with application of IPA to simulation optimization problems is that the modeler has to have a thorough knowledge of the simulation model and in some situations must have built it from scratch to be able to add additional tracking capabilities that are needed by IPA. Most practitioners build their simulation models using some kind of simulation language. With the advance of object oriented simulation methodology and languages, it will become even more difficult to build these additional tracking capabilities into a reusable simulation model.

5.1.3 Frequency Domain Analysis

Frequency domain analysis in estimating the sensitivity and gradients of the responses of simulation models was suggested by Schruben and Cogliano (1981). Additional work on the subject has been reported by Jacobson (1988) and Jacobson and Schruben (1988). The gradients are estimated by analyzing the power spectrum of the simulation output function which is affected by inducing specific sinusoidal oscillations to the input parameters. In a recent work, Jacobson and Schruben (1991) have used this in applying the Newton's method to simulation optimization. The frequency domain analysis suffers from the same difficulty as IPA because of the complexity of incorporating it with independently built simulation models.

5.1.4 Likelihood Ratio Estimators

Glynn (1987) presents an overview of Likelihood Ratio Estimators and their potential use in simulation optimization. He provides two algorithms by which the gradient of a simulation response function with respect to its parameters can be estimated. Rubenstein (1989) suggests a variation of this method and shows how it can be used in estimation of Hessians and higher level gradients to be incorporated in the Newton's method.

Once the method of estimating the gradients is decided upon, one of the available search techniques can be employed to search for the optimum. For a recent work using Quasi-Newton's method refer to Safizadeh (1992).

5.2 Stochasitic Approximation Methods (SAM)

Stochastic approximation methods refer to a family of recursive procedures that approach to the minimum or maximum of the theoretical regression function of a stochastic response surface using noisy observations made on the function. These are based on the original work by Robbins and Monro (1951) and Kiefer and Wolfowitz (1952). The original recursive formula is given for a single variable function and is stated as:

$$X_{n+1}=X_n+(a_n/2c_n)[f(X_n+c_n)-f(X_n-c_n)] \qquad (5.2.1)$$

where a_n and c_n are two series of real numbers that satisfy the following conditions:

$$\Sigma a_n<\infty,\ \mathrm{Lim}_{n\to\infty}(c_n)=0,\ \text{and}\ \mathrm{Lim}_{n\to\infty}(a_n/c_n)^2<\infty \quad (5.2.2)$$

It has been proven that as n approaches infinity X_n approaches to a solution such that the theoretical regression function of the stochastic response is maximized or minimized. This proof has been extended to multi-dimensional decision variables as well.

A neat characteristic of the stochastic approximation method when applied to simulation optimization is that the optimum of the expected value of the response could be reached using noisy observations. The difficulty is that a large number of iterations of the recursive formula will be required to obtain the optimum. Besides, for multi-dimensional decision vectors, p+1 observations will be needed for each iteration. Glynn (1986) has provided estimates of speed of convergence for some variations of this method. The other difficulty with these methods is the incorporation of the constraints into the optimization.

An earlier work in application of stochastic approximation method to simulation optimization is reported by Azadivar and Talavage (1980). In this work an automatic optimum seeking algorithm has been developed that could be interfaced with any independently built simulation model. In this algorithm the decision variables can be constrained by a set of linear deterministic constraints.

5.3 Response Surface Methodology (RSM)

Response surface methodology is the procedure of fitting a series of regression models to the responses of the simulation model evaluated at several points and trying to optimize the resulting regression function. The process usually starts with first order regression function and after reaching the vicinity of the optimum, higher degree regression functions are utilized. Among the earlier works in application of RSM to simulation optimization are those of Biles (1974) and Smith (1976). Additional work has been reported by Daugherty and Turnquist (1980), and Wilson (1987). Smith developed an automatic optimum seeking program based on RSM that could be interfaced with independently built simulation models. This program was developed for both constrained and unconstrained problems. Compared to many gradient based methods, RSM is a relatively efficient method of simulation optimization in terms of the number of simulation experiments needed. However, Azadivar and Talavage (1980) show that for complex functions with sharp ridges and flat valleys it does not provide good answers.

5.4 Heuristic Methods

There are two heuristic methods that have shown promise in application of simulation optimization. These are Box's (1965) Complex Search method and Simulated Annealing.

5.4.1 Complex Search

Complex search is an extension of Nelder and Mead's (1965) Simplex search that has been modified for constrained problems. The search starts with evaluation of points in a simplex consisting of p+1 vertices in the feasible region. It proceeds by continuously dropping the worst points from among the points in the simplex and adding new points determined by the reflection of this point through the centroid of the remaining vertices. The major issue in applying this procedure to simulation models is the determination of the worst point. Since the responses are stochastic, an apparently worst point may actually be one of the better points and dropping it may take the search away from the optimum region.

