Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

SIMULATION OPTIMIZATION OF STOCHASTIC SYSTEMS WITH INTEGER VARIABLES BY SEQUENTIAL LINEARIZATION

S. J. Abspoel
L. F. P. Etman
J. Vervoort
J. E. Rooda

Department of Mechanical Engineering
Eindhoven University of Technology
P.O. Box 513,
5600 MB, Eindhoven, THE NETHERLANDS

ABSTRACT

Discrete-event simulation is widely used to analyse and improve the performance of manufacturing systems. The related optimization problem often includes integer design variables and is defined by objective function and constraints that are expected values of stochastic functions. These stochastic functions have to be evaluated via the simulation model at the discrete levels of the integer design parameters. For such a simulation optimization problem with integer variables we have developed an optimization strategy that is based on a series of linear approximate subproblems. Each subproblem is built from the outcomes of simulation experiments. A D-optimal design of experiments is used to plan the simulation experiments. Stochasticity in constraint and objective functions is dealt with explicitly using safety indices. Two test problems will be presented to illustrate the optimization strategy. This includes a simulation based four-station production flow line problem.

1 INTRODUCTION

Industrial systems are becoming increasingly complex. Demands for flexibility, greater variety of products and advanced production processes, complicate the design of manufacturing systems. Within the Systems Engineering group, we use simulation models to estimate the behaviour and performance of manufacturing systems such as in semiconductor industry (e.g. Lemmen et al. 1999 and Rulkens et al. 1998).

Design optimization problems involving manufacturing systems have three characteristic properties which make them difficult to solve. Firstly, the design variables often have to be integer, for example: the number of machines in a work centre or the size of a batch. Secondly, performance measurements, such as lead time, throughput and work in process, are stochastic. Ignoring the effect of variability, for instance in process times, would largely limit the practical use of the outcome of a simulation (this is also stressed by Hopp and Spearman (1996)). Thirdly, the analysis of a design can be computationally expensive, especially for more complicated models.

Seeking help in designing complex industrial systems, we need a design optimization strategy that can deal with all three of these characteristics. Fu (1994) and Carson and Maria (1997) reviewed techniques for simulation optimization of stochastic systems. They identify gradient-based methods, genetic algorithms and statistical methods, mainly for continuous design variables. Fu concludes that the state of the art is far from satisfactory. Practical implementations have not attacked large problems, because of rising computational costs as the number of design variables increases.

In the field of structural optimisation, Thanedar and Vanderplaats (1995) and Arora, Huang, and Hsieh (1994) reviewed methods for optimizing deterministic non-linear problems that have discrete design variables. Thanedar and Vanderplaats (1995) state that if function evaluations are expensive, approximation methods can be used; however they do not guarantee a discrete optimum. Approximation methods (reviewed by Barthelemy and Haftka (1993)) have proved to be useful in finding practical solutions for structural optimization problems which involve computationally expensive analysis.

In this paper, we present a strategy using elements of both structural optimization and simulation optimization. An approximation concept, based on response surface methodology, avoids a direct coupling between simulation model and integer programming solver. Special attention is given to the combined treatment of integer design variables and stochastic response variables.

We implemented our optimization strategy using the numerical package Matlab and an external mixed integer programming solver. Two production flow line problems are used to illustrate the approach. The first is analytic and has two design variables. The second is simulation-based with four design variables.

2 OPTIMIZATION PROBLEM

A mathematical formulation of the optimization problem treated in this paper is given below:

$$\begin{array}{ll} \text{Problem} & \text{P} \\ \text{Minimize:} & E[F(\mathbf{x}, \boldsymbol{\omega})] \\ \text{subject to:} & E[G_j(\mathbf{x}, \boldsymbol{\omega})] \leq c_j, \quad j = 1, \ldots, m, \\ & x_i^l \leq x_i \leq x_i^u, \\ & x_i \in \mathbb{Z}, \qquad\qquad i = 1, \ldots, n. \end{array}$$

The objective is to find the design variable values $\mathbf{x}$ that will minimize the expected value of the stochastic objective function $F(\mathbf{x}, \boldsymbol{\omega})$, subject to expected value constraints on the stochastic functions $G_j(\mathbf{x}, \boldsymbol{\omega})$. An expected value constraint, $E[G_j(\mathbf{x}, \boldsymbol{\omega})] \leq c_j$, requires the expected value of G_j to be less than or equal to the bound c_j.

In the remainder of this paper, we assume that function evaluations of F and G_j are the results of a simulation experiment which is computationally expensive. Repeating a simulation experiment for the same design $\mathbf{x}$ will yield different values for F and G_j, because of the stochasticity in the simulation model (denoted by $\boldsymbol{\omega}$). For simulation models of industrial systems, the distributions of G_j and F are often unknown, can be not normal and can depend on the design variables values $\mathbf{x}$.

3 OPTIMIZATION STRATEGY

We have developed a sequential optimization strategy that uses a series of subproblems which are linear approximations of the original optimization problem. Starting from an initial design, in each cycle a linear approximate subproblem is created and solved. If its solution, called approximate optimum, improves the objective function value compared with the previous cycle, it becomes the next cycle's starting point.

