A comprehensive literature classification of simulation optimisation methods

Wafik Hachicha and Ahmed Ammeri and Faouzi Masmoudi and Habib Chachoub

Unit of Mechanic, Modelling and Production (U2MP), Unit of Logistic, Industrial and Quality Management (LOGIQ), University of Sfax

24. May 2010

Online at https://mpra.ub.uni-muenchen.de/27652/
MPRA Paper No. 27652, posted 26. December 2010 19:44 UTC
ABSTRACT: Simulation Optimization (SO) provides a structured approach to the system design and configuration when analytical expressions for input/output relationships are unavailable. Several excellent surveys have been written on this topic. Each survey concentrates on only a few classification criteria. This paper presents a literature survey with all classification criteria on techniques for SO according to the problem of characteristics such as shape of the response surface (global as compared to local optimization), objective functions (single or multiple objectives) and parameter spaces (discrete or continuous parameters). The survey focuses specifically on the SO problem that involves single performance measure.

KEYWORDS: Simulation Optimization, classification methods, literature survey

1 INTRODUCTION

Many real world problems in management and manufacturing are very complex and mathematically intractable so that simulation is the appropriate tool for system analysis and performance evaluation. Computer simulation requires developing a program that mimics the behavior of a system as it evolves over time and records the overall system performance. As the technology of computer hardware and software advances, simulation has emerged as an essential tool in research and real world applications. With the continuing developments in computer technology, simulation is receiving increasing attention as a decision-making tool.

Therefore, computer simulation is frequently used in evaluating complex systems and optimizing responses. However, simulation is not without pitfalls. Simulation output is subject to random errors and requires proper statistical analysis. Furthermore, simulation models are “run” rather than solved. It is traditionally viewed as a tool for performance evaluation instead of decision making or optimization. In recent years researchers have attempted to combine simulation and optimization procedures to provide a complete solution. However, Simulation is merely a tool for problem solving; by itself, it cannot provide an answer. In addition to a good model, one also needs a sound technique to utilize the information from a simulation to make a decision. One such technique is optimization via simulation.

Simulation Optimization (SO) provides a structured approach to system design and configuration when analytical expressions for input/output relationships are unavailable. According to (Carson and Maria, 1997) SO can be defined as the process of finding the best input variable values from among all possibilities without explicitly evaluating each possibility and the objective of SO is to minimize the resources spent while maximizing the information obtained in a simulation experiment. There has been a great deal of work on SO in the research literature. Many comprehensive reviews of literature have been written on this topic. (Fu, 1994) contributes a general review of techniques for optimizing stochastic discrete-event systems via simulation, focusing on the techniques for optimization from a finite set: multiple-comparison procedures and ranking-and-selection procedures. (Carson and Maria, 1997) provide knowledge about the area of SO, with an extensive reference list pointing to detailed treatment of specific techniques. (Andradottir, 1998) presents a review of methods for optimizing stochastic systems using simulation. The focus is on gradient-based techniques for optimization with respect to continuous decision parameters and on random search methods for optimization with respect to discrete decision parameters. (Paul & Chanev, 1998) present an attempt to apply genetic algorithms (GA) to the problem of optimizing an existing simulation model. They demonstrate the capability of GA to solve hard inverse problems even in the area of complex simulation model optimisation. (Swicher & Hyden, 2000) provides a brief survey of the literature on discrete-event SO over the past decade (1988 to 2000). (Swisher and al, 2003) present a survey of the literature on discrete-event SO published in recent years (1988 to 2003), with a particular focus on discrete in put parameter optimization. (Tekin and Sabuncuoglu, 2004) presented a comprehensive survey on techniques for simulation optimization. They classified the existing
techniques according to the characteristics of the problems such as objective functions (single or multiple objectives) and parameter spaces (discrete or continuous parameters) and shape of the response surface (global as compared to local optimization). They also discussed major advantages, drawbacks and, comparisons of these techniques in the paper.

