Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

PLAUSIBLE OPTIMA

Matthew Plumlee
Barry L. Nelson

Department of Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208-3119, USA

ABSTRACT

We propose a framework and specific algorithms for screening a large (perhaps countably infinite) space of feasible solutions to generate a subset containing the optimal solution with high confidence. We attain this goal even when only a small fraction of the feasible solutions are simulated. To accomplish it we exploit structural information about the space of functions within which the true objective function lies, and then assess how compatible optimality is for each feasible solution with respect to the observed simulation outputs and the assumed function space. The result is a set of *plausible optima*. This approach can be viewed as a way to avoid slow simulation by leveraging fast optimization. Explicit formulations of the general approach are provided when the space of functions is either Lipschitz or convex. We establish both small- and large-sample properties of the approach, and provide two numerical examples.

1 INTRODUCTION

Screening procedures, as we use the term here, eliminate sub-optimal feasible solutions, ideally with some assured level of statistical confidence, in a simulation optimization problem. Screening has been employed as the initial step in ranking-and-selection procedures to prune solutions; as a "clean up" step at the end of heuristic searches to discard inferior solutions; and as the objective in and of itself (Boesel et al. 2003; Gupta 1965; Nelson et al. 2001). This is *solution screening* rather than *factor screening*, the latter of which eliminates some components of all solutions because their effect on the simulation response is insignificant (Bettonvil and Kleijnen 1997; Wan et al. 2006).

The idea of this work is to locate plausible optima which enables screening a large, possibly unaccountably infinite, feasible space based on simulating a (perhaps very small) number of feasible solutions; see the illustration in Section 4. In traditional ranking-and-selection procedures that employ a screening step, the screening process is largely based on exhaustive simulation at every feasible solution. Simulation optimization methods, e.g., Andradóttir (1996), Chang et al. (2013), Salemi et al. (2018), and Wang et al. (2013), simulate only a small number of feasible solutions, but offer guarantees about a single returned solution after the algorithm completes. Compared to ranking-and-selection-type screening, our method provides similar inference without being exhaustive; compared to simulation optimization search methods, plausible optima provide global information about the feasible solution space, not just about the final selected solution.

Simulation, when viewed as a function of feasible solutions, often has underlying properties that can be thought of as structural information about the function. Examples include convexity or bounds. Existing screening methods have largely abstained from leveraging structural information, perhaps because when comparing a small number of solutions the difficulty of incorporating structural information can outweigh the potential benefits. However, if the goal is to screen hundreds, thousands or more feasible solutions, we show that structural information can improve screening by eliminating solutions with fewer runs of the simulation. This article proposes a method that leverages this type of structural information. Moreover, the

proposed method can eliminate solutions that were not simulated at all. Although this article presents the framework generally, we will illustrate the method by considering structural information of the Lipschitz and convex type.

Simulation optimization algorithms frequently employ structural information as they explore the space of feasible solutions, and after some stopping condition is satisfied they return a single solution that can be viewed as the approximation of the optimum. For instance, random search methods are predicated on some version of clustering of good solutions near the optimum; e.g., Andradóttir (1996). However, random search has no natural screening feature. There has been a recent surge in simulation optimization methods that can broadly be described as metamodel-based procedures; examples include regression-based methods such as STRONG (Chang et al. 2013) and R-SPLINE (Wang et al. 2013)—that assume the response function is locally linear or quadratic in the solutions—and Gaussian-process-based procedures that treat the unknown response as a realization of a random function (Salemi et al. 2018). A common feature of metamodel-based approaches is that data are collected, a metamodel is fit, and the metamodel is used to select the next approximation of the optimum. In principle the metamodel approach could be used for screening, but the production of good screening sets is quite difficult using these methods because of the potentially strong assumptions supporting the metamodel construction. Moreover, estimation uncertainties are not always quantified, making some assured level of statistical confidence difficult to attain. Thus, metamodel-based simulation optimization procedures have primarily been used for optimum seeking and no screening analog has been developed.

As a new direction, we introduce a framework for screening that returns a subset of feasible solutions that will include the optimal solution with a prespecified probability, even when far fewer than the total number of feasible solutions are simulated. We accomplish this by leveraging the structure of a space of functions where the true simulation response lies. In brief, our approach assesses how compatible optimality is for each feasible solution with respect to the observed simulation outputs over the assumed function space. Therefore we call our subset *plausible optima*. The general framework can accommodate discrete or continuous-valued decision variables.

The paper is organized as follows: We more fully describe the setting in which we do screening in Section 2. Section 3 presents our method for discovering plausible optima, and a stylized example in Section 4 illustrates it. Small- and large-sample properties are established in Section 5, followed by the computational methods needed to support them in Section 6. We close with two numerical examples in Section 7 and some concluding remarks in Section 8.

2 SETTING AND GOALS

This section outlines the mathematical setting for our approach. The feasible-solution space $\mathscr{X} \subseteq \Re^d$ represents all solutions, labeled x, from which we would like to discover the unique optimal $x^\star$. In this article we take "optimal" to imply minimum with respect to the objective function $\mu\colon \mathscr{X} \to \Re$; Thus, there exists $x^\star \in \mathscr{X}$ such that $\mu(x^\star) < \mu(x)$ for all $x \neq x^\star$. As a technical condition, we assume $\mu(\cdot)$ is bounded over $\mathscr{X}$.