3.1 Approximate Subproblem

Each subproblem consists of linear response surfaces for constraints and objective function. Since optimization problem P can behave non-linearly as a function of the design variables, the response surfaces only approximate F and G_j well in a small region – called search region – within the design space.

Within this search region, a design of experiments is created. Simulation experiments are carried out according to this design of experiments. Response surfaces, consisting of linear closed form terms, are built using the results of the simulation experiments. These response surfaces together form a linear deterministic approximation of the original stochastic optimization problem P, which is only valid within the search region.

Design point $\mathbf{x}^{(p,0)}$ denotes the starting point of the p-th cycle. The approximate optimum of the p-th cycle for the q-th iteration is denoted by $\mathbf{x}^{(p,q)}$. So, if two iterations are needed to find an improved design in cycle p, design $\mathbf{x}^{(p,2)}$ becomes the start design for the next cycle: $\mathbf{x}^{(p+1,0)}$. Using this notation, the approximate subproblem can be formulated as follows:

$$\begin{array}{ll} \text{Problem} & \tilde{\text{P}}^{(\text{p,q})} \\ \text{Minimize:} & \tilde{f}^{(p,q)} = a_0^{(p,q)} + \sum_{i=1}^{n} a_i^{(p,q)} \cdot x_i, \\ \text{subject to:} & \tilde{g}_j^{(p,q)} = b_{0j}^{(p,q)} + \sum_{i=1}^{n} b_{ij}^{(p,q)} \cdot x_i, \leqslant \tilde{c}_j^{(p,q)} \\ & x_i^{l(p,q)} \leqslant x_i \leqslant x_i^{u(p,q)}, \\ & j = 1, \ldots, m, \qquad i = 1, \ldots, n. \end{array}$$

3.2 Move Limit Strategy

Upper and lower bounds $\mathbf{x}^{l(p,q)}$ and $\mathbf{x}^{u(p,q)}$, called move limits, define the search region. In each cycle and iteration, the objective function value of the approximate optimum is compared with the previous designs. If it has improved, the approximate optimum becomes the starting point for the next cycle. If the approximate optimum has not improved, the search region is repositioned or reduced by the move limit strategy.

The move limit strategy uses two directional methods for positioning the search region. Directional method 1 (see Figure 1) uses the cycle start design $\mathbf{x}^{(p,0)}$ as a corner of the search region. The opposing corner is placed in the direction of the previous search direction, which is defined as the vector between the current cycle start design and the cycle start design of the previous cycle. Upper and lower limits on the design variable values for this method are calculated as follows:

$$\begin{aligned} x_i^{l(p,q)} &= x_i^{(p,0)} + \frac{1}{2} \cdot m_i^{(p,q)} \cdot \text{sign}\left(s_i^{(p-1)}\right) - \frac{1}{2} \cdot m_i^{(p,q)}, \\ x_i^{u(p,q)} &= x_i^{(p,0)} + \frac{1}{2} \cdot m_i^{(p,q)} \cdot \text{sign}\left(s_i^{(p-1)}\right) + \frac{1}{2} \cdot m_i^{(p,q)}, \\ s_i^{(p-1)} &= x_i^{(p,0)} - x_i^{(p-1,0)}, \\ i &= 1, \ldots, n. \end{aligned}$$

Parameter $m_i^{(p,q)}$ defines the size of the search region in scale units in direction i.

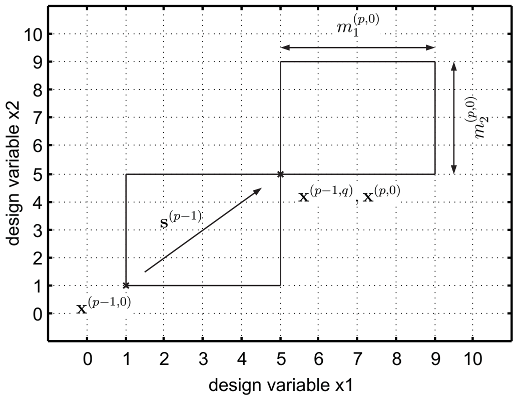


Figure 1: Directional Method 1

Directional method 2 (See Figure 2) makes the current cycle start design the centre of the search region. Upper- and lower bounds on the search region for this method are calculated as follows:

$$x_i^{l(p,q)} = x_i^{(p,0)} - \frac{1}{2} m_i^{(p,q)},$$
$$x_i^{u(p,q)} = x_i^{(p,0)} + \frac{1}{2} m_i^{(p,q)}.$$

