Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Plausible Screening Using Functional Properties for Simulations with Large Solution Spaces

David J. Eckman, Matthew Plumlee, Barry L. Nelson

To cite this article:
David J. Eckman, Matthew Plumlee, Barry L. Nelson (2022) Plausible Screening Using Functional Properties for Simulations with Large Solution Spaces. Operations Research

Published online in Articles in Advance 01 Feb 2022

. https://doi.org/10.1287/opre.2021.2206

http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Articles in Advance, pp. 1–17
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Methods

Plausible Screening Using Functional Properties for Simulations with Large Solution Spaces

David J. Eckman,[a] Matthew Plumlee,[b] Barry L. Nelson[b]

[a] Wm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, College Station, Texas 77843;
[b] Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, Illinois 60208

Contact: eckman@tamu.edu, https://orcid.org/0000-0002-6473-6434 (DJE); mplumlee@northwestern.edu, https://orcid.org/0000-0002-7820-3399 (MP); nelsonb@northwestern.edu, https://orcid.org/0000-0002-1325-2624 (BLN)

Received: October 9, 2020
Revised: June 5, 2021
Accepted: August 11, 2021
Published Online in Articles in Advance: February 1, 2022

Area of Review: Simulation

https://doi.org/10.1287/opre.2021.2206

Abstract. When working with models that allow for many candidate solutions, simulation practitioners can benefit from screening out unacceptable solutions in a statistically controlled way. However, for large solution spaces, estimating the performance of all solutions through simulation can prove impractical. We propose a statistical framework for screening solutions even when only a relatively small subset of them is simulated. Our framework derives its superiority over exhaustive screening approaches by leveraging available properties of the function that describes the performance of solutions. The framework is designed to work with a wide variety of available functional information and provides guarantees on both the confidence and consistency of the resulting screening inference. We provide explicit formulations for the properties of convexity and Lipschitz continuity and show through numerical examples that our procedures can efficiently screen out many unacceptable solutions.

Funding: This work was supported by the National Science Foundation [Grants DMS-1854562, DMS-1953111].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2021.2206.

Keywords: simulation optimization • screening • feasibility

1. Introduction

Operations researchers increasingly rely on stochastic simulations to understand complex systems. These simulation models are typically endowed with a vector of parameterized inputs we term a *solution*. Each solution has an associated performance that can be estimated by running replications of the simulation with the corresponding inputs. Motivating applications of this general approach arise in simulation optimization, feasibility determination, and model calibration.

When there are many candidate solutions, it can be difficult to thoroughly evaluate the performance of all solutions through exhaustive simulation. A more reasonable approach is to first screen out, meaning remove from consideration, solutions regarded as unacceptable based on initial experiments. However, obtaining even a single replication from all candidate solutions is sometimes impractical. Thus, our goal is to provide a method for screening solutions that can work even when simulating only a small subset of them. Screening procedures can be employed to efficiently remove unacceptable solutions before running a more intensive algorithm (Nelson et al. 2001) or for post hoc analysis (Boesel et al. 2003). This use of the term "screening" differs from "factor screening," which entails removing solution-defining variables having minimal impact on the performance (Bettonvil and Kleijnen 1997, Wan et al. 2006).

Although we discuss our methodological framework in general, we at times illustrate its uses for simulation optimization, in which typically the goal is to return a single solution. Within this setting, screening methods can be used in isolation to return a set of solutions believed to have optimal or near-optimal performances. The decision maker can then make a final selection from this set based on secondary performance measures or other practical considerations. Alternatively, the set of solutions returned by a screening procedure can be provided as input to a simulation-optimization algorithm designed for discrete feasible regions, such as COMPASS (Hong and Nelson 2006, Xu et al. 2010), nested partitions (Shi and Ólafsson 2000), or any number of ranking-and-selection (R&S) algorithms (Nelson et al. 2001). Screening methods can also be embedded within a simulation-optimization search, for example, nested partitions, adaptive random search (Andradóttir and Prudius 2010), and empirical stochastic branch and bound (Xu and Nelson 2013), to assess the plausible optimality of candidate solutions before simulating them. Our methods differ from many stand-alone simulation-optimization algorithms in that they deliver marginal (sometimes finite-sample) guarantees on a returned set

of solutions as opposed to asymptotic guarantees on a single returned solution.

Classical subset-selection methods (Gupta 1965, Nelson et al. 2001, Boesel et al. 2003) return a subset of solutions and guarantee that the optimal solution is retained with high probability. Although these methods are highly effective and have been extended to parallel computing environments (Ni et al. 2014), they do not solve our problem as posed as they still require simulating all candidate solutions. These methods treat the performances of solutions as being unrelated to their location in the solution space and, therefore, do not exploit any structural properties of the performance function.

A closely related technique that directly targets the performance function is simulation metamodeling. Here, one uses simulation outputs from replications obtained at a small set of solutions to build an approximate model of the performance function, often based on statistical or machine learning models. These metamodels allow one to predict performances at unsimulated solutions. Some metamodeling methods formalize functional properties as constraints and determine the metamodel that best fits the simulation outputs subject to those constraints, for example, convex and polynomial regression (Lim and Glynn 2012, Kleijnen 2015). Others impose a probabilistic structure, for example, Gaussian process regression (Ankenman et al. 2010) or Gaussian Markov random fields (Salemi et al. 2019b). Although metamodels are central to some simulation-optimization searches, for example, stochastic trust-region methods such as STRONG (Chang et al. 2013) and ASTRO-DF (Shashaani et al. 2018), to the best of our knowledge metamodels have not been used for screening. Furthermore, metamodels do not naturally lend themselves to probabilistic guarantees about the *relative* performance of a solution (e.g., optimality) without extremely strong assumptions (Wan et al. 2016). For example, the commonly used commercial software OptQuest employs neural networks to remove solutions from consideration (Laguna 2011), but the procedure lacks statistical guarantees on the screening inference. Our methods achieve the best of both: screening out unsimulated solutions while providing a statistical guarantee akin to that of subset selection.

Our framework converts general information about the performance function into a screening approach delivering statistical guarantees. This is valuable because, for some simulation models, it is possible to analytically or empirically establish properties of the performance function, such as Lipschitz continuity, convexity, or bounds. More specifically, we propose screening solutions by measuring the discrepancy between the observed data and the space of performance functions having certain known properties; our framework, thus, shares some concepts with constrained statistical inference (Silvapulle and Sen 2005). When further restricting the space of functions to those for which a particular solution is acceptable, this discrepancy measures the plausible acceptability of said solution. A very large discrepancy at a solution implies it is implausible that the solution is acceptable. Our methods accordingly remove from consideration solutions for which the discrepancy is sufficiently large—an act we term *plausible screening*. We prescribe reasonable discrepancies and cutoffs that achieve standard statistical properties desired in screening. With proper care, our methods can provide *confidence*—which can be thought of as the probability of correct selection guarantee from subset selection—and *consistency*—the concept that any unacceptable solution is screened out in the limit. Our results here substantially extend the preliminary results presented in Plumlee and Nelson (2018) and Eckman et al. (2020).

This article introduces the screening framework, details the computational implementation, and provides some numerical examples. In Section 2, we mathematically formulate the problem of screening unacceptable solutions, and in Section 3, we motivate our approach of exploiting available information about the performance function. Section 4 lays out the theoretical underpinning for assessing plausible acceptability and presents an algorithm for constructing a subset that attains asymptotic confidence and a weak form of consistency. We then present an alternative algorithm in Section 5 that can, in certain instances, more efficiently construct a relaxed subset of solutions. In Section 6, we test the algorithms on realistic simulation-optimization problems. We conclude in Section 7 with potential extensions of the framework and open research questions.

2. Setting and Goals

This section describes the setup for evaluating solutions via stochastic simulation, the general definition of acceptable solutions, and the statistical guarantees we desire in screening.

2.1. Stochastic Simulation

We lay out a mathematical framework for screening simulated solutions from a set of candidate solutions $\mathcal{X} \subseteq \mathbb{R}^d$, which can be discrete or continuous. Each solution $x \in \mathcal{X}$ has an associated scalar quantity of interest labeled $\mu(x)$, which is unknown but can be estimated by sampling replications of a stochastic simulation. We refer to $\mu(x)$ as the *performance* of solution x. For situations in which $\mathcal{X}$ is large, meaning either a large finite set, a countably infinite set, or a continuous space of solutions, estimating the performances of all candidate solutions is impractical or impossible. Thus, a decision

maker simulates only a subset of k solutions, $\mathsf{X} \equiv \{x_1, x_2, \ldots, x_k\} \subseteq \mathcal{X}$, termed the *experimental set*. Although we discuss the experimental set X generically, it may be chosen, for example, to fill $\mathcal{X}$ or to concentrate sampling around a region of interest. We find it convenient to consider the restriction of the function $\mu : \mathcal{X} \mapsto \mathbb{R}$ to X, denoted by $\mu(\mathsf{X}) \equiv (\mu(x_1), \mu(x_2), \ldots, \mu(x_k))^\top$, which is the vector of the performances of the simulated solutions. Although not directly observable, this vector can be estimated through simulation on the limited experimental set.

Let $Y_\ell(x)$ denote the (stochastic) output of the ℓth independent and identically distributed (i.i.d.) simulation replication at a solution x with $\mathbb{E}[Y_\ell(x)] = \mu(x)$ for all $\ell = 1, 2, \ldots$ and all solutions x in $\mathcal{X}$. For any pair of solutions x and x', $Y_\ell(x)$ and $Y_\ell(x')$ are related via a common covariance function $\Sigma : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ described by $\Sigma(x, x') \equiv \mathrm{Cov}(Y_\ell(x), Y_\ell(x'))$. The variance-covariance matrix denoted by $\Sigma(\mathsf{X})$ gives the covariance between outputs for all pairs of solutions in the experimental set, and we assume that $\Sigma(\mathsf{X})$ is positive definite. For a given ℓ, define $\mathbf{Y}_\ell \equiv (Y_\ell(x_1), Y_\ell(x_2), \ldots, Y_\ell(x_k))^\top$, the vector of outputs from the ℓth simulation replications at each solution in the experimental set. The vectors $\mathbf{Y}_1, \mathbf{Y}_2, \ldots,$ are assumed to be mutually independent and identically distributed. In our finite-sample results, we additionally assume that simulation outputs are jointly normally distributed, meaning

$$\mathbf{Y}_\ell \sim \mathcal{N}(\mu(\mathsf{X}), \Sigma(\mathsf{X})) \text{ for } \ell = 1, 2, \ldots. \tag{1}$$

We consider two ways of simulating replications across solutions:

Setting 1: Independent sampling: Outputs at different solutions are independent (i.e., $\Sigma(\mathsf{X})$ is diagonal), and the number of replications taken at each solution $x_i \in \mathsf{X}$ is n_i for $i = 1, 2, \ldots, k$, possibly unequal.

Setting 2: Dependent sampling: Outputs at different solutions are dependent—as would be the case if common random numbers (CRN) were used—and an equal number of replications is taken at each solution $x_i \in \mathsf{X}$, that is, $n_i = n$ for $i = 1, 2, \ldots, k$.

We estimate $\mu(\mathsf{X})$ by $\widehat{\mu} \equiv (\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_k)^\top$, where $\widehat{\mu}_i = n_i^{-1} \sum_{\ell=1}^{n_i} Y_\ell(x_i)$ for $i = 1, 2, \ldots, k$ and $\Sigma(\mathsf{X})$ by

$$\widehat{\Sigma} \equiv \begin{cases} \mathrm{diag}(\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \ldots, \widehat{\sigma}_k^2) \text{ where } \widehat{\sigma}_i^2 = (n_i - 1)^{-1} \\ \qquad \sum_{\ell=1}^{n_i} (Y_\ell(x_i) - \widehat{\mu}_i)^2 \text{for } i = 1, 2, \ldots, k \text{ in Setting 1,} \\ [\widehat{\sigma}_{ij}^2]_{k \times k} \text{ where } \widehat{\sigma}_{ij}^2 = (n-1)^{-1} \sum_{\ell=1}^{n} (Y_\ell(x_i) - \widehat{\mu}_i)(Y_\ell(x_j) - \widehat{\mu}_j) \\ \qquad \text{for } i, j = 1, 2, \ldots, k \text{ in Setting 2.} \end{cases}$$

We assume that $\widehat{\Sigma}$ is positive definite with probability one.

