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

SEQUENTIAL DETECTION OF CONVEXITY FROM NOISY FUNCTION EVALUATIONS

Nanjing Jian
Shane G. Henderson

Operations Research & Information Engineering
Cornell University
Ithaca, NY, 14853, U.S.A

Susan R. Hunter

School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, U.S.A.

ABSTRACT

Consider a real-valued function that can only be evaluated with error. Given estimates of the function values from simulation on a finite set of points, we seek a procedure to detect convexity or non-convexity of the true function restricted to those points. We review an existing frequentist hypothesis test, and introduce a sequential Bayesian procedure. Our Bayesian procedure applies for both independent sampling and sampling with common random numbers, with known or unknown sampling variance. In each iteration, we collect a set of samples and update a posterior distribution on the function values, and use that as the prior belief in our next iteration. We then approximate the probability that the function is convex based on the posterior using Monte Carlo simulation.

1 INTRODUCTION

Consider a real-valued function $f : \mathbb{R}^d \to \mathbb{R}$ that can only be evaluated through Monte Carlo simulation; that is, f can only be observed with error. Suppose we wish to determine whether the true function f, restricted to evaluation at a finite set of points, is convex or non-convex at those points. Since the Monte Carlo simulation yields only estimates of the function values on the finite set of points, we desire a probabilistic guarantee on the determination of convexity or non-convexity. We say that a function f, restricted to a finite set of points, is convex if a convex function exists that coincides with f at those points.

Knowing whether a function is convex or not is useful in several ways. First, if a function is convex over its entire domain, then any local minimum is a global minimum. Second, if a function is known to be convex over its entire domain, then one can apply specialized algorithms such as the level method (Nesterov 2004) for minimization. Third, one could employ a convexity test to attempt to identify "basins of attraction" around local minima in which the function is convex. Since the function, restricted to such a basin, is convex, one might then apply specialized techniques to obtain estimates of, or bounds on, (local) optimality gaps through a gradient-based cutting plane procedure similar to that developed in Glynn and Infanger (2013) for stochastic linear programs. Such estimates or bounds can then be used to define stopping rules for optimization algorithms. Fourth, convexity can suggest important qualitative properties, such as risk-averse or risk-seeking agent behavior (Abrevaya and Jiang 2005).

There is a vast body of literature on detecting the convexity of a function, thus we only discuss a few methods here. Suppose we frame the problem of detecting convexity into a hypothesis test. Then the literature can be divided into three categories based on how the null hypothesis is defined. In the first category, the null and alternative hypotheses are:

$$H_0 : f \in \mathcal{G} \text{ v.s. } H_a : f \notin \mathcal{G}, \tag{1}$$

where f is the function that we are interested in and $\mathcal{G}$ is a functional cone of all convex functions. In single dimension, Juditsky and Nemirovski (2002) use a nonparametric approach and focus on the case

where the noise is modeled as a Gaussian process. They employ the technique of estimating functionals and use the L^r distance between f and $\mathcal{G}$ as the test statistic.

The second category defines the null and alternative hypotheses as

$$H_0 : \boldsymbol{f} \in \mathbb{C} \text{ v.s. } H_a : \boldsymbol{f} \notin \mathbb{C}, \tag{2}$$

where $\boldsymbol{f}$ is the vector of function values on a finite set of points, and $\mathbb{C}$ is some appropriately-defined cone of convex functions on the finite sample points. Assuming the noise follows a multivariate Normal distribution, Silvapulle and Sen (2001) describe a test that uses the projected distance of $\boldsymbol{f}$ from $\mathbb{C}$ as the test statistic in a likelihood-ratio test of (2), and show that the test statistic follows a Chi-bar-square distribution whose parameters can be evaluated through simulation.

The third and most common category of literature focuses on testing structural properties by testing properties of regression estimators. In single dimension and under certain regularity conditions, Baraud, Huet, and Laurent (2005) show that a hypothesis test on a regression estimator is equivalent to the hypothesis test (2), where $\mathbb{C}$ is appropriately defined. Diack and Thomas-Agnan (1998), Wang and Meyer (2011) and Meyer (2012), use cubic regression splines to fit the observed samples and test if the model is convex based on the second-order derivative at knots. In multiple dimensions, Lau (1978) uses a parametric second order model to approximate the function and tests the convexity based on the model parameters. Lau (1978) also provides a useful survey of the early literature. For a small and localized set of data points, Abrevaya and Jiang (2005) define a simplex test statistic that counts all the possible convex and concave simplexes in the data points. Aside from the literature on convexity tests, a closely related field is convex regression, in which one fits a convex function to a given data set. There is also a vast body of literature on this topic (e.g. Judge and Takayama 1966, Allon et al. 2007, Seijo and Sen 2010, Hannah and Dunson 2011, Lim and Glynn 2012), among which Seijo and Sen (2010) provide a review of past work.

The work on testing (2), developed in the statistics and economics literature, mostly applies to the context where the number of samples is predetermined and fixed. However in our simulation context, we have the ability to draw samples sequentially, and thus choosing a sample size in advance may be undesirable: if the sample size at each sampled point is too small, then the test may fail to reject the null hypothesis and draws no conclusion. On the other hand, choosing a large sample size a priori could be expensive in both sampling cost and computational time. Hence, we wish to derive a sequential procedure that iteratively obtains samples of the function value at a fixed set of points until we are confident enough to conclude whether the function, restricted to those points, is convex or not. A frequentist sequential test based on the results in Silvapulle and Sen (2001) might be designed (Siegmund 1985), although determining the stopping region may prove to be challenging.

In this paper we use a Bayesian approach, providing a sequential procedure to estimate the posterior probability that the function, restricted to a finite set, is convex. Since we only ever evaluate the function at a finite set of points, we can never conclude that the function on its continuous domain is convex; the best we can hope for with our methods is to assert that with high probability there is a convex function consistent with the function values on the finite set of points we have tested, or to assert with high probability that the function is not convex. In a Bayesian setting, we always regard the function values at a finite set of points as a random vector, and we maintain a belief about its distribution. As we collect more information about the unknown function, a posterior distribution on the random vector is updated and used as a prior for later sampling. At any stage we can estimate the posterior probability that the function, restricted to the finite set of points, is convex using Monte Carlo. In this way, the Bayesian framework gives us a natural way to utilize the information from sequential sampling. Using standard Monte Carlo can be computationally burdensome, because at each generated point we need to solve a linear feasibility problem. We show how to re-use past samples using a change-of-measure technique that ensures that the resulting estimator of the posterior probability of convexity is unbiased.

