Generalized Integrated Brownian Fields for Simulation Metamodeling

Peter Salemi
Jeremy Staum
Barry Nelson

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

September 14, 2014

Abstract

We introduce a new class of Gaussian random fields (GRFs), which we call generalized integrated Brownian fields (GIBFs). We focus on the use of GIBFs for Gaussian process modeling in deterministic and stochastic simulation metamodeling. We build GIBFs from the well-known Brownian motion and discuss properties such as the Markov property and mean reversion; we provide a formal definition of the latter. We show that GIBFs have no mean reversion and differentiability that can vary in each coordinate, as well as the Markov property in one dimension. We also show how to implement stochastic kriging with GIBFs, covering trend modeling and parameter fitting. Lastly, we use tractable examples to demonstrate superior prediction ability as compared to the GRFs corresponding to the Gaussian and Matérn covariance functions.

1 Introduction

Stochastic simulations are often used to model complex systems in industrial engineering and operations research. Because simulation models are typically not limited by the complexity of the underlying system, simulation runs may be time-consuming to execute, especially when there are many scenarios that need to be evaluated. This limits the use of simulation models for supporting real-time decision making. When the simulation model can be run for a significant amount of time before decisions must be made, we can use the output from the simulation to build a statistical model of the response surface. We call this statistical model a simulation metamodel. Using the metamodel, we can predict the value of the response surface for any scenario, even if it has not been simulated.

A great deal of research has been directed towards fitting linear regression models to simulation output. However, we are particularly interested in general metamodeling approaches that assume less structure than linear models. In the deterministic computer experiments literature, the use of Gaussian process models has been remarkably successful for global metamodeling [Santner et al., 2010]. Following the introduction of Gaussian process models into the design and analysis of deterministic computer experiments, Mitchell and Morris [1992] introduced Gaussian process models

for representing the response surface in stochastic simulation. Since the predictions are made by fitting a Gaussian process, we are able to obtain a measure of uncertainty in predictions, which gives rise to confidence intervals. Furthermore, the measure of uncertainty in predictions facilitates sequential, adaptive experiment designs, and can provide statistical inference about the fitted model [Ankenman et al., 2010].

In simulation metamodeling using Gaussian process models, the response surface is modeled as a sample path of a Gaussian random field (GRF). A critical choice in fitting Gaussian process models is specifying the GRF. To obtain better prediction ability, the GRF should have desirable properties and be flexible enough to capture the characteristics of the response surface, such as the level of differentiability. A GRF is completely specified by its mean function (often assumed to be identically zero) and covariance function. Thus, selecting the proper covariance function is crucial for determining the prediction ability of the resulting Gaussian process model. Indeed, much research has been done that discusses the choice of covariance functions for Gaussian process modeling [Santner et al., 2010, Xie et al., 2010, Paciorek and Schervish, 2004].

Gaussian process models were initially used in geostatistics to predict the amount of gold in underground deposits [Krige, 1951]. For these applications, if we were interested in predicting the amount of gold underneath a region, knowing the amount of gold underneath the boundary of the region would not be sufficient information for our prediction. For example, if we knew there was a lot of gold nearby, but none necessarily underneath the boundary, we would still expect there to be gold underneath the region in which we were interested. We are mainly concerned with response surfaces in operations research, which are different than response surfaces in geostatistics. In operations research, if we are interested in predicting the value of the response surface in a region, then given sufficient information about the response surface on the boundary, information about the response surface outside of the region often would not assist in our predictions. By sufficient information, we mean the level of the response surface and perhaps some derivatives. For GRFs, this property is analogous to the Markov property: the GRF inside a region, given sufficient information (level and derivatives) on the boundary, is independent of the GRF outside [Pitt, 1971].

The ability to control the differentiability of the GRF is a characteristic that has received considerable attention in the literature [Santner et al., 2010]. A common class of GRFs that are used for metamodeling corresponds to the power exponential covariance function, for which the differentiability is controlled by a single parameter. However, these GRFs can only be non-differentiable or infinitely differentiable, depending on the value of the parameter. Another class of covariance functions is the class of Matérn covariance functions, which also has a single parameter that controls the differentiability of the GRF. In contrast to the power exponential covariance function, the GRFs corresponding to the Matérn class can have differentiability of any order. However, the differentiability cannot differ in each coordinate. See the beginning of Section 3 for further discussion.

Mean reversion is often an undesirable characteristic of the metamodel that arises from using a mean-reverting GRF. An example of mean reversion is given in Figure 1, which contains a plot of the response surface (the expected waiting time in an M/M/1 queue) and the metamodel built using a mean-reverting GRF. Mean reversion results because the covariances among the design points, the scenarios at which we run the simulation, and prediction points, the scenarios at which we want to make a prediction, gets weaker as the distance between them becomes greater, thus, the prediction reverts to the overall mean of the GRF. Any GRF in which the covariance between two points monotonically decreases to zero as the distance between the points increases exhibits mean reversion (see Section 5.1 for further discussion). Due to the poor predictions that can result from mean reversion, many methods have been proposed to reduce it for Gaussian process modeling (see, for example, Joseph et al. [2008], Joseph [2006], and Li and Sudjianto [2005]). Furthermore,

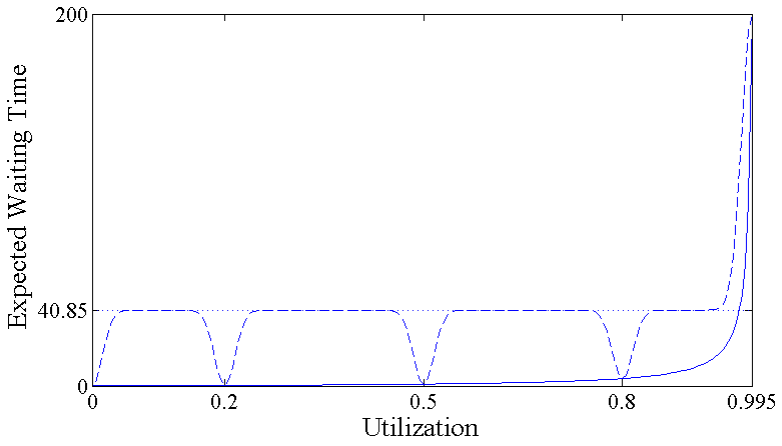


Figure 1: Mean reversion in a metamodel. The solid line is the expected waiting time for a customer in an M/M/1 queue, the dashed line is the metamodel, and the dotted line is the mean of the fitted Gaussian process. The data used to fit the metamodel are noiseless observations of the expected waiting time at the points 0, 0.2, 0.5, 0.8, and 0.995.

for these covariance functions, extrapolation causes severe mean reversion since the design points will only be on one side of the prediction point. As in the procedure in Liu and Staum [2010], it can sometimes be very expensive to avoid extrapolation, especially in high dimension, since an extremely large number of design points would be needed to cover a high-dimensional design space, the space of all possible design points. Thus, we would prefer to use a GRF which leads to a metamodel that does not exhibit mean reversion.

In this paper, we introduce a new class of GRFs, which we call generalized integrated Brownian fields (GIBFs), focusing on the use of GIBFs for Gaussian process modeling in deterministic and stochastic simulation metamodeling. There are two ways to construct GIBFs: using a probabilistic approach or the theory of reproducing kernels. In the latter, the covariance functions of GIBFs can be constructed using a novel parametrization of the reproducing kernel corresponding to the Sobolev-Hilbert space [Berlinet and Thomas-Agnan, 2004], which is a tensor-product Hilbert space. Although we will use the first construction and build GIBFs from Brownian motion, we will use the theory of reproducing kernels to prove a property of GIBFs. By placing the constuction of GIBFs in the probabilistic setting, we can discuss properties such as the Markov property and mean reversion; we provide a formal definition of the latter. We show that GIBFs have the Markov property in one dimension and no mean reversion. Furthermore, the differentiability of GIBFs can vary in each coordinate. We also show how to implement stochastic kriging with GIBFs, and use tractable examples to compare the prediction ability of GIBFs with the GRFs corresponding to the Gaussian and Matérn covariance functions.

Gaussian process modeling with GIBFs is a generalization of using smoothing splines with integrated Brownian motion in one dimension [Wahba, 1990]. Berlinet and Thomas-Agnan [2004] and Gu and Wahba [1992] provide very general guidelines for creating smoothing splines in a tensor-product Hilbert space. In addition to being very general, these guidelines also assume the user has performed a decomposition of the tensor-product Hilbert space, and has chosen which subspaces to include in the penalty term and which subspaces to disregard altogether by performing a model selection. Furthermore, each term in the penalty is weighted with a separate coefficient, which leads to many coefficients in high dimensions. The method presented in this paper is much easier

to implement; once the trend is chosen, the covariance function follows automatically, and the parameters are chosen from the simulation output using maximum likelihood estimation.

Although the use of Gaussian process models in simulation metamodeling has led to several different metamodeling techniques (see, for example, van Beers and Kleijnen [2003], Kleijnen and van Beers [2005], and Yin et al. [2011]), we will focus on the simulation metamodeling technique called stochastic kriging, which we discuss in Section 2. We then present GIBFs using a probabilistic approach in Section 3, and provide a guide to using these random fields with stochastic kriging in Section 4. We discuss several properties of GIBFs in Section 5 and conclude the paper with numerical experiments in Section 6 which show the improved prediction accuracy as compared to the well-known and highly-used Gaussian and Matérn covariance functions.

2 Stochastic Kriging

Gaussian process models have been used for approximating the output of deterministic computer experiments following the work of Sacks et al. [1989], which introduced kriging into the design and analysis of deterministic computer experiments. In kriging, the response surface $y(\cdot)$ at $\mathbf{x}$ is modeled as a realization of the random variable

$$\mathsf{Y}_{\mathsf{M}}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + \mathsf{M}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is a point in the design space $\mathcal{X}$ (the space of all possible design points), $\mathbf{f}(\cdot)$ is a $p \times 1$ vector of known functions, i.e., $\mathbf{f}(\cdot) = (f_1(\cdot), f_2(\cdot), \ldots, f_p(\cdot))^\top$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\mathsf{M}(\cdot)$ is a mean-zero GRF. In other words, sample paths of $\mathsf{M}(\cdot)$ can be thought of as being randomly sampled from a space of functions mapping $\mathbb{R}^d \to \mathbb{R}$, according to a Gaussian measure [Ankenman et al., 2010]. The GRF $\mathsf{M}(\cdot)$ is assumed to exhibit spatial covariance, which is determined by the covariance function $\Sigma_{\mathsf{M}}(\cdot, \cdot; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters. Specifically, the covariance between $\mathsf{M}(\cdot)$ at two points $\mathbf{x}$ and $\mathbf{x}'$ in the design space is given by

$$\mathrm{Cov}[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{x}')] = \Sigma_{\mathsf{M}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}).$$

For deterministic computer experiments where the output of the experiment contains no noise, the response surface can be observed exactly at each of the design points at which the computer experiment is run. Kriging results in an interpolation of the data, i.e., the metamodel is equal to the computer experiment output at each of the scenarios run, which fits the deterministic nature of the problem.

In the stochastic simulation case, we no longer observe the response surface without noise. Rather, we run the simulation model at k design points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ for a total of n_i replications at design point $\mathbf{x}_i$. Replication j at design point $\mathbf{x}_i$ is denoted by $\mathcal{Y}_j(\mathbf{x}_i)$. At design point $\mathbf{x}_i$ we collect the sample mean $\bar{\mathcal{Y}}(\mathbf{x}_i) = (1/n_i)\sum_{j=1}^{n_i} \mathcal{Y}_j(\mathbf{x}_i)$, and the sample standard deviation $s^2(\mathbf{x}_i) = (1/(n_i-1))\sum_{j=1}^{n_i}(\mathcal{Y}_j(\mathbf{x}_i) - \bar{\mathcal{Y}}(\mathbf{x}_i))^2$. Gaussian process modeling in stochastic simulation utilizes the sample means and sample standard deviations at the design points to build the Gaussian process model.

In stochastic kriging [Ankenman et al., 2010], the response surface is modeled as a sample path of the GRF $\mathsf{Y}_{\mathsf{M}}(\cdot)$, given by Equation (1), with mean function $\mathbf{f}(\cdot)^\top \boldsymbol{\beta}$ and covariance function $\Sigma_{\mathsf{M}}(\cdot, \cdot; \boldsymbol{\theta})$. The simulation output on replication j at design point $\mathbf{x}$ is modeled as a realization of the random variable

$$\mathsf{Y}_{\mathsf{M},\epsilon_j}(\mathbf{x}) = \mathsf{Y}_{\mathsf{M}}(\mathbf{x}) + \epsilon_j(\mathbf{x}),$$

where the mean-zero sampling noise in the replications $\epsilon_1(\mathbf{x}), \epsilon_2(\mathbf{x}), \ldots$ at a design point $\mathbf{x}$ is independent and identically distributed across replications. Following the recommendations of Chen

et al. [2012], we assume the sampling noise is independent across design points, i.e., we do not use Common Random Numbers (CRN), although our method will still work when CRN are used. The sampling noise is referred to as intrinsic uncertainty, since it is inherent in the stochastic simulation. The stochastic nature of M is called extrinsic uncertainty, since it is imposed on the problem to aid in the development of the metamodel [Ankenman et al., 2010].