2.2. Acceptable Solutions

For a given performance function μ, we define $\mathcal{A}$ as the set of solutions deemed *acceptable* by the decision maker, that is, those whose performances exhibit some quality of interest. Although $\mathcal{A}$ depends on the unknown function μ, we choose to suppress μ from the notation. Different definitions of acceptability arise in a variety of simulation applications and can be illustrated within the setting of production planning, such as semiconductor wafer fabrication (Liu et al. 2011). Discrete-event simulation models are used in this domain to study the costs associated with a given release plan—a schedule of batch jobs for different product types—subject to stochastic demand for the products. A decision maker may be interested in finding a release plan x whose expected total costs (defined as the sum of work-in-progress, inventory, and backlog costs) is within δ dollars of the smallest. Alternatively, the decision maker may wish to determine whether a given release plan satisfies a service requirement, for example, that the associated expected backlog cost is below μ_0. It may also be of interest to improve upon a *control* or default release plan x^c, such as the one suggested by a simple model. On the other hand, the decision maker may be interested in release plans whose expected work in progress is within ϵ units of $\mu^\dagger$ (Spearman et al. 1990).

Although we leave $\mathcal{A}$ purposely vague to demonstrate the versatility of the proposed framework, one can describe these common examples of $\mathcal{A}$ mathematically:

- Optimization: $\{x \in \mathcal{X} : \mu(x) \leq \min_{x' \in \mathcal{X}} \mu(x') + \delta\}$ for some optimality gap $\delta \geq 0$.
- Feasibility determination: $\{x \in \mathcal{X} : \mu(x) \leq \mu_0\}$ for some threshold μ_0.
- Comparison with a control: $\{x \in \mathcal{X} : \mu(x) \leq \mu(x^c)\}$ for some control solution $x^c \in \mathcal{X}$.
- Comparison with a target: $\{x \in \mathcal{X} : |\mu(x) - \mu^\dagger| \leq \epsilon\}$ for some tolerance $\epsilon \geq 0$ and target $\mu^\dagger$.

In the first three examples, it is assumed without loss of generality that smaller performance is preferable. A common feature is that determining whether a given solution belongs to $\mathcal{A}$ entails checking a (possibly infinite) system of linear inequalities with respect to the candidate solutions' performances. We later leverage this property to develop tractable methods for inferring whether an arbitrary solution is acceptable.

2.3. Statistical Guarantees in Screening

Ideally, the decision maker seeks to identify the full set of acceptable solutions and no others, to serve as the basis for some decision. In the presence of simulation error, the decision maker must settle for a subset of solutions having desirable statistical guarantees in terms of *screening*, that is, inferring which solutions are acceptable (Bechhofer et al. 1995). Let $\mathcal{S}_\mathsf{n}$ denote the subset of solutions returned after obtaining replications

at solutions in X as specified by $\mathsf{n} \equiv (n_1, n_2, \ldots, n_k)$ and screening all solutions in $\mathcal{X}$. Screening all solutions in a countably infinite or continuous solution space $\mathcal{X}$ is impossible. In such situations, one could discretize $\mathcal{X}$ and screen all solutions in the discretized set; however, this approach suffers in higher dimensions. Our screening machinery could also be employed to screen a smaller number of solutions that are of particular interest.

Our definitions of statistical guarantees of subsets suppose that the performance function μ belongs to some function space $\mathscr{M}$, which we specify in Section 3.2.

Definition 1 (Finite-Sample Confidence). A subset $\mathcal{S}_\mathsf{n}$ achieves *finite-sample confidence* $1-\alpha$ for $\alpha \in (0,1]$ if, for $\min_{i=1,2,\ldots,k} n_i \geq 2$ for Setting 1 (respectively, $\min_{i=1,2,\ldots,k} n_i \geq k+1$ for Setting 2) and any $\mu \in \mathscr{M}$, $\mathbb{P}(x_0 \in \mathcal{S}_\mathsf{n}) \geq 1-\alpha$ for all $x_0 \in \mathcal{A}$.

Finite-sample confidence states that, for any performance function in $\mathscr{M}$, each acceptable solution is correctly screened with marginal probability exceeding $1-\alpha$. For the most part, finite-sample confidence is unattainable unless the random outputs of the simulation replications come from a known family of distributions, as in Equation (1). A more widely achievable property is asymptotic confidence, which follows from designing methods for normally distributed outputs and applying the central limit theorem.

Definition 2 (Asymptotic Confidence). A subset $\mathcal{S}_\mathsf{n}$ achieves *asymptotic confidence* $1-\alpha$ for $\alpha \in (0,1]$ if, for any $\mu \in \mathscr{M}$, $\mathbb{P}(x_0 \in \mathcal{S}_\mathsf{n}) \gtrsim 1-\alpha$ as $\min_{i=1,2,\ldots,k} n_i \to \infty$ for all $x_0 \in \mathcal{A}$.

In Definition 2, the statement $\mathbb{P}(x_0 \in \mathcal{S}_\mathsf{n}) \gtrsim 1-\alpha$ means that, for any $\varepsilon > 0$, there exists an $n(\varepsilon, x_0)$ such that, for all n for which $\min_{i=1,2,\ldots,k} n_i \geq n(\varepsilon, x_0)$, $\mathbb{P}(x_0 \in \mathcal{S}_\mathsf{n}) \geq 1-\alpha-\varepsilon$. Finite-sample and asymptotic confidence are marginal guarantees, holding solution-wise. For finite solution spaces, one could conceivably deliver set-wise guarantees by splitting α over $|\mathcal{X}|$; we chose not to explore this further.

Finite-sample and asymptotic confidence describe a subset's ability to avoid screening out acceptable solutions with high probability but not its ability to screen out unacceptable solutions, that is, those that do not belong to $\mathcal{A}$. For this, we require the notion of consistency.

Definition 3 (Consistency). A subset $\mathcal{S}_\mathsf{n}$ achieves consistency if, for any $\mu \in \mathscr{M}$, $\mathbb{P}(x_0 \in \mathcal{S}_\mathsf{n}) \to 0$ as $\min_{i=1,2,\ldots,k} n_i \to \infty$ for all $x_0 \notin \mathcal{A}$.

Except in special cases, such as exhaustive simulation in which $\mathsf{X} = \mathcal{X}$, consistency is unachievable because, even with direct evaluation of $\mu(\mathsf{X})$, the rest of the performance function is indeterminable. We soon introduce a less exacting form of consistency that accounts for having simulated at only solutions in the experimental set.

3. Screening Using Functional Properties

In this section, we explain how known functional properties of the performance function can be combined with our screening framework. Section 3.3 develops the main ideas behind our methods in a simplified setting in which solutions are simulated without error.

3.1. Functional Properties of the Performance Function

Our goal is to use information obtained from a small experimental set to screen out a massive number of solutions that, given the data, could not plausibly be acceptable. Importantly, we seek the ability to screen out even unsimulated solutions. This task would be impossible without some means of relating an unsimulated solution's performance to those of simulated solutions from the experimental set. Our approach operates under the assumption that the decision maker possesses known or assumed properties of the performance function μ that enable such comparisons. Examples include knowledge that μ is convex (likewise concave, strongly convex, or almost convex) over $\mathcal{X}$, Lipschitz continuous (likewise Hölder continuous or second-order Lipschitz continuous) with a known or assumed upper bound on the associated constant or a polynomial in x with known or assumed degree. This type of information can also be augmented with auxiliary properties on the performances of individual solutions, such as bounds on μ or known performances of some solutions.

Our framework involves checking whether an arbitrary function lies in a function space described by the known properties. Some other methods instead assume a probabilistic structure for μ and leverage measures over function spaces. For instance, it is common in metamodeling (Santner et al. 2003, Salemi et al. 2019a) and Bayesian optimization (Frazier et al. 2009, Scott et al. 2011) to treat μ as a realization of a Gaussian process on $\mathcal{X}$. Gaussian processes postulate structural information in the form of a probability model that characterizes the performance function via correlation in the performances of neighboring solutions. Although this structure is never actually true, it provides a powerful paradigm for guiding a search. Our screening framework instead exploits actual properties of the performance function, and therefore, its conclusions are independent of any hypothesized prior distribution or estimates of artificial parameters. By the same token, our approach differs from that of estimating the posterior probability that μ satisfies certain functional

properties, for example, convexity (Jian and Henderson 2020).

Situations in which such knowledge of functional properties is available are not rare. In Section 6, we explicate two numerical examples in which Lipschitz continuity or convexity information is present. The functional properties in these examples are identified by studying continuous extensions of problems that originally feature discrete feasible regions. We call attention to two common, practical techniques for verifying functional properties of simulation models.

3.1.1. Inheritance from Sample-Path Functions. Many properties of sample-path functions are inherited when applying the expectation operator, for example, convexity (Shaked and Shanthikumar 1988), continuity (Shapiro and Wardi 1996), and bounds. Proving that the sample-path functions possess any such property with probability one implies that the performance function does as well (Kim et al. 2015). Examples include

Stochastic activity networks. The expected length of the longest path is a convex function in terms of the mean task durations; see appendix E of Plambeck et al. (1996) for a derivation.

Tandem production lines with unreliable machines. The steady-state throughput is a convex function in terms of the cycle times of the machines (Plambeck et al. 1996).

Inventory stocking under dynamic customer substitution. The expected profit can be shown to be a Hölder-continuous function in terms of the initial inventory levels (Mahajan and van Ryzin 2001).

3.1.2. Stochastic Orders. Some stochastic orders imply an inequality relating two expected values (Shaked and Shanthikumar 2007). This approach can be used to relate $\mu(x)$ and $\mu(x')$ for some $x \neq x'$ or to relate $\mu(x)$ to another expected value that is known, thereby providing a bound on $\mu(x)$. Examples include

$GI/GI/c$ queueing systems. Many random quantities of interest (e.g., sequential departure times) are stochastically ordered when comparing a $GI/GI/c$ queueing system with a first-in, first-out service discipline to another in which arrivals are arbitrarily assigned among c channels, independent of the service process (Wolff 1977).

Portfolio optimization. A risk-averse decision maker wishes to assemble a portfolio from a finite collection of assets to maximize the expected return rate while requiring that the portfolio's return rate stochastically dominate a benchmark rate (Dentcheva and Ruszczyński 2006).

3.2. Spaces of Performance Functions

We incorporate functional properties into our framework by characterizing how they restrict the set of functions to which μ can belong and, in turn, the values its restriction $\mu(\mathsf{X})$ can take. Let $\mathscr{F}$ denote the set of functions mapping from $\mathcal{X}$ to $\mathbb{R}$ and let $\mathscr{M} \subseteq \mathscr{F}$ denote the set of functions that possess the specified functional properties. Furthermore, for a given performance function $m \in \mathscr{M}$, let $\mathcal{A}(m)$ represent the corresponding set of acceptable solutions. Because screening takes place by examining individual solutions $x_0 \in \mathcal{X}$, we define $\mathscr{M}(x_0) \equiv \{m \in \mathscr{M} : x_0 \in \mathcal{A}(m)\}$, the set of functions in $\mathscr{M}$ for which solution x_0 is acceptable.

The set $\mathscr{M}(x_0)$ is difficult to work with because elements of $\mathscr{M}(x_0)$ are infinite dimensional. We achieve a more tractable formulation by projecting $\mathscr{M}(x_0)$ onto $\mathbb{R}^k$, with elements of this projected set corresponding to vectors of the performances of solutions in X. Recall that $\mu(\mathsf{X})$ represents the performances of the solutions in our experimental set, $x_1, x_2, \ldots, x_k$. We similarly use the notation $m(\mathsf{X})$ to denote the values an arbitrary function m takes at those same $x_1, x_2, \ldots, x_k$. The resulting projection of $\mathscr{M}(x_0)$ onto $\mathbb{R}^k$ is defined as

$$\mathsf{M}(x_0) \equiv \{\mathsf{m} \in \mathbb{R}^k : \text{there exists } m \in \mathscr{M}(x_0) \text{ such that } m(\mathsf{X}) = \mathsf{m}\},$$

the set of vectors of performances of the solutions $x_1, x_2, \ldots, x_k$ for which there exists an interpolating function m belonging to $\mathscr{M}(x_0)$. We illustrate these definitions for two problems: feasibility determination for a Lipschitz-continuous performance function and minimization of a convex performance function. Complete derivations are given in Online Appendix EC.1.

