Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

DISCRETE OPTIMIZATION VIA SIMULATION USING GAUSSIAN MARKOV RANDOM FIELDS

Peter Salemi
Barry L. Nelson
Jeremy Staum

Department of Industrial Engineering and Management Sciences
Northwestern University
2145 Sheridan Road
Evanston, IL, 60208-3119, U.S.A.

ABSTRACT

We construct a discrete optimization via simulation (DOvS) procedure using discrete Gaussian Markov random fields (GMRFs). Gaussian random fields (GRFs) are used in DOvS to balance exploration and exploitation. They enable computation of the expected improvement (EI) due to running the simulation to evaluate a feasible point of the optimization problem. Existing methods use GRFs with a continuous domain, which leads to dense covariance matrices, and therefore can be ill-suited for large-scale problems due to slow and ill-conditioned numerical computations. The use of GMRFs leads to sparse precision matrices, on which several sparse matrix techniques can be applied. To allocate the simulation effort throughout the procedure, we introduce a new EI criterion that incorporates the uncertainty in stochastic simulation by treating the value at the current optimal point as a random variable.

1 INTRODUCTION

Optimization plays a central role in the fields of industrial engineering and operations research. For many interesting optimization problems that occur in practice, the objective function cannot be evaluated explicitly and requires estimation using stochastic simulation. For example, suppose we are interested in minimizing the expected waiting time for an arbitrary customer calling into a call center with constraints on the service levels and number of servers in the call center. When the structure of the call center is too complicated to analyze analytically, this leads to a constrained optimization problem with an objective function that can only be estimated using stochastic simulation. Problems such as these are commonly referred to as optimization via simulation (OvS) problems. Furthermore, OvS problems in which the decision variables can only assume discrete values, such as those constrained to be integer valued or binary, are called discrete optimzation via simulation (DOvS) problems. In this paper, we focus on DOvS problems with integer-ordered decision variables.

DOvS has been an active area of research for many years, and the literature on DOvS is plentiful. The majority of DOvS methods fall into the category of adaptive random search (ARS) algorithms. These ARS algorithms contain a mixture of both local and global searching in an effort to explore favorable areas of the feasible regions in depth (local searching), while continuing to explore other areas of the feasible region to avoid getting trapped at a local optimum (global searching). Examples of ARS algorithms include the stochastic ruler method, simulated annealing, industrial-strength COMPASS, model reference adaptive search, and nested partitions (see, for example, Nelson (2010)).

We are interested in a procedure that automatically regulates exploration and exploitation. Furthermore, we want a procedure that utilizes a stopping criterion that considers the uncertainty at feasible points not

yet simulated. Several DOvS methods model the simulation output at the feasible points as a realization of a Gaussian random field (GRF). By imposing this uncertainty on the DOvS problem, concepts such as expected improvement (EI) and value of information can be used to determine which feasible point to simulate next. Furthermore, we can use this uncertainty to decide when we have sufficiently searched the feasible region and stop the procedure. Initially, GRF-based optimization methods were introduced for deterministic computer experiments in Jones, Schonlau, and Welch (1998), followed by Huang, Allen, Notz, and Zeng (2006) which considered computer experiments with noisy output. Recently, Frazier, Xie, and Chick (2011) and Quan, Yin, Ng, and Lee (2013) introduced GRF-based optimization methods into DOvS. Other GRF-based methods include Xu (2012), which combines an ARS method with stochastic kriging (Ankenman, Nelson, and Staum 2010) to help increase the efficiency of the ARS method, and Gaussian process-based search (Sun, Hong, and Hu 2011), which uses the GRF to create a sampling distribution that favors feasible points most likely to be optimal. All of these GRF-based methods use a GRF with a continuous domain and then project the problem into the discrete setting by only considering points that fall on the integer lattice. The use of a GRF with a continuous domain leads to a dense covariance matrix, which can be ill-suited for large-scale problems with many feasible solutions due to slow and ill-conditioned numerical computations.

We approach the DOvS problem by modeling the objective function values at the feasible points as a realization of a discrete Gaussian Markov random field (GMRF) (Rue and Held 2005). The Markov structure of GMRFs is intuitive for problems in industrial engineering and operations research (Salemi, Staum, and Nelson 2013). For example, if we were interested in predicting the value of the objective function at a feasible prediction point, then the values of the objective function at the feasible points in its neighborhood would typically be sufficient; others would provide very little information. GMRFs are typically defined on lattices, so the use of GMRFs in DOvS problems is more natural than using a GRF with a continuous domain. Most importantly, the Markov structure lends itself to more efficient and numerically stable calculations. A GMRF is defined by its precision matrix, which is the inverse of the covariance matrix. Using the Markov structure of GMRFs, the precision matrix of a GMRF can be constructed to be sparse. Thus, we can use several sparse matrix techniques to calculate expressions which involve the precision matrix. The size of the feasible region we are able to handle with the procedure presented in this paper depends on the structure of the feasible region of the DOvS problem, but should not be too large so that even sparse matrix techniques cannot handle them.

To allocate the simulation effort throughout the procedure, we utilize an EI criterion. Jones, Schonlau, and Welch (1998) first introduced an EI criterion for optimization problems in the field of deterministic computer experiments. In their case, the objective function could be observed without noise, so their EI criterion would not incorporate the uncertainty in the simulation output from a stochastic simulation. Huang, Allen, Notz, and Zeng (2006) and Quan, Yin, Ng, and Lee (2013) introduced other EI criteria for the case of noisy output; however, both criteria treat the value at the current optimal point as fixed instead of as a random variable. We introduce an EI criterion similar to that of Williams, Santner, and Notz (2000), which we call complete expected improvement (CEI). CEI incorporates the uncertainty in stochastic simulation by treating the value at the current optimal point as a random variable, as well as the feasible point being considered. The main difference between the EI criterion of Williams, Santner, and Notz (2000) and our CEI is the origin of the uncertainty. Williams, Santner, and Notz (2000) dealt with deterministic computer experiments with the uncertainty being introduced by inputs that contain noise, whereas our uncertainty is introduced by the stochastic noise in the simulation output. We use the CEI in our stopping criterion to determine when we have sufficiently searched the feasible region, by stopping when the CEI is too small to warrant more searching. Similar to EI criteria, the Knowledge Gradient (KG) policy (Frazier, Powell, and Dayanik 2009) is also used to guide the allocation of simulation effort and automatically regulate the exploration vs. exploitation trade-off. However, the KG policy determines the benefit of running another replication at a feasible point, whereas EI criteria estimate how much better the unknown objective function value at another feasible point is as compared to the current optimal.