The structure of this paper is as follows. Section 2 describes the problem more precisely. Sections 3 and 4 give preliminary details on updating the Bayesian posterior and testing for convexity. Section 5

contains the main algorithm, and Section 6 presents the change-of-measure technique. We provide some initial numerical results in Sections 7 and discuss them in Section 8.

In this paper, A^T denotes the transpose of the matrix A. For a set S, S° denotes its interior. We use $\Rightarrow$ for convergence in distribution. Let $\mathbb{1}\{\cdot\}$ denote an indicator function that takes the value of 1 if event $\{\cdot\}$ is true and 0 otherwise.

2 PROBLEM STATEMENT, ASSUMPTIONS, AND NOTATION

Consider a (deterministic) function $g : S \to \mathbb{R}$ where $S \subseteq \mathbb{R}^d$. We are given a set of points $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r\}$ at which we can obtain noisy estimates of g, where each $\boldsymbol{x}_i \in S^\circ$. The number of points, r, is finite and fixed. Let the (deterministic) vector of function values associated with the points in $\boldsymbol{x}$ be denoted $\boldsymbol{g} = (g(\boldsymbol{x}_1), g(\boldsymbol{x}_2), \ldots, g(\boldsymbol{x}_r)) \in \mathbb{R}^r$. We want to determine whether or not $\boldsymbol{g} \in \mathbb{C}$, where $\mathbb{C} = \mathbb{C}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r)$ is a set in $\mathbb{R}^r$ representing the set of convex functions evaluated on $\boldsymbol{x}$. (We define $\mathbb{C}$ precisely in Section 4.) A vector $\boldsymbol{y} \in \mathbb{C}$ if and only if there exists a convex function that goes through points $(\boldsymbol{x}_i, \boldsymbol{y}_i), i = 1, 2, \ldots, r$. We say that the vector of function values $\boldsymbol{y}$ is "convex" if $\boldsymbol{y} \in \mathbb{C}$. Also, let $\mathbb{C}^\circ$ be the set of strictly convex functions evaluated on $\boldsymbol{x}$. We say $\boldsymbol{y}$ is "strictly convex" if $\boldsymbol{y} \in \mathbb{C}^\circ$.

In our Bayesian setting, $\boldsymbol{g}$ is viewed as an unknown realization from the (prior) distribution of a random vector $\boldsymbol{f}$. When we say we obtain noisy estimates of $\boldsymbol{g}$ on $\boldsymbol{x}$, we mean that we obtain realizations of a random, r-by-1 vector $\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\xi}$, where $\boldsymbol{\xi} \in \mathbb{R}^r$ is random and represents the noise in function evaluations. Let Y_{ij} be a random variable representing the j^{th} (noisy) function evaluation at $\boldsymbol{x}_i$, for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots$, and denote a realization of Y_{ij} by y_{ij}.

We make the following assumptions:

1. Conditional on $\boldsymbol{f}$, the noise $\boldsymbol{\xi}$ is normally distributed with mean $\boldsymbol{0}$ and covariance matrix Γ, i.e., $\boldsymbol{Y} - \boldsymbol{f} \sim N(\boldsymbol{0}, \Gamma)$.
2. Conditional on $\boldsymbol{f}$, the sequence $(\boldsymbol{y}_j : j = 1, 2, \ldots)$ consists of i.i.d. random vectors, where $\boldsymbol{y}_j$ denotes the r-by-1 vector giving the samples $(Y_{ij} : i = 1, 2, \ldots, r)$ obtained at the jth stage of sampling.

Assumption 1 ensures that the observed noise is Gaussian. Assumption 2 ensures that the set of samples obtained for different iterations j are conditionally independent of one another. Importantly, this assumption does not require conditionally independent sampling across the points $\boldsymbol{x}_i, i = 1, 2, \ldots, r$. Thus Γ is not necessarily diagonal. In fact, Common Random Numbers (CRN) should probably be used to attempt to induce positive correlation amongst the function value estimates at each i, thereby reducing the variability in the estimated overall function "shape" (Chen, Ankenman, and Nelson 2012).

Let $\mathscr{A}_0$ denote the σ-field generated by $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r, \mu_0, \Lambda_0\}$, where μ_0 and Λ_0 represent our prior belief of the mean and covariance of $\boldsymbol{f}$. (We allow the points $\boldsymbol{x}$ to be random perhaps because they are sampled in some preliminary study, but from our perspective they are deterministic and hence included in $\mathscr{A}_0$.) Denote the σ-field generated by $\mathscr{A}_0$, together with the set of all collected data points to iteration n, $\{(\boldsymbol{x}_i, Y_{ij}), i = 1, \ldots, r, j = 1, \ldots, n\}$, $n = 1, 2, \ldots)$ by $\mathscr{A}_n$. Thus, $\{\mathscr{A}_n\}_{n=1,2,\ldots}$ defines a filtration, and $\mathscr{A}_n$ represents the information we have collected up to and including iteration n.

We will successively update the prior distribution, obtaining a posterior distribution that represents our belief about the function $\boldsymbol{g}$. In particular, we will be interested in computing $P(\boldsymbol{f} \in \mathbb{C} | \mathscr{A}_n)$, which is the posterior probability that $\boldsymbol{f}$ is convex, i.e., our belief that $\boldsymbol{g}$ is convex.

3 POSTERIOR UPDATES

We use a conjugate prior, which ensures that the posterior is easily computed, whether or not the sampling covariance matrix Γ is known. The updates given in this section are standard see, e.g., DeGroot (1970), Gelman et al. (2003), Bernardo and Smith (2008).

3.1 Conjugate Prior under Known Sampling Variance

Our (initial) prior belief on the function values is $\boldsymbol{f}|\mathscr{A}_0 \sim N(\mu_0, \Lambda_0)$, where $\mu_0 \in \mathbb{R}^r$ and $\Lambda_0 \in \mathbb{R}^{r\times r}$. When we have little or no prior information on $\boldsymbol{f}$ we adopt a non-informative prior by choosing μ_0 to be a constant and Λ_0 to be a diagonal matrix with huge diagonal values relative to the sampling variance, i.e. the diagonal of Γ. Other methods that carefully elicit a more informative prior are possible.