Example 1 (Feasibility Determination for a Lipschitz-Continuous Function). Let $\mathcal{A} = \{x \in \mathcal{X} : \mu(x) \leq \mu_0\}$ for some threshold μ_0 and suppose that μ is known to be Lipschitz continuous with constant γ. The set of γ-Lipschitz functions for which a given solution x_0 is feasible with respect to μ_0 is given by

$$\mathscr{M}(x_0) = \{m \in \mathscr{F} : |m(x) - m(x')| \leq \gamma \|x - x'\| \text{ for all } x, x' \in \mathcal{X} \text{ and } m(x_0) \leq \mu_0\},$$

where $\|\cdot\|$ denotes the Euclidean norm. Furthermore,

$$\mathsf{M}(x_0) = \{\mathsf{m} \in \mathbb{R}^k : \mathsf{m}_i - \mathsf{m}_j \leq \gamma \|x_i - x_j\| \text{ for all } i, j = 1, 2, \ldots, k \text{ and } \mathsf{m}_i \leq \mu_0 + \gamma \|x_i - x_0\| \text{ for all } i = 1, 2, \ldots, k\},$$

where m_i is the ith component of the vector m.

Example 2 (Minimizing a Convex Function). Let $\mathcal{A} = \{x \in \mathcal{X} : \mu(x) \leq \min_{x' \in \mathcal{X}} \mu(x')\}$ and suppose that μ is

known to be convex. One formulation of the set of convex functions for which a given solution x_0 is optimal is

$$\begin{aligned}\mathcal{M}(x_0) = \Big\{ & m \in \mathcal{F} : \text{for all } x \in \mathcal{X}, \text{ there exists } \xi(x) \in \mathbb{R}^d \\ & \text{such that} \\ & m(x) - m(x') \le (x - x')^\top \xi(x) \text{ for all } x' \in \mathcal{X} \text{ and} \\ & m(x_0) \le m(x) \Big\},\end{aligned}$$

where $\xi(x)$ represents a subgradient at the solution x. Furthermore,

$$\begin{aligned}\mathsf{M}(x_0) = \Big\{ & \mathsf{m} \in \mathbb{R}^k : \text{there exists } \mathsf{m}_0 \in \mathbb{R} \text{ and} \\ & \xi_1, \xi_2, \ldots, \xi_k \in \mathbb{R}^d \text{ such that} \\ & \mathsf{m}_i - \mathsf{m}_j + (x_j - x_i)^\top \xi_i \le 0 \text{ for all } i, j = 1, 2, \ldots, k \\ & \mathsf{m}_i - \mathsf{m}_0 + (x_0 - x_i)^\top \xi_i \le 0 \text{ for all } i = 1, 2, \ldots, k \\ & -\mathsf{m}_i + \mathsf{m}_0 \le 0 \text{ for all } i = 1, 2, \ldots, k \Big\}.\end{aligned}$$

Here, m_0 represents the performance of solution x_0 and $\xi_1, \xi_2, \ldots, \xi_k$ represent subgradients at solutions $x_1, x_2, \ldots, x_k$. (The term m_0 could be projected out to further simplify the formulation of $\mathsf{M}(x_0)$.)

It follows from these definitions that, for a given performance function $\mu \in \mathcal{M}$, its restriction $\mu(\mathsf{X})$ is in $\mathsf{M}(x_0)$ if x_0 is an acceptable solution. The converse, however, does not necessarily hold, because the restriction of μ to X does not determine the performances of solutions in the rest of the solution space. We next demonstrate the central role $\mathsf{M}(x_0)$ plays in screening.

3.3. Screening Solutions Without Estimation Error

Temporarily assume that solutions' performances can be calculated directly without simulation, that is, the decision maker can directly obtain $\mu(\mathsf{X})$. Given that only solutions in the experimental set have been evaluated, some solutions likely cannot be correctly screened with certainty. A reasonable approach is to classify as belonging to $\mathcal{A}$ any solution x_0 for which there exists a function in $\mathcal{M}(x_0)$ that interpolates $\mu(\mathsf{X})$. We denote the resulting subset of solutions by

$$\mathsf{S}(\mathsf{X}) \equiv \{x_0 \in \mathcal{X} : \mu(\mathsf{X}) \in \mathsf{M}(x_0)\}.$$

This notation reflects the dependence of the subset $\mathsf{S}(\mathsf{X})$ on X; a different experimental set would yield a different subset of solutions that are possibly acceptable.

Example 1 (Continued). For a γ-Lipschitz performance function μ, the set of solutions that are possibly feasible with respect to a threshold μ_0 is given by

$$\mathsf{S}(\mathsf{X}) = \left\{ x_0 \in \mathcal{X} : \max_{i=1,2,\ldots,k} \{\mu(x_i) - \gamma \|x_i - x_0\|\} \le \mu_0 \right\}.$$

Example 2 (Continued). For a convex performance function, the set of possibly optimal solutions is given by

$$\begin{aligned}\mathsf{S}(\mathsf{X}) = \Big\{ & x_0 \in \mathcal{X} : \text{there exists } \mathsf{m}_0 \in \mathbb{R} \text{ and} \\ & \xi_1, \xi_2, \ldots, \xi_k \in \mathbb{R}^d \text{ such that} \\ & (x_j - x_i)^\top \xi_i \le \mu(x_j) - \mu(x_i) \text{ for all } i, j = 1, 2, \ldots, k \\ & -\mathsf{m}_0 + (x_0 - x_i)^\top \xi_i \le -\mu(x_i) \text{ for all } i = 1, 2, \ldots, k \\ & \mathsf{m}_0 \le \mu(x_i) \text{ for all } i = 1, 2, \ldots, k \Big\}.\end{aligned}$$

The subset $\mathsf{S}(\mathsf{X})$ is the smallest subset that contains all acceptable solutions having only evaluated solutions in X and only knowing the given properties of μ. If an arbitrary solution x_0 is not in $\mathsf{S}(\mathsf{X})$, we conclude that there does not exist an interpolating function in $\mathcal{M}(x_0)$; hence, it is impossible that x_0 is an acceptable solution. Therefore, all acceptable solutions are included in this subset, that is, $\mathcal{A} \subseteq \mathsf{S}(\mathsf{X})$. In addition, if $\mathsf{X} = \mathcal{X}$, then all solutions can be correctly screened, meaning $\mathsf{S}(\mathsf{X}) = \mathcal{A}$. Thus, the gap between these two subsets of solutions comes from the fact that the experimental set comprises only a subset of the candidate solutions.

To determine if a solution x_0 belongs to $\mathsf{S}(\mathsf{X})$, one must check whether $\mu(\mathsf{X})$ belongs to $\mathsf{M}(x_0)$. If $\mathsf{M}(x_0)$ can be expressed as a polyhedron with an explicit constraint matrix and right-hand-side vector, as in Example 1, then checking if $\mu(\mathsf{X})$ is in $\mathsf{M}(x_0)$ is straightforward. More generally, if $\mathsf{M}(x_0)$ can be implicitly described as the projection of a polyhedron, as in Example 2, then checking if $\mu(\mathsf{X})$ is in $\mathsf{M}(x_0)$ involves solving a linear program. In Section 5, we exploit this fact to devise efficient methods for screening solutions. One can imagine further expanding this framework to nonlinear constraints, but this paper focuses on the potential in polyhedral representations.

As a setup for the following sections, we present a relaxed version of consistency featuring $\mathsf{S}(\mathsf{X})$.

Definition 4 ($\mathsf{S}(\mathsf{X})$ Consistency). A subset $\mathcal{S}_\mathsf{n}$ achieves $\mathsf{S}(\mathsf{X})$ consistency if, for any $\mu \in \mathcal{M}$, $\mathbb{P}(x_0 \in \mathcal{S}_\mathsf{n}) \to 0$ as $\min_{i=1,2,\ldots,k} n_i \to \infty$ for all $x_0 \notin \mathsf{S}(\mathsf{X})$.

As $\mathcal{A}$ is a subset of $\mathsf{S}(\mathsf{X})$, we conclude that $\mathsf{S}(\mathsf{X})$ consistency holds whenever consistency (Definition 3) holds. The property of $\mathsf{S}(\mathsf{X})$ consistency implies that, as the simulation effort at solutions in X increases to infinity, the probability that a given solution is in $\mathcal{S}_\mathsf{n}$ goes to zero for any solution that could be screened out if $\mu(\mathsf{X})$ were known. In other words, an $\mathsf{S}(\mathsf{X})$-consistent subset asymptotically screens out all solutions that—given the limited experimental set and known functional properties of the performance function—cannot possibly be acceptable.

4. Plausible Screening

In this section, we give an overview of our method of accounting for simulation error alongside theoretical results that justify its use.

4.1. Overview

When solutions in the experimental set are simulated without error, as in Section 3.3, a natural subset to return is $\mathsf{S}(\mathsf{X})$, which consists of all solutions x_0 for which $\mu(\mathsf{X}) \in \mathsf{M}(x_0)$. However, when there is simulation error, naively plugging in the estimator $\widehat{\mu}$ for the unknown $\mu(\mathsf{X})$ and retaining all solutions x_0 for which $\widehat{\mu} \in \mathsf{M}(x_0)$ does not produce a subset achieving confidence and $\mathsf{S}(\mathsf{X})$ consistency. Because the probability that $\widehat{\mu} \in \mathsf{M}(x_0)$ is not well controlled, this likely results in a set that eliminates too many solutions, thus violating the confidence guarantee through undercoverage. We properly account for the uncertainty about $\mu(\mathsf{X})$ by developing a subset comprising solutions x_0 for which $\widehat{\mu}$ is sufficiently close to $\mathsf{M}(x_0)$, where the precise meaning of "sufficiently close" ensures our guarantees of confidence and $\mathsf{S}(\mathsf{X})$ consistency are delivered.

To measure the distance between $\widehat{\mu}$ and $\mathsf{M}(x_0)$, we first introduce the *standardized discrepancy* between $\widehat{\mu}$ and a performance vector $\mathsf{m} = (\mathsf{m}_1, \mathsf{m}_2, \ldots, \mathsf{m}_k)$, denoted by $d_{\mathsf{n}}(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma})$. The vector of sample sizes, n, and sample variance-covariance matrix, $\widehat{\Sigma}$, appear in the standardized discrepancy for the purpose of scaling differences between performance vectors in line with the estimation error; specific examples are given as follows. We hereafter assume that the standardized discrepancy satisfies the following condition.

Condition 1. *$d_{\mathsf{n}}(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) \geq 0$ for all $\mathsf{m} \in \mathbb{R}^k$ and $d_{\mathsf{n}}(\widehat{\mu}, \widehat{\mu}, \widehat{\Sigma}) = 0$ with probability one.*

Additional conditions are introduced in Section 4.2 that are necessary for maintaining confidence and $\mathsf{S}(\mathsf{X})$ consistency.

Minimizing the standardized discrepancy over performance vectors in $\mathsf{M}(x_0)$ gives the *minimum standardized discrepancy* of x_0,

$$D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma}) \equiv \min_{\mathsf{m} \in \mathsf{M}(x_0)} d_{\mathsf{n}}(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}), \tag{2}$$

which can be interpreted as the distance between the sample mean vector $\widehat{\mu}$ and the set $\mathsf{M}(x_0)$. The minimum standardized discrepancy is an indication of how likely it is that, given the sample data, the true performance function μ belongs to $\mathscr{M}(x_0)$, the space of functions that possess the known functional properties and for which solution x_0 is acceptable. A smaller value of $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$ indicates stronger evidence that x_0 is an acceptable solution, and a larger value of $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$ indicates stronger evidence that x_0 is an unacceptable solution.