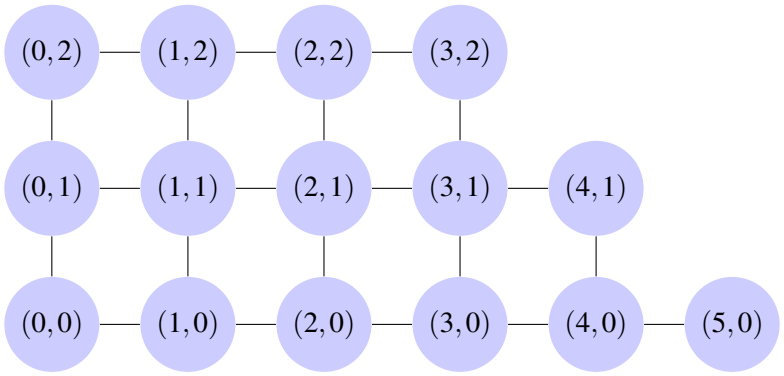


Figure 1: Example of a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ corresponding to a subset of the integer lattice $\mathbb{Z}^2$.

In the next section, we discuss GMRFs along with some of the methodology we will need for our procedure. In Section 3, we present an outline of our DOvS procedure and discuss several aspects of it in more detail, including parameter estimation and CEI. We present results of numerical experiments in Section 4.

2 DOvS USING GMRFs

A GMRF is a non-degenerate multivariate Gaussian random vector $\mathbb{Y} = (\mathbb{Y}_1, \mathbb{Y}_2, \ldots, \mathbb{Y}_n)^\top$ which is associated with an undirected and labelled graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ denotes the set of nodes and $\mathscr{E}$ denotes the set of edges. Each node in $\mathscr{V}$ is associated with a unique element of $\mathbb{Y}$, i.e., there is a one-to-one correspondence between the nodes of $\mathscr{V}$ and the elements of $\mathbb{Y}$. Two nodes in the graph are called neighbors if they are connected by an edge. If we denote the mean vector and precision matrix of $\mathbb{Y}$ by μ and $\mathbf{Q}$, respectively, then we can write the probability density function of the GMRF as

$$(2\pi)^{-n/2}|\mathbf{Q}|^{1/2}\exp\left(-\frac{1}{2}(\mathbf{y}-\mu)^\top\mathbf{Q}(\mathbf{y}-\mu)\right),$$

where the positive-definite precision matrix $\mathbf{Q}$ is the inverse of the covariance matrix. The entries of the precision matrix can be interpreted in terms of conditional precisions

$$\text{Prec}(\mathbb{Y}_i|\mathbb{Y}_{\mathscr{V}\setminus\{i\}}) = \mathrm{Q}_{ii}, \tag{1}$$

where $\mathbb{Y}_{\mathscr{V}\setminus\{i\}}$ is the set of nodes in $\mathscr{V}\setminus\{i\}$, and partial correlations

$$\text{Corr}(\mathbb{Y}_i, \mathbb{Y}_j|\mathbb{Y}_{\mathscr{V}\setminus\{i,j\}}) = -\frac{\mathrm{Q}_{ij}}{\sqrt{\mathrm{Q}_{ii}\mathrm{Q}_{jj}}}, \tag{2}$$

where $\mathbb{Y}_{\mathscr{V}\setminus\{i,j\}}$ is the set of nodes in $\mathscr{V}\setminus\{i,j\}$. The graph $\mathscr{G}$ determines the non-zero pattern of the precision matrix $\mathbf{Q}$, and vice versa, since for a GMRF $\mathbf{Q}_{ij} \neq 0$ if and only if $\{i,j\} \in \mathscr{E}$. Thus, the precision matrix will be sparse if the set of edges is small in size, and vice versa. GMRFs get the name "Markov" because they possess the local Markov property:

$$\mathbb{Y}_i \perp \mathbb{Y}_{\mathscr{V}\setminus\{i,\mathscr{N}_i\}}|\mathbb{Y}_{\mathscr{N}_i} \quad \text{for every} \quad i \in \mathscr{V},$$

where $\mathscr{N}_i$ is the set of neighbors of node i in $\mathscr{V}$. To better understand the local Markov property, consider the graph $\mathscr{G}$ in Figure 1. If we observe the value of the GMRF at $(1,0)$, $(3,0)$, and $(2,1)$, then our prediction of the value of the GMRF at $(2,0)$ is not affected by information about the GMRF at any other lattice point in the graph. The Markov property does not imply that nodes far away from one another are independent, but rather if we know the value of the GMRF at nodes close by, then we can ignore nodes further away. To completely specify a GMRF, we only need to specify the mean μ and the precision matrix $\mathbf{Q}$, whose non-zero pattern is associated with the structure of the graph $\mathscr{G}$.

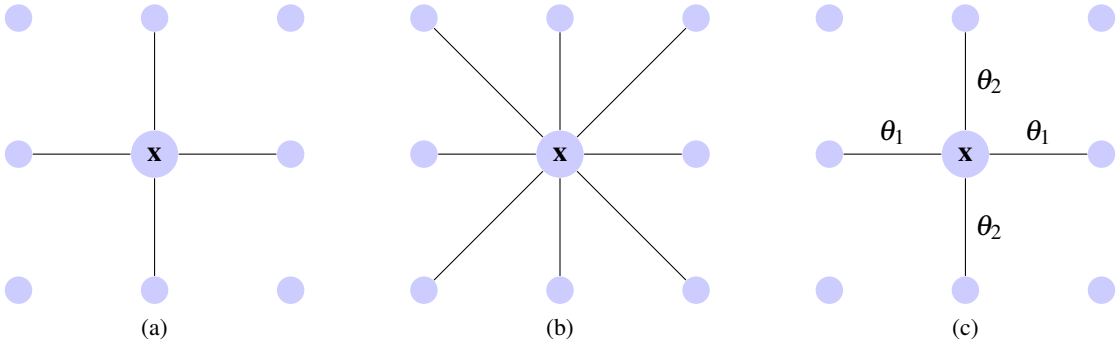


Figure 2: (a) and (b) are two possible neighborhood structures for a lattice point $\mathbf{x}$ in a two-dimensional integer lattice; (c) shows parameter values assigned to each edge in (a) using the function p.