Azadivar and Lee (1988) developed a program based on Complex Search that automatically applies this process to any given simulation model. The decision variables of these models can be constrained by deterministic as well as stochastic constraints that may be responses of the same or other simulation models. In order to avoid making a wrong decision regarding the worst point the values of the responses at vertices are compared statistically. If the result of the multiple comparison is conclusive and shows that one point is significantly worse than the others it is dropped. Otherwise additional simulation runs are made to reduce the variance and the comparison is repeated.

5.4.2 Simulated Annealing

Simulated annealing is a relatively new method that could be utilized for simulation optimization. A description of this procedure is presented by Eglese (1990). Simulated annealing is a gradient search method that attempts to achieve a global optimum. In order not to be trapped in a locally optimum region, this procedure sometimes accepts movements in directions other that steepest ascend or descend. The acceptance of an uphill rather that a downhill direction is controlled by a sequence of random variables with a controlled probability.

6 MULTI-CRITERIA OPTIMIZATION

In addition to the common difficulties with all other multi-criteria optimization problems, multi-criteria simulation optimization possesses its own complexities which are mostly due to the stochastic nature of the response functions. Most of the work done in this area are slight modifications of the techniques used in operations research for generic multi-objective optimization. Some of these approaches are:

- Using one of the responses as the primary response to be optimized subject to certain levels of achievement on the other objective functions. Biles (1975,1977) uses this approach in conjunction with a version of Box's complex method and alternatively with a variation of gradient and gradient projection method.

- Variations of goal programming approach as those reported by Biles and Swain (1980), Clayton et al (1982), and Rees et al (1985).

- Multi-attribute value function methods such as the one used by Mollaghasemi et al (1991) and Mollaghasemi and Evans (1992).

Among the procedures that have been developed specifically for simulation optimization Teleb and Azadivar (1992) use the stochastic nature of the responses to the advantage of optimization. They use the Complex search method but suggest an alternative way of comparing the responses at vertices. For each point in the complex they calculate a probability that the response vector belongs to the random vector representing the best value for all objective functions. The point with the lowest probability is dropped and its reflection with respect to the centroid of the rest of the points is added to the simplex.

7. NON-PARAMETRIC OPTIMIZATION

Many industrial, service, and other complex systems that are modeled by computer simulation need to be optimized in terms of their structural designs and operational policies. Mathematical programming techniques are not usually applicable in these situations. Examples of these systems are scheduling policies, layout problems and part routing policies. In order to address these problems, each function evaluation requires a new configuration of the simulation model. Furthermore, since the decision variables are not quantitative, regular hill climbing, infinitesimal perturbation analysis, and stochastic approximation methods are not quite applicable. To deal with these problems an automatic model generation and a new optimization procedure has to be developed.

Since this is a new area of attention not much work has been reported in the literature. An example of this approach is the work by Prakash and Shannon (1989). We believe this is a very important topic for the future of the simulation optimization. Developments in this area will provide the real answer to "how to" questions.

8 CONCLUSIONS AND RECOMMENDATIONS

The choice of the procedure to employ in simulation optimization depends on the analyst and the problem to be solved. However we believe the modeler is often not a good mathematician and the mathematician is not necessarily a good simulation modeler. When it comes time to model a complex system a team of experts will work on developing a valid simulation model. These models are usually rather complex and do not yield themselves to the type of tracking needed in perturbation analysis and frequency domain analysis. Until a significant progress is made in these areas, practitioners will treat their simulation model as a black box demanding instruction from the optimization routines should be such that they can directly interface with these black boxes and operate on them in an input output mode not putting too much demand on the modelers to modify them for each iteration.

We recommend, parallel to additional efforts spent on advancing theoretical concepts such as IPA and frequency domain analysis, researchers work on making simulation optimization procedures more suitable to be interfaced with independently built models. We believe intelligent frameworks to perform these interfaces will make this task more feasible. task.

REFERENCES

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

Azadivar, F. and Lee, Y. H., 1988. Optimization of discrete variable stochastic systems by computer simulation, *Mathematics and Computers in Simulation*, 30:331-345.

Biles, W. E., 1974. A gradient regression search procedure for simulation experimentation, *Proceedings of 1974 Winter Simulation Conference*, 491-497.

Biles, W. E., 1975. A response surface method for experimental optimization of multiresponse processes, *Industrial and Engineering Chemistry: Process Design and Development*, 14:152-158.

Biles, W. E., 1977. Optimization of multiple objective computer simulation: A non linear goal programming approach, *Proceedings of AIDS Conference*, Chicago, 250-252.

Biles, W. E., and Swain, J.J., 1980. Optimization and industrial experimentation, *Wiley Interscience*, New York.