Recall that we assumed $\boldsymbol{Y} - \boldsymbol{f}$ to be multivariate normal with mean $\boldsymbol{0}$ and covariance matrix Γ, where $\boldsymbol{0}$ is an $r \times 1$ vector of all zeros. If the sampling covariance matrix Γ is known, then our normal prior is conjugate. For the nth iteration we have prior belief $\boldsymbol{f}|\mathscr{A}_{n-1} \sim N(\mu_{n-1}, \Lambda_{n-1})$ and obtain s samples $(\boldsymbol{y}_j, j = 1,2,\ldots,s)$, where each $\boldsymbol{y}_j \in \mathbb{R}^r$ represents a set of noisy function evaluations at each of the r points. These samples have likelihood $\boldsymbol{y}_j|\boldsymbol{f} \sim$ i.i.d $N(\boldsymbol{f}, \Gamma), j = 1,2,\ldots,s$. The posterior distribution is

$$\boldsymbol{f}|\mathscr{A}_n \sim N(\mu_n, \Lambda_n).$$

The parameters are updated by

$$\begin{aligned} \Lambda_n^{-1} &= \Lambda_{n-1}^{-1} + s\Gamma^{-1} \\ \mu_n &= \Lambda_n(\Lambda_{n-1}^{-1}\mu_{n-1} + s\Gamma^{-1}\bar{\boldsymbol{y}}), \end{aligned} \tag{3}$$

where $\bar{\boldsymbol{y}}$ is the sample mean $s^{-1}\sum_{j=1}^{s}\boldsymbol{y}_j$. If only one new set of samples $\boldsymbol{y}$ is generated at iteration n, then $s = 1$ and $\bar{\boldsymbol{y}} = \boldsymbol{y}$. Although one can allow $s > 1$, we use $s = 1$ for this paper.

A careful implementation of the update (3) avoids matrix inversion and employs Cholesky factorization and/or the Sherman-Woodbury-Morrison formula (Golub and Van Loan 1996), but we omit the details due to space constraints.

3.2 Conjugate Prior under Unknown Sampling Variance

When the sampling variance Γ is unknown, the conjugate prior for Normal with unknown mean and unknown covariance is Normal-inverse-Wishart. A common non-informative prior used for this is the Jeffery's prior, in which the prior joint density $P(\boldsymbol{f}, \Gamma|\mathscr{A}_0)$ is proportional to $|\Gamma|^{-(r+1)/2}$ (Gelman et al. 2003). We transition this non-informative prior into the conjugate prior of Normal-inverse-Wishart by an initial sampling stage. Suppose in the initial stage, we collect a set of samples $\boldsymbol{y}_j \in \mathbb{R}^r, j = 1,2,\ldots,s$ of the function values $\boldsymbol{f}$ with sample average $\bar{\boldsymbol{y}}$, then the corresponding posterior density is Normal-inverse-Wishart (Gelman et al. 2003):

$$\begin{aligned} \Gamma|\mathscr{A}_0, \boldsymbol{y} &\sim \text{Inv-Wishart}_{\upsilon_0}(\Xi_0^{-1}) \\ \boldsymbol{f}|\Gamma, \mathscr{A}_0, \boldsymbol{y} &\sim N(\mu_0, \Gamma/\kappa_0), \end{aligned} \tag{4}$$

where

$$\mu_0 = \bar{\boldsymbol{y}}; \quad \kappa_0 = s; \quad \upsilon_0 = s-1; \quad \Xi_0 = (\sum_{j=1}^{s}(\boldsymbol{y}_j - \bar{\boldsymbol{y}})(\boldsymbol{y}_j - \bar{\boldsymbol{y}})^T)^{-1}.$$

Recall that a random $r \times r$ matrix Γ^{-1} has Wishart distribution with parameters υ and Ξ if the density function is proportional to $|\Xi|^{\upsilon/2}|\Gamma|^{-(\upsilon-r-1)/2}\exp\{-\text{tr}(\Xi\Gamma^{-1})/2\}$, where $\text{tr}(\cdot)$ is the trace of a matrix. The matrix Γ then follows a Inv-Wishart distribution with the same degrees of freedom υ and parameter Ξ^{-1}.

After the initial sampling stage, we perform the conjugate posterior updates as follows. At iteration n, with joint prior beliefs $\Gamma|\mathscr{A}_{n-1} \sim \text{Inv-Wishart}_{\upsilon_{n-1}}(\Xi_{n-1}^{-1})$, $\boldsymbol{f}|\mathscr{A}_{n-1}, \Gamma \sim N(\mu_{n-1}, \Gamma/\kappa_{n-1})$, and s samples $\boldsymbol{y}_j \sim$ i.i.d $N(\mu, \Gamma), j = 1,\ldots,s$ with each $\boldsymbol{y}_j \in \mathbb{R}^r$, the posterior parameters are (Gelman et al. 2003):

$$\begin{aligned} \mu_n &= \frac{\kappa_{n-1}}{\kappa_{n-1}+s}\mu_{n-1} + \frac{s}{\kappa_{n-1}+s}\bar{\boldsymbol{y}}; \quad \kappa_n = \kappa_{n-1} + s; \quad \upsilon_n = \upsilon_{n-1} + s; \\ \Xi_n &= \Xi_{n-1} + S + \frac{\kappa_{n-1}s}{\kappa_{n-1}+s}(\bar{\boldsymbol{y}} - \mu_{n-1})(\bar{\boldsymbol{y}} - \mu_{n-1})^T, \end{aligned} \tag{5}$$

where $S = \sum_{j=1}^{s} (\boldsymbol{y}_j - \bar{\boldsymbol{y}})(\boldsymbol{y}_j - \bar{\boldsymbol{y}})^T$. If only one sample $\boldsymbol{y}$ is generated in iteration n, then $s = 1$, and $\boldsymbol{y}_j = \bar{\boldsymbol{y}} = \boldsymbol{y}$ and $S = 0$. As noted in the previous section, while one can allow $s > 1$, we use $s = 1$.