2.1 Parameterization

In our DOvS problem with integer-ordered decision variables, the feasible region $\mathbb{X}$ is a finite subset of the d-dimensional integer lattice $\mathbb{Z}^d$. Thus, the straightforward construction of the graph $\mathscr{G}$ starts with defining the nodes of the graph to be the lattice points in the feasible region. To finish the construction of $\mathscr{G}$, we must specify when two nodes will be connected with an edge, which we do by specifying the neighborhood structure for any given node. Examples of two-dimensional neighborhood structures are given by Figure 2a and Figure 2b. The neighborhood represented in Figure 2a, given by the set of nodes $\{\mathbf{x}' \in \mathbb{X} : ||\mathbf{x}-\mathbf{x}'||_2 = 1\}$, has an upper bound of $2d$ neighbors in d dimensions. The neighborhood represented in Figure 2b, given by the set of nodes $\{\mathbf{x}' \in \mathbb{X} : ||\mathbf{x}-\mathbf{x}'||_\infty = 1\}$, has an upper bound of 2^{d+1} neighbors in d dimensions. Two nodes are connected by an edge if they are neighbors as determined by the neighborhood structure.

Since we are particularly focused on DOvS problems with large feasible regions, we will use the neighborhood structure represented by Figure 2a. Although we assume a specific neighborhood structure for efficient computations, our method can be used with any neighborhood structure. From the neighborhood structure represented by Figure 2a, the neighbors of a feasible point $\mathbf{x}$ are the set of feasible points $\{\mathbf{x}' \in \mathbb{X} : ||\mathbf{x}-\mathbf{x}'||_2 = 1\}$. This neighborhood structure leads to at most $2d|\mathbb{X}|$ edges in the set $\mathscr{E}$, and thus the fraction of non-zero entries in the precision matrix $\mathbf{Q}$ is bounded from above by $2d/|\mathbb{X}|$. Although we have defined the graph $\mathscr{G}$, and thus the non-zero pattern of the precision matrix, we have not yet specified the non-zero entries of the precision matrix. The non-zero entries of $\mathbf{Q}$ are determined by the function $p(\mathbf{x},\mathbf{x}';\theta)$, where $\theta = (\theta_0,\theta_1,\theta_2,\ldots,\theta_d)^\top$ is a vector of parameters, i.e., the ij^{th} entry of $\mathbf{Q}$ is given by $\mathrm{Q}_{ij} \triangleq p(\mathbf{x}_i,\mathbf{x}_j;\theta)$. We use the function

$$p(\mathbf{x},\mathbf{x}';\theta) = \begin{cases} \theta_0 & \text{if } \mathbf{x} = \mathbf{x}' \\ -\theta_0\theta_j & \text{if } |\mathbf{x}-\mathbf{x}'| = \mathbf{e}_j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

for $\mathbf{x},\mathbf{x}' \in \mathbb{Z}^d$, where $\mathbf{e}_j$ is the jth standard basis vector. Since the conditional precisions in Equation (1) must be positive, it follows that θ_0 must be positive. We also want neighbors to have non-negative partial correlations, given by Equation (2), so we restrict the values of $\theta_1,\theta_2,\ldots,\theta_d$ to be non-negative. Furthermore, we need $\theta_i \leq 1$ for $i = 1,2,\ldots,d$, since the partial correlations must be less than one. Lastly, we need $\mathbf{Q}$ to be positive-definite. We will use these restrictions later in Section 3.2 when we calculate the maximum likelihood estimates (MLEs) of the parameters. With this parameterization and these restrictions, $\mathbf{Q}$ is a non-singular M-matrix so its inverse is nonnegative, i.e., $[\mathbf{Q}^{-1}]_{ij} \geq 0$ for all i and j. In other words, there are no negative correlations among nodes in the GMRF, which is a property that we want. The function p, illustrated in Figure 2c, leads to a nonstationary GMRF; the variances at feasible points with

more neighbors are larger than the variances at feasible points with fewer neighbors. We are not concerned with the nonstationarity of the GMRF, since we only use the conditional variance, given the observed simulation output, in our procedure. We will write $\mathbf{Q}(\theta)$ to explicitly show the dependence of the precision matrix on the parameters.

2.2 Methodology

Consider the DOvS problem

$$\begin{aligned} &\text{minimize } y(\mathbf{x}) \triangleq \mathrm{E}[\mathrm{Y}(\mathbf{x})] \\ &\text{subject to } \mathbf{x} \in \mathbb{X}, \end{aligned}$$

where $\mathbb{X}$ is a finite subset of the d-dimensional integer lattice $\mathbb{Z}^d$ with $n \triangleq |\mathbb{X}|$. The distribution of the random variable $\mathrm{Y}(\mathbf{x})$, as a function of the feasible point $\mathbf{x}$, is unknown. However, the expectation $y(\mathbf{x}) \triangleq \mathrm{E}[\mathrm{Y}(\mathbf{x})]$ can be estimated using stochastic simulation.

Let $\mathbf{y}$ denote the vector of objective function values $(y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_n))^\top$. We model $\mathbf{y}$ as a realization of the GMRF

$$\mathbb{Y} \triangleq (\mathbb{Y}(\mathbf{x}_1), \mathbb{Y}(\mathbf{x}_2), \ldots, \mathbb{Y}(\mathbf{x}_n))^\top \sim \mathscr{N}(\mu, \mathbf{Q}(\theta)^{-1}) \tag{4}$$

with mean vector μ and precision matrix $\mathbf{Q}(\theta)$, as defined above. Assume that we have run the simulation at the feasible points in the set $\Xi_2 \subseteq \mathbb{X}$, which we call the design points. We partition $\mathbb{X}$ into the two disjoint sets Ξ_2 and $\Xi_1 = \mathbb{X} \setminus \Xi_2$. Thus, Ξ_2 is the set of design points at which we have run the simulation, and Ξ_1 is the set of feasible points at which we have not run the simulation. For simplicity, we use 1 as a subscript to denote quantities associated with the set Ξ_1 and 2 as a subscript to denote quantities associated with the set Ξ_2. Using these disjoint sets, we can partition the vectors $\mathbf{y}, \mathbb{Y}, \mu$, and the precision matrix $\mathbf{Q}(\theta)$ and rewrite expression (4) as

$$\begin{pmatrix} \mathbb{Y}_1 \\ \mathbb{Y}_2 \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \mathbf{Q}_{11}(\theta) & \mathbf{Q}_{12}(\theta) \\ \mathbf{Q}_{12}(\theta)^\top & \mathbf{Q}_{22}(\theta) \end{pmatrix}^{-1} \right).$$