We say that a function m is *plausible* with respect to an arbitrary solution x_0 if it belongs to $\mathscr{M}(x_0)$ and its restriction $m(\mathsf{X})$ is sufficiently close to $\widehat{\mu}$ in terms of the standardized discrepancy between the two vectors. From Definition 2, a solution x_0 admits a plausible function if and only if its minimum standardized discrepancy $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$ is sufficiently small. Our screening method, which we refer to as plausible screening (PS), returns the subset comprising solutions x_0 for which there exists a plausible function. To be precise, the PS subset $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ consists of solutions x_0 for which $\widehat{\mu}$ is within a distance D of $\mathsf{M}(x_0)$, that is,

$$\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}} \equiv \left\{ x_0 \in \mathcal{X} : D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma}) \leq \mathsf{D} \right\}.$$

Equivalently, $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ can be defined as all $x_0 \in \mathcal{X}$ such that $\widehat{\mu} \in \mathsf{R}(x_0)$, where

$$\mathsf{R}(x_0) \equiv \left\{ \widetilde{\mathsf{m}} \in \mathbb{R}^k : D_{\mathsf{n}}(x_0, \widetilde{\mathsf{m}}, \widehat{\Sigma}) \leq \mathsf{D} \right\}$$

is the set of performance vectors that are within a distance D to $\mathsf{M}(x_0)$. Just as $\mathsf{M}(x_0)$ is the performance set for which x_0 is possibly acceptable when $\mu(\mathsf{X})$ is directly observed, its random relaxation $\mathsf{R}(x_0)$ can be viewed as a performance set for which x_0 is plausibly acceptable in light of the uncertainty about $\mu(\mathsf{X})$.

4.2. Statistical Guarantees

From the definition of $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$, we can see that choosing D as the $1-\alpha$ quantile of the minimum standardized discrepancy $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$ leads to finite-sample confidence. However, the distribution of the minimum standardized discrepancy depends on the unknown quantities $\mu(\mathsf{X})$ and $\Sigma(\mathsf{X})$ in addition to sample sizes, functional constraints, and the definition of acceptability. For the cases we investigate, namely, Lipschitz continuity and convexity of μ, the associated quantile cannot be evaluated numerically or by Monte Carlo. We circumvent this by considering the statistic $d_{\mathsf{n}}(\mu(\mathsf{X}), \widehat{\mu}, \widehat{\Sigma})$, which first order stochastically dominates the minimum standardized discrepancy because, when x_0 is an acceptable solution, $\mu(\mathsf{X}) \in \mathsf{M}(x_0)$, and hence, $d_{\mathsf{n}}(\mu(\mathsf{X}), \widehat{\mu}, \widehat{\Sigma}) \geq \min_{\mathsf{m} \in \mathsf{M}(x_0)} d_{\mathsf{n}}(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) = D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$. We introduce standardized discrepancies for which $d_{\mathsf{n}}(\mu(\mathsf{X}), \widehat{\mu}, \widehat{\Sigma})$ is pivotal under a normality assumption; that is, its distribution is independent of $\mu(\mathsf{X})$ and $\Sigma(\mathsf{X})$. Its distribution is also independent of $\mathsf{M}(x_0)$ because setting $\mathsf{m} = \mu(\mathsf{X})$ avoids the minimization in Definition 2. This simplification allows us to derive a deterministic, uniform cutoff D that ensures $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ has the desired statistical properties.

We require that the pairing of $d_{\mathsf{n}}(\cdot, \widehat{\mu}, \widehat{\Sigma})$ and D satisfy three conditions for all $\mu(\mathsf{X}) \in \mathbb{R}^k$ and $\Sigma(\mathsf{X}) \in \mathbb{R}^{k\times k}$ positive definite:

Condition 2. $\mathbb{P}(d_{\mathsf{n}}(\mu(\mathsf{X}), \widehat{\mu}, \widehat{\Sigma}) \le \mathsf{D}) \ge 1-\alpha$ *for sufficiently large* $\min_{i=1,2,\ldots,k} n_i$.

Condition 3. $\mathbb{P}\left(d_{\mathsf{n}}(\mu(\mathsf{X}), \widehat{\mu}, \widehat{\Sigma}) \le \mathsf{D}\right) \to 1-\alpha$ *as* $\min_{i=1,2,\ldots,k} n_i \to \infty$.

Condition 4. $\max_{\mathsf{m}\in\mathbb{R}^k}\{\|\widehat{\mu}-\mathsf{m}\|\colon d_{\mathsf{n}}(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) \le \mathsf{D}\} \overset{w.p.1}{\to} 0$ *as* $\min_{i=1,2,\ldots,k} n_i \to \infty$, *where* $\|\cdot\|$ *again denotes the Euclidean norm.*

Although the choice of D satisfying the conditions depends on the values of k, n, and α, we choose to suppress this dependence in the notation.

Conditions 2 and 3 relate to finite-sample and asymptotic confidence, respectively, ensuring that D is sufficiently large. Condition 4, on the other hand, relates to consistency. It ensures that D remains sufficiently small as the sample sizes increase so that, for a solution to be included in $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$, the restriction of the best-fitting model to the solutions in the experimental set must more closely align with the observed sample means.

Theorems 1 and 2 establish that, under Conditions 2–4, $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ possesses the desired properties of confidence and $\mathsf{S}(\mathsf{X})$ consistency; proofs appear in Online Appendix EC.3.

Theorem 1. *If* $d_{\mathsf{n}}(\cdot, \widehat{\mu}, \widehat{\Sigma})$ *and* D *satisfy* Conditions 2 *and* 3, *then* $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ *achieves finite-sample confidence and asymptotic confidence.*

Theorem 2. *If* $d_{\mathsf{n}}(\cdot, \widehat{\mu}, \widehat{\Sigma})$ *and* D *satisfy* Condition 4, *then* $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ *achieves* $\mathsf{S}(\mathsf{X})$ *consistency.*

4.3. Standardized Discrepancies

Our screening framework can easily accommodate different choices of standardized discrepancies and cutoffs. We present several examples that satisfy Conditions 1–4 and provide a representative proof in Online Appendix EC.3. Condition 2 is established under the normality assumption stated in Equation (1).

In Setting 1, Conditions 1–4 are satisfied by

$$d_{\mathsf{n}}^1(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) \equiv \sum_{i=1}^{k} \frac{\sqrt{n_i}}{\widehat{\sigma}_i} |\widehat{\mu}_i - \mathsf{m}_i|$$

with D^1 defined as the $1-\alpha$ quantile of the sum of the absolute value of k independent t-distributed random variables, each with degrees of freedom $n_1-1, n_2-1, \ldots, n_k-1$, respectively; by

$$d_{\mathsf{n}}^2(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) \equiv \sum_{i=1}^{k} \frac{n_i}{\widehat{\sigma}_i^2} (\widehat{\mu}_i - \mathsf{m}_i)^2$$

with D^2 defined as the $1-\alpha$ quantile of the sum of k independent F-distributed random variables, each with numerator degrees of freedom one and denominator degrees of freedom $n_1-1, n_2-1, \ldots, n_k-1$, respectively; and by

$$d_{\mathsf{n}}^{\infty}(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) \equiv \max_{i=1,2,\ldots,k} \frac{\sqrt{n_i}}{\widehat{\sigma}_i} |\widehat{\mu}_i - \mathsf{m}_i|$$

with D^{∞} defined as the $1-\alpha$ quantile of the maximum of the absolute value of k independent t-distributed random variables, each with n_i-1 degrees of freedom. In our discussion, we find it convenient to refer to these standardized discrepancies by the shorthand d_{n}^1, d_{n}^2, and d_{n}^{∞}. Plumlee and Nelson (2018) focus on the choice of d_{n}^2 and D^2, and Eckman et al. (2020) explores connections to existing screening methods, such as the Screen-to-the-Best (STB) procedure (Nelson et al. 2001).

In Setting 2, Conditions 1–4 are satisfied by

$$d_{\mathsf{n}}^{\mathrm{CRN}}(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) \equiv n(\widehat{\mu}-\mathsf{m})^{\top} \widehat{\Sigma}^{-1} (\widehat{\mu}-\mathsf{m})$$

with $\mathsf{D}^{\mathrm{CRN}}$ defined as $k(n-1)/(n-k)$ times the $1-\alpha$ quantile of an F-distributed random variable with numerator degrees of freedom k and denominator degrees of freedom $n-k$ (Anderson 1984). In Setting 2, for $\widehat{\Sigma}$ to be invertible with probability one, a minimum of $k+1$ replications must be obtained from each solution, that is, $n \ge k+1$ (Anderson 1984).

As can be seen from these examples, a uniform cutoff D can be specified as the $1-\alpha$ quantile of the pivotal statistic $d_{\mathsf{n}}(\mu(\mathsf{X}), \widehat{\mu}, \widehat{\Sigma})$. The given cutoffs D^1, D^2, D^{∞}, and $\mathsf{D}^{\mathrm{CRN}}$ are the tightest uniform cutoffs for their respective standardized discrepancies that deliver finite-sample confidence irrespective of the properties of μ. To see this, consider the case in which the decision maker has complete knowledge of the performances of the solutions in the experimental set, that is, $\mathsf{M}(x_0) = \{\mu(\mathsf{X})\}$. Thus, $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma}) = d_{\mathsf{n}}(\mu(\mathsf{X}), \widehat{\mu}, \widehat{\Sigma})$ and the specified cutoffs are exactly the $1-\alpha$ quantiles of the minimum standardized discrepancies. The coverage of any acceptable solution x_0 is therefore exactly $1-\alpha$.

5. Computational Considerations and Relaxed Screening

Constructing $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ entails repeatedly solving the optimization problem described in Definition 2 and comparing its optimal value, $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$, to the cutoff D for each $x_0 \in \mathcal{X}$. Depending on the difficulty of the optimization problem and the number of candidate solutions, constructing $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ in this manner could be computationally expensive. In this section, we present

an alternative subset consisting of solutions for which $\widehat{\mu}$ belongs to a polyhedral relaxation of $\mathsf{R}(x_0)$. Screening a solution, therefore, involves solving a linear program which, in certain cases, can be substantially cheaper. Compared with the subset $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$, this approach results in a more conservative subset in the sense that it contains all of the solutions in $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ and possibly more.

5.1. Polyhedral Relaxation of $\mathsf{R}(x_0)$ via a Relaxation of $\mathscr{M}(x_0)$

We demonstrate this conservative approach for the situation in which $\mathsf{M}(x_0)$ can be described as the projection of a polyhedron.

Assumption 1. *For each solution* $x_0 \in \mathcal{X}$,

$$\mathsf{M}(x_0) = \{\mathsf{m} \in \mathbb{R}^k : \textit{there exists } \mathsf{w} \in \mathbb{R}^q \textit{ such that } A\mathsf{m} + C\mathsf{w} \le b\},$$

for some $A \in \mathbb{R}^{p\times k}$, $C \in \mathbb{R}^{p\times q}$, $b \in \mathbb{R}^p$, *where* A, C, *and* b *may depend on* x_0 *and* X.

For A, C, and b in Assumption 1, we suppress x_0 and X for notational convenience.

Assumption 1 depends on both the choice of function space and the definition of the set of acceptable solutions. This assumption holds for most combinations discussed in this paper, for example, feasibility determination for a Lipschitz-continuous function or minimization of a convex function. From the expressions for $\mathsf{M}(x_0)$ in Examples 1 and 2, one can directly obtain the corresponding A, C, and b.