The joint posterior distribution is:

$$\begin{aligned} \Gamma | \mathscr{A}_n &\sim \text{Inv-Wishart}_{\upsilon_n}(\Xi_n^{-1}) \\ \boldsymbol{f} | \Gamma, \mathscr{A}_n &\sim N(\mu_n, \Gamma / \kappa_n). \end{aligned} \tag{6}$$

From (6), the marginal distribution of the posterior mean $\boldsymbol{f}|\mathscr{A}_n$ is

$$\boldsymbol{f} | \mathscr{A}_n \sim t_{(\upsilon_n - r + 1)}(\mu_n, \Xi_n / (\kappa_n(\upsilon_n - r + 1))), \tag{7}$$

where $t_{(\upsilon_n - r+1)}(\mu_n, \Xi_n/(\kappa_n(\upsilon_n - r+1)))$ is a multidimensional Student-t distribution with $(\upsilon_n - r + 1)$ degrees of freedom, location parameter μ_n, and scale matrix $\Xi_n/(\kappa_n(\upsilon_n - r+1))$. The density function of $\boldsymbol{f}|\mathscr{A}_n$ is proportional to $|\Xi_n/(\kappa_n(\upsilon_n - r+1))|^{-1/2}\{1 + (\boldsymbol{f} - \mu_n)^T[\Xi_n/(\kappa_n(\upsilon_n - r+1))]^{-1}(\boldsymbol{f} - \mu_n)\}^{-(\upsilon+1)/2}$ (Gelman et al. 2003).

The Wishart distribution is the chi-squared distribution generalized to higher dimensions. Thus, random matrices with the Wishart distribution can be generated by summing the "squares" of multivariate normal random vectors. Recall that $\Gamma^{-1} \sim \text{Wishart}_\upsilon(\Xi)$ if and only if $\Gamma \sim \text{Inv-Wishart}_\upsilon(\Xi^{-1})$. To simulate $\Gamma^{-1} \sim \text{Wishart}_\upsilon(\Xi)$, one can simulate υ independent random vectors $W_1, \ldots, W_\upsilon$ from the (r-dimensional) multivariate $N(0, \Xi)$ distribution, then set $\Gamma = \sum_{i=1}^{\upsilon} W_i W_i^T$. This works as long as $\upsilon \geq r$, where r is the dimension of Ξ, which equals the number of sampled points in our problem.

4 CONVEXITY

Recall that for a function $g : \mathbb{R}^d \to \mathbb{R}$, we say that the vector of function values $\boldsymbol{g} = (g(\boldsymbol{x}_1), g(\boldsymbol{x}_2), \ldots, g(\boldsymbol{x}_r))$ is convex if and only if there is a convex function on the continuous domain that takes these values at the points in $\boldsymbol{x}$. This happens if and only if $\boldsymbol{g}$ lies in a certain convex closed polyhedral cone $\mathbb{C}$. We define this cone indirectly through the following equivalent condition; see p. 539 of Murty (1988) and Atlason, Epelman, and Henderson (2004): the vector $\boldsymbol{g}$ is convex if and only if we can fit a supporting hyperplane $\boldsymbol{a}_i^T \boldsymbol{x}_i + b_i$ at each point $(\boldsymbol{x}_i, g(\boldsymbol{x}_i))$ such that the hyperplane lies below all other points $((\boldsymbol{x}_j, g(\boldsymbol{x}_j)) : j \neq i)$. In other words, $\boldsymbol{g} \in \mathbb{C}$ if and only if all of the r linear systems (indexed by $i = 1, 2, \ldots, r$)

$$\begin{aligned} &\boldsymbol{a}_i^T \boldsymbol{x}_i + b_i = g(\boldsymbol{x}_i) \\ &\boldsymbol{a}_i^T \boldsymbol{x}_j + b_i \leq g(\boldsymbol{x}_j), \text{ for all } j \neq i \text{ and } j \in \{1, \ldots, r\}. \end{aligned} \tag{LS(i)}$$

are feasible in the variables $\boldsymbol{a}_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$.

The posterior probability $P(\boldsymbol{f} \in \mathbb{C} | \mathscr{A}_n)$ that $\boldsymbol{f}$ is convex is the probability that a random vector falls into a convex cone, which may be difficult to compute exactly. However, we can approximate this probability using Monte Carlo simulation. We simulate m random vectors $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m\}$ from the posterior distribution of $\boldsymbol{f}|\mathscr{A}_n$. For each $k = 1, 2, \ldots, m$, we then set $g(\boldsymbol{x}_i) = \boldsymbol{y}_k(i)$ for each $i = 1, 2, \ldots, r$ in turn and determine whether the systems (LS(i)) are feasible for all $i = 1, 2, \ldots, r$ or not. The proportion of $\boldsymbol{y}_k$'s for which all the systems are feasible is then our Monte-Carlo estimate of the desired posterior probability.

5 SEQUENTIAL ALGORITHM

The general idea of our sequential algorithm is as follows. At the beginning of each iteration of the method, we have a prior belief of the function values at $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ that comes from the information gathered so far. During the iteration, we obtain a set of new samples of the function values at these points using CRN and update the prior to a posterior distribution on the function values. Then we estimate the probability that the function is convex based on this posterior distribution using Monte Carlo, and decide whether to proceed to the next iteration. If another iteration is needed, the posterior will become our prior belief for the next iteration. In this section we present the detailed algorithmic form for this general idea, and show the asymptotic validity of our algorithm.

5.1 Main Algorithm

Algorithm 1 A sequential method for testing for convexity of the function

Require: Prior mean μ_0 and variance Λ_0 of the function values at $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$. Maximum number of iterations N.

1: Initialize $n = 1$.
2: **while** $n \leq N$ **do**
3: Obtain a new set of samples $\boldsymbol{y}_n$ at $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$.
4: Update the posterior distribution of $\boldsymbol{f}|\mathscr{A}_n$ from the new samples $\boldsymbol{y}_n$ using $\boldsymbol{f}|\mathscr{A}_{n-1}$ as the prior.
5: Estimate $p_n = P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$ from the distribution of $\boldsymbol{f}|\mathscr{A}_n$ obtaining a confidence interval $[\hat{p}_n - h_n, \hat{p}_n + h_n]$ by Algorithm 2.
6: Set $n = n + 1$.
7: **end while**
8: **return** A confidence interval $[\hat{p}_N - h_N, \hat{p}_N + h_N]$ of p from the last step.

In Step 4, when Γ is known, the posterior distribution of $\boldsymbol{f}|\mathscr{A}_n$ is multivariate normal, and its parameters are updated using (3); otherwise, the joint posterior distribution is normal-inverse-Wishart, and its parameters are updated using (5). For the estimation in Step 5, the Monte Carlo algorithm described in the previous section can be employed. A more detailed description is given in Section 5.3. We give a computationally more efficient method in Section 6. This algorithm loops through iterations until the runlength N is achieved, but one can stop after any number of iterations n with an interval estimate of $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$.