Let $\bar{\mathbb{y}}_2$ be the vector of simulation output at the design points. We model the simulation output $\bar{\mathbb{y}}_2$ as a realization of the GMRF $\mathbb{Y}_2^\varepsilon$, which we define by the conditional distribution

$$(\mathbb{Y}_2^\varepsilon - \mathbb{Y}_2) | (\mathbb{Y}_2 = \mathbf{y}_2) \sim \mathscr{N}(\vec{\mathbf{0}}_{n_2 \times 1}, \mathbf{Q}_\varepsilon^{-1}), \tag{5}$$

where the matrix $\mathbf{Q}_\varepsilon$, called the intrinsic precision matrix, is the precision matrix of the intrinsic noise inherent to the stochastic simulation. When we sample independently, $\mathbf{Q}_\varepsilon$ is a diagonal matrix, whereas when we sample with common random numbers (CRN), $\mathbf{Q}_\varepsilon$ is a dense matrix. Let $n_1 \triangleq |\Xi_1|$ and $n_2 \triangleq |\Xi_2|$. The conditional distribution $\mathbb{Y} | \mathbb{Y}_2^\varepsilon = \bar{\mathbb{y}}_2$ is given by

$$\begin{pmatrix} \mathbb{Y}_1 \\ \mathbb{Y}_2 \end{pmatrix} \big| (\mathbb{Y}_2^\varepsilon = \bar{\mathbb{y}}_2) \sim \mathscr{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \bar{\mathbf{Q}}(\theta)^{-1} \begin{pmatrix} \vec{\mathbf{0}}_{n_1 \times 1} \\ \mathbf{Q}_\varepsilon(\bar{\mathbb{y}}_2 - \mu_2) \end{pmatrix}, \bar{\mathbf{Q}}(\theta)^{-1} \right), \tag{6}$$

where $\vec{\mathbf{0}}_{n_1 \times 1}$ is the $n_1 \times 1$ vector of zeros and

$$\bar{\mathbf{Q}}(\theta) \triangleq \begin{pmatrix} \mathbf{Q}_{11}(\theta) & \mathbf{Q}_{12}(\theta) \\ \mathbf{Q}_{12}(\theta)^\top & \mathbf{Q}_{22}(\theta) \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{n_1 \times n_1} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_1 \times n_2}^\top & \mathbf{Q}_\varepsilon \end{pmatrix}$$

is the conditional precision matrix, where $\mathbf{0}_{n_i \times n_j}$ is the $n_i \times n_j$ matrix of zeros. The sparsity of $\bar{\mathbf{Q}}(\theta)$ is inherited directly from the sparsity of $\mathbf{Q}(\theta)$ and $\mathbf{Q}_\varepsilon$. Furthermore, when a direct method, such as the sparse Cholesky factorization, cannot be used to directly invert $\bar{\mathbf{Q}}(\theta)$, we can avoid direct inversion of $\bar{\mathbf{Q}}(\theta)$ by solving a sparse linear system whose coefficient matrix is the sparse matrix $\bar{\mathbf{Q}}(\theta)$.

3 GMRF-DOvS PROCEDURE

In this section, we provide an outline of our GMRF-DOvS procedure and give details for several aspects of the procedure. The outline is given in Section 3.1, followed by estimation of the parameters using the simulation output in Section 3.2. Afterwards, we define the CEI in Section 3.3 and discuss candidate feasible points, which are feasible points for which we calculate the CEI, in Section 3.4.

3.1 GMRF-DOvS Procedure Outline

Here we give a brief outline of the procedure, with details for each step discussed in later sections. Our procedure utilizes the CEI in the stopping criterion, by stopping when the CEI has been below a specified threshold δ for v times in a row. This stopping criterion has been suggested in other EI methods, such as the method in Huang, Allen, Notz, and Zeng (2006).

1. Set $u = 0$. Generate a set of k design points. Simulate r replications at each design point, and use the simulation output to calculate the MLEs for the GMRF parameters using the method in Section 3.2.
2. Calculate the vector of conditional means in the conditional distribution (6) and let $\tilde{\mathbf{x}}$ be the feasible point with the smallest conditional mean value.
3. Select c candidate feasible points, defined in Section 3.4. If the feasible region is not too large, and the CEI can be calculated for each feasible point, then all feasible points are candidates, i.e., $c = n$. If the feasible region is too large, and the CEI cannot be calculated for each feasible point, then select $c < n$ candidate feasible points.
4. Calculate the CEI, defined in Section 3.3, for each candidate feasible point. If the CEI at each candidate feasible point is less than δ, then set $u \leftarrow u + 1$. Otherwise, set $u = 0$. If $u < v$, go to Step 5. Otherwise, go to Step 7.
5. Simulate r replications at the candidate feasible point $\mathbf{x}^*_{\text{CEI}}$ that maximizes the CEI over the set of all candidate feasible points. Note that a feasible point can be selected more than once throughout the procedure, so the cumulative number of replications allocated to a feasible point can be more than r.
6. If $\mathbf{x}^*_{\text{CEI}}$ is a design point: update the simulation output at $\mathbf{x}^*_{\text{CEI}}$ with the r new replications and go to Step 2. If $\mathbf{x}^*_{\text{CEI}}$ is not a design point: add $\mathbf{x}^*_{\text{CEI}}$ to the set of design points, add the simulation output obtained at $\mathbf{x}^*_{\text{CEI}}$ to the entire collection of simulation output, and go to Step 2. *Optional at each iteration: re-calculate the MLEs using the entire collection of simulation output.*
7. Return the simulated feasible point with the smallest conditional mean value.

The inputs to the procedure that must be specified are k, r, c, δ, and v. Although the parameters of the GMRF must be fit during initialization, they do not need to be updated afterwards. We recommend updating the parameters more often during the beginning of the procedure and only sparingly later to reduce computation time.

3.2 Parameter Estimation

To implement GMRFs for DOvS, the mean vector μ needs more structure. We assume that $\mu = (\beta_0, \beta_0, \ldots, \beta_0)^\top$, i.e., we assume a constant trend, which is appropriate when we have no prior information about the objective function values at the feasible points.