We seek to extract a subset of $\mathscr{X}$ that includes $x^\star$ with a prescribed confidence by querying the simulation. This means we run the simulation model at x and observe a random variable $Y(x) = \mu(x) + e(x)$, where $e(x)$ is a mean-zero random variable with finite variance whose distribution depends on x. We assume that only a finite set of feasible solutions are simulated. We call this the *experimental set*, denoted by $\mathsf{X} = \{x_1, x_2, \ldots, x_K\}$, and it has cardinality K. The experimental set may or may not cover all of $\mathscr{X}$. The number of simulation replications at each x_k is n_k, which is assumed strictly larger than 1 and may differ by solution. Thus, for each $x_k \in \mathsf{X}$ there are simulation-generated replications $Y_1(x_k), Y_2(x_k), \ldots, Y_{n_k}(x_k)$. The associated sample mean and variance are

$$\widehat{\mu}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} Y_i(x_k) \text{ and } \widehat{\sigma}_k^2 = \frac{1}{n_k - 1}\sum_{i=1}^{n_k} \left(Y_i(x_k) - \widehat{\mu}_k\right)^2.$$

The total simulation budget for replications is N, therefore $\sum_{k=1}^{K} n_k = N$. For the purpose of examining performance as N gets large, the experimental set X remains fixed and n_k/N is bounded away from 0 as N grows. We assume all K solutions are simulated independently; i.e., we do not employ common random numbers.

Our set should guarantee some statistical properties, specifically the notions of confidence and consistency. A random set that depends on our data attains confidence if the probability of it containing $x^\star$ is sufficiently large. That is, let $\mathsf{S}_N \subset \mathscr{X}$ be a subset of feasible solutions based on N simulations. We would like this set to contain the optimal $x^\star$ with at least a prescribed probability $0 < 1-\alpha < 1$:

Definition 1 A subset S_N attains *finite-sample confidence* $1-\alpha$ if for all $N \geq 2K$, with $n_k > 1$ for all $k = 1,2,\ldots,K$, we have $\mathbb{P}(x^\star \in \mathsf{S}_N) \geq 1-\alpha$.

Finite-sample confidence is a difficult goal whose attainment is, for the most part, limited to situations in which our simulation produces random variables from specific distributions. The common case is when the outputs from the simulation are normally distributed, then $\widehat{\mu}_k$ and $\widehat{\sigma}_k$ have known distributions and are independent. A more widely achievable notion enables us to employ the Central Limit Theorem that lets $\widehat{\mu}_k$ be approximated as normally distributed.

Definition 2 A subset S_N attains *asymptotic confidence* $1-\alpha$ if for any $\delta > 0$ there exists an $N_0(\delta)$ such that $\mathbb{P}(x^\star \in \mathsf{S}_N) \geq 1-\alpha-\delta$ for all $N > N_0(\delta)$.

Confidence alone is not sufficient to distinguish good and bad screening procedures since an absurdly large subset could achieve confidence as defined above. Screening methods should also shrink the subset as the sample size N increases. We formalize this concept with the notion of *consistency*:

Definition 3 A subset S_N attains *consistency* if for all $x_0 \in \mathscr{X}$ such that $x_0 \neq x^\star$ we have $\lim_{N\to\infty} \mathbb{P}(x_0 \in \mathsf{S}_N) = 0$.

However, consistency is fundamentally unachievable in the prescribed setting when only $\mathsf{X} \subset \mathscr{X}$ solutions are simulated. Thus, we allow for a relaxed version of consistency:

Definition 4 Given some $\varepsilon > 0$, a subset S_N achieves ε-*consistency* when for all $x_0 \in \mathscr{X}$ such that $\mu(x_0) \geq \varepsilon + \mu(x^\star)$ we have $\lim_{N\to\infty} \mathbb{P}(x_0 \in \mathsf{S}_N) = 0$.

In other words, a subset achieving ε-consistency eventually excludes all solutions that are suboptimal by at least ε.

In the next section we provide a framework for generating subsets of plausible optima that have confidence and consistency.

Remark 1 There is a large literature on subset selection when $\mathscr{X}$ is finite, all feasible solutions are simulated, and the solutions are treated as categorical. The foundation is the work of Gupta (1965), but there have been many variations and extensions in statistics and simulation (Bechhofer et al. 1995; Kim and Nelson 2007). The objective described in this paper aligns closely with this literature. Subset selection inference may also be derived from multiple comparisons inference. For instance, multiple comparisons with the best (Hsu 1984) provides $1-\alpha$ simultaneous confidence intervals on

$$\mu(x_i) - \min_{j\neq i} \mu(x_j),\ i = 1,2,\ldots,K.$$

Therefore the optimal $x^\star$ is the single solution whose interval is entirely below 0, or is one of the solutions whose interval contains 0 and therefore cannot be distinguished from the best, with confidence $1-\alpha$. Due to the connection between some ranking-and-selection procedures and multiple comparisons with the best, ranking and selection procedures may also provide screening inference (Matejcik and Nelson 1995).

The methods described above are often based on the assumption of normally distributed output, and always assume that all feasible solutions are simulated. To the best of our knowledge there are no other methods designed to yield subset inference when $\mathscr{X}$ is not exhaustively simulated.

3 PLAUSIBLE OPTIMA

The proposed screening method will return what we term a subset of plausible optima, achieving either finite-sample or asymptotic confidence and ε-consistency even when some (typically most) feasible solutions are not simulated. This goal requires exploiting information about the structural link between simulated and not-simulated solutions.