Suppose that the simulation model has been run at the k design points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ yielding the vector of observed simulation output $\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{x}_1), \bar{\mathcal{Y}}(\mathbf{x}_2), \ldots, \bar{\mathcal{Y}}(\mathbf{x}_k))^\top$, and we now want to predict the response surface at $\mathbf{x}_0$. Let $\widehat{\Sigma}_\mathsf{M}$ be the $k \times k$ variance-covariance matrix with ijth entry $\Sigma_\mathsf{M}(\mathbf{x}_i, \mathbf{x}_j; \widehat{\boldsymbol{\theta}})$, where $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$, let $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \ldots, \mathbf{f}(\mathbf{x}_k))^\top$ be the $k \times p$ regression matrix, and let $\widehat{\Sigma}_\mathsf{M}(\mathbf{x}_0, \cdot)$ be the $k \times 1$ vector of spatial covariances between the design points and the prediction point, i.e., the ith entry of $\widehat{\Sigma}_\mathsf{M}(\mathbf{x}_0, \cdot)$ is $\Sigma_\mathsf{M}(\mathbf{x}_0, \mathbf{x}_i; \widehat{\boldsymbol{\theta}})$. Also, let $\widehat{\Sigma}_\epsilon = \mathrm{diag}\{s^2(\mathbf{x}_1)/n_1, s^2(\mathbf{x}_2)/n_2, \ldots, s^2(\mathbf{x}_k)/n_k\}$. For brevity, we write $\widehat{\Sigma} = \widehat{\Sigma}_\mathsf{M} + \widehat{\Sigma}_\epsilon$. The stochastic kriging prediction at $\mathbf{x}_0$ is given by

$$\widehat{\mathsf{Y}}_\mathsf{M}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \widehat{\boldsymbol{\beta}} + \widehat{\Sigma}_\mathsf{M}(\mathbf{x}_0, \cdot)^\top \widehat{\Sigma}^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}}), \tag{2}$$

where $\widehat{\boldsymbol{\beta}} = (\mathbf{F}^\top \widehat{\Sigma}^{-1} \mathbf{F})^{-1} \mathbf{F}^\top \widehat{\Sigma}^{-1} \bar{\mathcal{Y}}$. The mean-squared error ($\widehat{\mathrm{MSE}}$) of the prediction $\widehat{\mathsf{Y}}_\mathsf{M}(\mathbf{x}_0)$ is

$$\widehat{\mathrm{MSE}}(\widehat{\mathsf{Y}}_\mathsf{M}(\mathbf{x}_0)) = \Sigma_\mathsf{M}(\mathbf{x}_0, \mathbf{x}_0; \widehat{\boldsymbol{\theta}}) - \widehat{\Sigma}_\mathsf{M}(\mathbf{x}_0, \cdot)^\top \widehat{\Sigma}^{-1} \widehat{\Sigma}_\mathsf{M}(\mathbf{x}_0, \cdot) + \eta^\top (\mathbf{F}^\top \widehat{\Sigma}^{-1} \mathbf{F})^{-1} \eta, \tag{3}$$

where $\eta = \mathbf{f}(\mathbf{x}_0) - \mathbf{F}^\top \widehat{\Sigma}^{-1} \widehat{\Sigma}_\mathsf{M}(\mathbf{x}_0, \cdot)$. The last term arises because the $p \times 1$ vector of regression coefficients needs to be estimated, which inflates the MSE of the prediction. If $\boldsymbol{\beta}$ was known instead of estimated, the last term of $\widehat{\mathrm{MSE}}(\widehat{\mathsf{Y}}_\mathsf{M}(\mathbf{x}_0))$ would drop from the expression.

When derivatives of the response surface can be estimated using the simulation model, it has been shown that incorporating derivative information can substantially improve the prediction performance [Chen et al., 2013]. The model for the derivative of the response surface at the design point $\mathbf{x} \in \mathcal{X}$ is given by

$$\frac{\partial^{|\boldsymbol{\alpha}|} \mathsf{Y}_\mathsf{M}(\mathbf{x})}{\partial \mathbf{x}^{\boldsymbol{\alpha}}} = \left(\frac{\partial^{|\boldsymbol{\alpha}|} \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^{\boldsymbol{\alpha}}} \right)^\top \boldsymbol{\beta} + \frac{\partial^{|\boldsymbol{\alpha}|} \mathsf{M}(\mathbf{x})}{\partial \mathbf{x}^{\boldsymbol{\alpha}}},$$

and the corresponding model for the derivative estimate on simulation replication j at $\mathbf{x}$ is given by

$$\mathsf{D}_j^{\boldsymbol{\alpha}}(\mathbf{x}) = \frac{\partial^{|\boldsymbol{\alpha}|} \mathsf{Y}_\mathsf{M}(\mathbf{x})}{\partial \mathbf{x}^{\boldsymbol{\alpha}}} + \gamma_j^{\boldsymbol{\alpha}}(\mathbf{x}),$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)^\top$, $|\boldsymbol{\alpha}| = \sum_{i=1}^d \alpha_i$, $\partial^{|\boldsymbol{\alpha}|}/\partial \mathbf{x}^{\boldsymbol{\alpha}} = \partial^{|\boldsymbol{\alpha}|}/(\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d})$, and $\gamma_j^{\boldsymbol{\alpha}}(\mathbf{x})$ for $j = 1, 2, \ldots$ represent the mean-zero, independent and identically distributed sampling noise in the derivative estimates of the $\boldsymbol{\alpha}$th derivative. Since we are not using CRN, the sampling noise in the derivative estimates is independent across design points, i.e., $\gamma_k^{\boldsymbol{\alpha}_1}(\mathbf{x}_i)$ is independent of $\gamma_l^{\boldsymbol{\alpha}_2}(\mathbf{x}_j)$, for any $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, k$ and l and all $i \neq j$. Furthermore, $\gamma_k^{\boldsymbol{\alpha}}(\mathbf{x}_i)$ is independent of $\epsilon_l(\mathbf{x}_j)$, for any $\boldsymbol{\alpha}, k$, and l and all $i \neq j$. The covariance functions of the derivative processes are derived by taking the corresponding derivatives of the covariance function Σ_M, i.e., for $\mathbf{x}, \mathbf{y} \in \mathcal{X}$,

$$\mathrm{Cov}\left[\frac{\partial^{|\boldsymbol{\alpha}|} \mathsf{Y}_\mathsf{M}(\mathbf{x})}{\partial \mathbf{x}^{\boldsymbol{\alpha}}}, \frac{\partial^{|\boldsymbol{\beta}|} \mathsf{Y}_\mathsf{M}(\mathbf{y})}{\partial \mathbf{x}^{\boldsymbol{\beta}}} \right] = \left. \frac{\partial^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}{\partial \mathbf{s}^{\boldsymbol{\alpha}} \partial \mathbf{t}^{\boldsymbol{\beta}}} \Sigma_\mathsf{M}(\mathbf{s}, \mathbf{t}; \boldsymbol{\theta}) \right|_{(\mathbf{x}, \mathbf{y})}.$$

To illustrate stochastic kriging with derivative estimates, we present a one-dimensional, two-point example, and refer the reader to Chen et al. [2013] for details on implementation and results.

Consider a simulation model in which the design space is one-dimensional, with two design points x_1 and x_2. We run the simulation model at x_1 and x_2 for n_1 and n_2 replications, respectively. We obtain noisy estimates (the sample averages) of the response surface, $\bar{\mathcal{Y}}(x_1)$ and $\bar{\mathcal{Y}}(x_2)$, as well as a noisy estimate of the derivative at x_1, $\bar{\mathcal{D}}^1(x_1) = (1/n_1)\sum_{j=1}^{n_1} \mathcal{D}_j^1(x_1)$. The vector of observations $\bar{\mathcal{Y}}_+ = (\bar{\mathcal{Y}}(x_1), \bar{\mathcal{Y}}(x_2), \bar{\mathcal{D}}^1(x_1))^\top$ has mean $\mathbf{F}_+\boldsymbol{\beta}$, where

$$\mathbf{F}_+ = \left(\mathbf{f}(x_1), \mathbf{f}(x_2), \frac{\partial \mathbf{f}(x_1)}{\partial x}\right)^\top,$$

and variance-covariance matrix $\Sigma_+ = \Sigma_{\mathsf{M}_+} + \Sigma_{\epsilon_+}$, where

$$\Sigma_{\mathsf{M}_+} = \left(\begin{array}{cc|c} \Sigma_{\mathsf{M}}(x_1, x_1; \boldsymbol{\theta}) & \Sigma_{\mathsf{M}}(x_1, x_2; \boldsymbol{\theta}) & \frac{\partial \Sigma_{\mathsf{M}}(s,t;\boldsymbol{\theta})}{\partial t}|_{(x_1,x_1)} \\ \Sigma_{\mathsf{M}}(x_2, x_1; \boldsymbol{\theta}) & \Sigma_{\mathsf{M}}(x_2, x_2; \boldsymbol{\theta}) & \frac{\partial \Sigma_{\mathsf{M}}(s,t;\boldsymbol{\theta})}{\partial t}|_{(x_2,x_1)} \\ \hline \frac{\partial \Sigma_{\mathsf{M}}(s,t;\boldsymbol{\theta})}{\partial s}|_{(x_1,x_1)} & \frac{\partial \Sigma_{\mathsf{M}}(s,t;\boldsymbol{\theta})}{\partial s}|_{(x_1,x_2)} & \frac{\partial^2 \Sigma_{\mathsf{M}}(s,t;\boldsymbol{\theta})}{\partial s \partial t}|_{(x_1,x_1)} \end{array}\right)$$

and

$$\Sigma_{\epsilon_+} = \left(\begin{array}{cc|c} \mathrm{Var}\left[\bar{\epsilon}(x_1)\right] & 0 & \mathrm{Cov}\left[\bar{\gamma}(x_1), \bar{\epsilon}(x_1)\right] \\ 0 & \mathrm{Var}\left[\bar{\epsilon}(x_2)\right] & 0 \\ \hline \mathrm{Cov}\left[\bar{\epsilon}(x_1), \bar{\gamma}(x_1)\right] & 0 & \mathrm{Var}\left[\bar{\gamma}(x_1)\right] \end{array}\right).$$

Also, let $\Sigma_{\mathsf{M}_+}(\mathbf{x}_0, \cdot)$ be the vector of covariances between the prediction point $\mathbf{x}_0$ and the design points,

$$\Sigma_{\mathsf{M}_+}(x_0, \cdot) = \left(\begin{array}{c} \Sigma_{\mathsf{M}}(x_0, x_1; \boldsymbol{\theta}) \\ \Sigma_{\mathsf{M}}(x_0, x_2; \boldsymbol{\theta}) \\ \frac{\partial \Sigma_{\mathsf{M}}(s,t;\boldsymbol{\theta})}{\partial t}|_{(x_0,x_1)} \end{array}\right).$$

The stochastic kriging predictor, as well as the MSE of the prediction, is obtained by replacing $\bar{\mathcal{Y}}, \mathbf{F}, \widehat{\Sigma}$, and $\widehat{\Sigma}_{\mathsf{M}}(x_0, \cdot)$ by $\bar{\mathcal{Y}}_+, \mathbf{F}_+$, and the estimated versions of Σ_+ and $\Sigma_{\mathsf{M}_+}(x_0, \cdot)$, respectively, in Equations (2) and (3).

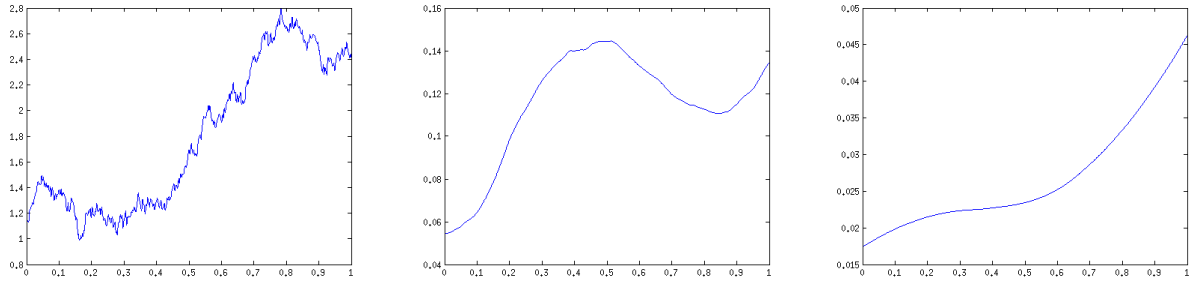
3 Generalized Integrated Brownian Fields

In stochastic kriging, the response surface is modeled as a sample path of the GRF $\mathsf{Y}_{\mathsf{M}}(\cdot)$, given by Equation (1), with mean function $\mathbf{f}(\cdot)^\top\boldsymbol{\beta}$ and covariance function $\Sigma_{\mathsf{M}}(\cdot, \cdot; \boldsymbol{\theta})$. The GRFs we construct in this section, i.e., GIBFs, have desirable properties such as the Markov property in one dimension, no mean reversion, and differentiability that can vary in each coordinate. We want to use GRFs with these properties in an effort to obtain better prediction ability.