The parameters β_0 and θ will be estimated from the simulation output using maximum likelihood estimation. Following Ankenman, Nelson, and Staum (2010), we use a plug-in estimator $\widehat{\mathbf{Q}}_\varepsilon$ for the intrinsic precision matrix $\mathbf{Q}_\varepsilon$. We assume independent sampling, i.e., a diagonal intrinsic precision matrix, since sampling with CRN leads to a dense intrinsic precision matrix and therefore $\bar{\mathbf{Q}}(\theta)$ will not be sparse. Denote the design points by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_2}$, and let $s^2(\mathbf{x}) = (1/r(\mathbf{x})) \sum_{i=1}^{r(\mathbf{x})} \left(\mathrm{Y}_i(\mathbf{x}) - \bar{\mathrm{Y}}(\mathbf{x})\right)^2$, where $r(\mathbf{x})$ is

the cumulative number of replications that have been allocated to design point $\mathbf{x}$, $\mathrm{Y}_i(\mathbf{x})$ is the simulation output at $\mathbf{x}$ on the ith replication, and $\bar{\mathrm{Y}}(\mathbf{x})$ is the sample average of the simulation output across the $r(\mathbf{x})$ replications at design point $\mathbf{x}$. The plug-in estimator for the intrinsic precision matrix is

$$\widehat{\mathbf{Q}}_\varepsilon = \operatorname{diag}\left(\frac{r(\mathbf{x}_1)}{s^2(\mathbf{x}_1)}, \frac{r(\mathbf{x}_2)}{s^2(\mathbf{x}_2)}, \ldots, \frac{r(\mathbf{x}_{n_2})}{s^2(\mathbf{x}_{n_2})}\right).$$

After substituting $\widehat{\mathbf{Q}}_\varepsilon$ for $\mathbf{Q}_\varepsilon$, we integrate the conditional distribution (5) over $\mathbf{y}_2$ to obtain the unconditional distribution

$$\mathbb{Y}_2^\varepsilon \sim \mathscr{N}(\beta_0 \mathbf{1}_{n_2\times 1}, \Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1}),$$

where $\Sigma_{22}(\theta) \triangleq (\mathbf{Q}_{22}(\theta) - \mathbf{Q}_{12}(\theta)^\top \mathbf{Q}_{11}(\theta)^{-1}\mathbf{Q}_{12}(\theta))^{-1}$, and $\mathbf{1}_{n_2\times 1}$ is the $n_2 \times 1$ vector of ones. Disregarding terms that do not depend on β_0 or θ, the log-likelihood function given the simulation output is

$$\begin{aligned}\mathscr{L}(\beta_0, \theta|\bar{\mathbf{y}}_2) \triangleq &\frac{1}{2}\log|(\Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1})^{-1}| \\ &- \frac{1}{2}(\bar{\mathbf{y}}_2 - \beta_0 \mathbf{1}_{n_2\times 1})^\top (\Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1})^{-1}(\bar{\mathbf{y}}_2 - \beta_0 \mathbf{1}_{n_2\times 1}),\end{aligned}$$

where $\bar{\mathbf{y}}_2$ is the vector of simulation output at the design points. Miller (1981) provides a method for efficiently calculating $(\Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1})^{-1}$ using only $\Sigma_{22}(\theta)^{-1}$ and matrix-vector products with the columns of $\widehat{\mathbf{Q}}_\varepsilon^{-1}$. Furthermore, we can calculate $\Sigma_{22}(\theta)^{-1} = \mathbf{Q}_{22}(\theta) - \mathbf{Q}_{12}(\theta)^\top \mathbf{Q}_{11}(\theta)^{-1}\mathbf{Q}_{12}(\theta)$ by solving the sparse linear system $\mathbf{Q}_{11}(\theta)\mathbf{W} = \mathbf{Q}_{12}(\theta)$ and substituting $\mathbf{W}$ for $\mathbf{Q}_{11}(\theta)^{-1}\mathbf{Q}_{12}(\theta)$. Fixing the parameters in θ, the MLE for β_0 is

$$\widehat{\beta}_0(\theta) = (\mathbf{1}_{n_2\times 1}^\top (\Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1})^{-1}\mathbf{1}_{n_2\times 1})^{-1}\mathbf{1}_{n_2\times 1}^\top (\Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1})^{-1}\bar{\mathbf{y}}_2,$$

where we have written $\widehat{\beta}_0$ as a function of θ to explicitly show dependence. Substituting the MLE $\widehat{\beta}_0(\theta)$ into $\mathscr{L}(\beta_0, \theta|\bar{\mathbf{y}}_2)$, we get the profile log-likelihood function

$$\begin{aligned}\mathscr{L}(\theta|\bar{\mathbf{y}}_2) \triangleq &\frac{1}{2}\log|(\Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1})^{-1}| \\ &- \frac{1}{2}(\bar{\mathbf{y}}_2 - \widehat{\beta}_0(\theta)\mathbf{1}_{n_2\times 1})^\top (\Sigma_{22}(\theta) + \widehat{\mathbf{Q}}_\varepsilon^{-1})^{-1}(\bar{\mathbf{y}}_2 - \widehat{\beta}_0(\theta)\mathbf{1}_{n_2\times 1}),\end{aligned}$$

which only depends on θ. The MLE $\widehat{\theta}$ for θ is the maximizer of the profile log-likelihood function $\mathscr{L}(\theta|\bar{\mathbf{y}}_2)$ over the possible values of θ, i.e.,

$$\widehat{\theta} \triangleq \arg\max_{\theta\in\Theta} \left\{\mathscr{L}(\theta|\bar{\mathbf{y}}_2) \,|\, 0 \le \theta_i \le 1 \text{ for } 1 \le i \le d, \theta_0 > 0\right\},$$

where Θ is the set of values of $\theta \in \mathbb{R}^{d+1}$ for which $\mathbf{Q}(\theta)$ is positive-definite.

3.3 Complete Expected Improvement

The EI criterion we introduce in this paper, the CEI, is similar to the EI criterion of Williams, Santner, and Notz (2000). The CEI takes into account both the uncertainty at a candidate feasible point $\mathbf{x}$ and the uncertainty at the current optimal point $\tilde{\mathbf{x}}$, the point with the smallest conditional mean value, by treating the value at both as random variables. Let $\mathbb{Y}(\mathbf{x})$ be the Gaussian random variable from the GMRF $\mathbb{Y}$ corresponding to the feasible point $\mathbf{x}$. Given the simulation output $\bar{\mathbf{y}}_2$, we have the conditional distribution