We adopt structural information in the form of a known space of functions $\mathscr{M}$ that contains the mean function $\mu(\cdot)$. That is, the unknown mean function is an element of this space. The space $\mathscr{M}$ should contain relevant structural information about the mean function; we assume this is known *a priori* and is not extracted from the simulation output data. Importantly, $\mathscr{M}$ is not a prior distribution on $\mu(\cdot)$; instead it characterizes properties that $\mu(\cdot)$ has.

Now for any $x_0 \in \mathscr{X}$ and our chosen space of functions, consider the smaller set of functions $\mathsf{M}(x_0) \subset \mathscr{M}$ which attain their optimum at x_0. If $\mathsf{M}(x_0)$ is empty then there are no plausible functions in $\mathscr{M}$ with x_0 as the optimum, and thus x_0 can be removed from our set. Of course, for any reasonable space of functions $\mathscr{M}$, no feasible solution x_0 can be eliminated in this way. But in light of simulation data obtained at X we may be able to remove feasible solutions with some confidence.

For a given $m(\cdot) \in \mathscr{M}$, a standardized empirical discrepancy between this function and the simulation output data is

$$\sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} \left(m(x_k) - \widehat{\mu}_k\right)^2.$$

One could interpret this standardized discrepancy as a constant away from twice the negative log-likelihood of $m(\cdot)$ when $\widehat{\mu}_k$ is Gaussian with mean $m(x_k)$ and variance $\widehat{\sigma}_k^2/n_k$. If $m(\cdot)$ is quite far away from the observed data at X, then it should have a high discrepancy. The standardized discrepancy can be associated with x_0 by measuring the minimum standardized discrepancy among the x_0-optimal functions:

$$L(x_0) = \min_{m \in \mathsf{M}(x_0)} \sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} \left(m(x_k) - \widehat{\mu}_k\right)^2. \tag{1}$$

Thus, $L(x_0)$ can be used to measure the agreement between our data and the hypothesis that x_0 is the optimal: An extremely large $L(x_0)$ presents evidence that x_0 is not the optimum. We therefore let the set of plausible optima be

$$\mathscr{S} = \{x \in \mathscr{X} : L(x) \leq \mathsf{L}\}.$$

The choice of L is based on the prescribed confidence. We let L be the $1-\alpha$ quantile of the sum of K independent F-distributed random variables each with numerator degrees of freedom 1 and denominator degrees of freedom $n_1-1, n_2-1, \ldots, n_K-1$, respectively. Computationally, we adopt Monte Carlo simulation to find this quantity. We will show this choice can achieve finite-sample confidence when the outputs from the simulation are normally distributed, and always achieves asymptotic confidence. Thus, with some loss of generality, we take this choice of L as the standard with the acknowledgment that other selections are available.

For illustrative purposes, this paper will consider three spaces of mean functions. The first, $\mathscr{M}_A$ is any mapping from $\mathscr{X}$ to $\mathfrak{R}$; $\mathscr{M}_A$ is clearly the most general. Second, we consider $\mathscr{M}_L$, the space of functions that are c-Lipschitz, with $c > 0$ being some upper bound on the Lipschitz constant of the unknown function $\mu(\cdot)$. Specifically:

Definition 5 The function $m(\cdot) \in \mathscr{M}_L$ if we have $|m(x) - m(x')| \leq c\|x - x'\|$ for all $x, x' \in \mathscr{X}$.

Lastly, we consider $\mathscr{M}_C$, the space of all convex functions on $\mathscr{X}$. Because we may have a discrete $\mathscr{X}$, we must pay careful attention to our definition of convexity. While convexity over discrete sets has a lengthy history (Murota 1998), we sidestep historical digressions in favor of the following definition:

Definition 6 The function $m(\cdot) \in \mathscr{M}_C$ if for all $x \in \mathscr{X}$ there exists a corresponding d-dimensional vector $\xi(x)$ such that

$$m(x') \geq m(x) + \xi(x)^\top (x' - x) \text{ for all } x' \in \mathscr{X}.$$

We employ this definition of convexity from here on. Notice that the more intuitive definitions of convexity satisfy this characterization. Namely, if λ is between zero and one, and x, x' and $\lambda x + (1-\lambda)x'$ are all in $\mathscr{X}$, then for all $m(\cdot) \in \mathscr{M}_C$ we have

$$m(\lambda x + (1-\lambda)x') \leq \lambda m(x) + (1-\lambda) m(x').$$

4 ILLUSTRATION

Consider a situation where solutions take values on the real line. We conduct an experiment by simulating n replications at two solutions x_1 and x_2, with $x_1 < x_2$, and receive data for which $\widehat{\mu}_1 < \widehat{\mu}_2$. For illustrative purposes, suppose also $\widehat{\sigma}_1^2 = \widehat{\sigma}_2^2 = 1$.

The traditional, exhaustive approach to screening implies we cannot eliminate any solutions other than one of these two. However, we can eliminate more solutions if we consider structural information. Functions that achieve the minimum standardized discrepancy for both the Lipschitz and convexity cases are drawn in Figure 1, but it should be noted that they are not unique. In the figure, each panel shows the standardized discrepancy minimizing function among all functions that have x_0 as their minima. The left panel shows this function for both the convex and Lipschitz cases when x_0 is chosen in the middle of the two solutions in the experimental set. Since the function interpolates the sample means corresponding to each solution, the standardized discrepancy is zero for the left panel.