A widely-used GRF corresponds to the so-called Gaussian covariance function, for which the covariance between the GRF at $\mathbf{x}$ and $\mathbf{x}'$ is given by $\Sigma_{\mathsf{M}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}) = \sigma^2 \exp\{-\sum_i \theta_i (x_i - x_i')^2\}$, where θ_i, x_i, and x_i' are the ith coordinates of $\boldsymbol{\theta}$, $\mathbf{x}$, and $\mathbf{x}'$, respectively, and σ^2 is the variance of the GRF. This GRF is mean-reverting, and is often criticized as being too smooth since it is infinitely differentiable in every coordinate. Another widely-used GRF corresponds to the Matérn covariance function, for which the covariance between the GRF at $\mathbf{x}$ and $\mathbf{x}'$ is given by

$$\Sigma_{\mathsf{M}}(\mathbf{x}, \mathbf{x}'; \theta) = \sigma^2 \frac{1}{\Gamma(v) 2^{v-1}} \left(\sqrt{2v}||\boldsymbol{\theta}^\top(\mathbf{x} - \mathbf{x}')||_2\right)^v K_v\left(\sqrt{2v}||\boldsymbol{\theta}^\top(\mathbf{x} - \mathbf{x}')||_2\right),$$

where $\Gamma(\cdot)$ is the gamma function and $K_v(\cdot)$ is the modified Bessel function of the second kind. This GRF is mean-reverting and the differentiability is controlled by the single parameter v. However, the differentiability cannot differ in each coordinate.



(a) 0-GIBM (b) 1-GIBM (c) 2-GIBM

Figure 2: Sample paths of GIBMs on the unit interval.

GIBFs are generalized versions of integrated Brownian fields [Fill and Torcaso, 2004], which are multivariate versions of integrated Brownian motions. We first construct one-dimensional generalized integrated Brownian motions and then construct multi-dimensional generalized integrated Brownian fields.

3.1 One-Dimensional Generalized Integrated Brownian Motions

Consider one-dimensional Brownian motion $B(\cdot;\theta)$ on the interval $[0,1]$ with volatility θ. This process is a real-valued, mean-zero Gaussian stochastic process with continuous, non-differentiable sample paths. The covariance between $B(\cdot;\theta)$ at $x, y \in [0,1]$ is given by $\Sigma_B(x,y;\theta) = \theta \min\{x,y\}$. An m-times differentiable stochastic process can be obtained by integrating $B(\cdot;\theta)$ for m times, which gives us m-integrated Brownian motion $B_m(\cdot;\theta)$ with volatility θ. The integral representation of m-integrated Brownian motion with volatility θ at x is

$$B_m(x;\theta) = \int_0^x B_{m-1}(u;\theta)du = \int_0^x \frac{(x-u)^m}{m!} dB(u;\theta), \tag{4}$$

where the first equality expresses $B_m(x;\theta)$ recursively with $B_0(\cdot;\theta) = B(\cdot;\theta)$, and the second equality follows from integration by parts, which expresses $B_m(x;\theta)$ as an integral with respect to Brownian motion with volatility θ. From the first integral in Equation (4), it is clear that the process $B_m(\cdot;\theta)$ and its m derivatives $B_m^{(i)}(\cdot;\theta)$, for $i = 1, 2, \ldots, m$, are zero at the boundary $x = 0$. These boundary conditions make $B_m(\cdot;\theta)$ unsuitable for metamodeling, since the response surface and its derivatives may not be zero at the boundary $x = 0$. We modify $B_m(\cdot;\theta)$ by adding a random polynomial whose coefficients are $m+1$ independent standard normal random variables $Z_0, Z_1, \ldots, Z_m$ scaled by some parameters. This process is denoted by $X_m(\cdot;\boldsymbol{\theta})$ and is called a generalized m-integrated Brownian motion (m-GIBM) and is defined by

$$X_m(x;\boldsymbol{\theta}) \triangleq \sum_{n=0}^{m} \frac{\sqrt{\theta_n} Z_n}{n!} x^n + B_m(x;\theta_{m+1}), \tag{5}$$

where θ has been relabelled as θ_{m+1} for convenience, $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{m+1})^\top$, and $B_m(\cdot;\theta_{m+1})$ is independent of Z_n for all $n = 1, 2, \ldots, m$. Figure 2 shows sample paths of a 0-GIBM, a 1-GIBM, and a 2-GIBM on the unit interval. Directly from the definitions of $B_m(\cdot;\theta)$ and $X_m(\cdot;\boldsymbol{\theta})$, it follows that the covariance between $X_m(\cdot;\boldsymbol{\theta})$ at $x, y \in [0,1]$ is given by

$$\Sigma_{X_m}(x,y;\boldsymbol{\theta}) = \sum_{k=0}^{m} \theta_k \frac{x^k y^k}{(k!)^2} + \theta_{m+1} \int_0^\infty \frac{(x-u)_+^m (y-u)_+^m}{(m!)^2} du.$$

For any m, the integral can be easily computed and has a convenient closed-form solution, composed of terms that are products of the functions min and max. If we let $x \wedge y \triangleq \min\{x, y\}$ and $x \vee y \triangleq \max\{x, y\}$, then the closed-form solutions for the cases $m = 0, 1, 2$ are given by

$$\int_0^\infty \frac{(x-u)_+^0 (y-u)_+^0}{(0!)^2} du = x \wedge y,$$
$$\int_0^\infty \frac{(x-u)_+^1 (y-u)_+^1}{(1!)^2} du = \frac{(x \wedge y)^2 (x \vee y)}{2} - \frac{(x \wedge y)^3}{6},$$
$$\int_0^\infty \frac{(x-u)_+^2 (y-u)_+^2}{(2!)^2} du = \frac{(x \wedge y)^3 (x \vee y)^2}{12} - \frac{(x \wedge y)^4 (x \vee y)}{24} + \frac{(x \wedge y)^5}{120},$$

respectively. The random polynomial given by the first term on the right hand side of Equation (5) is the linear combination of standard normal random variables with coefficients that are monomials of degree at most m. By specifying m, we are able to control the differentiability of the GIBM.

3.2 Multi-Dimensional Generalized Integrated Brownian Fields

In the multi-dimensional case, consider d-dimensional Brownian field $\mathbf{B}(\cdot; \boldsymbol{\theta})$ on $[0, 1]^d$ with volatility $\boldsymbol{\theta}$ [Holden et al., 2010], where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ is a vector of parameters. Here $\mathbf{B}(\cdot; \boldsymbol{\theta})$ is the tensor product of d independent copies of one-dimensional Brownian motions with varying volatilities. This field is a real-valued, mean-zero GRF with continuous, non-differentiable sample paths. The covariance between $\mathbf{B}(\cdot; \boldsymbol{\theta})$ at $\mathbf{x}, \mathbf{y} \in [0, 1]^d$ is given by $\Sigma_{\mathbf{B}}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = \prod_{i=1}^d \theta_i \min\{x_i, y_i\}$. Similar to the one-dimensional case, we can integrate Brownian field with volatility $\boldsymbol{\theta}$ over each coordinate to get a differentiable process. In the multi-dimensional case, each coordinate can be integrated a different number of times. If we integrate m_i times in the ith coordinate for $i = 1, 2, \ldots, d$, the resulting GRF is called $\mathbf{m}$-integrated Brownian field $\mathbf{B}_\mathbf{m}(\cdot; \boldsymbol{\theta})$ with volatility $\boldsymbol{\theta}$ [Fill and Torcaso, 2004], where $\mathbf{m} = (m_1, m_2 \ldots, m_d)^\top$. Using integration by parts, $\mathbf{B}_\mathbf{m}(\mathbf{x}; \boldsymbol{\theta})$ can be expressed as a multiple integral with respect to Brownian field with volatility $\boldsymbol{\theta}$,

$$\mathbf{B}_\mathbf{m}(\mathbf{x}; \boldsymbol{\theta}) \triangleq \int_0^{x_1} \cdots \int_0^{x_d} \prod_{i=1}^d \frac{(x_i - u_i)^{m_i}}{m_i!} d\mathbf{B}(\mathbf{u}; \boldsymbol{\theta}).$$

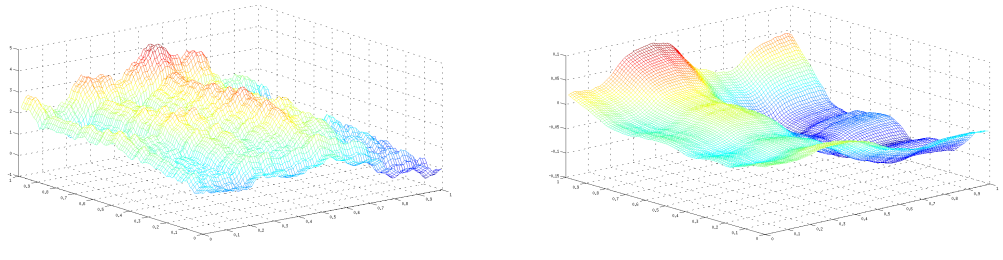
It follows immediately from this representation that the covariance between $\mathbf{B}_\mathbf{m}(\cdot; \boldsymbol{\theta})$ at $\mathbf{x}, \mathbf{y} \in [0, 1]^d$ is given by

$$\Sigma_{\mathbf{B}_\mathbf{m}}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = \prod_{i=1}^d \theta_i \int_0^\infty \frac{(x_i - u_i)_+^{m_i} (y_i - u_i)_+^{m_i}}{(m_i!)^2} du_i.$$

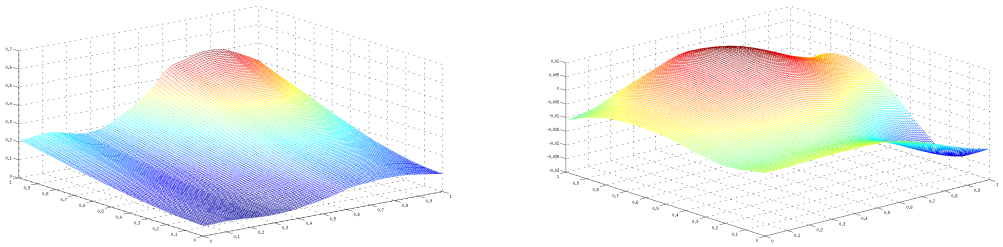
The covariance function $\Sigma_{\mathbf{B}_\mathbf{m}}(\cdot, \cdot; \boldsymbol{\theta})$ is the product of the covariance functions of the one-dimensional integrated Brownian motions $B_{m_1}(\cdot; \theta_1), B_{m_2}(\cdot; \theta_2), \ldots, B_{m_d}(\cdot; \theta_d)$. Similar to the one-dimensional case, $\mathbf{m}$-integrated Brownian field with volatility $\boldsymbol{\theta}$ has boundary conditions $\mathbf{B}_\mathbf{m}(\mathbf{x}; \boldsymbol{\theta}) = 0$ and $\mathbf{B}_\mathbf{m}^{|\mathbf{n}|}(\mathbf{x}; \boldsymbol{\theta}) = 0$ for all $\mathbf{n}$ such that $|\mathbf{n}| \leq |\mathbf{m}|$ and $\mathbf{x} \in [0, 1]^d$ such that $x_i = 0$ for some i. We define a new process whose covariance function is the product of covariance functions of d GIBMs, in the same way that the covariance function of $\mathbf{B}_\mathbf{m}(\cdot; \boldsymbol{\theta})$ is the product of the covariance functions of $B_{m_1}(\cdot; \theta_1), B_{m_2}(\cdot; \theta_2), \ldots, B_{m_d}(\cdot; \theta_d)$.