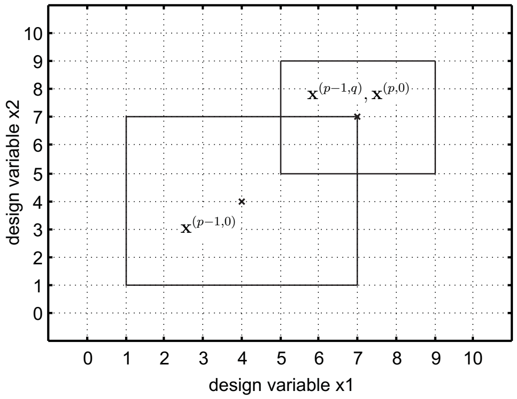


Figure 2: Directional Method 2

Search region bounds, calculated by both methods, can violate the bounds of the original design space $x_i^{l(p,q)}$ and $x_i^{u(p,q)}$. In that case, the search region is moved to lie entirely within the design space. Furthermore, if the calculated $x_i^{l(p,q)}$ or $x_i^{u(p,q)}$ values are not integer, $x_i^{l(p,q)}$ is decreased and $x_i^{u(p,q)}$ is increased to the first integer.

3.3 Design of Experiments

To calculate the approximate subproblem, N simulation experiments are planned within the search region. A D-optimal design for a linear model, without interaction, is used to plan these simulation experiments. Contrary to factorial and fractional factorial designs, a D-optimal design can be created for any number of experiments, that is larger than or equal to the number of parameters in the model. Therefore, a linear relation between the number of design variables and the number of experiments exists. This helps to keep the number of simulation experiments at a manageable level for increasing numbers of design variables.

3.4 Accepting Designs

An approximate optimum design $\mathbf{x}^{(p,q)}$ is accepted as the start design for the next cycles, if the following conditions hold:

- $\mathbf{x}^{(p,q)}$ is feasible.
- $\mathbf{x}^{(p,q)}$ was not previously found.
- Objective function value $F(\mathbf{x}^{(p,q)})$ has not increased compared with $F(\mathbf{x}^{(p,0)})$.

3.4.1 Stochastic Constraints

If the constraint function G_j is deterministic, the feasibility of a design $\mathbf{x}$ can be determined comparing constraint value $G_j(\mathbf{x})$ with constraint boundary c_j. However, this approach is not applicable, if the constraints are stochastic.

Consider a stochastic constraint function G_j. Carrying out M simulation experiments for the same design variable values $\mathbf{x}^{(p,q)}$ results in M different constraint values, $g_{j\ k}^{(p,q)}$ with $(k = 1, \ldots, M)$, varying around the mean value $\bar{g}_j^{(p,q)}$. This sample mean will be an unbiased estimate of the expected constraint value for sufficiently large numbers of M. If the mean constraint value is close to the constraint boundary, the constraint values will be both bigger and smaller than the constraint boundary. Therefore, the same design can be considered infeasible or feasible, because of the stochasticity.

The feasibility of $\mathbf{x}^{(p,q)}$ for constraint $E(G_j) \leqslant c_j$ depends on the difference between $\bar{g}_j^{(p,q)}$ and the constraint bound c_j. Since, the constraint values are stochastic, this difference is stochastic as well. We consider a design feasible if it is statistically likely to be feasible.

For normally distributed constraint values, a statistical hypothesis test shows with which probability the expected constraint value $E(G_j)$ is smaller than the constraint value. The difference between mean constraint value and boundary divided by the standard deviation of the constraint values follows a student-t distribution. We call this quotient the safety index for constraints, and compare it with the student-t distribution to find the probability with which the design is feasible. The safety index for constraint G_j in cycle p and iteration q is defined as follows:

$$\beta_{g_j}^{(p,q)} = \frac{\bar{g}_j^{(p,q)} - c_j}{s_{g_j}^{(p,q)} / \sqrt{M}} \quad j = 1, \ldots, m \tag{1}$$

where the mean $\bar{g}^{(p,q)}$ and the standard deviation $s_{g_j}^{(p,q)}$ are given by:

$$\bar{g}_j^{(p,q)} = \sum_{k=1}^{M} g_{jk}^{(p,q)}/M,$$

$$s_{g_j}^{(p,q)} = \sqrt{\sum_{k=1}^{M}(g_{jk}^{(p,q)} - \bar{g}_j^{(p,q)})^2/(M-1)}.$$

The numerator in (1) is the difference between mean and constraint bound. The denominator is the error with which the mean $\bar{g}_j^{(p,q)}$ is known. This error depends on the standard deviation of the constraint values and the number of replications M. For larger M, the estimate of $\bar{g}_j^{(p,q)}$ becomes more accurate.

For each solution $\mathbf{x}^{(p,q)}$ of the approximate optimization problem, M simulation experiments are carried out to calculate the constraint values and the safety index. We consider the constraint, $E(G_j) \leqslant c_j$, inactive if the safety index is smaller than some specified margin: $\beta_{gj}^{(p,q)} < -\beta_g^{\text{spec}}$. Similarly, the constraint is considered violated if $\beta_{gj}^{(p,q)} > \beta_g^{\text{spec}}$. In the remaining case, $-\beta_g^{\text{spec}} \leq \beta_{g_j}^{(p,q)} \leq -\beta_g^{\text{spec}}$, we consider the constraint active. A design is thus considered feasible, if $\beta_{gj}^{(p,q)} \leqslant -\beta_g^{\text{spec}} \forall j \in \{1, \ldots, m\}$.