5.2 Asymptotic Validity

We would like the posterior probability of convexity to converge to 1 in the event that the realized value of $\boldsymbol{f}$ from the prior distribution is convex, and to converge to 0 when it is nonconvex. In the "knife-edge" case where the realized value of $\boldsymbol{f}$ is convex but not strictly convex, we cannot hope for convergence to 0 or 1 because the realized value of $\boldsymbol{f}$ lies on the boundary of $\mathbb{C}$. We establish this convergence in the case where the sampling covariance matrix Γ is known in Theorem 1 below. We conjecture that convergence also holds in the unknown sampling case.

Theorem 1 Assume Γ is known and positive definite. Let $p_n = P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$ be the n-iteration posterior probability that $\boldsymbol{f}$ is convex as in Algorithm 1. As the number of iterations $n \to \infty$, $p_n - \mathbb{1}\{\boldsymbol{f} \in \mathbb{C}\} \to 0$ a.s.

Proof. (Sketch.) We use the fact that $\Lambda_n^{-1/2}(\mu_n - \boldsymbol{f}) \Rightarrow N(0, I)$ as $n \to \infty$, along with the fact that Λ_n is asymptotically Γ/n to conclude that $\mu_n - \boldsymbol{f} \to 0$ in probabilty as $n \to \infty$. Conditional on $\mathscr{A}_n$, the posterior probability, p_n, is the probability that a normal random vector with mean μ_n and covariance Λ_n is contained in $\mathbb{C}$. The probability that $\boldsymbol{f}$ is on the boundary of the convex cone $\mathbb{C}$ is 0, so it suffices to consider the two cases $\boldsymbol{f} \in \mathbb{C}^\circ$ and $\boldsymbol{f} \notin \mathbb{C}$. In the first case, the normal random vector is eventually mostly concentrated in a ball around μ_n that is wholly contained in $\mathbb{C}$, and so $p_n \to 1$ in probability as $n \to \infty$. In the second case, $\boldsymbol{f}$ can be strictly separated from $\mathbb{C}$ and since the normal distribution eventually concentrates outside $\mathbb{C}$, $p_n \to 0$ as $n \to \infty$ in probability. Since $(p_n : n \geq 0)$ is a uniformly integrable martingale, it converges almost surely, and hence the almost-sure limit is $\mathbb{1}\{\boldsymbol{f} \in C\}$. □

5.3 Algorithm for Convexity Test

We can use the following method for estimating $p_n = P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$. The method generates m samples of the function based on the current posterior belief and approximates the probability using the Monte-Carlo method.

Algorithm 2 Method for obtaining an estimation $\hat{p}_n$ to $p_n = P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$.

Require: Posterior distribution of $\boldsymbol{f}|\mathscr{A}_n$ obtained from Algorithm 1; Number of Monte Carlo samples m allowed.

Initialize an $m \times 1$ array $I_n = [I_n^1, I_n^2, \ldots, I_n^m] = [1, \ldots, 1]$.

for $k = 1, \ldots, m$ **do**

Generate an $r \times 1$ Normal random vector $\boldsymbol{y}_n = (y_n(\boldsymbol{x}_1), y_n(\boldsymbol{x}_2), \ldots, y_n(\boldsymbol{x}_r))^T$ from the distribution of $\boldsymbol{f}|\mathscr{A}_n$.

for $i = 1, \ldots, r$ **do**

Solve the feasibility problem (LS(i)) with $g(\boldsymbol{x}_j) = y_n(\boldsymbol{x}_j)$, $j = 1, 2, \ldots, r$.

if LS(i) is infeasible **then**

$I_n^k = \mathbb{1}\{\boldsymbol{y}_n \in \mathbb{C}\} = 0$;

BREAK the inner loop and go to next k.

end if

end for

end for

Calculate the mean p_n^m and standard deviation s_n^m of I_n.

return $\hat{p}_n = p_n^m$ as an estimator of $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$, along with the half-width $h_n = 1.96 * \frac{s_n^m}{\sqrt{m}}$ of a 95% confidence interval.

Step 5 of Algorithm 2 can be achieved using a linear program solver. We also present the following corollary regarding the convergence of the estimator of the probability that $\boldsymbol{f}$ is convex. The proof is omitted due to space limitations.

Corollary 2 Assume Γ is known and positive definite. Let p_n^m be the m-sample estimator of $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$ from Algorithm 2. As the number of iterations of Algorithm 1 $n \to \infty$ and the number of Monte Carlo samples in Algorithm 2 $m \to \infty$, $p_n^m - \mathbb{1}\{\boldsymbol{f} \in \mathbb{C}\} \to 0$ in probability.

6 CHANGE OF MEASURE

In Algorithm 2, the probability that $\boldsymbol{f}$ is convex is estimated by solving up to m feasibility problems LS(i), which is more computationally expensive than the Bayesian updates. Here we introduce a change-of-measure idea to avoid generating new samples from the posterior and solving the linear systems every time we update the posterior.

Suppose for iteration n, with inputs μ_n, Λ_n, and m, Algorithm 2 generates m i.i.d. samples $\{\boldsymbol{Y}_n^k : k = 1, 2, \ldots, m\}$ of $\boldsymbol{Y}_n \sim N(\mu_n, \Lambda_n)$, giving $\{I_n^k = \mathbb{1}\{\boldsymbol{Y}_n^k \in \mathbb{C}\} : k = 1, 2, \ldots, m\}$ after solving the m feasibility problems. Then, at iteration $n+\ell$, for $\ell \geq 0$, $p_{n+\ell} = P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_{n+\ell})$ can be estimated by a sample average of iid replicates of $\hat{p}_{n+\ell} = \mathbb{1}\{\boldsymbol{Y}_n \in \mathbb{C}\}L_n$, where the likelihood ratio L_n is a ratio of normal densities with parameters $\mu_{n+\ell}, \Lambda_{n+\ell}$ (numerator) and μ_n, Λ_n (denominator). In other words, we can re-use the m samples from $N(\mu_n, \Lambda_n)$, thereby avoiding solving m new feasibility problems. Such an estimator has finite variance in the case where Γ is known, as we show in Theorem 3 below. We conjecture that this is also true for the unknown Γ case.

Given this result, one might be tempted to simply generate a single sample at the outset of Algorithm 1, and re-use that sample from then on in every iteration. While such an estimator does have finite variance, it is likely that the variance grows as ℓ increases, so we recommend instead that this estimator be used only for small ℓ, say $\ell \leq 5$. (We are exploring the quality of the estimators as a function of ℓ.)

Theorem 3 Suppose Γ is known and positive definite. Then the estimator $\hat{p}_{n+\ell}$ of $p_{n+\ell} = P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_{n+\ell})$ is unbiased and has finite variance.