Definition 3.1. *The mean-zero Gaussian random field $\mathbf{X}_\mathbf{m}(\cdot; \boldsymbol{\theta})$ on $[0, 1]^d$ whose covariance at $\mathbf{x}, \mathbf{y} \in [0, 1]^d$ is given by*

$$\Sigma_{\mathbf{X}_\mathbf{m}}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = \prod_{i=1}^d \left(\sum_{k=0}^{m_i} \theta_{i,k} \frac{x_i^k y_i^k}{(k!)^2} + \theta_{i,m_i+1} \int_0^\infty \frac{(x_i - u_i)_+^{m_i} (y_i - u_i)_+^{m_i}}{(m_i!)^2} du_i \right), \tag{6}$$



(a) $(0,0)$-GIBF (b) $(1,1)$-GIBF



(c) $(1,2)$-GIBF (d) $(2,2)$-GIBF

Figure 3: Sample paths of GIBFs on the unit square.

where $\boldsymbol{\theta} = (\theta_{1,0}, \ldots, \theta_{1,m_1+1}, \theta_{2,0}, \ldots, \theta_{d,m_d+1})^\top$ *is a vector of parameters, is called a generalized* $\mathbf{m}$*-integrated Brownian field (*$\mathbf{m}$*-GIBF).*

Figure 3 shows sample paths of a $(0,0)$-GIBF, a $(1,1)$-GIBF, a $(1,2)$-GIBF, and a $(2,2)$-GIBF on the unit square $[0,1]^2$. We have defined $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ in terms of its covariance function, which is all that is required to fully define a mean-zero GRF. To get a better understanding of an $\mathbf{m}$-GIBF, an equivalent formulation of $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ at $\mathbf{x}$ is given by

$$\mathbf{X}_\mathbf{m}(\mathbf{x};\boldsymbol{\theta}) \triangleq \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{m}} C_\mathbf{n}(\boldsymbol{\theta})\mathbf{x}^\mathbf{n} Z_\mathbf{n} + \sum_{\substack{1\le i\le d \\ 1\le j_1<j_2<\ldots<j_i\le d}} \sum_{\substack{0\le k_j\le m_j \\ j\neq j_l,\forall l}} C_{\mathbf{j}_i,\mathbf{k}_{d-i}}(\mathbf{x},\boldsymbol{\theta})\mathbf{B}^{\mathbf{j}_i,\mathbf{k}_{d-i}}_{(m_{j_1},\ldots,m_{j_i})}(x_{j_1},\ldots,x_{j_i};\mathbf{1}), \tag{7}$$

where $\mathbf{j}_i = \{j_1, j_2, \ldots, j_i\}$, $\mathbf{k}_{d-i} = \{k_j : j \neq j_l, \forall l\}$, and the multi-dimensional sum is over all $\mathbf{n} = (n_1, n_2, \ldots, n_d)$ such that $\mathbf{0} \le \mathbf{n} \le \mathbf{m}$. The functions $C_\mathbf{n}(\cdot)$ and $C_{\mathbf{j}_i,\mathbf{k}_{d-i}}(\cdot,\cdot)$ are deterministic functions of $\mathbf{x}$ and $\boldsymbol{\theta}$, and although closed-form expressions can be obtained for each, they are not needed for implementation and do not add any insight into the process. Equation (7) is the multi-dimensional analog of Equation (5). The first term in Equation (7) is a random polynomial of degree $\mathbf{m}$, which is the linear combination of standard normal random variables with coefficients that are monomials of degree at most $\mathbf{m}$. The second term is the sum of integrated Brownian fields over every i-face of $[0,1]^d$, for $i = 1, 2, \ldots, d$. In other words, we sum integrated Brownian fields over each edge, face, cell, 4-face, 5-face, etc. of $[0,1]^d$. Since the functions $C_\mathbf{n}(\cdot)$ and $C_{\mathbf{j}_i,\mathbf{k}_{d-i}}(\cdot,\cdot)$ are deterministic functions of $\mathbf{x}$ and $\boldsymbol{\theta}$, the randomness in $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ is due to the standard normal random variables $Z_\mathbf{n}$ and the integrated Brownian fields $\mathbf{B}^{\mathbf{j}_i,\mathbf{k}_{d-i}}_{(m_{j_1},\ldots,m_{j_i})}(\cdot,\mathbf{1})$, which are all independent from each other. From the formulation of $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ given by Equation (7), it is clear that $\mathbf{m}$-GIBF does not have any boundary conditions. Furthermore, we are able to control the differentiability in each coordinate by specifying each entry of the vector $\mathbf{m} = (m_1, m_2, \ldots, m_d)$.

Remark: The covariance function of $\mathbf{X}_{\mathbf{m}}(\cdot;\boldsymbol{\theta})$ on $[0,1]^d$ is the reproducing kernel of the tensor product Hilbert space $\mathcal{H} = \bigotimes_{i=1}^{d} H^{m_i+1}[0,1]$, where

$$H^{m_i+1}[0,1] = \{\phi : \phi, \phi^{(1)}, \ldots, \phi^{(m_i)} \text{absolutely continuous}, \ \phi^{(m_i+1)} \in \mathbb{L}^2[0,1]\},$$

endowed with the inner product

$$\langle \phi_1, \phi_2 \rangle = \sum_{k=0}^{m_i} \frac{1}{\theta_k} \phi_1^{(k)}(0)\phi_2^{(k)}(0) + \frac{1}{\theta_{i,m_i+1}} \int_0^1 \phi_1^{(m_i+1)} \phi_2^{(m_i+1)} d\mu \tag{8}$$

for $\phi_1, \phi_2 \in H^{m_i+1}[0,1]$. Therefore, the same results can be obtained by considering multi-dimensional tensor product smoothing splines on the space $\mathcal{H}$ endowed with the inner product with our parameterization, given by Equation (8).

4 Stochastic Kriging with Generalized Integrated Brownian Fields

For stochastic kriging with **m**-GIBF, the response surface $y(\cdot)$ at $\mathbf{x}$ is modeled as a realization of the random variable

$$\mathsf{Y}_{\mathbf{X}_{\mathbf{m}}}(\mathbf{x};\boldsymbol{\theta}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + \widetilde{\mathbf{X}}_{\mathbf{m}}(\mathbf{x};\boldsymbol{\theta}), \tag{9}$$

where $\mathbf{f}(\cdot)$ and $\boldsymbol{\beta}$ are as before, and $\widetilde{\mathbf{X}}_{\mathbf{m}}(\cdot;\boldsymbol{\theta})$ is a modified version of **m**-GIBF, discussed in Section 4.1, which accounts for the basis functions in $\mathbf{f}(\cdot)$. To implement stochastic kriging with GIBFs, we need to choose the vector of basis functions $\mathbf{f}(\cdot)$ to be used for trend modeling and values for the parameters $\mathbf{m}$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$. This section discusses these aspects of fitting GIBFs, including trend modeling in Section 4.1, followed by maximum likelihood estimation of the parameters in Section 4.2, assuming that the vector of basis functions has been fixed. The properties of metamodels built using **m**-GIBF are given in Section 5.

4.1 Trend Modeling

To maintain the differentiability of the metamodel, we assume that each basis function in the $p \times 1$ vector of basis functions $\mathbf{f}(\cdot)$ is m_i times continuously differentiable in the ith coordinate. Any function can be a basis function as long as it satisfies this differentiability condition.

For certain basis functions, the covariance function needs to be modified. For stochastic kriging with **m**-GIBF, when a basis function is the monomial $\mathbf{x}^{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)^\top$ and $\alpha_i \leq m_i$ for $i = 1, 2, \ldots, d$, we need to subtract $\prod_{i=1}^{d} \theta_{i,\alpha_i} x_i^{\alpha_i} y_i^{\alpha_i}/(\alpha_i!)^2$ from the covariance function given by Equation (6). The need for this modification of the covariance function is the following. For stochastic kriging with the GRF $\mathsf{Y}_{\mathsf{M}}(\cdot)$, given by Equation (1), the difference $y(\cdot) - \mathbf{f}(\cdot)\boldsymbol{\beta}$ is modeled as a sample path of the mean-zero GRF $\mathsf{M}(\cdot)$. When $\mathbf{x}^{\boldsymbol{\alpha}}$ is included in $\mathbf{f}(\cdot)$, the variability of the simulation output $\bar{\mathcal{Y}}$ associated with the subspace spanned by $\mathbf{x}^{\boldsymbol{\alpha}}$ is eliminated by taking the difference $\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}}$. To avoid redundancy when we use an **m**-GIBF as the mean-zero GRF $\mathsf{M}(\cdot)$, we remove the term $C_{\boldsymbol{\alpha}}(\boldsymbol{\theta})\mathbf{x}^{\boldsymbol{\alpha}} Z_{\boldsymbol{\alpha}}$ in the random polynomial in Equation (7). This term is the GRF whose covariance at $\mathbf{x}, \mathbf{y} \in [0,1]^d$ is given by $\prod_{i=1}^{d} \theta_{i,\alpha_i} x_i^{\alpha_i} y_i^{\alpha_i}/(\alpha_i!)^2$.

Another explanation for the modification of the covariance function can be given in terms of boundary conditions. The formulation of **m**-GIBF given by Equation (7) is the sum of a term involving an **m**-integrated Brownian field and other terms that compensate for its boundary conditions. When the basis function $\mathbf{x}^{\boldsymbol{\alpha}}$ is included in $\mathbf{f}(\cdot)$, we do not need the term $C_{\boldsymbol{\alpha}}(\boldsymbol{\theta})\mathbf{x}^{\boldsymbol{\alpha}} Z_{\boldsymbol{\alpha}}$ in the random polynomial $\sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{m}} C_{\mathbf{n}}(\boldsymbol{\theta})\mathbf{x}^{\mathbf{n}} Z_{\mathbf{n}}$, since the corresponding boundary condition is compensated

by the term involving $\mathbf{x}^{\boldsymbol{\alpha}}$ in the trend function. For example, consider the GRF $\mathsf{Y}_{\mathbf{X_m}}(\cdot;\boldsymbol{\theta})$ with $\mathbf{f}(\cdot) = (1)^\top$, whose value at $\mathbf{x}$ is given by

$$\mathsf{Y}_{\mathbf{X_m}}(\mathbf{x};\boldsymbol{\theta}) = \beta_0 + \widetilde{\mathbf{X}}_{\mathbf{m}}(\mathbf{x};\boldsymbol{\theta}),$$

where $\widetilde{\mathbf{X}}_{\mathbf{m}}(\mathbf{x};\boldsymbol{\theta}) = \mathbf{X}_{\mathbf{m}}(\mathbf{x};\boldsymbol{\theta}) - C_{\mathbf{0}}(\boldsymbol{\theta})Z_{\mathbf{0}}$. Although $\widetilde{\mathbf{X}}_{\mathbf{m}}(\cdot;\boldsymbol{\theta})$ has the boundary condition $\widetilde{\mathbf{X}}_{\mathbf{m}}(\mathbf{0};\boldsymbol{\theta}) = 0$, $\mathsf{Y}_{\mathbf{X_m}}(\cdot)$ has no boundary conditions since the constant trend compensates for the boundary condition of $\widetilde{\mathbf{X}}_{\mathbf{m}}(\cdot;\boldsymbol{\theta})$ at the origin, i.e., $\mathsf{Y}_{\mathbf{X_m}}(\mathbf{0};\boldsymbol{\theta}) = \beta_0$. In general, we define the GRF $\widetilde{\mathbf{X}}_{\mathbf{m}}(\cdot;\boldsymbol{\theta})$ to be the mean-zero GRF whose covariance function is the covariance function of the $\mathbf{m}$-GIBF with the proper terms subtracted. We denote the covariance function of $\widetilde{\mathbf{X}}_{\mathbf{m}}(\cdot;\boldsymbol{\theta})$ by $\Sigma_{\widetilde{\mathbf{X}}_{\mathbf{m}}}(\cdot,\cdot;\boldsymbol{\theta})$.