We assumed the constraint values g_{jk} to be symmetrically distributed. If the distribution is not symmetrical, the left and right margins on the safety index should be chosen to be unequal. Furthermore, the value of β_g^{spec} should be chosen in accordance with the distribution of the constraint function G_j, and the probability with which feasibility is required. For sufficient large M and a normal distributed constraint value, $\beta_g^{\text{spec}} = 2$ yields a probability of 2.5% or less that $E(G_j) \leqslant c_j$ is not satisfied.

3.4.2 Correcting Constraints

When solving the approximate optimization problem, the objective is to find a solution as close to the constraint boundaries as possible. However, because of the safety index, a certain distance between the design and the constraint boundaries is needed for feasibility. To increase the probability of finding feasible designs, the constraint boundaries $\mathbf{c}$ in the approximated optimization problem $\tilde{P}^{(p,q)}$ are tightened with a correction factor. The corrected constraint bound is defined as follows:

$$\tilde{c}_j = c_j - \beta_g^{\text{spec}} \cdot s_{g_j}(\mathbf{x}^{(p,q-1)})/\sqrt{M}, \quad j = 1, \ldots, m.$$

The correction factor on the constraint boundaries equals the specified margin on the safety index multiplied by the standard deviation of the mean. We assume that this standard deviation can depend on the design variable values. Therefore, the standard deviation which is used to correct the constraint boundaries should be equal to the standard deviation in the approximate optimum solution $\mathbf{x}^{(p,q)}$. Since this solution is still unknown, the standard deviation calculated for the previous solution $\mathbf{x}^{(p,q-1)}$ is used as an estimate.

3.4.3 Stochastic Objective Functions

Stochastic objective functions are treated similarly to stochastic constraint functions. Stochasticity prevents improvements in the objective function value to be calculated from the difference between the current and the previous value. The current mean objective function value $\bar{f}_2$ is compared with the previous mean objective function value $\bar{f}_1$ instead. Therefore, we define the following safety index for objective functions:

$$\beta_f = \frac{\bar{f}_2 - \bar{f}_1}{s_f\sqrt{(2/M)}}, \tag{2}$$

with

$$s_f = \sqrt{\frac{s_{f_1}^2 + s_{f_2}^2}{2}},$$

$$s_{f_{1,2}} = \sqrt{\sum_{k=1}^{M}(f_{1,2\ k} - \bar{f}_{1,2})^2/(M-1)}.$$

Mean objective function values ($\bar{f}_1$ and $\bar{f}_2$) and corresponding standard deviations (s_{f1} and s_{f2}) are both based on the M experiments carried out in each of the two design points.

We consider the objective function decreased, if $\beta_f < -\beta_f^{\text{spec}}$. If $-\beta_f^{\text{spec}} \leqslant \beta_f \leqslant \beta_f^{\text{spec}}$, the objective function value is considered unchanged. The objective function value is considered increased if $\beta_f > \beta_f^{\text{spec}}$.

If the objective function values, f_1 and f_2, are normally distributed, the index β_f follows a student-t distribution. In that case, β_f is related to the probability with which the mean objective function values are different, given the number of replications M. For large enough M and an index value smaller than -2, the probability with which the objective function has not decreased is less than 2.5%. For a value between -2 and 2, the objective function values have to be considered equal, with a 5% probability of error. If the index value is larger than 2, the probability with which the objective function has not increased is less than 2.5%. Similarly to selecting β_g^{spec}, care should be taken, since the condition of normality may not hold.

3.4.4 Using Safety Indices to Accept Designs

Combining the safety indices for both objective function and constraints, a design is accepted if the following conditions hold:

- $\max_{j=1,\ldots,m}(\beta_{gj}^{(p,q)}) < \max_{j=1,\ldots,m}(\beta_{g}^{\text{spec}})$
- the design was not previously found
- $\beta_f^{(p,q)} < \beta_f^{\text{spec}}$.

If the initial design is infeasible, the approximate optimization problem may not have a feasible solution in the search region. Slack variables then relax the constraint bounds until the design closest to the feasible domain is found. This (still infeasible) approximate optimum design will be accepted, if the following conditions are satisfied.

- $\max_{j=1,\ldots,m}(\beta_{gj}^{(p,q)}) < \max_{j=1,\ldots,m}(\beta_{gj}^{(p,0)})$
- the design was not previously found.

3.5 Stopping Rules

Optimization should stop once the optimum solution has been found. Since the true optimum is often unknown, optimization stops when it is unlikely that further improvement will be found. Concretely, optimization stops, if $\mathbf{x}^{(p,q)}$ has been accepted and one of the following stopping criteria is met:

- $\max(\mathbf{m}^{(p,q)}) < 2$ and directional method $= 2$
- $-\beta_f^{\text{spec}} < \beta_f^{(p,q)} < \beta_f^{\text{spec}}$, $-\beta_f^{\text{spec}} < \beta_f^{(p,0)} < \beta_f^{\text{spec}}$ and $\mathbf{x}^{(p,q)}$ is feasible.