$$\begin{pmatrix}\mathbb{Y}_{|\bar{\mathbf{y}}_2}(\tilde{\mathbf{x}}) \\ \mathbb{Y}_{|\bar{\mathbf{y}}_2}(\mathbf{x})\end{pmatrix} \triangleq \begin{pmatrix}\mathbb{Y}(\tilde{\mathbf{x}}) \\ \mathbb{Y}(\mathbf{x})\end{pmatrix} \Big| (\mathbb{Y}_2^\varepsilon = \bar{\mathbf{y}}_2) \sim \mathscr{N}\left(\begin{pmatrix}\mu_{|\bar{\mathbf{y}}_2}(\tilde{\mathbf{x}}) \\ \mu_{|\bar{\mathbf{y}}_2}(\mathbf{x})\end{pmatrix}, \begin{pmatrix}\sigma^2_{|\bar{\mathbf{y}}_2}(\tilde{\mathbf{x}}) & \rho\,\sigma_{|\bar{\mathbf{y}}_2}(\tilde{\mathbf{x}})\sigma_{|\bar{\mathbf{y}}_2}(\mathbf{x}) \\ \rho\,\sigma_{|\bar{\mathbf{y}}_2}(\tilde{\mathbf{x}})\sigma_{|\bar{\mathbf{y}}_2}(\mathbf{x}) & \sigma^2_{|\bar{\mathbf{y}}_2}(\mathbf{x})\end{pmatrix}\right), \tag{7}$$

where $\mathbf{x} \neq \tilde{\mathbf{x}}$ is another feasible point, and ρ is the correlation coefficient between the random variables $\mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}})$ and $\mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$, i.e., the conditional correlation coefficient between the random variables $\mathbb{Y}(\tilde{\mathbf{x}})$ and $\mathbb{Y}(\mathbf{x})$. We use the conditional distribution (6) to calculate the terms in the conditional distribution (7). Specifically, we calculate the vector of conditional means

$$\widehat{\beta}_0(\widehat{\theta})\mathbf{1}_{n\times 1} + \bar{\mathbf{Q}}(\widehat{\theta})^{-1}\begin{pmatrix} \vec{\mathbf{0}}_{n_1\times 1} \\ \widehat{\mathbf{Q}}_\varepsilon(\bar{\mathcal{Y}}_2 - \widehat{\beta}_0(\widehat{\theta})\mathbf{1}_{n_2\times 1}) \end{pmatrix}$$

to obtain $\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}})$ and $\mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$, and the columns of the conditional covariance matrix $\bar{\mathbf{Q}}(\widehat{\theta})^{-1}$ corresponding to $\tilde{\mathbf{x}}$ and $\mathbf{x}$ to obtain ρ, $\sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}})$, and $\sigma^2_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$. Note that we can calculate the jth column $[\bar{\mathbf{Q}}(\widehat{\theta})^{-1}]_{\cdot j}$ of $\bar{\mathbf{Q}}(\widehat{\theta})^{-1}$ by solving the sparse linear system $\bar{\mathbf{Q}}(\widehat{\theta})[\bar{\mathbf{Q}}(\widehat{\theta})^{-1}]_{\cdot j} = \mathbf{e}_j$. Let

$$\sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x}) \triangleq \sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) + \sigma^2_{|\bar{\mathcal{Y}}_2}(\mathbf{x}) - 2\rho\sigma_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}})\sigma_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$$

be the variance of the difference $\mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$. Then, the CEI at the feasible point $\mathbf{x}$, denoted by $\mathrm{CEI}(\mathbf{x})$, is defined as

$$\begin{aligned}\mathrm{CEI}(\mathbf{x}) &\triangleq \mathrm{E}[\max(\mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\mathbf{x}), 0)] \\ &= (\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x}))\Phi\left(\frac{\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})}{\sigma_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})}\right) + \sigma_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})\phi\left(\frac{\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})}{\sigma_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})}\right), \end{aligned} \tag{8}$$

where $\sigma_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x}) \triangleq \sqrt{\sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})}$, Φ is the cumulative distribution function of a standard normal random variable, and ϕ is the density function of a standard normal random variable. We can use the expression for EI in Jones, Schonlau, and Welch (1998) to derive expression (8), by noting that the difference $\mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mathbb{Y}_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$ is a Gaussian random variable with mean $\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$ and variance $\sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})$. Furthermore, from the analysis in Jones, Schonlau, and Welch (1998), we know that $\mathrm{CEI}(\mathbf{x})$ is increasing in both $\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$ and $\sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})$. The point $\tilde{\mathbf{x}}$ is the feasible point with the smallest conditional mean value given the simulation output, so $\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) \leq \mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{X}$. Since $\mathrm{CEI}(\mathbf{x})$ is increasing in $\mu_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}}) - \mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$, $\mathrm{CEI}(\mathbf{x})$ is decreasing in $\mu_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$. Therefore, points with smaller conditional means will tend to have larger CEIs. As the conditional correlation coefficient ρ decreases, with $\sigma_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}})$ and $\sigma_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$ held constant, $\sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})$ increases. Similarly, as $\sigma_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$ increases with ρ and $\sigma_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}})$ held constant, $\sigma^2_{|\bar{\mathcal{Y}}_2}(\tilde{\mathbf{x}},\mathbf{x})$ increases. Feasible points that are far away from $\tilde{\mathbf{x}}$ will tend to have smaller values of ρ, and feasible points that are far away from the design points will tend to have larger values of $\sigma_{|\bar{\mathcal{Y}}_2}(\mathbf{x})$. Thus, feasible points that are far away from $\tilde{\mathbf{x}}$ or far away from the design points will tend to have larger CEIs. We use these tendencies in Section 3.4 to help us choose candidate feasible points that we expect to have the largest CEIs.

By treating the value at the current optimal point $\tilde{\mathbf{x}}$ and the value at the candidate feasible point $\mathbf{x}$ as random variables, we take into account their conditional correlation ρ. With the other terms held constant, increasing ρ decreases the CEI. Thus, using the CEI criterion encourages exploration of the feasible region, since ρ increases as we move further away from the current optimal point.

3.4 Selecting Candidate Feasible Points

A candidate feasible point is a feasible point for which we calculate the CEI during an iteration of the procedure. When the number n of feasible points is not too large, we can calculate the CEI for all feasible points. When this is possible, the candidate feasible points consist of the entire set of feasible points, i.e.,

there are $c = n$ candidates. However, for DOvS problems with large feasible regions, calculating the CEI for every feasible point is too computationally expensive since we need to solve a sparse linear system to calculate each column of the conditional covariance matrix $\bar{\mathbf{Q}}(\widehat{\theta})^{-1}$. Instead, we select a subset of $c < n$ feasible points as candidates, and choose to simulate at the candidate with the largest CEI. We divide the candidate feasible points into three groups. The first and second groups exploit good areas of the feasible region, where we might expect to find the feasible point with the largest CEI, by utilizing the tendencies given at the end of Section 3.3. The first group consists of feasible points that have small conditional mean values, and the second group consists of feasible points that are far away from the design points. Since these two groups are based on tendencies, the third group is composed of randomly sampled feasible points to help ensure the feasible region is sufficiently searched. Let $s \leq c$ denote the number of candidate feasible points that will be randomly sampled from the feasible region.