(Piera and al, 2004) described a new approach to integrate evaluation (simulation) methods with search methods (optimization) based not only on simulation results but also information from the simulation model. (Fu, 2006) develop a simulation optimization algorithm for determining the traffic light signal timings for an intersection of two one-way street traffic flows modelled as single-server queues. The system performance is estimated via stochastic discrete-event simulation, and gradient-based search based on stochastic approximation is applied. (Rosen and al, 2007) propose a SO method that involves a preference model, specifically adapted for decision making with simulation models. The proposed SO method is evaluated against two SO methods with embedded deterministic, multiple criteria decision making strategies. (Almeder and Preusser, 2007) presented a new approach that combines the advantages of complex simulation models and abstract optimization models. More recently (Fu and al, 2008) give a tutorial introduction to SO, beginning by classifying the problem setting according to the decision variables and constraints, putting the setting in the simulation context, and then summarize the main approaches to SO. (Alrefaei and Diabat, 2009) have focused on simulated annealing algorithm for solving multi-objective SO problems. The algorithm is based on the idea of simulated annealing with constant temperature, and uses a rule for accepting a candidate solution that depends on the individual estimated objective function values. For more comprehensive studies on this topic, the reader can referred to some others reviews papers including (Glynn, 1989), (Metketon, 1987), (Azadivar, 1992), (Fu, 1994), (Kleijn, 1995) and references therein.

Applications of the SO approach exist in many fields. With respect to operations research problems we mention above all inventory models (Kochel and Nieland, 2005), logistic systems (Kochel and al, 2003; Yoo and al, 2010), and manufacturing systems (Kochel and Nieland, 2002; Kampf and Kochel, 2006).

As a result of this literature survey, there are several ways SO problems can be classified. We can classify SO problems regarding their input variables (quantitative variables and qualitative variables), output variables (a single objective SO problems or multi-objective problems), parameter spaces (discrete or continuous parameters), the shape of the response surface (global as compared to local optimization) or by their Optimization Procedure. Our main contribution in this paper is to provide a literature survey with all classification criteria and to propose a global classification scheme of SO methods.

The remainder of the paper is organized as follows: Section 2 we establish a common framework for simulation optimization problems and present the notation to be used. Section 3 presents a literature survey with all classification criteria on techniques for simulation optimization. Finally Section 4 draws conclusions.

2. PROBLEM DESCRIPTION

SO refers to the process for finding optimal system design whose performance is estimated by simulation (Kabirian and Olafsson, 2007), and the problem setting thus contains the usual optimization components:

✓ Decision variables,
✓ objective function, and
✓ Constraints.

The SO can be defined as the latter case: repeated analysis of the simulation model with different values of design parameters, in an attempt to identify best simulated system performance. The design parameters of the real system are set to the ‘optimal’ parameter values determined by the SO exercise, rather than in an ad hoc manner based on qualitative insights gained from experimenting the simulation model. A very general formulation of the above SO problem is to minimize the expected value of the objective function with respect to its constraint set as:

$$\min_{\theta} f(\theta),$$

where \(\theta\) is a p-dimensional vector of all the decision variables, \(\Theta\) is the feasible region and \(f: \Theta \rightarrow \mathbb{R}\) is the objective function. If \(f(\theta)\) is a one-dimensional vector, the problem is single objective optimization, whereas if its dimension is more than one, the problem becomes multiobjective. The optimum is denoted by \(\theta^\ast\). Without loss of generality, we will consider the minimization problem throughout the paper.