The move limits are reduced, if the search region is positioned according to directional method 2. If the move limit parameters $\mathbf{m}^{(p,q)}$ become smaller than 2, the plan points of the experimental design cannot be placed with the start design as the centre any more. The boundaries of the search region then lay between two integer points. So, we assume that further improvement of the current design is not possible and the optimization stops.

Secondly, optimization is stopped if the objective function value does not significantly decreases or increases. We assume this is the case if the indices, $\beta_f^{(p,q)}$ of the cycle optimum and $\beta_f^{(p,0)}$ of the cycle start design, both lay in the range $< -\beta_f^{\text{spec}}, \beta_f^{\text{spec}} >$. Precondition is that $\mathbf{x}^{(p,q)}$ is feasible and has been accepted.

4 TEST PROBLEMS

We will illustrate the optimization approach, described in the previous sections, using two small production line problems: one analytic example with two design variables and one simulation-based example with four variables. The optimization problems contain stochastic constraint functions. The objective function is in both cases deterministic.

For stochastic problems, the optimization strategy may find multiple solutions to be optimal. To analyse these solutions, the following definitions from Loh and Papalambros (1991) for a deterministic discrete non linear programming problem are helpful. For the stochastic case, the expressions ‘feasible’ and ‘smaller than or equal to’ should be interpreted in the sense of subsection Accepting Designs.

Definition 1 *The discrete neighbourhood of a point* $\mathbf{x}$ *is defined as the set of all points* $\mathbf{y}$*, whose discrete components differ +1, 0, or -1 discrete units from the corresponding components of* $\mathbf{x}$ *,* $\mathbf{x}$ *itself being excluded from its own discrete neighbourhood. Formally, this set is then defined as*

$$DN(\mathbf{x}) = \{\mathbf{y} : |y_i - x_i| = 1 \text{ or } 0 \text{ discrete units}, \quad i = 1, \ldots, n; \mathbf{y} \neq \mathbf{x}\}$$

Definition 2 *The point* $\mathbf{x}$ *is said to be a local optimum for a problem* $DDNLP$*, if* $\mathbf{x}$ *is feasible for problem* $DDNLP$ *and* $f(\mathbf{x}) \leqslant f(\mathbf{y})$ *for all feasible* $\mathbf{y}$ *contained in* $DN(\mathbf{x})$.

4.1 Two-Station Flow Line (Analytic)

A production flow line consists of two workstations W_1 and W_2. See Figure 3. Workstation W_1 has x_1 identical machines with a mean process time of $t_{e1} = 0.12$ hour and a squared coefficient of variation of $c_{e1}^2 = 4.0$. The second workstation W_2 has x_2 machines with $t_{e2} = 0.18$ hour and $c_{e2}^2 = 2.25$. Each workstation has a single infinite buffer to temporarily store jobs that need to be processed. Jobs arrive at the first workstation with a mean inter arrival time of $t_a = 0.05$ hour and a squared coefficient of variation of $c_a^2 = 0.25$ hour.

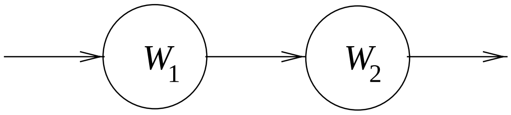


Figure 3: Two-Station Flow Line

The optimization problem is defined as:

$$\begin{array}{ll} \text{Minimize} & f = 5.0 + 1.6x_1 + x_2 \\ \text{subject to} & E(G_1) \leqslant 0.5, \\ & E(G_2) \leqslant 2.0, \\ & x_1, x_2 \in \mathbb{N}^+, \\ & x_1 > 2.4, x_2 > 3.6, \end{array} \tag{3}$$

$$\text{with}\quad \begin{aligned} G_1 &= \phi_{tot} + \epsilon_1, \\ G_2 &= w_2 + \epsilon_2, \\ \epsilon_1 &\in N(0, 0.05\phi_{tot}), \\ \epsilon_2 &\in N(0, 0.05w_2). \end{aligned}$$

Herein, objective function f represents the installation costs being the sum of fixed and variable cost. The first constraint function constrains the average total throughput time to a maximum of 0.5 hour. The second constraint limits the maximum average number of jobs waiting in the queue of the second workstation. Both the total throughput time ϕ_{tot} and mean queue-length w_2 are computed using Little's Law and the following two relations from Hopp and Spearman (1996) for a G/G/m-queueing system:

$$\phi_q = \frac{c_a^2 + c_e^2}{2} \frac{u^{(\sqrt{2(m+1)}-1)}}{m(1-u)} t_e,$$

$$c_d^2 = 1 + (1-u^2)(c_a^2 - 1) + \frac{u^2}{\sqrt{m}}(c_e^2 - 1),$$

with m being the number of machines in a station, u the utilization, and c_d the departure coefficient of variation. Stochastic responses G_1 and G_2 are obtained by adding a normally distributed error term to ϕ_{tot} and w_2. The standard deviations are 5% of the deterministic ϕ_{tot} and w_2 values, which means that the standard deviations depend on $\mathbf{x}$.