To explain our approach, define the polyhedron $\mathsf{P} \equiv \{(\mathsf{m},\mathsf{w}) \in \mathbb{R}^k \times \mathbb{R}^q : A\mathsf{m} + C\mathsf{w} \le b\}$ such that the projection of P onto $\mathbb{R}^k$ is $\mathsf{M}(x_0)$. We obtain a relaxation of $\mathsf{M}(x_0)$ by first relaxing P, by increasing its right-hand-side vector b, and then projecting the enlarged polyhedron onto $\mathbb{R}^k$. To compensate for the uncertainty about $\mu(\mathsf{X})$, we offset b by defining

$$b_j' = b_j + \max_{\mathsf{m}\in\mathbb{R}^k} \left\{a_j^\top(\widehat{\mu} - \mathsf{m}) : d_{\mathsf{n}}(\mathsf{m},\widehat{\mu},\widehat{\Sigma}) \le \mathsf{D}\right\} \text{ for all } j = 1,2,\ldots,p,$$

where a_j is the jth row of A, expressed as a column vector. This offset to b_j is the maximum amount by which the left-hand side of the constraint, $a_j^\top\mathsf{m} + c_j^\top\mathsf{w}$, could increase if we plugged in a model m that is sufficiently close to $\widehat{\mu}$ as measured by the standardized discrepancy. Condition 1 implies that, for any $a \in \mathbb{R}^k$, $\max_{\mathsf{m}\in\mathbb{R}^k}\left\{a^\top(\widehat{\mu} - \mathsf{m}) : d_{\mathsf{n}}(\mathsf{m},\widehat{\mu},\widehat{\Sigma}) \le \mathsf{D}\right\} \ge 0$; thus, $b_j' \ge b_j$ with probability one for all $j = 1,2,\ldots,p$.

For the four standardized discrepancies outlined in Section 4.2,

$$\max_{\mathsf{m}\in\mathbb{R}^k}\left\{a_j^\top(\widehat{\mu} - \mathsf{m}) : d_{\mathsf{n}}^1(\mathsf{m},\widehat{\mu},\widehat{\Sigma}) \le \mathsf{D}^1\right\} = \mathsf{D}^1 \max_{i=1,2,\ldots,k} \frac{\widehat{\sigma}_i}{\sqrt{n_i}} |a_{ji}|,$$

$$\max_{\mathsf{m}\in\mathbb{R}^k}\left\{a_j^\top(\widehat{\mu} - \mathsf{m}) : d_{\mathsf{n}}^2(\mathsf{m},\widehat{\mu},\widehat{\Sigma}) \le \mathsf{D}^2\right\} = \sqrt{\mathsf{D}^2 \sum_{i=1}^k \frac{\widehat{\sigma}_i^2}{n_i} a_{ji}^2},$$

$$\max_{\mathsf{m}\in\mathbb{R}^k}\left\{a_j^\top(\widehat{\mu} - \mathsf{m}) : d_{\mathsf{n}}^\infty(\mathsf{m},\widehat{\mu},\widehat{\Sigma}) \le \mathsf{D}^\infty\right\} = \mathsf{D}^\infty \sum_{i=1}^k \frac{\widehat{\sigma}_i}{\sqrt{n_i}} |a_{ji}|, \text{ and}$$

$$\max_{\mathsf{m}\in\mathbb{R}^k}\left\{a_j^\top(\widehat{\mu} - \mathsf{m}) : d_{\mathsf{n}}^{\mathrm{CRN}}(\mathsf{m},\widehat{\mu},\widehat{\Sigma}) \le \mathsf{D}^{\mathrm{CRN}}\right\} = \sqrt{\frac{\mathsf{D}^{\mathrm{CRN}}}{n} a_j^\top \widehat{\Sigma} a_j},$$

for all $j = 1,2,\ldots,p$; derivations appear in Online Appendix EC.2. In the case of $d_{\mathsf{n}}^{\mathrm{CRN}}$ and $\mathsf{D}^{\mathrm{CRN}}$, adjusting the right-hand-side vector in this way follows the approach of Anderson (1984) for constructing simultaneous confidence intervals for linear combinations of the components of $\mu(\mathsf{X})$; see equation (15) therein. From these expressions, it is apparent that $b' \equiv (b_1', b_2', \ldots, b_p')^\top$ is a random vector whose components are functions of $\widehat{\Sigma}$ and n but not $\widehat{\mu}$.

5.2. Relaxed Plausible Screening

The projection of the relaxation of P onto $\mathbb{R}^k$ is given by

$$\mathsf{R}'(x_0) \equiv \{\mathsf{m} \in \mathbb{R}^k : \text{there exists } \mathsf{w} \in \mathbb{R}^q \text{ such that } A\mathsf{m} + C\mathsf{w} \le b'\}.$$

The polyhedron $\mathsf{R}'(x_0)$ is a random relaxation of $\mathsf{M}(x_0)$, and Lemma 1 further shows that it is also a relaxation of $\mathsf{R}(x_0)$; the proof of Lemma 1 can be found in Online Appendix EC.3.

Lemma 1. *If Assumption 1 holds, then* $\mathsf{R}(x_0) \subseteq \mathsf{R}'(x_0)$ *with probability one for all* $x_0 \in \mathcal{X}$.

Our more conservative screening method, which we refer to as relaxed plausible screening (RPS), returns a subset $\mathcal{S}_{\mathsf{n}}^{\mathrm{RPS}}$ defined as

$$\mathcal{S}_{\mathsf{n}}^{\mathrm{RPS}} \equiv \{x_0 \in \mathcal{X} : \widehat{\mu} \in \mathsf{R}'(x_0)\},$$

the conservatism of which is made clear in Corollary 1.

Corollary 1. *If Assumption 1 holds, then* $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}} \subseteq \mathcal{S}_{\mathsf{n}}^{\mathrm{RPS}}$ *with probability one.*

Theorems 3 and 4 establish that, under Conditions 2–4, $\mathcal{S}_n^{\mathrm{RPS}}$ possesses the desired properties of confidence and $\mathsf{S}(\mathsf{X})$ consistency.

Theorem 3. *If $d_n(\cdot, \widehat{\mu}, \widehat{\Sigma})$ and D satisfy* Conditions 2 *and* 3*, then $\mathcal{S}_n^{\mathrm{RPS}}$ achieves finite-sample confidence and asymptotic confidence.*

Theorem 4. *If $d_n(\cdot, \widehat{\mu}, \widehat{\Sigma})$ and D satisfy* Condition 4*, then $\mathcal{S}_n^{\mathrm{RPS}}$ achieves $\mathsf{S}(\mathsf{X})$ consistency.*

The relaxation $\mathsf{R}'(x_0)$ that is used to construct $\mathcal{S}_n^{\mathrm{RPS}}$ depends on the representation of $\mathsf{M}(x_0)$ in Assumption 1. Hence, a different representation of $\mathsf{M}(x_0)$—meaning a different choice of A, C, and b—can result in a different relaxation $\mathsf{R}'(x_0)$ and, thus, different solutions being included in $\mathcal{S}_n^{\mathrm{RPS}}$. The extreme case of this would be the elimination of C altogether as shown in Theorem 5. This result demonstrates that the representation $\mathsf{M}(x_0) = \{\mathsf{m} \in \mathbb{R}^k : \overline{A}\mathsf{m} \le \overline{b}\}$ for some $\overline{A} \in \mathbb{R}^{\overline{p} \times k}$ and $\overline{b} \in \mathbb{R}^{\overline{p}}$ yields a tighter polyhedral relaxation of $\mathsf{R}(x_0)$ and, thus, a smaller subset.

Theorem 5. *Suppose Assumption 1 holds and that, for a fixed $x_0 \in \mathcal{X}$,*

$$\begin{aligned}\mathsf{M}(x_0) &= \{\mathsf{m} \in \mathbb{R}^k : \textit{there exists } \mathsf{w} \in \mathbb{R}^q \textit{ such that} \\ &\qquad A\mathsf{m} + C\mathsf{w} \le b\} \\ &= \{\mathsf{m} \in \mathbb{R}^k : \overline{A}\mathsf{m} \le \overline{b}\},\end{aligned}$$

for some $A \in \mathbb{R}^{p \times k}$, $C \in \mathbb{R}^{p \times q}$, $b \in \mathbb{R}^p$, $\overline{A} \in \mathbb{R}^{\overline{p} \times k}$, and $\overline{b} \in \mathbb{R}^{\overline{p}}$. Then, for any $\mu \in \mathcal{M}$,

$$\overline{\mathsf{R}}'(x_0) \equiv \left\{\mathsf{m} \in \mathbb{R}^k : \overline{A}\mathsf{m} \le \overline{b}'\right\} \subseteq \mathsf{R}'(x_0) \textit{ with probability one,}$$

where

$$\overline{b}'_j = \overline{b}_j + \max_{\mathsf{m} \in \mathbb{R}^k} \{\overline{a}_j^\top(\widehat{\mu} - \mathsf{m}) : d_n(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma}) \le \mathsf{D}\} \textit{ for all } j = 1, 2, \ldots, \overline{p}.$$

In some cases, for example, optimizing a Lipschitz-continuous function, deriving an explicit polyhedral representation of $\mathsf{M}(x_0)$ is relatively straightforward, and in other cases, for example, optimizing a convex function, it is challenging. Projecting out some or all components of w has the potential to yield a less conservative subset $\mathcal{S}_n^{\mathrm{RPS}}$ but can come at the cost of an increase in the number of constraints implicitly describing $\mathsf{M}(x_0)$. Although classical techniques for eliminating variables, for example, Fourier–Motzkin elimination, can cause an explosion in the number of constraints, many of them redundant, recent advances are more promising (Jing et al. 2018).

Remark 1. Both $\mathcal{S}_n^{\mathrm{PS}}$ and $\mathcal{S}_n^{\mathrm{RPS}}$ exhibit an appealing, intuitive trait: given the same observed simulation outputs, knowing additional functional properties of μ leads to a smaller subset. That is, adding constraints that further shrink $\mathsf{M}(x_0)$ results in more solutions being screened out. This assertion is made mathematically precise in Online Theorems EC.1 and EC.2.

5.3. Optimization Problems

Checking whether $\widehat{\mu} \in \mathsf{R}'(x_0)$ amounts to checking the feasibility of a system of linear equations—namely, does there exist a $\mathsf{w} \in \mathbb{R}^q$ such that $C\mathsf{w} \le b' - A\widehat{\mu}$? This is equivalent to determining the sign of the optimal value of a related linear program:

$$z_n \equiv \max_{\mathsf{w}, \eta} \eta \text{ s.t. } C\mathsf{w} + \eta \mathbf{1}_p \le b' - A\widehat{\mu}, \tag{3}$$

where $\mathbf{1}_p$ is a p-vector of ones. The notation z_n reflects the dependence of the parameters of the optimization problem on the sample sizes; it is also convenient in the proofs of the asymptotic guarantees delivered by $\mathcal{S}_n^{\mathrm{RPS}}$. If $z_n \ge 0$, the solution x_0 is included in $\mathcal{S}_n^{\mathrm{RPS}}$; otherwise, it is excluded.

On the other hand, constructing $\mathcal{S}_n^{\mathrm{PS}}$ requires evaluating $D_n(x_0, \widehat{\mu}, \widehat{\Sigma}) \equiv \min_{(\mathsf{m}, \mathsf{w}) \in \mathsf{P}} d_n(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma})$. Definition 3 therefore reduces the number of decision variables by roughly k, relative to optimizing over P. If $\mathsf{M}(x_0)$ can be expressed as a projection with few extra variables (small q), then solving the problem in Definition 3 may be appreciably faster than solving $\min_{(\mathsf{m}, \mathsf{w}) \in \mathsf{P}} d_n(\mathsf{m}, \widehat{\mu}, \widehat{\Sigma})$ with greater savings as the size of the experimental set increases. Furthermore, if a large number of solutions are to be screened, the computational savings from working with $\mathcal{S}_n^{\mathrm{RPS}}$ can be substantial. Table 1 summarizes properties of the optimization problems associated with screening solutions via the PS and RPS methods for the four standardized discrepancies.

Remark 2. Example 1 features an explicit polyhedral representation of $\mathsf{M}(x_0)$, that is, $q = 0$. Thus, for feasibility determination for Lipschitz performance functions, $\mathcal{S}_n^{\mathrm{RPS}}$ can be constructed without optimization by simply checking whether $A\widehat{\mu} \le b'$ for each solution.