In the right panel of Figure 1, where $x_0 > x_2$, there is a non-zero discrepancy for both function spaces. In the Lipschitz case, if x_0 is the optimum, then $m(x_1) \geq m(x_2) - c|x_0 - x_2|$; and if $\widehat{\mu}_1 < \widehat{\mu}_2 - c(x_0 - x_2)$ then there must be a discrepancy associated with x_0. In the convex case, for x_0 to be optimal and to the right of x_2 requires $m(x_1) \geq m(x_2)$. Thus, when we leverage $\mathscr{M}_L$ then

$$L(x_0) = \begin{cases} 0 & \text{if } |x_0 - x_2| > c^{-1}(\widehat{\mu}_2 - \widehat{\mu}_1), \\ \frac{n}{2}(\widehat{\mu}_1 - \widehat{\mu}_2 + c|x_0 - x_2|)^2 & \text{if } |x_0 - x_2| \leq c^{-1}(\widehat{\mu}_2 - \widehat{\mu}_1), \end{cases}$$

while when we leverage $\mathscr{M}_C$ then

$$L(x_0) = \begin{cases} 0 & \text{if } x_0 < x_2, \\ \frac{n}{2}(\widehat{\mu}_1 - \widehat{\mu}_2)^2 & \text{if } x_0 \geq x_2. \end{cases}$$

These closed forms are not generally available when more solutions are included in the experiment.

For these two function spaces, there is some evidence that x_0 is not optimal when it is either near x_2 in the Lipschitz case or to the right of x_2 in the convex case. The decision to include x_0 in the set is controlled by the cutoff, L, to the set. But since L is decreasing in n, we will eventually be able to eliminate everything within a $c^{-1}(\widehat{\mu}_2 - \widehat{\mu}_1)$ ball around x_2 for the Lipschitz case, and equal to or larger than x_2 for the convex case. Thus even a very limited experiment can eliminate an infinite number of feasible solutions if we properly use information about the true mean function.

5 CONFIDENCE AND CONSISTENCY

In this section we present results on confidence and consistency; only the proof of Theorem 1 is included to illustrate the main ideas. The fact that $\mathscr{S}$ attains both finite and asymptotic $1-\alpha$ confidence is an immediate consequence of the construction of $\mathscr{S}$.

Theorem 1 If outputs from the simulation are normally distributed, then $\mathscr{S}$ achieves finite-sample confidence.

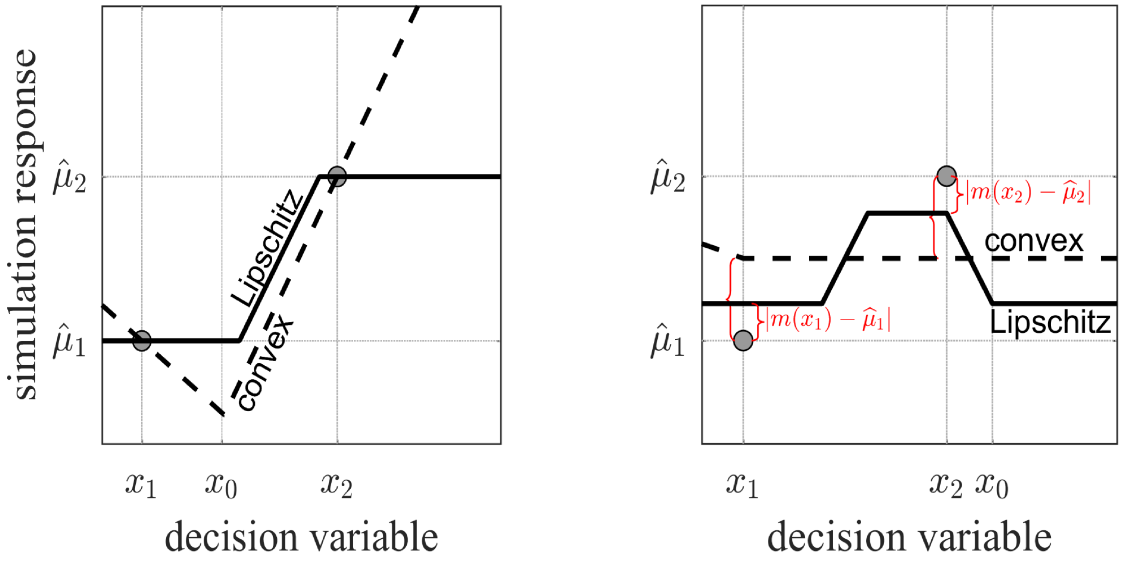

Figure 1: Illustration of plausible optima for the $K = 2$ case where solutions lie on the real line.

Proof of Theorem 1. Recall that we assume that the true mean function $\mu \in \mathsf{M}(x^\star)$. Thus,

$$L(x^\star) = \min_{m \in \mathsf{M}(x^\star)} \sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} \left(m(x_k) - \widehat{\mu}_k\right)^2 \leq \sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} \left(\mu(x_k) - \widehat{\mu}_k\right)^2$$

which implies

$$\mathbb{P}(x^\star \in \mathscr{S}) = \mathbb{P}\left(L(x^\star) \leq \mathsf{L}\right) \geq \mathbb{P}\left(\sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} \left(\mu(x_k) - \widehat{\mu}_k\right)^2 \leq \mathsf{L}\right).$$

For each k, $\frac{n_k}{\widehat{\sigma}_k^2} \left(\mu(x_k) - \widehat{\mu}_k\right)^2$ follows an F_{1,n_k-1} distribution. Therefore, by construction of L,

$$\mathbb{P}\left(\sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} \left(\mu(x_k) - \widehat{\mu}_k\right)^2 \leq \mathsf{L}\right) = 1 - \alpha.$$

□

Theorem 2 If outputs from the simulation have a bounded variance, $\mathscr{S}$ achieves asymptotic confidence.