We assume that the system under consideration is complex enough that the expected performance \(f(\theta)\) of each system design \(\theta \in \Theta\) cannot be determined exactly, but is instead estimated through simulation. The optimization model response function is represented by \(f(\theta)\) which is usually the expected value (long-term average) of some simulated system performance measure \(Y\) as a function of the design parameter vector \(\theta\). That is

$$f(\theta) = E(Y(\theta, \varepsilon))$$

where \(\varepsilon\) represents the stochastic effects in the system. The form of \(f\) is not known. Its value is estimated using \(n\) runs of the simulation model under the design scenario specified by \(\theta\).
Simulation optimization methods

Local optimization

- $\Theta$ a discrete set

$\Theta$ continuous

- $f$ continuous

Metamodel Methods

$\Theta$ is finite, small

- Statistical Selection Methods
  - Ranking and Selection
  - Multiple Comparaison
  - Others

Statistical Selection Methods

$\Theta$ is large or $\infty$

- Random Search
- General Search Strategies
- Ordinal Optimization
- Nested Partitions Method
- Others

Global optimization

- Bayesian/Sampling Algorithms
- Gradient Surface Methods
- Simulated Annealing
- Tabu Search

- Evolutionary Algorithms
  - Genetic Algorithm
  - Evolutionary Programming
  - Evolution Strategy

- Gradient Approaches
  - Perturbation Analysis (PA)
  - Likelihood Ratios (LR) $\equiv$ Score Function (SF)
  - Finite Difference (FD)
  - Frequency Domain Analysis (FDA)
  - Harmonic Analysis (HA)
  - Others

- No Gradient Approaches
  - Sample path optimization
  - Simplex search method
  - Hooke and Jeeves method
  - Others

Legend

- Condition
- Method group

Figure 1: Proposed classification scheme of simulation optimization methods
\[ f(\theta) = \sum_{i=1}^{n} \frac{Y_i}{n} \]  

In Eq. (3) the dependence of Y on the value of \( \theta \) has been suppressed. While \( f(\theta) \) is deterministic, its estimate is stochastic, since the simulation run time must be finite (so \( n < \infty \)).

It should be noted that, the general formulation (1) subsumes the usual mathematical programming settings (which prefers \( x \) to \( \theta \) for its decision variables):

- If \( f(\theta) \) is a scalar function, the problem is single objective optimization; whereas if it is a vector, the problem becomes multi-objective.
- \( f(\theta) \) is linear in \( \theta \) and \( \Theta \) can be expressed as a set of linear equations in \( \theta \) corresponds to linear programming, or mixed integer linear programming if part of the \( \Theta \) space involves an integer (e.g., \{0,1\} binary) constraint.

3. THE PROPOSED CLASSIFICATION

As pointed out by (Tekin and Sabuncuoglu, 2004) the existing studies can be classified under two main headings: local optimization; and global optimization. Local optimization techniques are further classified in terms of discrete and continuous parameter spaces the discrete case can also further classified into finite parameter space and infinite parameter space. The main goal of this research is to propose a global classification scheme of SO problems. This structure is shown in Figure 1. For more details on the major techniques, we have presented, in table 1, table 2, and table 3 other SO references.

### 3.1 Local optimization

Local optimization problems are discussed in terms of discrete and continuous decision spaces. In a discrete space, decision variables take a discrete set of values such as the number of machines in the system, alternative locations of depots, different scheduling rules or policies, etc. Several techniques have been developed for simulation optimization when the input parameter values (i.e., the size of \( \Theta \)) is discrete. If the set \( \Theta \) is finite and small, ranking and selection and multiple comparisons procedures are appropriate. If the set \( \Theta \) is infinite but very large, then techniques such as ordinal optimization, random search, nested partitions method and general search strategies have been adapted for the simulation environment. However in the continuous case, gradient-based methods or metamodel based optimization can be used.

#### 3.1.1 Discrete input parameter methods

The discrete input parameter case differentiates techniques appropriate for small and for large numbers of feasible input parameter values. As discussed in (Swisher and al, 2000) several techniques have been developed. If the set \( \Theta \) is finite and small, ranking and selection and multiple comparison procedures are appropriate. If the set \( \Theta \) is infinite or very large, then techniques such as ordinal optimization, random search, nested partitions method and general search strategies have been adapted for the simulation environment. However in the continuous case, gradient-based methods or metamodel based optimization can be used.