1. $\lfloor \frac{c-s}{2} \rfloor$ candidates are the feasible points with the $\lfloor \frac{c-s}{2} \rfloor$ smallest conditional mean values.
2. $\lceil \frac{c-s}{2} \rceil$ candidates are selected using the Voronoi tesselation (Preparata and Shamos 1985). We first select $t \triangleq \lceil \frac{c-s}{2} \rceil$ feasible points $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_t$ from a space-filling point-set. Let $\mathscr{N}(\mathbf{s}_i)$ be the set of $d+1$ nearest-neighbors to the feasible point $\mathbf{s}_i$. Calculate the vertex $\mathbf{v}_i$ of the Voronoi tesselation for each of the $\mathscr{N}(\mathbf{s}_i)$. If necessary, project any $\mathbf{v}_i$ back into the feasible region. The vertices $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t$ are t candidate feasible points.
3. s candidate points are randomly sampled feasible points, or feasible points from a space-filling point-set. Fishman (2003) provides a method for sampling approximately uniformly from a polyhedron.

Instead of using the Voronoi tesselation to choose candidates in the second group, when it is not too burdensome, we can partition the feasible region into blocks and keep track of how many design points fall into each block. Then we can select feasible points from the blocks that have the lowest numbers of design points as candidates in the second group.

4 NUMERICAL EXPERIMENTS

For our experiments, we use the (s,S) inventory example from Koenig and Law (1985). The objective function for this (s,S) inventory system is the expected average cost per period of the inventory system over 30 periods. We assume the demand for inventory in each period is a sequence of i.i.d. Poisson random variables with a common mean of 25. The order and unit shipping costs are \$32 and \$3, repsectively. Furthermore, the holding cost is \$1 per unit, per period, and the back-order cost is \$5 per unit, per period. The decision variables are s and S, and our DOvS problem is to minimize the expected average cost per period, subject to the obvious constraint $s < S$. We also assume that the largest inventory we would consider having is 100 units; this leads to a feasible region with 4950 feasible points. The optimal solution to this DOvS problem occurs at $s = 18$ and $S = 53$ with an optimal expected average cost per period of \$106.18.

We ran our procedure 50 times, each time starting with $k = 20$ randomly chosen feasible points as initial design points. For each of the 20 initial design points, we simulated $r = 20$ replications and used the simulation output to calculate the MLEs for the parameters. We only calculated the MLEs once at the beginning of each run, and did not recalculate the MLEs throughout the procedure. At each iteration, we calculated the CEI for every feasible point in the feasible region, i.e., $c = 4950$, and ran the simulation for 20 replications at the feasible point that maximized the CEI. The procedure stopped when the maximum CEI over all feasible points was less than one for 10 times in a row, i.e., $\delta = 1$ and $v = 10$ in our procedure. The vector of conditional means and the conditional covariance matrix, both of which were needed to calculate the CEI for every feasible point, were obtained using the Cholesky factorization. Specifically, we calculated $\mathbf{Q}(\theta)^{-1}$ by calculating the Cholesky factorization $\mathbf{L}(\theta)$ of $\mathbf{Q}(\theta)$ and then $\mathbf{Q}(\theta)^{-1} = (\mathbf{L}(\theta)^{\top})^{-1}(\mathbf{L}(\theta)^{-1})$. Although our DOvS procedure can handle problems with more than 4950 feasible points, we want to observe how our procedure performs when we can calculate the CEI for all feasible points, along with the corresponding runtimes to perform all of the calculations.

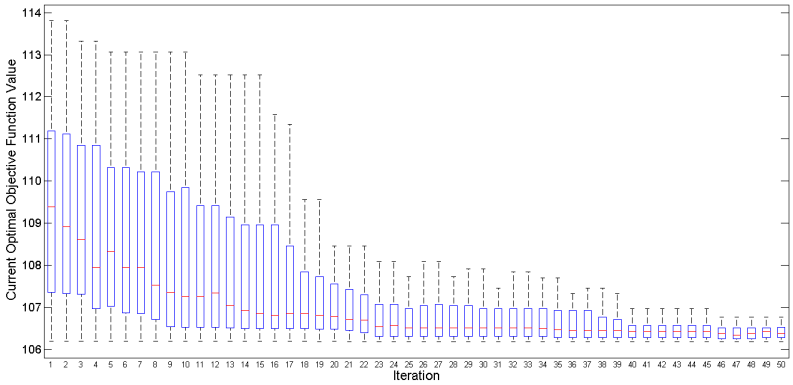


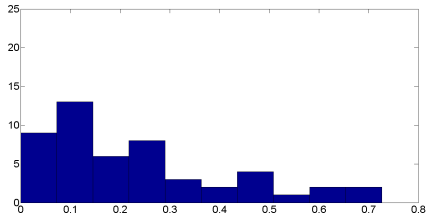
Figure 3: Boxplots of experiment results for 50 runs of our procedure on the (s,S) inventory system. The whiskers in each boxplot extend to 1.5 times the interquartile range.

Figure 3 displays boxplots of the first 50 iterations of the procedure with three outliers not shown, although they were used in the calculation of the quartiles, with the x-axis giving the iteration number and the y-axis giving the actual objective function value at the current optimal point for the corresponding iteration. The actual objective function value was obtained by simulating 10,000 replications of the (s,S) inventory system. As can be seen from Figure 3, our procedure is consistently converging to the global optimal solution within 50 iterations, even when the 20 initial design points are far away from the global optimal solution. Calculating the Cholesky factorization $\mathbf{L}(\theta)$ of $\mathbf{Q}(\theta)$ took 0.04 seconds, on average, and using the Cholesky factorization to calculate the inverse of the precision matrix took 4.54 seconds, on average. Furthermore, calculating the vector of conditional means took 0.02 seconds, on average. In total, each iteration took 4.67 seconds, on average.

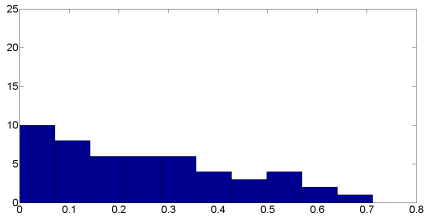
Although each run required more than 50 iterations before the stopping criterion was satisfied, the behavior of the boxplots remained the same after the first 50 iterations. The reason for this behavior may be that, for the (s,S) inventory problem, the feasible points with the largest CEIs were feasible points that had not yet been simulated. Thus, a feasible point was never chosen more than once, and at most 20 replications were allocated to it. However, the GMRF procedure still returned a near optimal feasible point on every run.