As noted previously, consistency is difficult in the fixed experimental setting where data are only observed at K feasible solutions. Instead, we will look to ε-consistency to evaluate the method. Let G be the space of functions within $\mathscr{M}$ that interpolate the true function on the experimental set X, i.e.

$$\mathsf{G} = \{m \in \mathscr{M} \text{ such that } m(x_k) = \mu(x_k) \text{ for } k = 1, 2, \ldots, K\}.$$

The set G is non-empty because μ must lie in it. Consider now the set of all solutions such that there is an interpolating function that is also an x-optimal function

$$\mathscr{S}_o = \{x \in \mathscr{X} \text{ such that } \mathsf{G} \cap \mathsf{M}(x) \text{ is non-empty}\}.$$

Then define

$$\varepsilon_o = \max_{x \in \mathscr{S}_o} \mu(x) - \mu(x^\star).$$

The fundamental consistency result is then:

Theorem 3 $\mathscr{S}$ has ε_o-consistency.

Remark 2 Deriving specific values of, or bounds on, ε_o requires joint analysis of $\mathscr{X}$, X, and $\mathscr{M}$. But there is at least one transparent result about ε_o. If $\mathscr{X}$ is a finite set of discrete points and $\mathsf{X} = \mathscr{X}$, then $\varepsilon_o = 0$ since $\mathscr{S} = \{x^\star\}$. Thus, if we are in an exhaustive simulation setting then our method produces the stronger version of consistency. Further research is needed to explore the properties of ε_o for other settings.

Remark 3 Although beyond the scope of this article, we conjecture that the set of plausible optima can be consistent if the experimental set X gets larger with N. The trade-off between the number of solutions in the experimental set and the replications at each solution is expected to play a critical role in developing such a consistency result (Ensor and Glynn 1997).

6 COMPUTATIONAL METHODS

If we are only interested in the theoretical construction of sets, then the framework underlying plausible optima is complete. However, the construction thus far is not practical: there needs to be some way to compute $L(x_0)$ for a feasible solution x_0, simulated or not. Since $L(x_0)$ is the result of an optimization problem, we view this as trading simulation for optimization at x_0. The relative computational burdens of simulation versus optimization are, of course, problem dependent. However, since we will potentially solve this optimization problem for each feasible x_0, speed of optimization is a critical consideration. The three function spaces we have considered are all large enough that optimization to compute $L(x_0)$ takes place over an infinite set; see (1). Moreover, the number of constraints in (1) is on the order of the cardinality of $\mathscr{X}$, which can also be infinite. This naturally evokes the question: can we actually compute these plausible optima sets?

We now show that our three function spaces lead to a formulation for $L(x_0)$ that is a quadratic program with a finite number of decision variables and constraints. The simplest version of this occurs when we have the most general function information:

Theorem 4 Suppose $\mathscr{M} = \mathscr{M}_A$. Then for all $x_0 \notin \mathsf{X}$, $L(x_0) = 0$. Further, for all $l \in \{1,2,\ldots,K\}$,

$$L(x_l) = \min_{m_1,\ldots,m_K} \sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} (m_k - \widehat{\mu}_k)^2$$
$$\text{subject to } m_l \leq m_k \text{ for all } k = 1,2,\ldots,K.$$

Thus, when no function information is used we revert to something analogous to the exhaustive simulation case: No feasible solutions outside of the experimental set X can be eliminated. More elegant reformulations exist in the latter two cases.

Theorem 5 Suppose $\mathscr{M} = \mathscr{M}_L$. Then

$$\begin{aligned} L(x_0) = \min_{m_0,m_1,\ldots,m_K} & \sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} (m_k - \widehat{\mu}_k)^2 \\ \text{subject to } & m_k - m_l \leq c\|x_k - x_l\| \text{ for } k,l = 1,2,\ldots,K \\ & m_k - m_0 \leq c\|x_k - x_0\| \text{ for } k = 1,2,\ldots,K \\ & m_0 \leq m_k \text{ for all } k = 1,2,\ldots,K. \end{aligned} \tag{2}$$

Theorem 6 Suppose $\mathscr{M} = \mathscr{M}_C$. Let $\xi_1, \xi_2, \ldots, \xi_K$ be d dimensional decision variables. Then

$$L(x_0) = \min_{m_0, m_1, \ldots, m_K, \xi_1, \ldots, \xi_K} \sum_{k=1}^{K} \frac{n_k}{\widehat{\sigma}_k^2} (m_k - \widehat{\mu}_k)^2$$
$$\text{subject to } m_k - m_l \leq \xi_k^\top (x_k - x_l) \text{ for } k, l = 1, 2, \ldots, K,$$
$$m_k - m_0 \leq \xi_k^\top (x_k - x_0) \text{ for } k = 1, 2, \ldots, K.$$
$$m_0 \leq m_k \text{ for } k = 1, 2, \ldots, K$$

We now offer a proof of the first result. Our proof of the second result is quite similar.

Proof of Theorem 5. Consider (2) the "new problem" and (1) the "original problem." Let $\dot{m}(\cdot)$ be a feasible solution to the original problem. This maps to a feasible solution in the new problem with the same objective function value with $\ddot{m}_k = \dot{m}(x_k)$ for $k = 0, 1, \ldots, K$. Clearly, $(\ddot{m}_0, \ddot{m}_1, \ldots, \ddot{m}_K)$ is feasible for the new problem. Conclude the new problem has a solution that is less than or equal to the original problem.