The deterministic optimization problem is visualized in Figure 4. The feasible domain is the upper right part of the plot. The optimum design is (4,5). This deterministic problem is solved starting from each discrete point in the design space $3 \leqslant x_1 \leqslant 13$, $4 \leqslant x_1 \leqslant 14$, with initial size of the search region $m_{1,2}^{(0,0)} = 4$. The number of experiments N to build to linear approximations is four, which gives an experiment in each corner. About 88% of the runs converged to (4,5), the other runs stopped in (4,6). Ending in (4,6) instead of (4,5) is explained by the local inaccuracy of the multipoint linear constraint approximations and the fact that (4,5) lies close to the bound of the second constraint. Depending on the position of (the corners of) the search region the linear constraint approximations may exclude (4,5) from the feasible domain.

The local inaccuracy of the multipoint linear approximations becomes worse when the functions behave more non-linearly. This influences the optimization performance of the sequential linearization approach. This can be illustrated using the two-station problem by taking the mean process time of the first work station equal to 0.14 hour instead of 0.12 hour. Then the discrete value of $x_1 = 3$ comes much closer to the bound of $x_1 > 2.8$ where the utilization of W_1 approaches one and the curve of the throughput time increases in a highly nonlinear fashion. For this slightly different problem the true optimum solution becomes (4,6) (see Figure 5). Running the sequential linearization approach starting from the same grid of initial designs gives for 32% of the runs optimum (4,6), 44% gives (4,7), 12% design (4,8), and the other 12% ends in (4,9), (4,10), (4,11), (4,12), or (4,13). An optimization run that has one of these points as intermediate approximate optimum solution will probably end there because, according to the linear approximation of the throughput time constraint, the point is infeasible, while it is actually feasible.

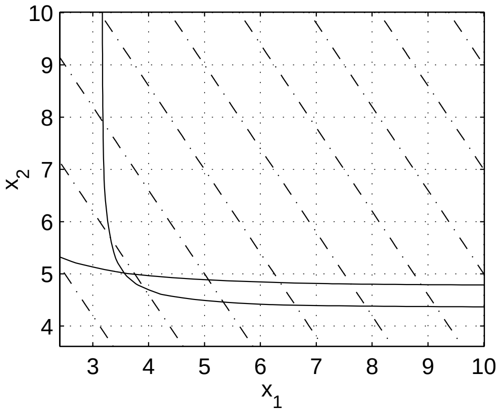


Figure 4: Deterministic Two-Station Flow Line Problem for $t_{e1} = 0.12$ hour

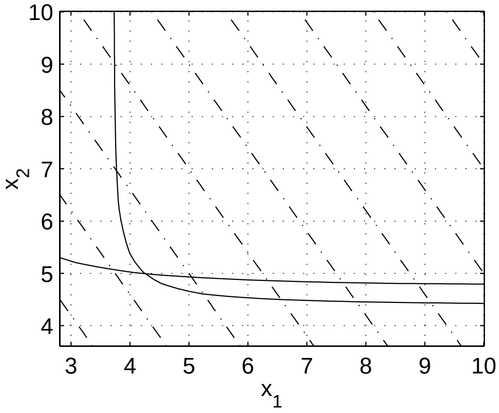


Figure 5: Deterministic Two-Station Flow Line Problem for $t_{e1} = 0.14$ hour

Stochastic optimization problem (4.1) with $t_{e1} = 0.12$ hour has been solved for several values of N and M. Twenty optimization runs are started from the previously mentioned grid of initial designs. Parameter β_g^{spec} is set to 2; the objective function is deterministic. The initial size of the search region is 4. The outcome of the optimization runs is summarized in Table 1.

Increasing N improves the linear approximation. By increasing N the frequency with which the true optimum (4,5) is found instead of its neighbor (4,6) increases. For $N = 32$ and $M = 25$ the results compare with the outcome of the deterministic experiment. Parameter M has main influence on the number of designs other than (4,5) and (4,6)

Table 1: Results Stochastic Two-Station Problem (4.1).

No.		No. replications M		
points	Design	5	10	25
N		[%]	[%]	[%]
4	(4,5)	66.2	79.7	81.6
	(4,6)	22.8	17.1	17.3
	other	11.0	3.2	1.1
8	(4,5)	73.0	84.5	86.8
	(4,6)	18.3	13.6	12.8
	other	8.7	1.9	0.4
32	(4,5)	77.9	87.3	88.4
	(4,6)	13.0	11.6	11.5
	other	9.1	1.1	0.1

that is found. Increasing M decreases the number of other designs found. For higher M the standard deviations are better estimated, and the required safety margins decrease.

4.2 Four-Station Production Line (Simulation)

We consider a four-station production flow line (see Figure 6) with a negative exponential arrival pattern of 2.5 jobs an hour. Each work centre consists of x_i identical machines with single infinite buffers. Objective is to determine the number of machines in each work centre that yields the minimum cost solution and a maximum average throughput time of 6.0 hours.