4.2 Parameter Estimation

Assuming that the vector of basis functions has been fixed, let $\widehat{\mathbf{m}}, \widehat{\boldsymbol{\beta}}$, and $\widehat{\boldsymbol{\theta}}$ denote the maximum likelihood estimators (MLEs) for $\mathbf{m}, \boldsymbol{\beta}$, and $\boldsymbol{\theta}$, respectively. Finding the MLEs involves solving an optimization problem with both continuous and integer decision variables, since $\mathbf{m}$ must be a vector of integers. Here we describe parameter estimation when the simulation output does not include derivative estimates. When derivative estimates are used in stochastic kriging, parameter estimation is exactly the same as below with the proper matrices substituted [Chen et al., 2013]. Given fixed values for $\mathbf{m}$ and $\boldsymbol{\theta}$, the MLE of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}(\mathbf{m},\boldsymbol{\theta}) = \left(\mathbf{F}^\top \Sigma(\mathbf{m},\boldsymbol{\theta})^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^\top\Sigma(\mathbf{m},\boldsymbol{\theta})^{-1}\bar{\mathcal{Y}},$$

where $\Sigma(\mathbf{m},\boldsymbol{\theta}) = \Sigma_{\widetilde{X}_{\mathbf{m}}}(\boldsymbol{\theta}) + \widehat{\Sigma}_\epsilon$, $\Sigma_{\widetilde{X}_{\mathbf{m}}}(\boldsymbol{\theta})$ is the $k \times k$ variance-covariance matrix with ijth entry $\Sigma_{\widetilde{\mathbf{X}}_{\mathbf{m}}}(\mathbf{x}_i,\mathbf{x}_j;\boldsymbol{\theta})$, and $\widehat{\boldsymbol{\beta}}$ and Σ have been written as functions of $\mathbf{m}$ and $\boldsymbol{\theta}$ to explicitly show dependence. If we profile over the MLE of $\boldsymbol{\beta}$ and ignore constants, then the profile log-likelihood function [Shao, 2010] is given by

$$\mathcal{L}(\mathbf{m},\boldsymbol{\theta}|\bar{\mathcal{Y}}) = -\frac{1}{2}\log(|\Sigma(\mathbf{m},\boldsymbol{\theta})|) - \frac{1}{2}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}}(\mathbf{m},\boldsymbol{\theta}))^\top\Sigma(\mathbf{m},\boldsymbol{\theta})^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}}(\mathbf{m},\boldsymbol{\theta})),$$

where $\bar{\mathcal{Y}}$ is the vector of simulation output. The MLEs of $\mathbf{m}$ and $\boldsymbol{\theta}$ are given by

$$(\widehat{\mathbf{m}},\widehat{\boldsymbol{\theta}}) = \arg\min_{\mathbf{m},\boldsymbol{\theta}}\left\{-\mathcal{L}(\mathbf{m},\boldsymbol{\theta}|\bar{\mathcal{Y}})\,\middle|\,\mathbf{m}\in\mathbb{Z}_+^d, \boldsymbol{\theta}\in\mathbb{R}_+^M\right\},$$

where $M = \sum_{i=1}^d (m_i + 2)$, and $\mathbb{Z}_+^d$ and $\mathbb{R}_+^M$ are the feasible values for $\mathbf{m}$ and $\boldsymbol{\theta}$, respectively. To solve this optimization problem, we first fix $\mathbf{m}$ to some value $\mathbf{m}' \in \mathbb{Z}_+^d$ and search for the MLE of $\boldsymbol{\theta} \in \mathbb{R}_+^M$ given the fixed value $\mathbf{m}'$ of the order. Then, we choose a different value of $\mathbf{m} \neq \mathbf{m}'$ in $\mathbb{Z}_+^d$ and repeat the process until we are satisfied with our solution, i.e., we do not exhaust the search space $\mathbb{Z}_+^d$ of $\mathbf{m}$.

Instead of searching over the unbounded space $\mathbb{Z}_+^d$ for the MLE of $\mathbf{m}$, we limit our search to the bounded set $\{1,2\}^d$, which has 2^d elements. We only search the bounded set $\{1,2\}^d$ for the MLE of $\mathbf{m}$, since we have found in our practical experience with metamodeling of engineering simulations that it is sufficient to only consider GIBFs that are at least once-differentiable in each coordinate and at most twice-differentiable in each coordinate, i.e., these GIBFs are flexible enough for most response surfaces. When the order of the GIBF is increased in a coordinate, the computational cost of finding the MLEs of the parameters increases since the number of parameters increases. When m_i is at most two in each coordinate, the number of parameters is manageable.

For a fixed value $\mathbf{m}'$ of $\mathbf{m}$, instead of searching over the unbounded space $\mathbb{R}_+^M$ for the MLE of $\boldsymbol{\theta}$, we add a dummy parameter τ which allows $\boldsymbol{\theta}$ to be restricted to the unit hypercube $[0,1]^M$. In other words, only the magnitudes of the parameters in $\boldsymbol{\theta}$ relative to each other are important since the actual magnitude is absorbed in τ. The re-parameterized covariance function for $\mathbf{m}'$-GIBF is

$$\Sigma_{\mathbf{X}_{\mathbf{m}'}}(\mathbf{x},\mathbf{y};\boldsymbol{\theta},\tau) = \tau \prod_{i=1}^{d} \left(\sum_{k=0}^{m_i'} \theta_{i,k} \frac{x_i^k y_i^k}{(k!)^2} + \theta_{i,m_i'+1} \int_0^1 \frac{(x_i-u_i)_+^{m_i'}(y_i-u_i)_+^{m_i'}}{(m_i'!)^2} du_i \right),$$

where now $\boldsymbol{\theta} \in [0,1]^M$ and $\tau \geq 0$. Given $\mathbf{m}'$, the MLE of $\boldsymbol{\theta}$ is now given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ -\mathcal{L}(\mathbf{m}', \boldsymbol{\theta}, \tau^*(\mathbf{m}', \boldsymbol{\theta}) | \bar{\mathcal{Y}}) \,\middle|\, \boldsymbol{\theta} \in [0,1]^M \right\}, \tag{10}$$

where

$$\mathcal{L}(\mathbf{m}, \boldsymbol{\theta}, \tau | \bar{\mathcal{Y}}) = -\frac{1}{2}\log(|\Sigma(\mathbf{m}, \boldsymbol{\theta}, \tau)|) - \frac{1}{2}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}}(\mathbf{m}, \boldsymbol{\theta}, \tau))^\top \Sigma(\mathbf{m}, \boldsymbol{\theta}, \tau)^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}}(\mathbf{m}, \boldsymbol{\theta}, \tau))$$

is the re-parameterized profile log-likelihood function, $\widehat{\boldsymbol{\beta}}$ and Σ have been written as functions of $\mathbf{m}, \boldsymbol{\theta}$, and τ to explicitly show dependence, and $\tau^*(\mathbf{m}, \boldsymbol{\theta})$ is the value of τ that minimizes $\mathcal{L}(\mathbf{m}, \boldsymbol{\theta}, \tau | \bar{\mathcal{Y}})$ with $\mathbf{m}$ and $\boldsymbol{\theta}$ fixed. Finding $\tau^*(\mathbf{m}, \boldsymbol{\theta})$ can be done efficiently using a line search method, and supplying the solver with the gradient $\partial\mathcal{L}(\mathbf{m}, \boldsymbol{\theta}, \tau | \bar{\mathcal{Y}})/\partial\tau$, which can be easily computed using matrix calculus. We can now solve the constrained optimization problem (10) by evaluating $-\mathcal{L}(\mathbf{m}', \boldsymbol{\theta}, \tau^*(\boldsymbol{\theta}) | \bar{\mathcal{Y}})$ at a low-discrepancy point-set in $[0,1]^M$ and use the point that minimizes this quanitity as the starting solution for a non-linear optimization algorithm.

5 Properties

In this section, we discuss several properties of GIBFs and metamodels built using stochastic kriging with GIBFs. We focus particularly on the differentiability of the metamodel, the non-stationarity and lack of mean reversion of GIBFs, as well as the Markov property of GIBFs.

To analyze the differentiabilty of metamodels built using stochastic kriging with GIBFs, we rewrite the stochastic kriging predictor as the affine combination of the k basis functions $\Sigma_{\widetilde{\mathbf{X}}_{\widehat{\mathbf{m}}}}(\cdot, \mathbf{x}_i; \widehat{\boldsymbol{\theta}})$, for $i = 1, 2, \ldots, k$. Indeed, we rearrange Equation (2) for $\mathbf{x} \in \mathcal{X}$ as

$$\widehat{\mathrm{Y}}(\mathbf{x}) = \sum_{i=1}^{p} f_i(\mathbf{x})\widehat{\beta}_i + \sum_{i=1}^{k} c_i \Sigma_{\widetilde{\mathbf{X}}_{\widehat{\mathbf{m}}}}(\mathbf{x}, \mathbf{x}_i; \widehat{\boldsymbol{\theta}}),$$

where $\mathbf{c} = \Sigma(\widehat{\mathbf{m}}, \widehat{\boldsymbol{\theta}})^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}})$. Using this formulation of the stochastic kriging predictor, we can see that $f_i(\cdot)$ and $\Sigma_{\widetilde{\mathbf{X}}_{\widehat{\mathbf{m}}}}(\cdot, \mathbf{x}_i; \widehat{\boldsymbol{\theta}})$ are the only terms that depend on $\mathbf{x}$ in this expression, so the differentiability of the metamodel is determined by $\mathbf{f}(\cdot)$ and $\Sigma_{\widetilde{\mathbf{X}}_{\widehat{\mathbf{m}}}}(\cdot, \cdot; \widehat{\boldsymbol{\theta}})$. Since we are assuming each function in the trend vector is $\widehat{m}_i$ times differentiable in the ith coordinate and the covariance function $\Sigma_{\widetilde{\mathbf{X}}_{\widehat{\mathbf{m}}}}(\cdot, \cdot; \widehat{\boldsymbol{\theta}})$ is $\widehat{m}_i$ times differentiable in the ith coordinate, the metamodel is $\widehat{m}_i$ times differentiable in the ith coordinate.

In contrast to the stationary GRFs corresponding to the Gaussian and Matérn covariance functions, GIBFs are non-stationary. However, we are not concerned with the non-stationarity of GIBFs since we only use the conditional distribution, given the simulation output at the design points. The conditional distribution for any GRF, stationary or non-stationary, is always non-stationary; as we move away from the design points, the conditional variance increases. Consider

kriging with Brownian motion: although Brownian motion is non-stationary (as we move away from the origin, the variance of Brownian motion increases), when we condition on the simulation output at two design points, the resulting process between the design points is a Brownian bridge. The variance of the Brownian bridge will be largest in the center of the design points, and decrease as we get closer to either design point. This property is exactly what we want for metamodeling.

5.1 No Mean Reversion

As mentioned in the introduction, a well-known problem of Gaussian process models is the presence of mean reversion. The concept of mean reversion is well-defined for stochastic processes that are parameterized on the time domain: the process is mean-reverting if it tends to drift towards its long-term mean over time. A well-known example of a mean-reverting stochastic process is the Ornstein-Uhlenbeck process. In contrast, the concept of mean reversion has not been defined in terms of random fields parameterized on a multi-dimensional spatial domain. In this case, we no longer have a concept of time. We make the following definition of mean reversion for random fields parameterized on a multi-dimensional spatial domain:

Definition 5.1. *Let* $\mathsf{M}(\cdot)$ *be a random field with mean function* $m(\cdot)$ *defined on a convex cone* $\mathcal{C}$*. The random field* $\mathsf{M}(\cdot)$ *is mean-reverting if*

$$\mathrm{E}[\mathsf{M}(\lambda\mathbf{x}) - m(\lambda\mathbf{x})|\mathsf{M}(\tilde{\mathbf{x}}_1), \mathsf{M}(\tilde{\mathbf{x}}_2), \ldots, \mathsf{M}(\tilde{\mathbf{x}}_k)] \xrightarrow{d} 0 \tag{11}$$

as $\lambda \to \infty$*, for any* $k \geq 1$ *points* $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_k \in \mathcal{C}$ *and any* $\mathbf{x} \neq \vec{\mathbf{0}}$ *in* $\mathcal{C}$*.*

Essentially, we can think of the points $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_k$ as being the design points at which we are able to observe the value of the random field $\mathsf{M}(\cdot)$. $\mathrm{E}[\mathsf{M}(\lambda\mathbf{x})|\mathsf{M}(\tilde{\mathbf{x}}_1), \mathsf{M}(\tilde{\mathbf{x}}_2), \ldots, \mathsf{M}(\tilde{\mathbf{x}}_k)]$ is the kriging predictor at the point $\lambda\mathbf{x}$, based on the observations at the design points. The difference $\mathrm{E}[\mathsf{M}(\lambda\mathbf{x})|\mathsf{M}(\tilde{\mathbf{x}}_1), \mathsf{M}(\tilde{\mathbf{x}}_2), \ldots, \mathsf{M}(\tilde{\mathbf{x}}_k)] - m(\lambda\mathbf{x})$ is the difference between the kriging predictor and the unconditional mean. Thus, a random field is mean-reverting if the difference between the kriging predictor and the unconditional mean converges in distribution to zero as we move away from the design points. The next theorem shows that this definition of a mean-reverting random field is consistent with the behavior of GRFs that correspond to the Gaussian and Matérn covariance functions.