Now let $\dot{m}_0, \dot{m}_1, \ldots, \dot{m}_K$ be a feasible solution in the new problem. We show that this maps to a feasible solution to the original problem with the same objective function value. Consider

$$m^+(x) = \min_{k=0,1,\ldots,K} \dot{m}_k + c\|x - x_k\|, \quad m^-(x) = \max_{k=0,1,\ldots,K} \dot{m}_k - c\|x - x_k\|,$$

$$k^+(x) = \operatorname*{arg\,min}_{k=0,1,\ldots,K} \dot{m}_k + c\|x - x_k\|, \text{ and } k^-(x) = \operatorname*{arg\,max}_{k=0,1,\ldots,K} \dot{m}_k - c\|x - x_k\|.$$

Construct $\ddot{m}(x) = (m^+(x) + m^-(x))/2$. We first observe that $m^+(x_k) = \dot{m}_k$ and $m^-(x_k) = \dot{m}_k$ for $k = 1, 2, \ldots, K$, and thus the objective function is unchanged. Also, $\ddot{m}(\cdot)$ is Lipschitz. Lastly, we need to demonstrate preservation of the optimality of x_0 in the new construction. We have by construction that for any $x \in \mathscr{X}$ $\ddot{m}(x_0) = \dot{m}_0 \leq \dot{m}_{k^+(x)} = \ddot{m}(x_{k^+(x)})$. Also by construction

$$\frac{1}{2}\ddot{m}(x_{k^+(x)}) - \frac{c}{2}\|x - x_{k^+(x)}\| \leq \frac{1}{2}\ddot{m}(x_{k^-(x)}) - \frac{c}{2}\|x - x_{k^-(x)}\|.$$

Adding these inequalities together gives

$$\ddot{m}(x_0) \leq \frac{1}{2}\left(\ddot{m}(x_{k^+(x)}) + \ddot{m}(x_{k^-(x)}) + c\|x - x_{k^+(x)}| - c\|x - x_{k^-(x)}\|\right) = \ddot{m}(x).$$

Therefore, we conclude $\ddot{m}(\cdot)$ is feasible for the original problem and thus the original problem's value is less than or equal to the new problem's value. □

The two reformulations in Theorems 5–6 yield convex quadratic programs. For the Lipschitz case, there are $K+1$ scalar decision variables and order K^2 inequality constraints. For the convex case, there are $K+1+dK$ scalar decision variables and also order K^2 inequality constraints. Thus, these will not be large problems for modern quadratic programming algorithms unless we simulate a huge number of feasible solutions or the dimension of the feasible-solution space is large.

Remark 4 At present we are unsure what other versions of $\mathscr{M}$ might yield tractable optimization formulations for $L(x_0)$. Certainly the intersection of $\mathscr{M}_L$ and $\mathscr{M}_C$ would produce nice formulations. Slight modifications would also yield nice formulations for Hölder spaces and strongly convex functions. These can be considered somewhat trivial extensions. There are many other function spaces that have not been considered here. If we generalized the function spaces, perhaps the constraints would not be linear, creating additional computational concerns.

Remark 5 Since many quadratic programs will have to be solved over many feasible solutions, it may prove useful to investigate the computation of these quadratic programs in a sequential form, perhaps using ideas from Wright and Nocedal (1999).

7 NUMERICAL EXAMPLES

This section presents the results after implementation of the proposed screening approach on classical simulation optimization problems. The two examples are designed to highlight the potential for the method when there are many solutions available and limited simulation budget.

7.1 *M/M/x* Staffing Problem

Consider a first-come-first-served x-server queue staffing problem. The variant we consider has a Poisson arrival process with an arrival rate of 1 and exponentially distributed service times with mean 1000. Our objective is to minimize the sum of the cost of staffing the queue, $0.01x$, and the average square root of the total time in system. This square root transformation is designed to emulate a concave utility function with respect to time in system. Our simulation was run for 10,000 time units and initialized from the steady state conditions of the queue. The optimal solution can be analytically computed as $x^\star = 1036$.

We obtain the subset $\mathscr{S}$ of plausible optima by simulating n replications at solutions $\mathsf{X} = \{1020, 1025, \ldots, 1115\}$. For our numerical illustration we take $n = 5, 10, 25$, implying total budgets of $N = 100, 200, 500$, respectively. We exploit either convexity information or Lipschitz information with $c = 0.03$. As a comparison, we also applied a classical subset selection algorithm that uses an exhaustive simulation at all 100 feasible solutions x between 1020 to 1119; specifically, Boesel et al. (2003). For all methods, we used $\alpha = 0.05$. We generated 100 macroreplications of the entire experiment and recorded the number of times a solution is in the subset for all three approaches.

The results are shown in Figure 2. Each panel shows the frequency that a particular x value is included in the set. The horizontal solid line at 0.95 in all panels represents the desired confidence level. The dashed line shows the frequency of inclusion of the optimal solution $x^\star = 1036$.

All methods, as expected and by design, appear to have good confidence properties. The distinguishing results are with respect to consistency. As a general trend, the classical subset selection algorithm does worse than the plausible optima approach. We imagine this is because the classical approach does not leverage structural information. The convex case provides better consistency properties compared to the Lipschitz case. The most striking results happen when there is a small total budget, i.e., $N \leq 200$. When $N = 100$ the classical subset selection procedure cannot be applied because there is only a single observation at each feasible solution, thus all solutions remain in the set. The plausible optima approach, by contrast, does well in this small-data environment.