6. Numerical Experiments

To illuminate the theoretical developments thus far in a more practical light, we implement the PS and RPS approaches on two simulation-optimization problems. Both examples illustrate how prior knowledge of functional properties can assist in screening out swathes of unacceptable solutions using only a limited experimental set. In each case, a discrete simulation-optimization problem is posed, but functional properties are established by studying extensions of the performance function or sample-path functions to a continuous solution space. Our first example in Section 6.1 illustrates how the behavior of PS varies depending on

Table 1. Properties of the Plausible Screening and Relaxed Plausible Screening Optimization Problems

Subset	Discrepancy	Linear/quadratic	Number of decision variables	Number of constraints
$\mathcal{S}_n^{PS}$	d_n^1	Linear	$2k+q$	$p+2k$
	d_n^2	Quadratic	$k+q$	p
	d_n^∞	Linear	$k+q+1$	$p+2k$
	d_n^{CRN}	Quadratic	$k+q$	p
$\mathcal{S}_n^{RPS}$	All	Linear	$q+1$	p

Note. k is the number of solutions in X, and p and q are the number of constraints and extra variables in the description of $\mathsf{M}(x_0)$ as in Assumption 1.

the standardized discrepancy and demonstrates the advantages over subset-selection procedures. In a much larger example described in Section 6.2, PS and RPS screen out hundreds of thousands of solutions using an experimental set consisting of only 100 solutions.

We implement our methods in MATLAB using the software's built-in optimization algorithms with their default settings: linprog (dual-simplex method) for linear programs and quadprog (interior-point method) for quadratic programs. Source code is available at https://github.com/daveckman/plausible-screening. We ran our experiments on a high-performance computing cluster using eight cores on a compute node with 256 GB of RAM. For the first example, we ran independent macroreplications of our methods in parallel to study the differences between methods, and for the larger second example, we classified solutions in parallel to mirror a reasonable implementation in practice.

6.1. Newsvendor Problem

The first problem is a modified version of the classical newsvendor problem (Porteus 1990). Here, a vendor orders inventory of a given product in discrete quantities at a per-unit order cost c_{order}, observes a realization of stochastic demand V for a continuous quantity of the product, and sells it at a per-unit sales price p_{sales}. For example, consider a gas station operator who orders gasoline in truckloads but sells it in continuous quantities at the pump. At the end of the sales period, leftover inventory is salvaged at a per-unit price $p_{salvage}$, and unmet demand incurs a fixed per-unit cost of $c_{shortage}$.

The vendor's objective is to determine the order quantity that maximizes the expected profit or, equivalently, minimizes the expected loss over the following sales period. For a fixed realization of demand, V, the loss associated with an order quantity x is given by

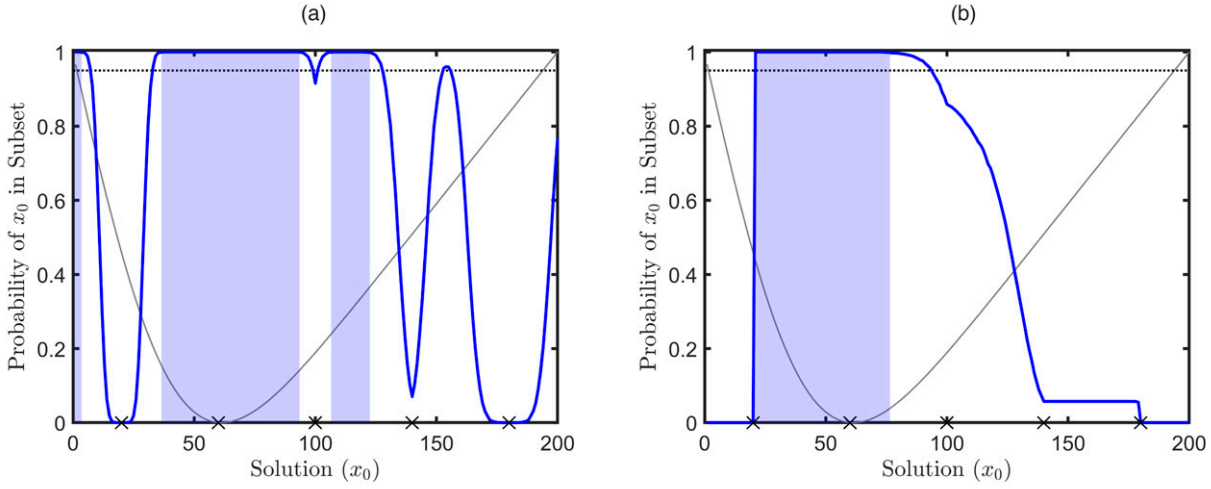
$$Y(x,V) = c_{order}x - p_{sales}\min\{V,x\} - p_{salvage}\max\{x-V,0\} + c_{shortage}\max\{V-x,0\}. \quad (4)$$

The sample-path function $Y(\cdot,V)$ is convex in x provided $p_{sales} \geq p_{salvage}$. Furthermore, $Y(\cdot,V)$ is γ-Lipschitz continuous with constant $\gamma = \max\{p_{sales} + c_{shortage} - c_{order}, c_{order} - p_{salvage}\}$. The expected loss function $\mu(x) := \mathbb{E}_V[Y(x,V)]$ inherits these properties from the sample-path functions as discussed in Section 3.1. In our experiments, we set $c_{order} = 3$, $p_{sales} = 9$, $p_{salvage} = 1$, and $c_{shortage} = 1$ with V being Weibull distributed with scale parameter 50 and shape parameter two.

We considered a feasible region $\mathcal{X} = \{1,2,\ldots,200\}$ and tested our methods by simulating at five evenly spaced solutions (20, 60, 100, 140, 180) with a total sample size of 400 replications. Though not presented in this article, we varied the experimental set and arrived at similar conclusions as the ones presented in this article. We tested the PS method with the d_n^1, d_n^2, and d_n^∞ standardized discrepancies and $1-\alpha = 0.95$ when either exploiting the properties that μ is convex or Lipschitz continuous with $\gamma = 7$. As a benchmark, we applied the STB subset-selection procedure of Nelson et al. (2001), which takes an equal number of i.i.d. replications from all solutions in $\mathcal{X}$ and achieves both finite-sample confidence (under the normality assumption) and asymptotic confidence. Given the same total sample size of 400, the STB procedure took two replications at each feasible solution. We ran 3,000 macroreplications of each procedure.

With a small total sample size spread thinly over the solution space, the STB procedure struggled to eliminate solutions, failing to screen out any solutions on 93.8% of the macroreplications and never screening out more than four solutions. In addition, each feasible solution was retained on at least 99.5% of the macroreplications. Figure 1 shows the empirical probability that individual solutions were included in $\mathcal{S}_n^{PS}$ for d_n^2; curves for d_n^1 and d_n^∞ were similar. In both the Lipschitz and convex cases, our method retained the optimal solution, $\mathcal{A} = \{x^*\} = \{61\}$, on all macroreplications, indicating conservatism. The two instances of functional properties led to interesting features in the geometry of the retained solutions. In the Lipschitz case, PS screened out solutions near clearly suboptimal solutions in the experimental set, namely, $x = 20$, $x = 140$, and $x = 180$, and in the convex case, it screened out those on the periphery of the feasible region. Because

Figure 1. (Color online) Empirical Probability of Including Individual Solutions in $\mathcal{S}_n^{PS}$ for the d_n^2 Standardized Discrepancy with 80 Replications Taken at $k = 5$ Equally Spaced Solutions When Separately Using Knowledge That the Objective Function Is Lipschitz Continuous or Convex

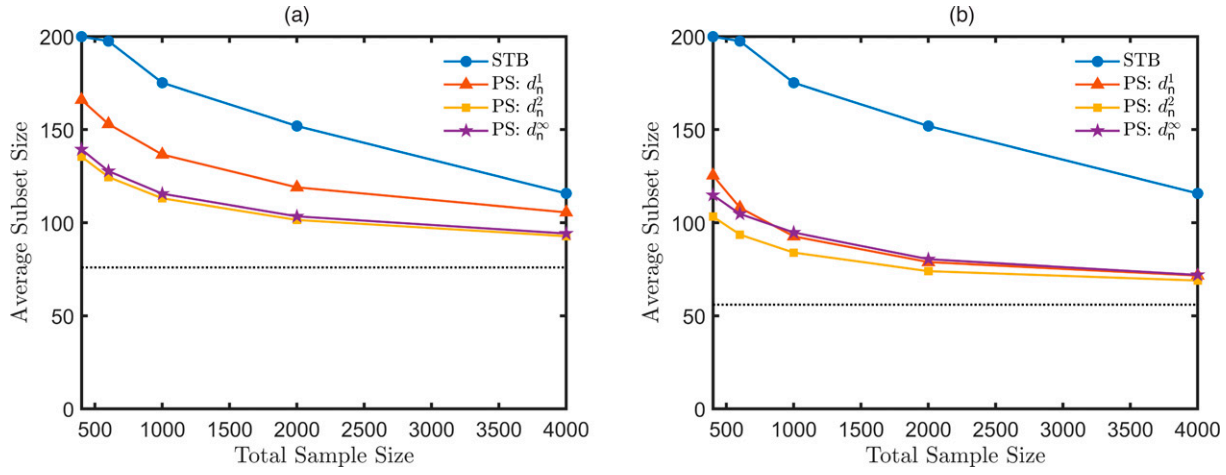


Notes. The thin gray line depicts the (shifted and scaled) objective function, the black dotted line indicates the desired coverage of $1 - \alpha = 0.95$, the black Xs indicate the solutions in the experimental set, and the shaded regions indicate the solutions in $\mathsf{S}(\mathsf{X})$. (a) Lipschitz. (b) Convex.

the probability of being in $\mathcal{S}_n^{PS}$ is neither zero nor one for many solutions, the composition of $\mathcal{S}_n^{PS}$ varied from macroreplication to macroreplication even with 80 replications taken at each solution in X. The subset $\mathcal{S}_n^{PS}$ also differed from $\mathsf{S}(\mathsf{X})$, the subset of solutions that would be returned by an oracle who can observe $\mu(\mathsf{X})$ without simulation error, implying that more solutions could be screened out if the number of replications were increased.

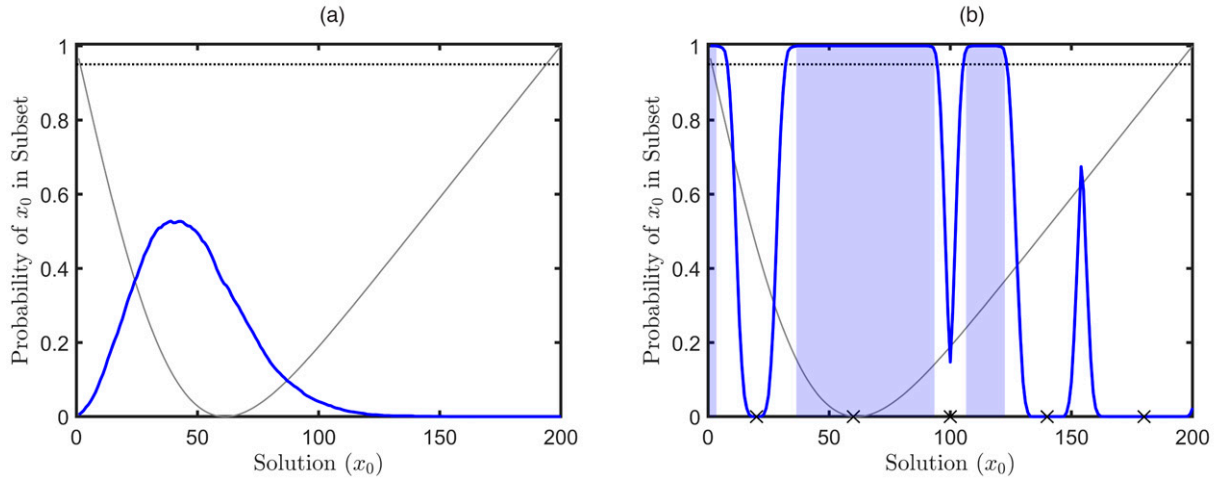
We also varied the total sample size, testing budgets of 400, 600, 1,000, 2,000, and 4,000 replications. Figure 2 shows the average subset sizes for the four procedures when fixing $k = 5$ and increasing the total sample size. All methods returned smaller subsets on average when taking more samples with STB reducing the gap relative to PS. This is a consequence of the limited inference PS can make, having simulated only a fixed experimental set. Specifically, as the total sample size increases, $\mathcal{S}_n^{PS}$ achieves $\mathsf{S}(\mathsf{X})$ consistency—the cardinality of which is shown in Figure 2—while STB eventually screens out all strictly suboptimal solutions. Figure 2 demonstrates that knowing μ is convex leads to more

Figure 2. (Color online) Average Subset Sizes for the STB Procedure and PS with the d_n^1, d_n^2, and d_n^∞ Standardized Discrepancies for $k = 5$ and Different Total Sample Sizes



Notes. The black dotted line indicates the cardinality of $\mathsf{S}(\mathsf{X})$. All average sample sizes are individually precise to within ±1 with 95% confidence. (a) Lipschitz. (b) Convex.