Theorem 5.2. *Let* $\mathsf{M}(\cdot)$ *be a GRF defined on* $\mathbb{R}^d_+$ *with mean function* $m(\cdot)$ *and covariance function* $\mathrm{Cov}[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{x}')] = \tau^2 r(\mathbf{x} - \mathbf{x}'; \boldsymbol{\theta})$*, for some scalar* τ *and function* $r(\cdot; \boldsymbol{\theta})$ *such that* $r(\mathbf{x}; \boldsymbol{\theta}) \to 0$ *as* $||\mathbf{x}|| \to \infty$ *and* $r(\mathbf{0}; \boldsymbol{\theta}) = 1$*. Then,* $\mathsf{M}(\cdot)$ *is mean-reverting.*

Proof. Let $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_k$ be any $k \geq 1$ points in $\mathbb{R}^d_+$ and $\mathbf{x}$ be any point in $\mathbb{R}^d_+$ not equal to the zero vector. Furthermore, let $r(\lambda\mathbf{x}, \cdot)$ be the column vector whose ith entry is $r(\lambda\mathbf{x} - \tilde{\mathbf{x}}_i; \boldsymbol{\theta})$ and $\mathbf{R}$ be the correlation matrix whose ijth entry is $r(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j; \boldsymbol{\theta})$. Using this notation, the conditional expectation in Equation (11) can be written as

$$\mathrm{E}[\mathsf{M}(\lambda\mathbf{x}) - m(\lambda\mathbf{x})|\mathsf{M}(\tilde{\mathbf{x}}_1), \mathsf{M}(\tilde{\mathbf{x}}_2), \ldots, \mathsf{M}(\tilde{\mathbf{x}}_k)] = r(\lambda\mathbf{x}, \cdot)^\top \mathbf{R}^{-1} \begin{pmatrix} \mathsf{M}(\tilde{\mathbf{x}}_1) - m(\tilde{\mathbf{x}}_1) \\ \vdots \\ \mathsf{M}(\tilde{\mathbf{x}}_k) - m(\tilde{\mathbf{x}}_k) \end{pmatrix}, \tag{12}$$

which is an affine combination of mean-zero Gaussian random variables, and thus also follows a Gaussian distribution

$$\mathrm{E}[\mathsf{M}(\lambda\mathbf{x}) - m(\lambda\mathbf{x})|\mathsf{M}(\tilde{\mathbf{x}}_1), \mathsf{M}(\tilde{\mathbf{x}}_2), \ldots, \mathsf{M}(\tilde{\mathbf{x}}_k)] \overset{d}{=} \mathcal{N}(0, \tau^2 r(\lambda\mathbf{x}, \cdot)^\top \mathbf{R}^{-1} r(\lambda\mathbf{x}, \cdot)). \tag{13}$$

Since $\mathsf{M}(\cdot)$ is a GRF, the variance τ^2 of $\mathsf{M}(\cdot)$ is finite. Thus, the variance of the conditional expectation converges to zero iff $r(\lambda\mathbf{x},\cdot)^\top\mathbf{R}^{-1}r(\lambda\mathbf{x},\cdot) \to 0$ as $\lambda \to \infty$. The vector $r(\lambda\mathbf{x},\cdot)$ converges to a vector of zeros as $\lambda \to \infty$ since $||\lambda\mathbf{x} - \tilde{\mathbf{x}}_i|| \to \infty$ for $i = 1, 2, \ldots, k$. Furthermore, the inverse of the correlation matrix $\mathbf{R}^{-1}$ is positive-definite, so the quadratic form defined by $f(\mathbf{x}) = \mathbf{x}^\top\mathbf{R}^{-1}\mathbf{x}$ is a continuous, strictly convex function with $f(\mathbf{0}) = 0$. Therefore, the variance of the conditional expectation $\tau^2 r(\lambda\mathbf{x},\cdot)^\top\mathbf{R}^{-1}r(\lambda\mathbf{x},\cdot)$ converges to zero as $\lambda \to \infty$, which implies that the conditional expectation converges in distribution to zero. □

The main property of the covariance function $\tau^2 r(\mathbf{x} - \mathbf{x}';\boldsymbol{\theta})$ on which the proof above relies is that $r(\mathbf{x};\boldsymbol{\theta})$ decays to zero as $||\mathbf{x}|| \to \infty$. As we will see in the next theorem, GIBFs do not exhibit mean reversion because their covariance functions do not have this property. In our definition of mean reversion, the domain of the random field must be a convex cone. However, we defined GIBFs on the unit hypercube $[0,1]^d$ so that the covariance function of $\mathbf{X}_\mathbf{m}(\cdot,\cdot;\boldsymbol{\theta})$ on $[0,1]^d$ is the reproducing kernel for the tensor product Hilbert space $\mathcal{H}$. Thus, for the following theorem, we consider GIBFs on the convex cone $\mathbb{R}^d_+$. Besides the domain of the random field, nothing in our definition of GIBFs changes. Furthermore, we need the notion of a non-trivial GIBF: at least one θ_i in the covariance function, given by Equation (6), is strictly positive.

Theorem 5.3. *Let $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ be a non-trivial $\mathbf{m}$-GIBF on $\mathbb{R}^d_+$. Then, $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ is not mean-reverting.*

Proof. Let $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_k$ be any $k \geq 1$ points in the interior of $\mathbb{R}^d_+$. Furthermore, let $\mathbf{x}$ be any point in the interior of $\mathbb{R}^d_+$. The conditional expectation in Equation (11) can be written as

$$\mathrm{E}[\mathbf{X}_\mathbf{m}(\lambda\mathbf{x};\boldsymbol{\theta})|\mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_1;\boldsymbol{\theta}), \mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_2;\boldsymbol{\theta}), \ldots, \mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_k;\boldsymbol{\theta})] = \boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)^\top\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_1;\boldsymbol{\theta})\\ \vdots \\ \mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_k;\boldsymbol{\theta})\end{pmatrix},$$

where $\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)$ is the column vector whose ith entry is $\mathrm{Cov}[\mathbf{X}_\mathbf{m}(\lambda\mathbf{x};\boldsymbol{\theta}), \mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_i;\boldsymbol{\theta})]$ and $\boldsymbol{\Sigma}$ is the covariance matrix whose ijth entry is $\mathrm{Cov}[\mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_i;\boldsymbol{\theta}), \mathbf{X}_\mathbf{m}(\tilde{\mathbf{x}}_j;\boldsymbol{\theta})]$. The conditional expectation is an affine combination of mean-zero Gaussian random variables, so its distribution is Gaussian with mean zero and variance $\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)$.

Since $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ is non-trivial and $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_k$ and $\mathbf{x}$ lie in the interior of $\mathbb{R}^d_+$, each entry in the vector $\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)$ is non-decreasing and at least one is strictly increasing in λ, so we have

$$\liminf_{\lambda\to\infty} ||\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)|| > \epsilon$$

for some $\epsilon > 0$. Using this fact, we will show that the variance of the conditional expectation does not converge to zero as $\lambda \to \infty$. Since $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_k$ lie in the interior of $\mathbb{R}^d_+$ and $\mathbf{X}_\mathbf{m}(\cdot;\boldsymbol{\theta})$ is non-trivial, the covariance matrix $\boldsymbol{\Sigma}$ is positive definite. Thus, the maximum eigenvalue $\lambda_{\max}$ of $\boldsymbol{\Sigma}$ is strictly positive. For the quadratic form defined by $f(\mathbf{x}) = \mathbf{x}^\top\boldsymbol{\Sigma}^{-1}\mathbf{x}$, a standard result for positive-definite matrices gives us the inequality

$$\frac{1}{\lambda_{\max}}\mathbf{x}^\top\mathbf{x} \leq \mathbf{x}^\top\boldsymbol{\Sigma}^{-1}\mathbf{x}. \tag{14}$$

Using (14) we have

$$\frac{1}{\lambda_{\max}}||\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)||^2 \leq \boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot),$$

and since $\liminf_{\lambda\to\infty} ||\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)|| > \epsilon$ as $\lambda \to \infty$,

$$\frac{\epsilon^2}{\lambda_{\max}} < \frac{1}{\lambda_{\max}} \liminf_{\lambda\to\infty} ||\boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)||^2 \leq \liminf_{\lambda\to\infty} \boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}(\lambda\mathbf{x},\cdot)$$

and the result follows. □

5.2 Markov Property

Consider a GRF $\mathsf{M}(\cdot)$ defined on a set $\mathcal{S} \subseteq \mathbb{R}^d$ with continuous derivatives $\mathsf{M}^{(\boldsymbol{\alpha})}(\cdot)$, for all $\boldsymbol{\alpha} \leq \mathbf{m}$, where $\mathsf{M}^{(\boldsymbol{\alpha})}(\mathbf{x}) = \partial^{|\boldsymbol{\alpha}|}\mathsf{M}(\mathbf{x})/\partial\mathbf{x}^{\boldsymbol{\alpha}}$. Roughly speaking, we call $\mathsf{M}(\cdot)$ a Markov random field if, for every bounded open set $\mathcal{O} \subseteq \mathcal{S}$,

$$\mathrm{E}[\mathsf{M}(\mathbf{x})|\mathsf{M}^{(\boldsymbol{\alpha})}(\mathbf{z}),\, \forall \mathbf{z} \in \mathcal{O}^{\complement}, \forall \boldsymbol{\alpha} \leq \mathbf{m}] = \mathrm{E}[\mathsf{M}(\mathbf{x})|\mathsf{M}^{(\boldsymbol{\alpha})}(\mathbf{z}),\, \forall \mathbf{z} \in \partial\mathcal{O}, \forall \boldsymbol{\alpha} \leq \mathbf{m}]$$

for any $\mathbf{x} \in \mathcal{O}$, where $\partial\mathcal{C}$ is the boundary of $\mathcal{C}$ [Pitt, 1971]. For a more precise, measure-theoretic definition, see Pitt [1971] or Künsch [1979]. The formal proof of the Markov property for a one-dimensional m-GIBM involves the reproducing kernel Hilbert space whose reproducing kernel is the covariance function of the m-GIBM. However, for an intuitive argument for why an m-GIBM is Markov, consider a time-based interpretation of the Markov property: given sufficient information (level and derivatives) at the present, the past and future are independent. The value of an m-GIBM at some time t can be written as its initial value plus the integral of an $(m-1)$-GIBM:

$$X_m(t;\boldsymbol{\theta}) = X_m(0;\boldsymbol{\theta}) + \int_0^t X_{m-1}(u;\boldsymbol{\theta}_{m-1})du,$$

where we have used the parameter t to denote time and $\boldsymbol{\theta}_{m-1}$ denotes the parameters of the $(m-1)$-GIBM that make the equation true. Let t_0 denote the present time. Using this integral representation of $X_m(t;\boldsymbol{\theta})$, we have

$$\begin{aligned}&\mathrm{E}[X_m(t;\boldsymbol{\theta})|X_m^{(v)}(s;\boldsymbol{\theta}), s \leq t_0, v \leq m] \\ &\qquad = X_m(t_0;\boldsymbol{\theta}) + \mathrm{E}\left[\int_{t_0}^t X_{m-1}(u;\boldsymbol{\theta}_{m-1})du \,\middle|\, X_m^{(v)}(s;\boldsymbol{\theta}), s \leq t_0, v \leq m\right].\end{aligned}$$

The information up to time t_0 is the sigma-algebra generated by $\{X_m^{(v)}(s;\boldsymbol{\theta}), s \leq t_0, v \leq m\}$ and is equal to the sigma-algebra generated by $\{X_v(s;\boldsymbol{\theta}_v), s \leq t_0, v \leq m\}$, since $X_m^{(v)}(s;\boldsymbol{\theta}) = X_{m-v}(s;\boldsymbol{\theta}_{m-v})$, where $\boldsymbol{\theta}_{m-v}$ are the parameters of the corresponding $(m-v)$-GIBM that make the equality true and $\boldsymbol{\theta}_m = \boldsymbol{\theta}$. Furthermore, the sigma-algebra generated by $\{X_v(s;\boldsymbol{\theta}_v), s \leq t_0, v \leq m\}$ is equal to the sigma-algebra generated by $\{X_m(0;\boldsymbol{\theta}), X_v(s;\boldsymbol{\theta}_v), s \leq t_0, v \leq m-1\}$, since $X_m(s;\boldsymbol{\theta})$ for $0 < s \leq t_0$ is completely determined by $X_m(0;\boldsymbol{\theta})$ and $X_v(s;\boldsymbol{\theta}_v)$, $s \leq t_0$, $v \leq m-1$. Thus,

$$\begin{aligned}&\mathrm{E}[X_m(t;\boldsymbol{\theta})|X_m^{(v)}(s;\boldsymbol{\theta}), s \leq t_0, v \leq m] \\ &\qquad = X_m(t_0;\boldsymbol{\theta}) + \mathrm{E}\left[\int_{t_0}^t X_{m-1}(u;\boldsymbol{\theta}_{m-1})du \,|X_m(0;\boldsymbol{\theta}), X_v(s;\boldsymbol{\theta}_v), s \leq t_0, v \leq m-1\right].\end{aligned}$$

Finally, since $X_m(0;\boldsymbol{\theta})$ and $X_{m-1}(s;\boldsymbol{\theta}_{m-1})$ are independent for $0 \leq s \leq t$, we have

$$\begin{aligned}&\mathrm{E}[X_m(t;\boldsymbol{\theta})|X_m^{(v)}(s;\boldsymbol{\theta}), s \leq t_0, v \leq m] \\ &\qquad = X_m(t_0;\boldsymbol{\theta}) + \mathrm{E}\left[\int_{t_0}^t X_{m-1}(u;\boldsymbol{\theta}_{m-1})du \,\middle|\, X_{m-1}^{(v)}(s;\boldsymbol{\theta}_{m-1}), s \leq t_0, v \leq m-1\right].\end{aligned}$$

Thus, $X_m(\cdot;\boldsymbol{\theta})$ will be Markov if $X_{m-1}(\cdot;\boldsymbol{\theta}_{m-1})$ is Markov. Using induction, we only need to show that $X_0(\cdot;\boldsymbol{\theta})$ is Markov. This can easily been shown since $X_0(\cdot;\boldsymbol{\theta})$ is a translation and scaling of Brownian motion, i.e., $X_0(x;\boldsymbol{\theta}) = \theta_2' B(x+\theta_1')$ for some $\theta_1', \theta_2' \geq 0$, which is a Markov process. We now give a formal proof of the Markov property for one-dimensional GIBMs.