Proof. The conditional distribution of $\boldsymbol{f}$, conditional on $\mathscr{A}_n$, which we write as $\boldsymbol{f}|\mathscr{A}_n$, is multivariate normal for any n, and so $\boldsymbol{f}|\mathscr{A}_n$ is absolutely continuous with respect to $\boldsymbol{f}|\mathscr{A}_{n+\ell}$. It immediately follows that $E(\hat{p}_{n+\ell}|\mathscr{A}_{n+\ell}) = p_{n+\ell}$, so $\hat{p}_{n+\ell}$ is unbiased. It remains to show that $E(\hat{p}^2_{n+\ell}|\mathscr{A}_{n+\ell}) < \infty$, so that the estimator has finite second moment and hence variance.

Since Γ is known, we have $\Lambda_n^{-1} = \Lambda_0^{-1} + n\Gamma^{-1}$ for every n from (3). Hence, assuming $\boldsymbol{Y}$ is sampled from $\mathrm{N}(\mu_n, \Lambda_n)$,

$$\begin{aligned}
E(\hat{p}^2_{n+\ell}|\mathscr{A}_{n+\ell}) &= E(\mathbb{1}\{\boldsymbol{Y} \in \mathbb{C}\}L_n^2|\mathscr{A}_{n+\ell}) \\
&= \frac{|\Lambda_n|^r}{|\Lambda_{n+\ell}|^r} E(\mathbb{1}\{\boldsymbol{Y} \in \mathbb{C}\} \exp\{-[(\boldsymbol{Y}-\mu_{n+\ell})^T(\Lambda_0^{-1} + (n+\ell)\Gamma^{-1})(\boldsymbol{Y}-\mu_{n+\ell}) \\
&\qquad - (\boldsymbol{Y}-\mu_n)^T(\Lambda_0^{-1} + n\Gamma^{-1})(\boldsymbol{Y}-\mu_n)]\}|\mathscr{A}_{n+\ell}) \\
&= \frac{|\Lambda_n|^r}{|\Lambda_{n+\ell}|^r} E\left(\mathbb{1}\{\boldsymbol{Y} \in \mathbb{C}\} e^{\phi(\boldsymbol{Y})}|\mathscr{A}_{n+\ell}\right),
\end{aligned}$$

where we define $\phi(\boldsymbol{Y})$ implicitly in the last step. Thus if we can show $E[e^{\phi(\boldsymbol{Y})}|\mathscr{A}_{n+\ell}] < \infty$, then the proof is complete. To that end,

$$\begin{aligned}
\phi(\boldsymbol{Y}) &= -(\boldsymbol{Y}-\mu_{n+\ell})^T\Lambda_0^{-1}(\boldsymbol{Y}-\mu_{n+\ell}) - (n+\ell)(\boldsymbol{Y}-\mu_{n+\ell})^T\Gamma^{-1}(\boldsymbol{Y}-\mu_{n+\ell}) \\
&\qquad + (\boldsymbol{Y}-\mu_n)^T\Lambda_0^{-1}(\boldsymbol{Y}-\mu_n) + n(\boldsymbol{Y}-\mu_n)^T\Gamma^{-1}(\boldsymbol{Y}-\mu_n) \\
&= -\ell\boldsymbol{Y}^T\Gamma^{-1}\boldsymbol{Y} + a_n^T\boldsymbol{Y} + c_n,
\end{aligned}$$

for some $a_n \in \mathbb{R}^r$ and $c_n \in \mathbb{R}$ that can be expressed in terms of $\Lambda_0, \Gamma, \mu_n, \mu_{n+\ell}$. Hence a_n and c_n can be treated as constant when we condition on $\mathscr{A}_{n+\ell}$. It immediately follows that $\phi(\cdot)$ is bounded above, by M say, where M is a random variable that is measurable with respect to $\mathscr{A}_{n+\ell}$, because the bound above is a negative-definite quadratic function. Hence $0 \le e^{\phi(\boldsymbol{Y})} \le e^M$ and so $E[e^{\phi(\boldsymbol{Y})}|\mathscr{A}_{n+\ell}] < \infty$. □

7 NUMERICAL RESULTS

We implemented Algorithms 1 and 2 in Matlab and tested them on strictly convex, strictly concave, and linear functions. For each case, we observed the behavior of the estimated $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$ and its confidence interval as n grows.

For functions in dimension d, we sample $r = 2d+1$ points in space. The choice of r is somewhat arbitrary, at least at this stage in our work, although we must have $r > d+1$, since a linear function fits any set of up to $d+1$ points. The sampled points are generated by Latin Hypercube Sampling to ensure that the samples are evenly spread in each dimension. Again, this is a somewhat arbitrary choice and something that requires further work. As the sampling variance is usually unknown in practice, all of the following tests are based on the normal-inverse-Wishart updates with a non-informative prior. When the number of iterations of Algorithm 1 grows large, we use the change-of-measure method to reduce the computation. The Matlab linear program solver *linprog*() is used to solve the problems LS(i) in Algorithm 2.

7.1 Strictly Convex and Concave Functions

Consider the squared norm function $g_1(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = \sum_{i=1}^d \boldsymbol{x}(i)^2, \boldsymbol{x} \in [-1,1]^d$ that is strictly convex. We let $d = 30$ and use a (symmetric) sampling covariance matrix that has independent $U(0,1)$ off-diagonal entries and 10^4 on the diagonal. The number of Monte Carlo samples in each iteration generated by Algorithm 2 is set to $m = 100$. Note that m controls a major part of the computational time since it is the number of feasibility problems we need to solve in each iteration of the posterior update.

In one realization of this example with design points chosen by Latin Hypercube sampling, it appears that Algorithm 1 and Algorithm 2 behave well. The trajectory of p_n^m is given in Figure 1. The solid line gives the estimator and the dashed lines are the upper and lower 95% confidence bounds for $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$.

We can see that the confidence interval becomes 1 ± 0 after 30 iterations. (Of course, for sufficiently large m the estimator will never *exactly* equal 1 since $p_n < 1$ for all n, but for modest m the estimator can indeed exactly equal 1.) The computational time for this experiment is 1307 seconds.