Comparison with the KN Ranking and Selection Procedure

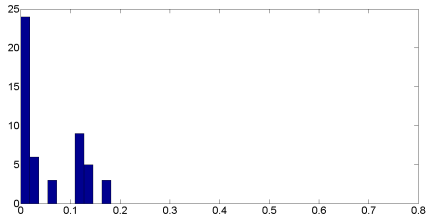
Since we calculated the CEI at every feasible point during each iteration of the procedure, we were implicitly considering all feasible points. In other words, even though we did not simulate at every feasible point, each feasible point was considered since we calculated the CEI using the simulation output at the design points. In this section, we compare our procedure to the KN ranking and selection procedure (Kim and Nelson 2001) which explicitly considers every feasible point by simulating at each of them. The KN procedure sequentially chooses smaller and smaller subsets of feasible points that are deemed better than the rest with respect to some criterion. For the KN procedure, we use CRN and set the number of common first-stage replications at each feasible point to $n_0 = 10$, with $\alpha = 0.05$ and $\delta = 1$. As with our procedure, we ran the KN procedure for 50 runs. The role of δ in our procedure can be compared to the role of δ in the KN procedure. For the KN procedure, the probability of selecting a feasible point whose objective function value is within δ of the optimal is at least $1-\alpha$. Thus, δ specifies the set of feasible points that are acceptable for the procedure to select as optimal; we are indifferent to which feasible point in this set is chosen. For our procedure, we stop searching for better feasible points when the improvement we expect in the objective function value at every feasible point is less than δ for v times in a row. Figures



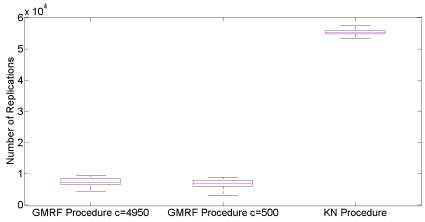
(a) GMRF procedure using $c = 4950$ candidates



(b) GMRF procedure using $c = 500$ candidates



(c) KN procedure



(d) Number of replications

Figure 4: (a), (b), (c) Histograms of the difference between actual optimal objective value and actual objective values returned by each procedure. (d) Boxplots of number of replications required for each procedure. The whiskers in each boxplot extend to 1.5 times the interquartile range.

4(a)–(c) give the results for our procedure when we calculate the CEI at every feasible point ($c = 4950$ candidates), our procedure when we calculate the CEI at $c = 500$ candidate feasible points (with $s = 0$), and the KN procedure, respectively. From these figures, we can see that all three procedures locate near optimal feasible points, if not the optimal feasible point, before stopping. Between the three procedures, the KN procedure found the better solution on most runs. However, as can be seen in Figure 4(d), the KN procedure used more than five times as many replications as the GMRF procedure, using both $c = 500$ and $c = 4950$ candidates, before it stopped. The GMRF procedure using $c = 4950$ candidates and the GMRF procedure using $c = 500$ gave similar results, used nearly the same number of replications, and visited nearly the same number of feasible points on each run. Thus, we can see the benefit of using $c < 4950$ candidates; we implicitly consider all feasible points even without calculating the CEI at every one.

ACKNOWLEDGMENTS

This article is based upon work supported by the National Science Foundation under Grant No. CMMI-0900354. We would also like to thank Håvard Rue for helping us with the conditional distribution.

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58 (2): 371–382.

Fishman, G. S. 2003. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research. NY: Springer.

Frazier, P., W. Powell, and S. Dayanik. 2009. "The Knowledge-Gradient Policy for Correlated Normal Beliefs". *INFORMS Journal on Computing* 21 (4): 599–613.

Frazier, P. I., J. Xie, and S. E. Chick. 2011. "Value of Information Methods for Pairwise Sampling with Correlations". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 3979–3991. Piscataway, NJ: IEEE.

Huang, D., T. T. Allen, W. I. Notz, and N. Zeng. 2006. "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Metamodels". *Journal of Global Optimization* 34 (3): 441–466.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13:455–492.

Kim, S., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-Zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation* 11 (3): 251–273.

Koenig, L. W., and A. M. Law. 1985. "A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations, With Applications to Simulation". *Communications in Statistics* B14:719–734.

Miller, K. S. 1981. "On the Inverse of the Sum of Matrices". *Mathematics Magazine* 54 (2): 67–72.

Nelson, B. L. 2010. "Optimization via Simulation Over Discrete Decision Variables". *Tutorials in Operations Research*:193–207.

Preparata, F. P., and M. I. Shamos. 1985. *Computational Geometry: An Introduction*. Texts and Monographs in Computer Science. New York: Springer-Verlag.

Quan, N., J. Yin, S. H. Ng, and L. H. Lee. 2013. "Simulation Optimization via Kriging: A Sequential Search Using Expected Improvement with Computing Budget Constraints". *IIE Transactions* 45:763–780.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields Theory and Applications*. Chapman and Hall/CRC, New York.

Salemi, P., J. Staum, and B. L. Nelson. 2013. "Generalized Integrated Brownian Fields for Simulation Metamodeling". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 543–554. Piscataway, NJ: IEEE.

Sun, L., L. J. Hong, and Z. Hu. 2011. "Optimization via Simulation Using Gaussian Process-based Search". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 4139–4150. Piscataway, NJ: IEEE.

Williams, B. J., T. J. Santner, and W. I. Notz. 2000. "Sequential Design of Computer Experiments to Minimize Integrated Response Functions". *Statistica Sinica* 10:1133–1152.

Xu, J. 2012. "Efficient Discrete Optimization via Simulation Using Stochastic Kriging". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 466–477. Piscataway, NJ: IEEE.

AUTHOR BIOGRAPHIES

PETER SALEMI received his Ph.D. from the Department of Industrial Engineering and Management Sciences at Northwestern University. His research interests are in simulation metamodeling and optimization via simulation. His email address is peter.salemi@gmail.com.

BARRY L. NELSON is the Walter P. Murphy Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University, and a Fellow of INFORMS. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. His email and web addresses are nelsonb@northwestern.edu and www.iems.northwestern.edu/~nelsonb/.

JEREMY STAUM is an Associate Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He coordinated the Risk Analysis track of the 2007 and 2011 Winter Simulation Conferences and serves as department editor for financial engineering at *IIE Transactions* and as an associate editor at *Management Science*. His email and website are j-staum@northwestern.edu and users.iems.northwestern.edu/~staum.