Figure 3. (Color online) Empirical Probability of Including Individual Solutions in the STB Subset and $\mathcal{S}_n^{PS}$ with the d_n^{CRN} Standardized Discrepancy with 80 Replications Taken at $k = 5$ Equally Spaced Solutions When Using Knowledge That the Objective Function Is Lipschitz Continuous



Notes. (a) STB with CRN. (b) PS with d_n^{CRN}.

powerful screening than knowing a universal Lipschitz constant. In both cases, PS screened out anywhere from 15% to 65% of the feasible solutions on average while simulating only 2.5% of them, and more solutions could be screened out if the decision maker were willing to accept more risk as represented by the nominal confidence level. A separate analysis measuring the average average-optimality gap of the solutions in the returned subsets yielded the same conclusions. Figure 2 also gives a sense of how the average subset sizes of PS are affected by the variability of the outputs because the expressions for the standardized discrepancies indicate that increasing n_i is tantamount to decreasing $\widehat{\sigma}_i^2$.

We also compared PS with the d_n^{CRN} standardized discrepancy to a version of the STB procedure that accommodates the use of CRN; see section 3 of Nelson et al. (2001) for STB details. We again took $k = 5$ with a total sample size of 400 replications, and each procedure generated its replications using CRN across solutions. Figure 3 shows the empirical probability that individual solutions were included in the returned subset for STB with CRN and PS when exploiting knowledge that μ is Lipschitz continuous. The STB procedure with CRN was more liberal in screening out solutions—returning an average subset of size 28—but severely undercovered the optimal solution, retaining it on only 36% of the macroreplications. This behavior is a consequence of the severe nonnormality of the outputs and the use of CRN with a small sample size per solution. To be precise, the STB procedure with CRN obtains two sample-path functions $Y(\cdot, V_1)$ and $Y(\cdot, V_2)$ and performs pairwise comparisons based on the variance of $Y(x, V_1) - Y(x', V_1)$ and $Y(x, V_2) - Y(x', V_2)$ for solutions $x, x' \in \mathcal{X}$. From Equation (4), it can be seen that, for $x, x' \notin [\min\{V_1, V_2\}, \max\{V_1, V_2\}]$, the variance of the two differences is zero, implying that any solution $x_0 \notin [\min\{V_1, V_2\}, \max\{V_1, V_2\}]$ will be screened out. Because the mode of the Weibull distribution from which V_1 and V_2 is generated is about 35.4, the STB subsets are biased to the left of $x^* = 61$.

PS with d_n^{CRN} screened out similar solutions to its counterparts that use independent sampling but returned somewhat smaller subsets with an average size of 104 solutions. (In the convex case, PS similarly returned smaller subsets when using CRN with an average size of 70.) This additional screening power should be weighed against the increased difficulty of the underlying optimization problems, that is, the need to solve quadratic programs with dense Hessian matrices.

6.2. Tandem Production Line Problem

The second problem is a resource-allocation problem for a production line with manufacturing blocking (e.g., buffers) adapted from Plambeck et al. (1996). The decision maker is tasked with allocating discrete resources across five single-server stations arranged in a tandem (serial) configuration. Each station processes products using a first-in, first-out service discipline. If station i is allocated a_i resources, its cycle (processing) time for a given product is assumed to be exponentially distributed with rate parameter $\rho_i = \overline{\rho}_i(1 + a_i)$, where $\overline{\rho}_i$ is a base processing rate. We set $\overline{\rho}_1 = 3$, $\overline{\rho}_2 = 5$, $\overline{\rho}_3 = 2$, $\overline{\rho}_4 = 5$, and $\overline{\rho}_5 = 1$.

There is a buffer in front of each machine for products awaiting processing. If the buffer is full, upstream stations can become blocked, whereas if it is empty, downstream stations can become starved. We assume that there is an infinite supply of products immediately available to process at station 1 and an infinite-capacity buffer in front of that station; that is, there is no external arrival process. The buffer capacities in front of stations 2–5 are fixed at 4, 6, 8, and 4, respectively.

The decision maker's objective is to allocate 50 resources to minimize the expected completion time of the 100th product. Under the preceding assumptions, a continuous extension of the objective function is convex in the allocation $x \equiv (a_1, a_2, a_3, a_4, a_5)$; see section IV.B of Shanthikumar and Yao (1989) for a complete derivation. We restrict attention to solutions that allocate all available resources, that is, $a_1 + a_2 + a_3 + a_4 + a_5 = 50$, where $a_i \in \mathbb{Z}$ for $i = 1, 2, \ldots, k$, resulting in a total of 316,251 feasible solutions. Because of this tight constraint, the feasible region can be reduced to a four-dimensional space.

Shanthikumar and Yao (1989) provide dynamic recursion equations for simulating the completion times of all products, thereby avoiding the need to run a full discrete-event simulation of the system. Even so, we consider this problem to be representative of large-scale discrete simulation-optimization problems for which simulating all feasible solutions is impractical, but properties of the objective function may be known. In such cases, the available computational budget may permit only a small fraction of feasible solutions to be simulated. We fixed a total sample size of 10,000 replications, which is enough to simulate one replication from about 3% of the feasible solutions. We ran a single macroreplication of the PS and RPS methods with d_{n}^1, d_{n}^2, and d_{n}^∞. An experimental set consisting of $k = 100$ reasonably space-filling solutions was determined using the type of designs employed in Mak and Joseph (2018); hence, 100 replications were generated at each solution in X.

Screening and timing results for each method are given in Table 2. All three versions of PS screened out more than 60% of the feasible solutions while simulating only 0.03% of them. The efficacy of RPS varied depending on the standardized discrepancy. For d_{n}^∞, the same subset of solutions was returned by PS and RPS (i.e., $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}} = \mathcal{S}_{\mathsf{n}}^{\mathrm{RPS}}$), yet for d_{n}^1, no solutions were screened out by RPS. We suspect that this disparity between PS and RPS for d_{n}^1 is due to the sparsity of A and the rapid growth rate of D^1 with increasing k. In particular, the offset $\mathsf{D}^1 \max_{i=1,2,\ldots,k}(\widehat{\sigma}_i/\sqrt{n_i})\,|a_{ji}|$ is likely much larger than $\mathsf{D}^\infty \sum_{i=1}^k (\widehat{\sigma}_i/\sqrt{n_i})\,|a_{ji}|$, leading to a more conservative subset.

Remark 3. In all of our experiments for PS and RPS with d_{n}^∞, we observed that, on all macroreplications, $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}} = \mathcal{S}_{\mathsf{n}}^{\mathrm{RPS}}$ for both the Lipschitz and convex cases. Theorem EC.3 formalizes this observation and proves that it holds with probability one for the Lipschitz case. We were unable to prove an analogous result for the convex case.

All together, the results in Table 2 illustrate the diverse performance of the various methods. PS with d_{n}^2, which required the solution of quadratic programs, was the most computationally intensive procedure. At the other extreme, PS with d_{n}^1 was roughly 20 times faster, but retained about twice as many solutions. The most effective and efficient procedure was RPS with d_{n}^∞; it removed more than 80% of the feasible solutions with an overall run time of about eight core hours. As a practical recommendation, for either d_{n}^2 or d_{n}^∞, the faster RPS method can be run first, followed by the PS method on the solutions in the returned subset $\mathcal{S}_{\mathsf{n}}^{\mathrm{RPS}}$. This approach notably does not require splitting α to preserve the statistical guarantee as is sometimes the case with multistage selection procedures (Nelson et al. 2001).

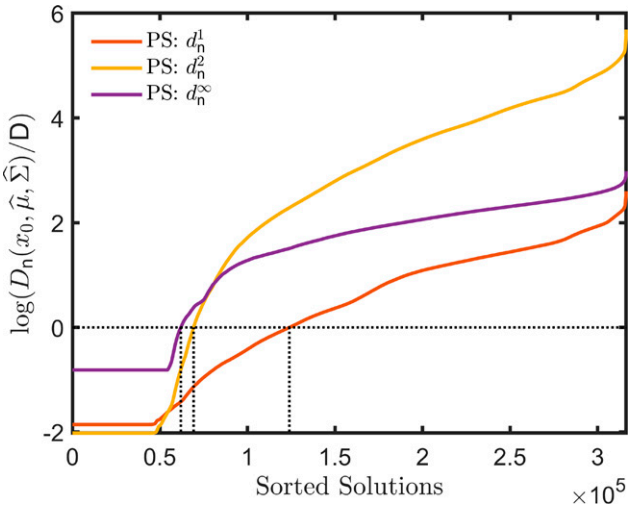
Figure 4 shows the sorted minimum standardized discrepancies of the feasible solutions, $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$, relative to the cutoff, D, for the three versions of PS. The minimum standardized discrepancies were divided by the cutoffs and log-transformed to produce a clear, standardized comparison. The flat stretches on the left-hand side of Figure 4 correspond to solutions x_0 for which there exists an x_0-optimal convex function that coincides with the best-fitting convex function (with respect to the standardized discrepancy) at solutions in X. More solutions can be screened out if a tighter statistically valid cutoff value is used, especially for the d_{n}^1 standardized discrepancy—the potential gains for the d_{n}^2 and d_{n}^∞ standardized discrepancies are more limited.

Without an oracle for evaluating the true objective function, we took 500 replications at each feasible

Table 2. Times and Subset Sizes for a Single Macroreplication on the Tandem Production Line Problem

Method and discrepancy	Time per solution, s	Subset size ($\lvert\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}\rvert$ / $\lvert\mathcal{S}_{\mathsf{n}}^{\mathrm{RPS}}\rvert$)	Fraction screened
PS with d_{n}^1/RPS with d_{n}^1	0.08/0.07	123,904/316,251	60.8%/0%
PS with d_{n}^2/RPS with d_{n}^2	1.63/0.09	69,198/83,748	78.1%/73.5%
PS with d_{n}^∞/RPS with d_{n}^∞	0.40/0.09	61,897/61,897	80.4%/80.4%

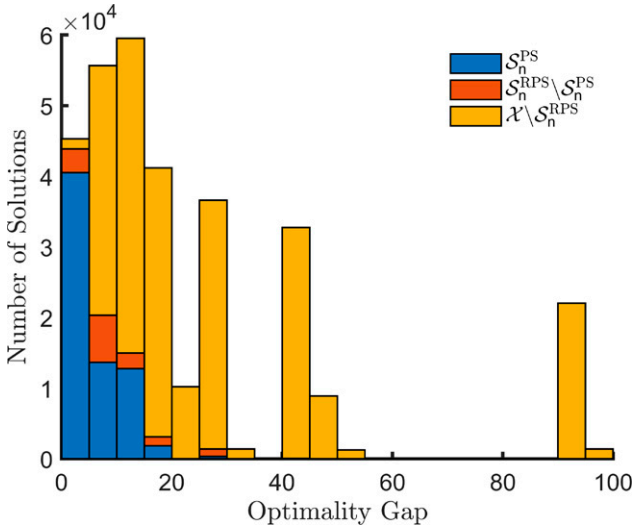
Figure 4. (Color online) Sorted Logarithm of Scaled Minimum Standardized Discrepancies of Feasible Solutions for a Single Macroreplication of PS with the d_n^1, d_n^2, and d_n^∞ Standardized Discrepancies



Notes. The horizontal black dotted line differentiates solutions that are retained (below) and screened out (above). The vertical black dotted lines indicate subset sizes.

solution (using CRN) and estimated the optimality gaps—the differences in performance between each solution and the optimal solution—based on the sample means. Figure 5 shows the optimality gaps for the feasible solutions as well as those in $\mathcal{S}_n^{PS}$ and $\mathcal{S}_n^{RPS}$ with d_n^2. The results demonstrate that PS and RPS can screen out a large portion of the inferior solutions while retaining high-quality solutions.