7.2 Inventory Problem

Consider an (s, S) inventory problem where we select the minimum inventory that triggers a new order s and the amount we should be ordering up to, S, when an order is placed. There are costs associated with stocking out or holding extra product, and the numerical values of these costs are exactly as in Koenig and Law (1985). The simulation will take place over 100 months. We consider a design region where the minimum inventory is in $[10, 80]$ and the order-up-to quantity is in $[10, 100]$. Additionally, we assume the order-up-to quantity is at least 15 more than the minimum inventory. In total, this gives us 2911 feasible solutions.

Suppose we are limited to a total budget of $N = 250$ simulations. In this case exhaustive simulation is not possible; we can only simulate fewer than 10% of the feasible solutions if we have one replication per solution. We instead choose 25 solutions in the space to simulate $n = 10$ replications each. The mean function is not convex, thus we will use the plausible optima approach leveraging Lipschitz information with $c = 3$. We obtain 100 macroreplications of the entire experiment and record the occurrence of each solution in the plausible optima subset.

The results are shown in Figure 3. The solid lines represent the boundaries of the feasible region. The $*$ marks the true optimal solution. The diamonds mark the experimental set X. The size of the circles

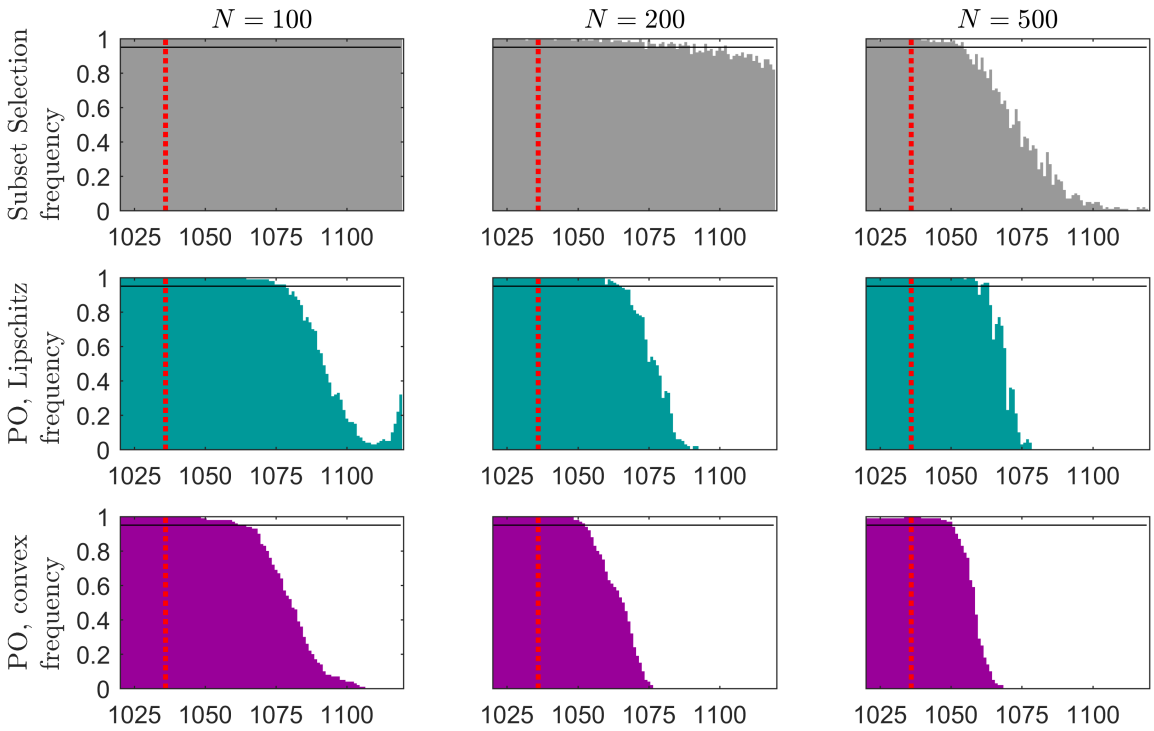

Figure 2: Numerical results for the $M/M/x$ experiment.

indicates how often in the 100 macroreplications that a solution remained in the set of plausible optima. A circle of diameter 1 implies it was in the set every attempt and no circle implies it was never in the set.

The results show that we can eliminate roughly half of the feasible solutions using the proposed approach. Given a larger simulation budget, an exhaustive approach with a generic subset selection algorithm would give a similar result. Thus, there are two ways to assess the performance for this example. First, the proposed plausible optima can be used to halve the number of feasible solutions before applying an exhaustive simulation optimization algorithm. Alternatively, if one was limited to only these 250 simulations because of a computational budget, the plausible optima can be used with the understanding that some poorer solutions have been eliminated. In either case, it appears there is some advantage to considering plausible optima.

8 CONCLUDING REMARKS

This article introduced the foundation for a new approach for constructing screening sets after a limited simulation experiment. By using underlying structural function information we are able to eliminate unsimulated solutions. The method was illustrated using two reasonable cases of structural function information, Lipschitz bounds and convexity, and more cases are foreshadowed in Remark 4 at the end of Section 6. The numerical examples demonstrate that even when the budget for simulation is much less than the total number of feasible solutions, many solutions can be screened out, even those with no data attached to them. The theory underpinning these sets relates to the classical notions of statistical confidence and pointwise consistency.