<table>
<thead>
<tr>
<th>Optimization - simulation methods</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finite space</strong></td>
<td></td>
</tr>
<tr>
<td>Ranking and Selection</td>
<td>(Koenig and Law, 1985), (Chen and al, 1996; 1997), (Morrice and al, 1998; 1999), (Hyden and Schruben, 1999), (Branke and al, 2007)</td>
</tr>
<tr>
<td><strong>Infinite space</strong></td>
<td></td>
</tr>
<tr>
<td>General Search Strategies</td>
<td>Eglese, 1990), (Fleischer, 1995), (Leipins and Hilliard, 1989), (Muilenbein, 1997), (Glover and Laguna, 1997), (Manz and al, 1989), (Haddock and Mittenhall, 1992), (Baretto and al, 1999), (Zeng and Wu, 1993), (Yucesan and Jacobson, 1996), (Stuckman and al, 1991), (Brady and Mc Garvey, 1998), (Dummiller, 1999), (Facenda and Tenga, 1992), (Tomkins and Azadivar, 1995), (Glover et al, 1996), (Hall and al, 1996), (Hall and Bowden, 1997).</td>
</tr>
<tr>
<td>Random Search</td>
<td>(Andradottir, 2005), (Fu, 2007), (Chen et al, 2008).</td>
</tr>
<tr>
<td>Nested Partitions Methods</td>
<td>(Shi and Olafsson, 1997), (Shi et al, 1999).</td>
</tr>
</tbody>
</table>
3.1.1.1 Finite space

When the optimization involves selecting the best of a few alternatives, that is $\Theta = \{\theta_1, \theta_2, m\}$, where $m$ is relatively small, then it may be possible to evaluate every solution and compare the performance. The two most popular methodologies for the class of the problem are: Raking and Selection (R&S) and Multiple Comparison Procedures (MCPs). (R&S) focuses on selecting the optimal input parameter values. In MCPs the idea is to run a number of replications and make conclusions on a performance measure by constructing confidence intervals. MCPs approach the optimization problem as a statistical inference problem and, unlike R&S procedures, do not guarantee a decision. For excellent review of these techniques, one can refer to (Kim and Nelson, 2006)

3.1.1.2 Infinite space

When it is not possible to evaluate every solution using a statistical selection procedure, and the set $\Theta$ is very large, then different other methods must be applied. Such as Ordinal optimization reduces the search for an optimal solution to sampling over a very large set of solutions to sampling over a smaller, more manageable set of good solutions. A more detailed treatment of this topic can found in (Ho and Deng, 1994), (Lee et al, 1999).

General search strategies such as simulated annealing (Egleins, 1990); (Fleischer, 1995), genetic algorithms (Leipins and Hilliard, 1989); (Muhlenbein, 1997) and tabu search (Glover and Laguna, 1997). (Glover and Laguna, 1993) have been adapted for the stochastic environment associated with discrete-event simulation optimization.

Random search (RS) was first developed for deterministic optimization, but has been extended to the stochastic setting. (RS) can work on an infinite parameter space. Inputs of upper and lower bounds on each of the controllable factors define an overall search region. For more details see on random search methods in simulation, see (Andradottir, 2005)

Nested Partitions (NP) is a randomized method for global optimization. The motivation for this method is that some parts of the feasible region may be most likely to contain the global optimum. Hence it is efficient to concentrate the computational effort in these regions. The Nested Partitions Method takes a global perspective and combines global and local search techniques. For more details, see (Shi and Olafsson, 1997), (Shi and al, 1999).

3.1.2 Continuous input parameter methods

The feasible region, $\Theta$, is uncountable and infinite when the set of in put parameters are continuous. Continuous simulation optimization problems fall into two categories: metamodel methods and stochastic gradient estimation.