Figure 5. (Color online) Histogram of the Optimality Gaps of Solutions Retained in $\mathcal{S}_n^{PS}$, the Optimality Gaps of Additional Solutions in $\mathcal{S}_n^{RPS}$, and the Optimality Gaps of All Remaining Solutions for the d_n^2 Standardized Discrepancy



7. Conclusions and Discussions

This article describes a novel but nascent framework for screening solutions whose performances can be evaluated via stochastic simulation. In contrast to traditional subset-selection procedures, our methods can screen out unsimulated solutions, making them appealing statistical-inference techniques for large-scale simulation-optimization problems on which such procedures are otherwise unworkable. For the plausible screening method, solutions are screened by minimizing a standardized discrepancy—a function measuring the distance between the sample means and a given vector—over a feasible region characterized by known properties of the performance function. For the relaxed plausible screening method, solutions are screened by checking the feasibility of a system of linear equations. Both methods return subsets of solutions that attain typical statistical properties of confidence and consistency. Experimental results demonstrate the power of exploiting known functional properties with varying degrees of effectiveness for different standardized discrepancies. When the functional properties and definition of acceptability entail comparing the performances of only a few solutions, that is, the constraint matrix A is sparse, as is the case when optimizing Lipschitz continuous or convex performance functions, we recommend the d_n^∞ standardized discrepancy as an efficient and powerful choice.

The proposed methodology can be extended well beyond the initial treatment in this paper. Other, more sophisticated, forms of functional properties can be incorporated, such as local Lipschitz continuity, quasiconvexity, or unimodality. One could also imagine employing stochastic gradient estimators to further enhance screening. Answering the question of how one acquires functional information is critical to convert this idea into a practical tool. One direction could pair this methodology with existing tests for functional properties (Juditsky and Nemirovski 2002, Lim 2020) or schemes for estimating Lipschitz constants (Calliess 2017). Another tact is to explicitly leverage our minimal discrepancy to test for functional properties of performance functions, though we have not fully developed these ideas. We conjecture that there are many classes of simulation problems with functional information available upon careful examination. Developing new methods for analytically verifying and empirically detecting functional properties is an important direction for future research.

The asymptotic guarantees delivered by our methods—confidence and consistency—are predicated on the estimator of $\mu(x)$ being asymptotically normal and consistent. It should be possible to develop PS methods for cases in which $\mu(x)$ is a functional other than the mean, such as a quantile, for which the maximum likelihood estimator satisfies these conditions.

Another area of future research is how the choice of the experimental set, X, and the number of simulation replications allocated to solutions in it, n, dictate the effectiveness of our methods. There are many relevant practical questions that can be addressed: Given a fixed budget, is it better to obtain few replications at many solutions or more replications at fewer solutions? Given the known properties of μ, how should the solutions in X be spread over $\mathcal{X}$? The answers to these questions might be informed by an asymptotic analysis of our methods as k and n increase together. Extending our methods to allow for sequential experimentation has great potential. Adaptively identifying solutions in X at which to obtain more replications or new solutions to add to X can lead to more efficient and powerful screening. However, preserving the statistical guarantees of such procedures requires careful attention.

Combining PS methods with simulation-optimization algorithms is another worthy topic for further investigation. Important issues in the design of new combined procedures include how best to use the subsets and inferences provided by plausible screening methods. The statistical guarantees offered by our screening methods pair especially well with those delivered by R&S algorithms, such as ensuring that a near-optimal solution is ultimately selected with high probability.

Acknowledgments

The authors thank the reviewers and associate editor for their careful reading of the paper and helpful suggestions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Anderson TW (1984) *An Introduction to Multivariate Statistical Analysis*, 2nd ed. (John Wiley & Sons, New York).

Andradóttir S, Prudius AA (2010) Adaptive random search for continuous simulation optimization. *Naval Res. Logist.* 57(6): 583–604.

Ankenman B, Nelson BL, Staum J (2010) Stochastic kriging for simulation metamodeling. *Oper. Res.* 58(2):371–382.

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

Bettonvil B, Kleijnen JP (1997) Searching for important factors in simulation models with many factors: Sequential bifurcation. *Eur. J. Oper. Res.* 96(1):180–194.

Boesel J, Nelson BL, Kim SH (2003) Using ranking and selection to "clean up" after simulation optimization. *Oper. Res.* 51(5): 814–825.

Calliess JP (2017) Lipschitz optimisation for Lipschitz interpolation. *2017 Amer. Control Conf.* (IEEE, Red Hook, NY), 3141–3146.

Chang KH, Hong LJ, Wan H (2013) Stochastic trust-region response-surface method (STRONG)—A new response-surface framework for simulation optimization. *INFORMS J. Comput.* 25(2):230–243.

Dentcheva D, Ruszczyński A (2006) Portfolio optimization with stochastic dominance constraints. *J. Banking Finance* 30(2):433–451.

Eckman DJ, Plumlee M, Nelson BL (2020) Revisiting subset selection. Bae KH, Feng B, Kim S, Lazarova-Molnar S, Zheng Z, Roeder T, Thiesing R, eds. *Proc. 2020 Winter Simulation Conf.* (IEEE, Piscataway, NJ), 2972–2983.

Frazier P, Powell W, Dayanik S (2009) The knowledge-gradient policy for correlated normal beliefs. *INFORMS J. Comput.* 21(4):599–613.

Gupta SS (1965) On some multiple decision (selection and ranking) rules. *Technometrics* 7(2):225–245.

Hong LJ, Nelson BL (2006) Discrete optimization via simulation using COMPASS. *Oper. Res.* 54(1):115–129.

Jian N, Henderson SG (2020) Estimating the probability that a function observed with noise is convex. *INFORMS J. Comput.* 32(2): 376–389.

Jing R, Maza MM, Talaashrafi D (2018) Complexity estimates for Fourier-Motzkin elimination. Preprint, submitted November 5, https://arxiv.org/abs/1811.01510.

Juditsky A, Nemirovski A (2002) On nonparametric tests of positivity/monotonicity/convexity. *Ann. Statist.* 30(2):498–527.

Kim S, Pasupathy R, Henderson SG (2015) A guide to sample average approximation. Fu M, ed. *Handbook of Simulation Optimization* (Springer-Verlag, New York), 207–243.

Kleijnen JPC (2015) *Design and Analysis of Simulation Experiments* (Springer, Cham, Switzerland).

Laguna M (2011) Optquest: Optimization of complex systems. Technical report. Accessed May 8, 2020, https://www.opttek.com/sites/default/files/pdfs/OptQuest-Optimization%20of%20Complex%20Systems.pdf.

Lim E (2020) The limiting behavior of isotonic and convex regression estimators when the model is misspecified. *Electronic J. Statist.* 14(1):2053–2097.

Lim E, Glynn PW (2012) Consistency of multidimensional convex regression. *Oper. Res.* 60(1):196–208.

Liu J, Li C, Yang F, Wan H, Uzsoy R (2011) Production planning for semiconductor manufacturing via simulation optimization. Jain S, Creasey RR, Himmelspach J, White KP, Fu M, eds. *Proc. 2011 Winter Simulation Conf.* (IEEE, Piscataway, NJ), 3612–3622.

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Oper. Res.* 49(3):334–351.

Mak S, Joseph VR (2018) Minimax and minimax projection designs using clustering. *Comput. Graphical Statist.* 27(1):166–178.

Nelson BL, Swann J, Goldsman D, Song W (2001) Simple procedures for selecting the best simulated system when the number of alternatives is large. *Oper. Res.* 49(6):950–963.

Ni EC, Henderson SG, Hunter SR (2014) A comparison of two parallel ranking and selection procedures. Tolk A, Diallo SY, Ryzhov IO, Yilmaz L, Buckley S, Miller JA, eds. *Proc. 2014 Winter Simulation Conf.*, (IEEE, Piscataway, NJ), 3761–3772.

Plambeck EL, Fu BR, Robinson SM, Suri R (1996) Sample-path optimization of convex stochastic performance functions. *Math. Programming* 75(2):137–176.

Plumlee M, Nelson BL (2018) Plausible optima. Rabe M, Juan AA, Mustafee N, Skoogh A, Jain S, Johansson B, eds. *Proc. 2018 Winter Simulation Conf.* (IEEE, Piscataway, NJ), 1981–1992.

Porteus EL (1990) Stochastic inventory theory. Heyman DP, Sobel MJ, eds. *Stochastic Models*, Handbooks in Operations Research and Management Science, vol. 2 (Elsevier, New York), 605–652.

Salemi P, Staum J, Nelson BL (2019a) Generalized integrated Brownian fields for simulation metamodeling. *Oper. Res.* 67(3):874–891.

Salemi P, Song E, Nelson BL, Staum J (2019b) Gaussian Markov random fields for discrete optimization via simulation: Framework and algorithms. *Oper. Res.* 67(1):250–266.

Santner TJ, Williams BJ, Notz WI (2003) *The Design and Analysis of Computer Experiments*, 2nd ed. (Springer, New York).

Scott W, Frazier P, Powell W (2011) The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression. *SIAM J. Optim.* 21(3):996–1026.

Shaked M, Shanthikumar JG (1988) Stochastic convexity and its applications. *Adv. Appl. Probab.* 20(2):427–446.

Shaked M, Shanthikumar JG (2007) *Stochastic Orders* (Springer Science+Business Media, New York).

Shanthikumar JG, Yao DD (1989) Second-order stochastic properties in queueing systems. *Proc. IEEE.* 77(1):162–170.

Shapiro A, Wardi Y (1996) Convergence analysis of gradient descent stochastic algorithms. *J. Optim. Theory Appl.* 91(2):439–454.

Shashaani S, Hashemi FS, Pasupathy R (2018) ASTRO-DF: A class of adaptive sampling trust-region algorithms for derivative-free stochastic optimization. *SIAM J. Optim.* 28(4):3145–3176.

Shi L, Ólafsson S (2000) Nested partitions method for global optimization. *Oper. Res.* 48(3):390–407.

Silvapulle MJ, Sen PK (2005) *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions* (John Wiley & Sons, Inc., Hoboken, NJ).

Spearman ML, Woodruff DL, Hopp WJ (1990) CONWIP: A pull alternative to kanban. *Internat. J. Production Res.* 28(5):879–894.

Wan H, Ankenman BE, Nelson BL (2006) Controlled sequential bifurcation: A new factor-screening method for discrete-event simulation. *Oper. Res.* 54(4):743–755.

Wan F, Liu W, Bretz F, Han Y (2016) Confidence sets for optimal factor levels of a response surface. *Biometrics* 72(4):1285–1293.

Wolff RW (1977) An upper bound for multi-channel queues. *J. Appl. Probab.* 14(4):884–888.

Xu WL, Nelson BL (2013) Empirical stochastic branch-and-bound for optimization via simulation. *IIE Trans.* 45(7):685–698.

Xu J, Nelson BL, Hong JL (2010) Industrial strength COMPASS: A comprehensive algorithm and software for optimization via simulation. *ACM Trans. Model. Comput. Simulation* 20(1):1–29.

David J. Eckman is an assistant professor in the Wm Michael Barnes '64 Department of Industrial and Systems Engineering at Texas A&M University. His research interests deal with optimization and output analysis for stochastic simulation models.

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.

Barry L. Nelson is the Walter P. Murphy Professor of the Department of Industrial Engineering and Management Sciences at Northwestern University. His research focus is on the design and analysis of computer simulation experiments on models of discrete-event, stochastic systems. He is a Fellow of INFORMS and IIE.