For each work centre the following is given (see Table 2): fixed costs FC_i, unit costs UC_i, mean process times MPT_i and their coefficients of variance SCV_i.

Table 2: Data of the Line Design Problem (taken from Hopp and Spearman (1996))

Station	Fixed cost	Unit cost	MPT	SCV
	[\$1,000]	[\$1,000]	[hrs]	[-]
1	225	100	1.50	1.00
2	150	155	0.78	1.00
3	200	90	1.10	3.14
4	250	130	1.60	0.10

Utilization of each work centre should remain below 1 to avoid stock piling. Thus, the installed machine capacity x_i/MPT_i should be larger than the arrival rate $r_a = 2.5$ jobs/hour. This leads to the constraint:

$$u_i = \frac{r_a MPT_i}{x_i} < 1 \; \forall \; i \in \{1, 2, 3, 4\}$$

Table 3: Minimum Cost Utilization Feasible Configuration

Station	Machines	Utilization	Cost [\$1,000]
1	4	0.94	625
2	2	0.98	460
3	3	0.92	470
4	5	0.80	900
Total			2,455

Using this constraint, the minimum number of machines needed at each work centre can be calculated. This minimum cost utilization feasible solution is presented in Table 3. However, this solution may not satisfy the throughput time constraint. The optimization problem can now be formulated as follows:

$$\begin{array}{ll} \text{Minimize:} & F = \sum_{i=1}^{4} FC_i + \sum_{i=1}^{4} UC_i x_i \qquad x_i \in \mathbb{Z}^+ \\ \text{subject to:} & E(THT) \leq 6.0 \\ & x_1 \geq 4, \quad x_2 \geq 2, \quad , x_3 \geq 3, \quad , x_4 \geq 5. \end{array}$$

A simulation model is used to estimate the expected value of the throughput time $E(THT)$. In this model, gamma distributions, with means and variances in accordance with Table 2, model the process times.

An estimate of the mean throughput time is made by calculating the average throughput after running the simulation model for a sufficient length of time. The effect of simulation run length on the average throughput time is studied using only two design variables. We introduce the following variables: $x'_1 = x_1 = x_2$ and $x'_2 = x_3 = x_4$. Figure 7 shows the contour lines of the average throughput time as a function of x'_1 and x'_2 for run lengths of: 1,000, 10,000 and 50,000 jobs.

Contour lines in Figure 7 are lines of constant average throughput time. A constraint that requires the average throughput time to be smaller than some constant value looks like a contour line. Comparing Figures 7(a), 7(b) and 7(c), we see the contour lines and the constraints become smoother as the run length increases. For the optimization we chose a run length of 50,000 jobs.

Fifty optimization runs were carried out, starting from the minimum cost utilization feasible design $x_0 = (4, 2, 3, 5)$ with initial move limit parameter $m_i^{(0,0)} = 4$. Both the number of experiments N and the number of replications M were taken equal to 15. The margin on the safety index β_g^{spec} was set to 2.

Table 4 shows the solutions that resulted from these optimization runs. On average, the optimization run converges within four to five cycles. Nine different solutions have been found. The design variable values of these solutions and their frequencies of occurrence are shown in Table 5.

Combining the most frequently occurring design variable values (Table 5), the optimum solution is $\mathbf{x} = (6, 3, 6, 6)$. This is also the most often found solution

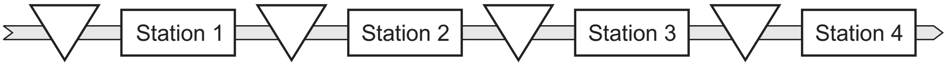
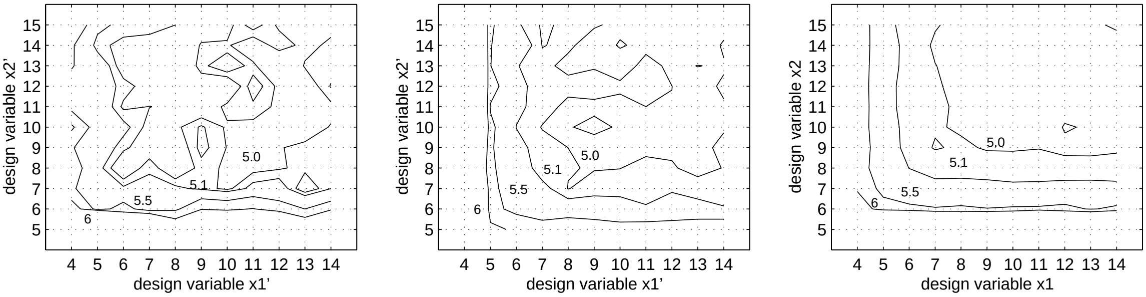


Figure 6: Four Station Production Flow Line



(a) Run length = 1,000 jobs (b) Run length = 10,000 jobs (c) Run length = 50,000 jobs

Figure 7: Contour Lines of the Calculated Mean Cycle Time of the Four-Station Production Flow Line with $x_1' = x_1 = x_2$ and $x_2' = x_3 = x_4$ for Several Simulation Run Lengths.