3.1.2.1 Metamodel Methods

Metamodeling, which was first described by (Blanning, 1975) is a process of developing a mathematical relationship between a response measure of interest and a set of input variables. Metamodels Methods are often constructed in two primary stages. The first stage uses a factorial experimental design to collect a structured data set that is then used to find the functional relationship between decision variables and responses. After a metamodel is constructed, the what-if analysis can be obtained without the further consumption of expensive computing resources. See (Hurrrion and Birgil, 1999), (Kleijn, 1987) (Yesilyurt and Patera, 1995), (Barton and Meekesheimer, 2006).

3.1.2.2 Stochastic gradient estimation.

The goal of stochastic gradient estimation is to estimate the gradient of the performance measure with respect to the parameters. An extensive body of research exists for SO problems of this type. Input parameter methods may be classified as either gradient-based or non gradient-based.

a) Gradient based Approaches

A considerable amount of research has focused specifically on techniques for gradient estimation, such as, finite difference estimation, perturbation analysis (see Ho, 1984; Glynn, 1989; Suri, 1989), likelihood ratio estimators (see Glynn, 1987; Rubinstein and Shapiro, 1993) and frequency domain experimentation (Schruben and Cogliano, 1991)

Perturbation analysis (PA) estimates all gradients of an objective function from a single simulation run (Bettonvil 1989), (Glasserman 1991), (Ho and Cao, 1991). The idea is that in a system, if a decision parameter is perturbed by an infinitesimal amount, the sensitivity of the output variable to the parameter can be estimated by tracing its pattern of propagation.

The likelihood ratio (LR), or the score functions, method involves expressing the performance measure as an integral involving the product of the densities of the underlying random variables. In the (LR) the expressing gradient of the expected value of an output variable is expressed as the expected value of a function. Since its introduction to simulation field a significant volume of work on this topic has been reported in the literature. A sample of these works can be found in (Glynn, 1989a), (Glynn, 1989b), (Reiman and Weiss, 1986), (Glynn 1987), (Rubinstein 1991), and (Rubinstein and Shapiro 1993).

As pointed by (Carson and Maria, 1997) Frequency Domain Methods is one in which selected input parameters are oscillated sinusoidally at different frequencies during one long simulation run.
Frequency domain experiments involve addressing three questions: how does one determine the unit of the experimental or oscillation index, how does one select the driving frequencies, and how does one set the oscillation amplitudes. For more details on this technique, see (Schruben and Cogliano, 1981), (Morrice and Schruben, 1989), (Hazra et al, 1997), (Heidergott, 1995), *Harmonic analysis* (Jacobson and Schruben 1999), have been studied with the objective of developing efficient gradient estimators applicable to a broad class of discrete-event simulation models. These gradient estimators are then imbedded in optimization algorithms which control the step size taken in the gradient direction at each iteration.

b) Non-gradient based approaches

Non-gradient approaches provide an alternative to gradient estimation-based procedures. These methods include the Nelder-Mead (simplex) method, the Hooke and Jeeves method and Sample Path Method. The main idea of these methods is to take a large enough set of samples so that the stochastic problem is basically turned back into a deterministic problem to which the tools of nonlinear programming could be applied; for more details, see (Haddock and Bengu, 1987), (Barton and Ivey, 1991; 1996), (Gurkan et al, 1994), (Robinson, 1996), (Humphrey and Wilson, 1998), (Kleywegt et al, 2001)

### 3.2 Global search methods

Different methods can be used, such as evolutionary algorithms, simulated annealing, tabu search, Bayesian/sampling algorithms, and gradient surface method. Several of them are iterative.