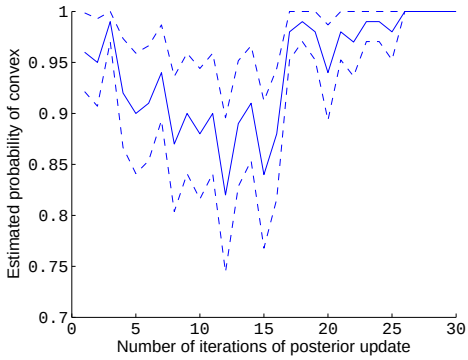

Figure 1: The probability estimation within 30 iterations for a 30-dimensional squared norm function.

Now with the same level of noise, consider the function $g_2(\boldsymbol{x}) = -g_1(\boldsymbol{x})$ that is strictly concave. We found that Algorithm 2 sometimes converges erroneously to 1 when the $r = 2d+1$ randomly sampled points $\boldsymbol{x}$ happen to be the vertices of their convex hull. In this case, connecting the points $(\boldsymbol{x}_i, g_2(\boldsymbol{x}_i))$ forms a convex polytope, whose faces lie on the hyperplanes that are defined by $(\boldsymbol{a}_i, b_i)$ feasible to LS(i). If we increase the number of sampled points, the chance of this happening can be reduced. Figure 2 shows an example in dimension $d = 10$ with $r = 100$ sampled points where p_n^m slowly converges to 0 after 100 iterations. An alternative way is to ensure that at least one point is sampled in the interior of the convex hull of the exisiting points. For example, in Figure 3, we subsequently added the origin $\boldsymbol{x}_{r+1} = \mathbf{0}$ to the set of sampled points $\boldsymbol{x}$ and re-ran the algorithm with the same stream of random numbers. The resulting graph is given in Figure 3. With the new set of sampled points, our algorithm was able to detect the structure more easily. The test took only 20 iterations to converge and required only $r = 2d+2 = 22$ points, which is a significant saving in computational effort. With this improvement, further research is needed to determine other appropriate sampling methods and the associated sampling quantity.

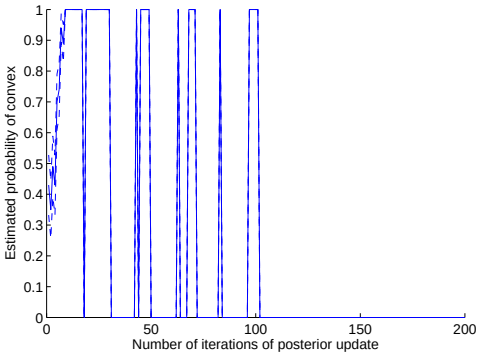

Figure 2: The estimation with increased sample size.

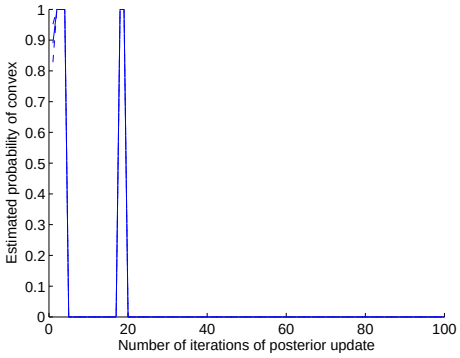

Figure 3: The estimation with origin sampled.

7.2 A Linear Function

We now test our algorithm on $g(\boldsymbol{x}) = 0, \boldsymbol{x} \in [-1,1]^d$, $d = 2$, with noise $N(\mathbf{0}, \Gamma)$ where Γ is equal to 4 on the diagonal and uniform between 0 and 1 off the diagonal. This function is convex, but it lies on the boundary of the cone $\mathbb{C}$, and therefore we cannot expect the posterior probability of convexity to converge to 0 or 1.

In this example, our algorithm is inconclusive, as shown by the trajectory of p_n^m in Figure 4. This result is consistent with the notion that a linear function is the least favorable configuration in a convexity test.

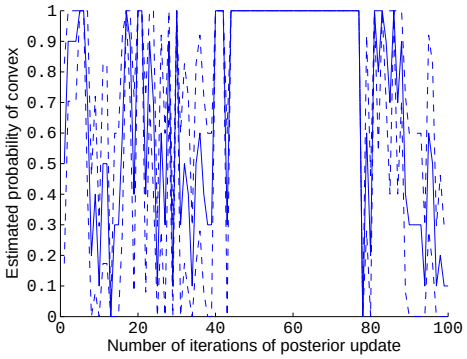

Figure 4: The probability estimation for a 2-dimensional linear function.

7.3 A Function of Unknown Structure

We also test our algorithm on a function with unknown structure — a modified version of the Ambulance Bases problem from SimOpt (Pasupathy and Henderson 2007). In this problem there is a single ambulance in the unit square $[0,1]^2$. Calls arrive independently according to a Poisson process at a random location that follows a density function that is independent of the arrival time. For each call, the ambulance travels at a constant rate to the call and spends a gamma distributed scene time at the call location. The ambulance responds to the calls in FIFO order. The objective of this problem is to choose the base location of the ambulance such that the long-run average response time is minimized. The more detailed problem statement and suggested parameter values can be found at the SimOpt website.

We chose 5 points randomly in $[0,1]^2$ and ran the algorithm for $n = 50$ iterations, with $m = 100$ Monte Carlo samples within each iteration. The resulting confidence interval for $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_{50})$ is 1 ± 0. Figure 5 shows the posterior mean μ_n after $n = 50$ iterations. It can be seen that the estimated function values at the 6 sampled points form a convex basin. This suggests that a good guess for a local minimizer lies inside the convex hull of these 6 sampled points, and would probably be near the sample point $(0.50, 0.50)$.

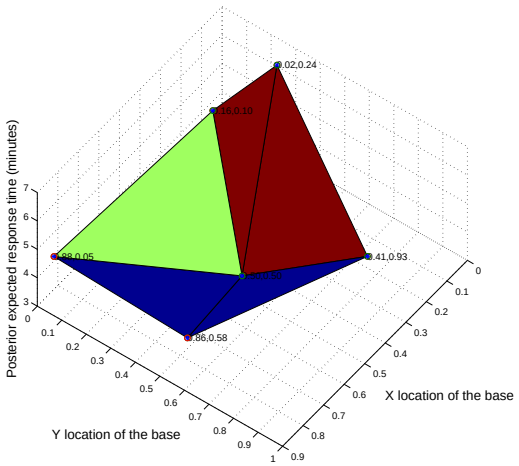

Figure 5: The posterior expected response time as a function of the location of the base.