Box, E. P., 1956. A new method for constrained optimization and a comparison with other methods, *The Computer Journal*, 8:42-52.

Clayton, E. R., W.E. Weber, and B.W. Taylor, 1982. A goal programming approach to the optimization of multiresponse simulation models, *IIE Transactions*, 14: 282-287.

Daugherty, A. F., and Turnquist, M. A., 1980. Simulation optimization using response surfaces based on spline approximations, *Proceedings of the Winter Simulation Conference*, 183-193.

Eglese, R. W., 1990. Simulated annealing: a tool for operational research, *European Journal of Operational Research*, 46:271-281.

Glynn, P. W., 1986. Optimization of stochastic systems, *Proceedings of 1986 Winter Simulation Conference*, 52-59.

Glynn, P. W., 1987. Likelihood ratio gradient estimation: an overview, *Proceedings of 1987 Winter Simulation Conference*, 366-375.

Glynn, P. W., 1986. Stochastic approximation for Monte Carlo optimization, *Proceedings of Winter Simulation Conference*, 356-364.

P. Heidelberger, 1986. Limitations of infinitesimal perturbation analysis, *IBM Research Report RC 11891*, IBM Watson Research Labs, Yorktown Heights, NY.

Ho, Y. C., 1984. Perturbation analysis methods for discrete event dynamical systems and simulations, *Proceedings of Winter Simulation Conference*, 171-173.

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

Ho, Y. C., Cao, X. R., and Cassandras, C., 1983. Infinitesimal and finite perturbation analysis of queuing networks, *Automatica*, 19:439-445.

Ho, T. C., Suri, R., Cao, X. R., Diehl, G. W., Dille, J. W., and Zazanis, M., 1984. Optimization of large multiclass (non-product-form) queuing networks using perturbation analysis, *Large Scale Systems*, 7:165-180.

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

Jacobson, S. H., and Schruben, L. W., 1991. A simulation optimization procedure using harmonic analysis, working paper.

Kiefer, J., and Wolfowitz, J., 1952. Stochastic estimation of the maximum of a regression function, *Annals of Mathematical Statistics*, 23:462-446.

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

Mollaghasemi, M., Evans, G.W., and Biles, W. E., 1991. An approach for optimizing multiple response simulation models, *Proceedings of 13th Annual Conference in Computers and Industrial Engineering*.

Mollaghasemi, M., and Evans, G.W., 1992. A methodology for stochastic optimization of multiple response simulation models, working paper.

J.A. Nelder and R. Mead, 1976. A simplex method for function minimization, *Computer Journal*, 7, 308-313 (1965).9.

Pegden, C. D., and M. P. Gately, Decision Optimization for GASP IV Simulation Models, *Proceedings of the 1977 Winter Simulation Conference*, 125-133.

Prakash, S., and Shannon, R. E., 1989. Intelligent back end of a goal oriented simulation environment for discrete-part manufacturing, *Proceedings of 1989 Winter Simulation Conference*, 883-891.

Rees, L. P., E. R. Calyton, and B. W. Taylor, 1985. Solving multiple response simulation models using response surface methodology within a lexicographic goal programming framework, *IIE Transactions*, 17:47-57.

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

Rubinstein, R. Y., 1989. Sensitivity analysis and performance extrapolation for computer simulation methods, *Operations Research*, 37: No.1.

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

Safizadeh, M. H. and Signorila, R., 1992. Optimization of simulation via quasi-newton methods, working paper.

Schruben, L. W., 1986. Simulation optimization using frequency domain methods, *Proceedings of the Winter Simulation Conference*, 366-369.

Schruben, L. W., and Cogliano, V. J., 1991. Simulation sensitivity analysis: a frequency domain approach, *Proceedings of the Winter Simulation Conference*, 455-459.

D.E. Smith, Automatic optimum-seeking program for digital simulation, *Simulation 27*:27-32.

Suri, R., 1983. Infinitesimal perturbation analysis of discrete even dynamic systems: a general theory, *Proceedings of the Twenty-Second IEEE Conference on Decision and Control*, 1030-1038.

Teleb, R., and F. Azadivar, 1992. A methodology for solving multi-objective simulation-optimization problems, *European Journal of Operational Research*, to appear.

Wilson, J. R., 1987. Future direction in response surface methodology for simulation, *Proceedings of the Winter Simulation Conference*, 1987, 378-381.

AUTHOR BIOGRAPHY

FARHAD AZADIVAR is a professor in the Department of Industrial Engineering at Kansas State University. He is also the director of the Advanced Manufacturing Institute of the university which is a Center of Excellence in research on manufacturing for the state of Kansas. He received his B.S. degree in mechanical engineering and his M.S. degree in systems engineering. He received his Ph.D. in industrial engineering from Purdue University in 1980. His areas of research are in modeling and optimization of manufacturing systems.