#### 3.2.1 Evolutionary Algorithms

Evolutionary *Algorithms* (EAs) are population-based metaheuristic optimization algorithms that use biology-inspired mechanisms like mutation, crossover, natural selection, and survival of the fittest in order to refine a set of solution candidates iteratively. The most popular (EA) are Genetic Algorithms (GAs), Evolutionary Programming (EP) and Evolution strategy (ES). For more material on the (EAs) see for example: (Back and al, 1997), (Hall et al, 1996), (Schweifel, 1995), (Biehahn and Nissen, 1994), (Maria, 1995), (Pierraval and Tautou, 1997), (Muhlebein, 1997), (Pirreval and Paris, 2000), (Cassady et al, 2000), (Yuan, 2009), (Farzanegan and Vahidipour, 2009).

#### 3.2.2 Simulated Annealing

*Simulated Annealing* (SA) is a stochastic optimization method analogous to the physical annealing process where an alloy is cooled gradually so that a minimal energy state is achieved. Overviews of this heuristic can be found in (Kirkpatrick et al, 1983), (Van Laarhoven and Aarts, 1987), (Johnson et al, 1989), (Eglese, 1990) and (Koulamas et al, 1994), (Fleisher, 1995) and (Koulamas et al, 1994), (Fleisher, 1995), (Alrefaei et al, 1995).

#### 3.2.3 Tabu Search

*Tabu Search* (TS) developed by (Fred Glover, 1989) is distinguished by introducing adaptive memory into meta-
heuristic search, together with associated strategies for exploiting such memory, equipping it to penetrate complexities that often confound other approaches.

Table 3: Methods based on Global search

<table>
<thead>
<tr>
<th>Simulation Optimization methods</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian Sampling algorithms</td>
<td>(Lorenzen, 1985), (Easom, 1990), (Stuckman and Easom, 1992)</td>
</tr>
<tr>
<td>Gradient Surface Methods</td>
<td>(Ho et al, 1992), (Fu, 2007).</td>
</tr>
</tbody>
</table>

TS is an adaptive procedure with the ability to utilize many other methods which it directs to overcome their limitations of getting stuck in local optima. For more details for this method the reader can referred to (Osman, 1993), (Glover and Laguna, 1997), (Hu, 1992), (Osman, 1993), (Lutz et al, 1998), (Martin et al, 1998), (Dengis and Alabas, 2000), (April et al, 2003), (Hedar and Fukushima, 2006).

3.2.4 Bayesian/sampling algorithms

The Bayesian/Sampling (B/S) methodology is a search strategy where at each iteration; the next guess is chosen to be the point that maximizes the probability of not exceeding the previous value by some positive constant. See (Lorenzen, 1985) and (Easom, 1990) for more details.

3.2.5 Gradient surface method

Gradient surface method (GSM) combines the advantages of response surface methodology (RSM) and efficient derivative estimation techniques like perturbation analysis (PA) or likelihood ratio method (LR). In GSM, the gradient estimation is obtained by PA (or LR), and the performance gradient surface is obtained from observations at various points in a fashion similar to the RSM. Zero points of the successive approximating gradient surface are then taken as the estimates of the optimal solution. GSM is characterized by several attractive features: it is a single-run method and more efficient than RSM; it uses at each iteration step the information from all data points rather than just the local gradient; it tries to capture the global features of the gradient surface and thereby quickly arrives at the vicinity of the optimal solution. See (Ho et al, 1992) and (Fu, 2007) for more details.

4. CONCLUSION

Simulation Optimization (SO) is an optimization itself; it is required if one wants to find the best steady-state values of important process variables. This is an active research area that has sparked as much interest in the academic world as in practical settings. The most exciting developments are usually reported annually in the Winter Simulation Conference. In this paper we provide a general overview of the different approaches for simulation optimization with all classification criteria found in the research literature and we propose a global classification scheme of SO methods. Comparative studies based on performances measure between these approaches will be our research perspective.

REFERENCES

Blanning W.R (1975), The construction and implementation of metamodels, Simulation 24–25 (6) pp. 177–184


- 11 -