Table 4: Calculated Optimum Designs for the Four-Station Flow Line

Design				Frequency
x_1	x_2	x_3	x_4	
6	3	6	6	18
6	4	5	5	10
7	3	5	6	6
6	4	5	7	4
7	5	6	5	4
7	3	6	5	3
5	3	6	6	2
5	4	5	6	2
6	4	6	5	1

Table 5: Frequencies of Design Variable Values.

Value	Frequency			
	x_1	x_2	x_3	x_4
3	0	29	0	0
4	0	17	0	0
5	4	4	22	18
6	33	0	28	28
7	13	0	0	4

according to Table 4. Nearly all other solutions, that were found, are members of its discrete neighbourhood.

5 CONCLUSIONS

We have presented an optimization strategy to solve simulation-based optimization problems, using a series of approximate subproblems. Optimal design of experiments is used to plan the simulation experiments, carried out at the discrete points of the search region. Safety indices for constraints and the objective function deal with the stochasticity present in simulation models. A move limit strategy redefines the size and position of the search region during the optimization.

The number of experiments N planned in the search region controls the accuracy of the approximations. However, linear approximations of objective and constraint functions are used, and the experiments are limited to integer values for the design variables. Therefore the approximations are unable to follow the local curvature of non-linear objective and constraint functions. As a result, an optimization run can end at a non-optimal point in the deterministic case, because of the limited accuracy of the approximation. Stochasticity also causes multiple solutions to be found.

Optimization results of two test problems have been presented. The analytical two-station flow line shows the effect of the linear approximations and the role of parameters N and M. The four-station production line is a successful application of the optimization strategy to a simple (stochastic) simulation model. In about thirty six per cent of the optimization runs, the same final optimum solution is found; almost all other solutions found are discrete neighbors. For this problem it can be shown that variance changes as a function of the design variable values. This underlines the importance of recalculating the safety indices for each approximate optimum solution.

REFERENCES

Arora, J. S., M. W. Huang, and C. C. Hsieh. 1994. Methods for optimization of nonlinear problems with discrete design variables: a review. *Structural Optimization* 8: 69–85.

Barthelemy, J. -F. M., and R. T. Haftka. 1993. Approximation concepts for optimum structural design. *Structural Optimization* 5: 129–144.

Carson, Y., and A. Maria. 1997. Simulation optimization: methods and applications. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 118–126. Institute of Electrical and Electronics Engineers, Piscataway, New Yersey.

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

Hopp, W. J., and M. L. Spearman. 1996. *Factory Physics - Foundations of Manufacturing Management*. Chicago: Irwin.

Lemmen, B., E. J. J. van Campen, H. Roede, and J. E. Rooda. 1999. Clustertool optimization through scheduling rules. In *Proceedings of the Eight International Symposium on Semiconductor Manufacturing*, ed. B. Sohn et al., 89–92.

Loh, H. T., and P. Y. Papalambros. 1991. A sequential linearization approach for solving mixed discrete nonlinear design optimization problems. *Journal of Mechanical Design* 113: 325–334.

Rulkens, H. J. A., E. J. J. van Campen, J. van Herk, and J. E. Rooda. 1998. Batch size optimization of a furnace and pre-clean area by using dynamic simulations. In *Proceedings of the 1998 SEMI/IEEE Advanced Semiconductor Manufacturing Conference*, ed. S. McClure, J. Goodman, and K. Conway, 439–444.

Thanedar, P. B. and G. N. Vanderplaats. 1995. Survey of discrete variable optimization for structural design. *Journal of Structural Design* 121: 301–306.

AUTHOR BIOGRAPHIES

S.J. ABSPOEL currently is a trainee of the International Program in Logistics Management Systems at Eindhoven University of Technology. He received a M.S. degree in mechanical engineering from Eindhoven University of Technology. His email address is `<s.j.abspoel@tm.tue.nl>`.

L.F.P. ETMAN (corresponding author) is assistant professor in the Department of Mechanical Engineering at Eindhoven University of Technology. He received M.S. and Ph.D. degrees in mechanical engineering from Eindhoven University of Technology. His email and web addresses are `<l.f.p.etman@tue.nl>` and `<se.wtb.tue.nl/˜pascal/>`.

J. VERVOORT is a graduate student with the Systems Engineering Group in the Department of Mechanical Engineering at Eindhoven University of Technology. His email and web addresses are `<j.vervoort@stud.tue.nl>` and `<se.wtb.tue.nl/˜vervoort/>`.

J.E. ROODA is professor in Systems Engineering in the Department of Mechanical Engineering at Eindhoven University of Technology. He received a M.S. degree in food processing technology from Wageningen University of Agriculture, and a Ph.D. degree in mechanical engineering from Twente University. His email and web addresses are `<j.e.rooda@tue.nl>` and `<se.wtb.tue.nl/>`.