We acknowledge there is currently a gap between the concepts in this work and the black box approach to simulation optimization in modern software. To implement the method of plausible optima in practice, one must address the assumption that the user knows the function space. In some cases it is reasonable that convexity information is available. In other cases it is not known *a priori* if convexity holds. So perhaps

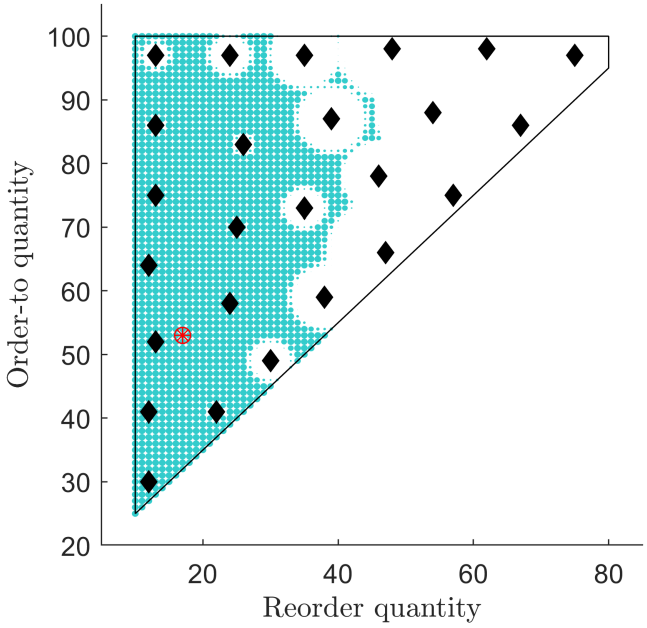

Figure 3: Numerical results for the inventory experiment.

a separate investigation of convexity is required; an example of one such approach is Jian et al. (2014). In the case of Lipschitz information, the user may not know the Lipschitz constant or a reasonable upper bound on it. Practical methods should include some estimation of this constant. Initial experiments by the authors' have found that an overestimated Lipschitz constant still results in screening out very poor performing regions of the design space but causes drastic reductions in screening power near the optima. One idea is to leverage bounds on an estimated derivative (Fu 2006). Ideally there will be some type of integrated approach that provides guidance on reasonable function spaces and the estimation of constants. Research is ongoing on these matters.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Grant Number CMMI-1634982.

REFERENCES

Andradóttir, S. 1996. "A Global Search Method for Discrete Stochastic Optimization". *SIAM Journal on Optimization* 6(2):513–530.

Bechhofer, R., T. Santner, and D. Goldsman. 1995. *Designing Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: Wiley & Sons.

Bettonvil, B., and J. P. Kleijnen. 1997. "Searching for Important Factors in Simulation Models with Many Factors: Sequential Bifurcation". *European Journal of Operational Research* 96(1):180–194.

Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. "Using Ranking and Selection to Clean Up after Simulation Optimization". *Operations Research* 51(5):814–825.

Chang, K.-H., L. J. Hong, and H. Wan. 2013. "Stochastic Trust-Region Response-Surface Method (STRONG)—A New Response-Surface Framework for Simulation Optimization". *INFORMS Journal on Computing* 25(2):230–243.

Ensor, K. B., and P. W. Glynn. 1997. "Stochastic Optimization via Grid Search". Volume 33 of *Lectures in Applied Mathematics-American Mathematical Society*, 89–100. American Mathematical Society.

Fu, M. C. 2006. "Chapter 19 Gradient Estimation". In *Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*, 575–616. New York: Elsevier.

Gupta, S. S. 1965. "On Some Multiple Decision (Selection and Ranking) Rules". *Technometrics* 7(2):225–245.

Hsu, J. C. 1984. "Constrained Simultaneous Confidence Intervals for Multiple Comparisons with the Best". *Annals of Statistics* 12(3):1136–1144.

Jian, N., S. G. Henderson, and S. R. Hunter. 2014. "Sequential Detection of Convexity from Noisy Function Evaluations". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk et al., 3892–3903. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Kim, S.-H., and B. L. Nelson. 2007. "Recent Advances in Ranking and Selection". In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson et al., 162–172. Piscataway, New Jersey: IEEE.

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–Simulation and Computation* 14(3):719–734.

Matejcik, F. J., and B. L. Nelson. 1995. "Two-Stage Multiple Comparisons with the Best for Computer Simulation". *Operations Research* 43(4):633–640.

Murota, K. 1998. "Discrete Convex Analysis". *Mathematical Programming* 83(1-3):313–371.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 2001. "Simple Procedures for Selecting the Best Simulated System when the Number of Alternatives is Large". *Operations Research* 49(6):950–963.

Salemi, P., E. Song, B. L. Nelson, and J. Staum. 2018. "Gaussian Markov Random Fields for Discrete Optimization via Simulation: Framework and Algorithms". *Operations Research*. In press.

Wan, H., B. E. Ankenman, and B. L. Nelson. 2006. "Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation". *Operations Research* 54(4):743–755.

Wang, H., R. Pasupathy, and B. W. Schmeiser. 2013. "Integer-Ordered Simulation Optimization using R-SPLINE: Retrospective Search with Piecewise-Linear Interpolation and Neighborhood Enumeration". *ACM Transactions on Modeling and Computer Simulation* 23(3):17:1–17:24.

Wright, S., and J. Nocedal. 1999. *Numerical Optimization*. New York: Springer-Verlag.

AUTHOR BIOGRAPHIES

MATTHEW PLUMLEE is an Assistant Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He primarily researches uncertainty quantification methods for computational models of systems. His e-mail address is mplumlee@northwestern.edu.

BARRY L. NELSON is the Walter P. Murphy Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IISE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail address is nelsonb@northwestern.edu.