Theorem 5.4. *The Gaussian stochastic process $X_m(\cdot;\boldsymbol{\theta})$ is Markov.*

Proof. We use Theorem 5.1 of Künsch [1979] and the fact that the covariance function of $X_m(\cdot;\boldsymbol{\theta})$ restricted to $[0,1]$ is the reproducing kernel of the reproducing kernel Hilbert space

$$H^{m+1}[0,1] = \{\phi : \phi, \phi^{(1)}, \ldots, \phi^{(m_i)} \text{absolutely continuous, } \phi^{(m_i+1)} \in \mathbb{L}^2[0,1]\},$$

endowed with the inner product

$$\langle \phi_1, \phi_2 \rangle = \sum_{k=0}^{m} \theta_k \phi_1^{(k)}(0)\phi_2^{(k)}(0) + \theta_{m+1} \int_0^1 \phi_1^{(m+1)} \phi_2^{(m+1)} d\mu.$$

for $\phi_1, \phi_2 \in H^{m+1}[0,1]$. By Theorem 5.1 of Künsch [1979], we only need to show two things: if $\phi_1, \phi_2 \in H^{m+1}[0,1]$ with $\operatorname{supp}\phi_1 \cap \operatorname{supp}\phi_2 = \varnothing$, then $\langle \phi_1, \phi_2 \rangle = 0$, and if $\phi = \phi_1 + \phi_2 \in H^{m+1}[0,1]$ with $\operatorname{supp}\phi_1 \cap \operatorname{supp}\phi_2 = \varnothing$, then ϕ_1 and $\phi_2 \in H^{m+1}[0,1]$.

We first show that if $\phi_1, \phi_2 \in H^{m+1}[0,1]$, with $\operatorname{supp}\phi_1 \cap \operatorname{supp}\phi_2 = \varnothing$, then $\langle \phi_1, \phi_2 \rangle = 0$. Indeed, let $\phi_1, \phi_2 \in H^{m+1}[0,1]$ be such that $\operatorname{supp}\phi_1 \cap \operatorname{supp}\phi_2 = \varnothing$. Since $\operatorname{supp}\phi_1 \cap \operatorname{supp}\phi_2 = \varnothing$, the interval $[0,\epsilon)$, for some $\epsilon > 0$, is contained in at most one of $\operatorname{supp}\phi_1$ or $\operatorname{supp}\phi_2$. Assume without loss of generality that $[0,\epsilon) \nsubseteq \operatorname{supp}\phi_1$. Then, $\phi_1^{(0)}(0), \phi_1^{(1)}(0), \ldots, \phi_1^{(m)}(0) = 0$. Thus, we have $\sum_{k=0}^{m} \theta_k \phi_1^{(k)}(0)\phi_2^{(k)}(0) = 0$. To finish showing that $\langle \phi_1, \phi_2 \rangle = 0$, we only need to show that the integral in the inner product is equal to zero. Since $\operatorname{supp}\phi_1^{(m+1)} \subseteq \operatorname{supp}\phi_1$ and $\operatorname{supp}\phi_2^{(m+1)} \subseteq \operatorname{supp}\phi_2$, we have that $\operatorname{supp}\phi_1^{(m+1)} \cap \operatorname{supp}\phi_2^{m+1} = \varnothing$. Thus, their product $\phi_1^{(m+1)}\phi_2^{(m+1)}$ is zero and

$$\int_0^1 \phi_1^{(m+1)} \phi_2^{(m+1)} d\mu = 0.$$

Together with $\sum_{k=0}^{m} \theta_k \phi_1^{(k)}(0)\phi_2^{(k)}(0) = 0$, we have $\langle \phi_1, \phi_2 \rangle = 0$.

Next, we show that if $\phi \in H^{(m+1)}[0,1]$ is decomposed as $\phi = \phi_1 + \phi_2$, with $\operatorname{supp}\phi_1 \cap \operatorname{supp}\phi_2 = \varnothing$, then $\phi_1, \phi_2 \in H^{(m+1)}$. We first show that $\phi_1^{(m+1)}$ and $\phi_2^{(m+1)} \in \mathbb{L}^2[0,1]$. Since $\phi \in H^{(m+1)}[0,1]$, we have that $\phi^{(m+1)} \in \mathbb{L}^2[0,1]$. Thus,

$$\int_0^1 \left(\phi^{(m+1)}\right)^2 d\mu < \infty.$$

The fact that $\phi_1^{(m+1)}$ and $\phi_2^{(m+1)}$ are also in $\mathbb{L}^2[0,1]$ follows directly from the inequality

$$\int_0^1 \left(\phi_i^{(m+1)}\right)^2 d\mu \leq \int_0^1 \left(\phi^{(m+1)}\right)^2 d\mu,$$

for $i = 1, 2$. We have left to show that if $\phi^{(i)}$, for $i = 1, 2, \ldots, m$, is absolutely continuous, then $\phi_1^{(i)}$ and $\phi_2^{(i)}$ are absolutely continuous. We only need to prove that $\phi_1^{(i)}$ is absolutely continuous, since the absolute continuity of $\phi_2^{(i)}$ follows from the fact that the difference of two absolutely continuous functions is absolutely continuous. Since ϕ_1 vanishes on $(\operatorname{supp}\phi_1)^{\complement}$ and ϕ_2 vanishes on $(\operatorname{supp}\phi_2)^{\complement}$, we have $\operatorname{supp}\phi_1^{(i)} \subseteq \operatorname{supp}\phi_1$ and $\operatorname{supp}\phi_2^{(i)} \subseteq \operatorname{supp}\phi_2$. Thus, $\operatorname{supp}\phi_1^{(i)} \cap \operatorname{supp}\phi_2^{(i)} = \varnothing$. Since $\phi^{(i)}$ is

absolutely continuous, for any $\epsilon > 0$ there exists a $\delta > 0$ such that whenever a finite set of disjoint intervals $\{[x_k, y_k]\}_k$ of $[0,1]$, with $x_k \le y_k$ for all k, satisfies

$$\sum_k (y_k - x_k) < \delta, \quad \text{then} \quad \sum_k |\phi^{(i)}(y_k) - \phi^{(i)}(x_k)| < \epsilon.$$

Each interval $[x_k, y_k]$ can be classified into one of the following three cases:

i both $\phi_1^{(i)}(y_k) = \phi_1^{(i)}(x_k) = 0$,

ii exactly one of $\phi_1^{(i)}(y_k)$ or $\phi_1^{(i)}(x_k)$ is 0 and unequal to the corresponding $\phi^{(i)}(y_k)$ or $\phi^{(i)}(x_k)$,

iii both $\phi_1^{(i)}(y_k) = \phi^{(i)}(y_k)$ and $\phi_1^{(i)}(x_k) = \phi^{(i)}(x_k)$.

For an interval $[x_k, y_k]$ satisfying case (i), we have $|\phi_1^{(i)}(y_k) - \phi_1^{(i)}(x_k)| = 0$. Similarly, for an interval $[x_k, y_k]$ satisfying case (iii), we have $|\phi_1^{(i)}(y_k) - \phi_1^{(i)}(x_k)| = |\phi^{(i)}(y_k) - \phi^{(i)}(x_k)|$.

For an interval $[x_k, y_k]$ satisfying case (ii), we can assume without loss of generality that $\phi_1^{(i)}(x_k) = 0$. Furthermore, we can assume that $\phi_1^{(i)}(y_k) \neq 0$, otherwise if $\phi_1^{(i)}(y_k) = 0$ we would have case (i). Since $\operatorname{supp} \phi_1^{(i)} \cap \operatorname{supp} \phi_2^{(i)} = \varnothing$, there exists a $z_k \in [x_k, y_k]$ such that $\phi^{(i)}(z_k) = 0$. Let Π be the set of indices for which case (ii) holds. Since $|z_k - x_k| + |y_k - z_k| \le |y_k - x_k|$ for any $k \in \Pi$, we have

$$\sum_{k \in \Pi^{\complement}} |y_k - x_k| + \sum_{k \in \Pi} (|z_k - x_k| + |y_k - z_k|) < \delta.$$

From the absolute continuity of $\phi^{(i)}$,

$$\sum_{k \in \Pi^{\complement}} |\phi^{(i)}(y_k) - \phi^{(i)}(x_k)| + \sum_{k \in \Pi} (|\phi^{(i)}(z_k) - \phi^{(i)}(x_k)| + |\phi^{(i)}(y_k) - \phi^{(i)}(z_k)|) < \epsilon.$$

Since $\phi^{(i)}(z_k) = 0$ and $\phi^{(i)}(y_k) = \phi_1^{(i)}(y_k)$ for $k \in \Pi$, we have

$$\sum_{k \in \Pi^{\complement}} |\phi^{(i)}(y_k) - \phi^{(i)}(x_k)| + \sum_{k \in \Pi} (|\phi^{(i)}(x_k)| + |\phi_1^{(i)}(y_k)|) < \epsilon. \tag{15}$$

Using (15), we have

$$\begin{aligned} \sum_k |\phi_1^{(i)}(y_k) - \phi_1^{(i)}(x_k)| &= \sum_{k \in \Pi^{\complement}} |\phi_1^{(i)}(y_k) - \phi_1^{(i)}(x_k)| + \sum_{k \in \Pi} |\phi_1^{(i)}(y_k)| \\ &\le \sum_{k \in \Pi^{\complement}} |\phi^{(i)}(y_k) - \phi^{(i)}(x_k)| + \sum_{k \in \Pi} (|\phi^{(i)}(x_k)| + |\phi_1^{(i)}(y_k)|) < \epsilon. \end{aligned}$$

Thus, ϕ_1 is absolutely continuous and the proof is complete. □

For the Markov property of multi-dimensional GIBFs, we make a conjecture.

Conjecture 5.1. *The Gaussian random field $\mathbf{X_m}(\cdot; \boldsymbol{\theta})$ is Markov.*

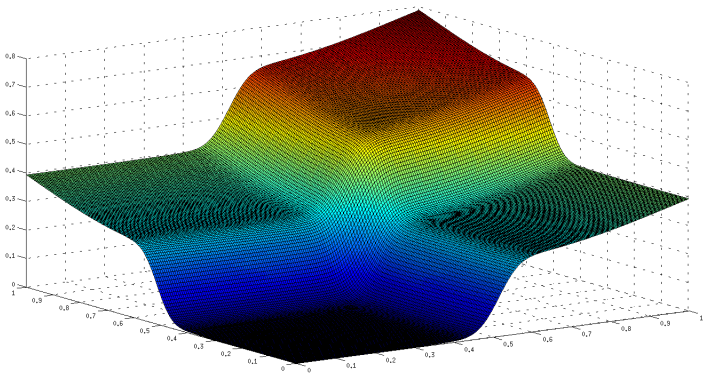


Figure 4: The credit risk response surface.

6 Numerical Experiments

The purpose of our experiments is to assess the prediction ability of stochastic kriging with GIBFs. We are mainly concerned with how different types of response surfaces and whether or not we incorporate derivative information effect our predictions. Furthermore, we will experiment with different levels of Monte Carlo noise, including experiments with no noise in the simulation output, i.e., we are able to observe the actual response surface at the design points. In our experiments, we compare stochastic kriging with GIBFs to stochastic kriging with the GRFs corresponding to the Gaussian and Matérn covariance functions. The GRFs corresponding to the Gaussian and Matérn covariance functions can result in metamodels with mean reversion. We wish to demonstrate the superiority of GIBFs to the GRFs corresponding to the Gaussian and Matérn covariance functions even when the latter do not cause mean reversion, so we use experiment designs with sufficiently many design points.