8 DISCUSSION

The numerical experiments reported in the previous section provide some interesting observations. First, the choice of sampling points $\boldsymbol{x}$ at which to observe the function value plays an important role. In our experiment on the function $g_2(\cdot) = -\|\boldsymbol{x}\|^2$, it is clear that sampling at the origin and a few points away from the origin is the best strategy for detection of nonconvexity. However, when the function is unknown, as is typical, our algorithm may provide the wrong conclusion when the function values at the sampled points do not reflect the structure well. This observation suggests that we should improve our sequential method to accommodate a "poor" selection of sampled points. For example, we might dynamically expand the set $\boldsymbol{x}$ of sampling points based on what we observe in the iterations.

Second, the number of points in the set $\boldsymbol{x}$ affects the result of the test. As we add more points, the distance between points shrinks, and on smaller length scales even strongly convex functions can appear to be approximately linear. In such a situation, if the sampling variance is sufficiently large relative to the "signal" we are trying to detect, we might struggle to detect convexity. Therefore we might need to choose a modest number of points wisely to obtain an accurate result.

Third, linear functions are the least favorable configuration in the test, and in that case the resulting estimated probability oscillates. This is not surprising because as our function gets closer to linear, or more generally close to the boundary of the cone $\mathbb{C}$, any small fluctuation in the function values at each point can easily change the feasibility of $\mathrm{LS}(i)$. To help deal with near-linear functions, in future we might attempt to modify the feasibility problems $\mathrm{LS}(i)$ to output a "distance" of the input function values to the boundary of the convex cone $\mathbb{C}$ instead of just a 0-1 indicator.

Finally, given that computing an estimator of the posterior probability of convexity is the computational bottleneck in our procedure, it is worth searching for more efficient approaches to estimate this probability beyond the change-of-measure approach given here.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation Grant CMMI-1200315.

REFERENCES

Abrevaya, J., and W. Jiang. 2005. "A Nonparametric Approach to Measuring and Testing Curvature". *Journal of Business & Economic Statistics* 23:1–19.

Allon, G., M. Beenstock, S. Hackman, U. Passy, and A. Shapiro. 2007. "Nonparametric Estimation of Concave Production Technologies by Entropic Methods". *Journal of Applied Econometrics* 22 (4): 795–816.

Atlason, J., M. A. Epelman, and S. G. Henderson. 2004. "Call Center Staffing with Simulation and Cutting Plane Methods". *Annals of Operations Research* 127:333–358.

Baraud, Y., S. Huet, and B. Laurent. 2005. "Testing Convex Hypotheses on the Mean of a Gaussian Vector. Application to Testing Qualitative Hypotheses on a Regression Function". *The Annals of Statistics* 33 (1): 214–257.

Bernardo, J. M., and A. F. M. Smith. 2008. *Bayesian Theory*, 240–376. John Wiley & Sons, Inc.

Chen, X., B. E. Ankenman, and B. L. Nelson. 2012. "The Effects of Common Random Numbers on Stochastic Kriging Metamodels". *ACM TOMACS* 22 (2): Article 7.

DeGroot, M. 1970. *Optimal Statistical Decisions*. New York, NY: McGraw-Hill.

Diack, C. A. T., and C. Thomas-Agnan. 1998. "A Nonparametric Test of the Non-convexity of Regression". *Nonparametric Statistics* 9:335–362.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2003, July. *Bayesian Data Analysis*. 2nd ed. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC.

Glynn, P., and G. Infanger. 2013. "Simulation-based Confidence Bounds for Two-stage Stochastic Programs". *Mathematical Programming* 138 (1-2): 15–42.

Golub, G. H., and C. F. Van Loan. 1996, October. *Matrix Computations*. 3rd ed. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press.

Hannah, L. A., and D. B. Dunson. 2011, May. "Multivariate Convex Regression with Adaptive Partitioning". *ArXiv e-prints*.

Judge, G. G., and T. Takayama. 1966. "Inequality Restrictions in Regression Analysis". *Journal of the American Statistical Association* 61 (313): pp. 166–181.

Juditsky, A., and A. Nemirovski. 2002. "On Nonparametric Tests of Positivity/Monotonicity/Convexity". *The Annals of Statistics* 30 (2): 498–527.

Lau, L. J. 1978. "Testing and Imposing Monoticity, Convexity, and Quasi-convexity Constraints". *Electronic Journal of Statistics* 1:409–453.

Lim, E., and P. W. Glynn. 2012. "Consistency of Multidimensional Convex Regression.". *Operations Research* 60 (1): 196–208.

Meyer, M. C. 2012. "Constrained penalized splines". *Canadian Journal of Statistics* 40 (1): 190–206.

Murty, K. G. 1988. *Linear Complementarity, Linear and Nonlinear Programming*. Berlin: Heldermann Verlag.

Nesterov, Y. 2004. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic.

Pasupathy, Raghu and Henderson, Shane G. 2007. "Ambulance Bases". Accessed May. 15, 2014. http://simopt.org/wiki/index.php?title=Ambulances_in_a_square.

Seijo, E., and B. Sen. 2010, March. "Nonparametric Least Squares Estimation of a Multivariate Convex Regression Function". *ArXiv e-prints*.

Siegmund, D. 1985. *Sequential Analysis: Tests and Confidence Intervals*. Springer Series in Statistics. Springer.

Silvapulle, M. J., and P. K. Sen. 2001. *Constrained Statistical Inference: Order, Inequality, and Shape Restrictions*, Chapter 3, 59–141. John Wiley & Sons, Inc.

Wang, J. C., and M. C. Meyer. 2011. "Testing the Monotonicity or Convexity of a Function Using Regression Splines". *Canadian Journal of Statistics* 39 (1): 89–107.

AUTHOR BIOGRAPHIES

NANJING JIAN is a PhD student in the School of Operations Research and Information Engineering at Cornell University. She received her B.S. in Industrial and Systems Engineering at University of Wisconsin - Madison in 2012. Her research interest is in simulation optimization. Her email address is nj227@cornell.edu.

SHANE G. HENDERSON is a professor in the School of Operations Research and Information Engineering at Cornell University. His research interests include discrete-event simulation and simulation optimization, and he has worked for some time with emergency services. He co-edited the Proceedings of the 2007 Winter Simulation Conference. His web page is http://people.orie.cornell.edu/∼shane.

SUSAN HUNTER is an assistant professor in the School of Industrial Engineering at Purdue University. Her research interests include Monte Carlo methods and simulation optimization. Her email address is susanhunter@purdue.edu, and her webpage is http://web.ics.purdue.edu/∼hunter63/.