6.1 Credit Risk Example

In this example, the response surface is the expected loss of a credit portfolio, given values of latent variables that trigger the default of the obligors, normalized by the number of obligors [Glasserman et al., 2008]. Consider a credit portfolio with m obligors, and let Y_k be the default indicator ($= 1$ for default, $= 0$ otherwise) for the kth obligor, p_k be the marginal probability that the kth obligor defaults, and l_k be the deterministic loss resulting from default of the kth obligor. The dependence among the default indicators Y_k is modeled by a multifactor Gaussian copula model with a finite number of types. In other words, $Y_k = \mathbf{1}\{\tilde{Z}_k > \Phi^{-1}(1-p_k)\}$, where Φ is the cumulative normal distribution, and $\tilde{Z}_1, \tilde{Z}_2, \ldots$ are correlated standard normal random variables. To model the correlation of the standard normal random variables $\tilde{Z}_1, \tilde{Z}_2, \ldots$, we assume that there are d factors and t types of obligors. If obligor k is of type j, then the latent variable is given by $\tilde{Z}_k = \mathbf{a}_j^\top \mathbf{Z} + b_j \epsilon_k$, where $\mathbf{a}_j \in \mathbb{R}^d$, $\mathbf{Z}$ is a d-dimensional standard normal random vector that represents market risk factors which affect all of the obligors, $b_j = (1 - \mathbf{a}_j^\top \mathbf{a}_j)^{1/2}$, and the ϵ_k are independent standard normal random variables. The total loss from defaults is $L_m = \sum_{k=1}^m l_k Y_k$, which is a discrete random variable. The closed-form expression for the response surface is

$$y_{cr}(\mathbf{x}) = \frac{1}{m}\mathrm{E}[L_m|\mathbf{Z} = \mathbf{x}] = \frac{1}{m}\sum_{k=1}^{m} l_k \, \Phi\left(\frac{\mathbf{a}_j^\top \mathbf{z} + \Phi^{-1}(p_k)}{b_j}\right), \tag{16}$$

and is plotted in Figure 4. We use the closed-form expression to obtain noiseless observations of the response surface, as well as to determine the accuracy of the predictions. To obtain noisy observations (simulation output) of the response surface, we use the importance sampling method of Glasserman et al. [2008] to estimate the expected loss $\mathrm{E}[L_m|\mathbf{Z} = \mathbf{x}]$ of the credit portfolio. Specifically, for each replication $r = 1, 2, \ldots, R$ we use the importance sampling method to get an estimate $P^r(L_m > i|\mathbf{Z} = \mathbf{x})$ of $P(L_m > i|\mathbf{Z} = \mathbf{x})$, $i = 1, 2, \ldots, m-1$. Then, our estimate of $\mathrm{E}[L_m|\mathbf{Z} = \mathbf{x}]$ on the rth replication is

$$\sum_{i=0}^{m-1} P^r(L_m > i|\mathbf{Z} = \mathbf{x}),$$

and our estimate of the expected loss of the credit portfolio at design point $\mathbf{x}$ is

$$\frac{1}{R}\sum_{r=1}^{R}\sum_{i=0}^{m-1} P^r(L_m > i|\mathbf{Z} = \mathbf{x}).$$

Our estimate of the gradient at a design point is obtained using the method of finite-differences with CRN and the same number of replications R used to obtain the estimates of the response surface.

For our particular example, consider the case with two factors and four types of obligors: $\mathbf{a}_1^\top = (0.85, 0), \mathbf{a}_2^\top = (0.25, 0), \mathbf{a}_3^\top = (0, 0.85)$, and $\mathbf{a}_4^\top = (0, 0.25)$. Each type has three obligors, i.e., $m = 12$, with $l_k = 1$ and $p_k = 0.01$ for every obligor. The design space for this example is the square $[-5, 10]^2$. The design points are the first k points from a scrambled Sobol point-set rescaled to fit within the design space. The prediction points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{1023}$ are the first 1023 points from the Halton point-set rescaled to fit inside the square $[-3.5, 8.5]^2$, so the metamodels are assessed within the interior of the design space. We repeat each experiment 50 times to get 50 metamodels $\hat{y}_{cr}^1, \hat{y}_{cr}^2, \ldots, \hat{y}_{cr}^{50}$, and use the Root Empirical Mean Squared Error

$$\mathrm{REMSE} = \sqrt{\frac{1}{51150}\sum_{j=1}^{50}\sum_{i=1}^{1023}\left(\hat{y}_{cr}^j(\mathbf{p}_i) - y_{cr}(\mathbf{p}_i)\right)^2}$$

as our measure of prediction ability, where $\hat{y}_{cr}^j(\mathbf{p}_i)$ is the value of the jth simulation metamodel $\hat{y}_{cr}^j$ at $\mathbf{p}_i$, and $y_{cr}(\mathbf{p}_i)$ is the actual value of the response surface at $\mathbf{p}_i$. The actual value of the response surface is computed using the closed-form expression given above. We use a constant trend model for GIBFs and the GRFs corresponding to the Gaussian and Matérn covariance functions.

Experiment Results

The experiment results for the credit risk example are given in Tables 1 and 2 for varying amounts of design points and Monte Carlo noise. An interesting characteristic of the credit risk response surface y_{cr} occurs in regions of the design space where there is a change in the number of types of obligors that are likely to default. In these regions of the design space, there is an abrupt change in the response surface. Although the response surface remains twice differentiable in these areas, the first partial derivatives change very rapidly. We can see from the experiment results that using GRFs whose differentiability can be controlled, i.e., the GRF corresponding to the Matérn covariance function and GIBFs, resulted in better predictions than the GRF corresponding to the Gaussian covariance function, which is infinitely differentiable. When we use a GRF whose differentiability can be controlled, the order of differentiability can be adjusted to account for the

Table 1: REMSE for the credit risk example, $k = 31$

GMRF	no noise	$R = 100$	$R = 25$
$(1, 1)$-GIBF	0.019	0.021	0.024
$(2, 2)$-GIBF	0.031	0.032	0.035
Matérn	0.04	0.041	0.046
Gaussian	0.045	0.048	0.05
$(1, 1)$-GIBF-gradient	0.012	0.014	0.016
$(2, 2)$-GIBF-gradient	0.018	0.025	0.026
Matérn-gradient	0.024	0.027	0.031
Gaussian-gradient	0.033	0.036	0.041

Table 2: REMSE for the credit risk example, $k = 63$

GMRF	no noise	$R = 100$	$R = 25$
$(1, 1)$-GIBF	0.0033	0.0039	0.0042
$(2, 2)$-GIBF	0.0041	0.0046	0.0053
Matérn	0.016	0.019	0.023
Gaussian	0.032	0.041	0.046
$(1, 1)$-GIBF-gradient	0.002	0.0023	0.003
$(2, 2)$-GIBF-gradient	0.0023	0.0027	0.0034
Matérn-gradient	0.009	0.011	0.014
Gaussian-gradient	0.018	0.021	0.025

lack of smoothness of the response surface. In this example, the MLEs for the parameters of both the GRF corresponding to the Matérn covariance function and GIBFs resulted in GRFs with lower orders of differentiability. Between the GRF corresponding to the Matérn covariance function and GIBFs, the metamodels constructed using GIBFs resulted in better predictions. When gradient information was included, the GRFs with controllable differentiability still gave better predictions than the GRF corresponding to the Gaussian covariance function, and the metamodels constructed using GIBFs still outperformed the GRF corresponding to the Matérn covariance function.

Although Tables 1 and 2 give results for both $(1, 1)$-GIBF and $(2, 2)$-GIBF, the MLE for the order of GIBF in each experiment was $(1, 1)$. We also give the results for $(2, 2)$-GIBF to see how using a GIBF with a higher order of differentiability affects the performance on a problem where a lower order of differentiability is beneficial. Even though the credit risk response surface is twice differentiable, as is $(2, 2)$-GIBF, the once differentiable $(1, 1)$-GIBF was a better choice for this experiment design. Although $(2, 2)$-GIBF resulted in worse predictions than $(1, 1)$-GIBF, it still outperformed the GRFs corresponding to the Matérn and Gaussian covariance functions. Finding the MLEs for the parameters of $(1, 1)$-GIBF and $(2, 2)$-GIBF required more effort than the GRFs corresponding to the Gaussian and Matérn covariance functions, since $(1, 1)$-GIBF and $(2, 2)$-GIBF have more parameters, but was still fast and efficient.

6.2 Jackson Network Example

The response surface in the previous example had the same order of differentiability in each coordinate and was twice differentiable. Our next response surface has varying orders of differentiability in each coordinate and is non-differentiable in several regions of the design space.

In this example, we use a 3-station Jackson network with deterministic routing. In this network, 3 types of products arrive to the first station of a system of 3 single-server stations according to a Poisson Process with a total arrival rate of λ. Let α_1, α_2 and α_3 be the fraction of each product, with the obvious constraint $\alpha_1 + \alpha_2 + \alpha_3 = 1$, and let $\rho < 1$ be the utilization of the system, i.e., the utilization of the bottleneck station. The response surface is the expected steady-state cycle time of product 2, as a function of the fraction α_i of each product type and the utilization ρ. At the jth station, the service time for product i is exponentially distributed with rate μ_j, i.e., the service rate only depends on the station and not the product type. Furthermore, products of type i make δ_{ij} visits to the jth station. For given values of $\alpha_1, \alpha_2, \alpha_3$, and ρ, we adjust the arrival rate λ so that the utilization of the system is indeed equal to the given ρ. Thus, we keep the service rate at each station fixed, and only adjust the arrival rate λ to get the specified utilization ρ of the system.

The closed-form expression for the response surface is

$$y_{jn}(\mathbf{x}) = \sum_{j=1}^{3} \frac{\delta_{1j}}{\mu_j \left[1 - \rho \left(\frac{\sum_{k=1}^{3} \alpha_k \delta_{kj}/\mu_j}{\max_h \sum_{k=1}^{3} \alpha_k \delta_{kh}/\mu_h} \right) \right]}, \tag{17}$$

where $\mathbf{x} = (\alpha_1', \alpha_2', \rho) \in [0,1]^2 \times [0.7, 0.95]$. We use the method in Borkowski and Piepel [2009] to transform $(\alpha_1', \alpha_2', \rho)$ into $(\alpha_1, \alpha_2, \alpha_3, \rho)$ with $\alpha_1 + \alpha_2 + \alpha_3 = 1$, which we then use to calculate the right hand side of Equation (17). As in the credit risk example, we use the closed-form expression to obtain noiseless observations of the response surface, as well as to determine the accuracy of the predictions. To obtain noisy observations (simulation output) of the response surface, we simulate the Jackson network for a specified run-length with a varied number of replications R at each design point. To reduce the bias of the estimator for the expected steady-state cycle time, we delete the initial transient.

For our particular example, the service rate at each station is given by the vector $(\mu_1, \mu_2, \mu_3)^\top = (4, 3, 2.8)^\top$ and the number of visits by product i to station j is given by the matrix

$$[\delta]_{ij} = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

The design points are the first k points from a scrambled Sobol point-set rescaled to fit within the design space, and the prediction points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{1023}$ are the first 1023 points from the Halton point-set rescaled to fit inside the hyper-rectangle $[0.1, 0.9]^2 \times [0.725, 0.925]$, so the metamodels are assessed within the interior of the design space. As in the credit risk example, we repeat each experiment 50 times to get 50 metamodels and use the REMSE as our measure of prediction ability. The actual value of the response surface is computed using the closed-form expression given above. We use a constant trend model for GIBFs and the GRFs corresponding to the Gaussian and Matérn covariance functions.

Experiment Results

The experiment results for the Jackson network example are given in Tables 3 and 4 for varying amounts of design points and Monte Carlo noise. The response surface y_{jn} is differentiable in the coordinate corresponding to ρ. However, some areas of y_{jn} are nondifferentiable in the coordinates corresponding to α_1' and α_2'. These areas of nondifferentiability arise because, for a fixed utilization ρ, as we change the fraction of each product, the bottleneck station of the Jackson network may

Table 3: REMSE for the Jackson network example, $k = 100$

GMRF	no noise	$R = 100$	$R = 50$
$(1, 1, 1)$-GIBF	0.27	0.35	0.41
$(1, 1, 2)$-GIBF	0.26	0.32	0.37
Matérn	0.44	0.51	0.62
Gaussian	0.57	0.68	0.80

Table 4: REMSE for the Jackson network example, $k = 200$

GMRF	no noise	$R = 100$	$R = 50$
$(1, 1, 1)$-GIBF	0.19	0.25	0.32
$(1, 1, 2)$-GIBF	0.16	0.20	0.26
Matérn	0.31	0.39	0.53
Gaussian	0.34	0.44	0.60

switch, creating a sudden change in the behavior of the system. From the experiment results in Tables 3 and 4, we can see that using a GRF whose differentiability can be controlled, i.e., the GRF corresponding to the Matérn covariance function and GIBFs, resulted in better predictions than the infinitely differentiable GRF corresponding to the Gaussian covariance function. Furthermore, GIBFs gave better predictions than the GRF corresponding to the Matérn covariance function.

The MLE for the order of the GIBF differed for different experiments; the MLE was either $(1, 1, 1)$ or $(1, 1, 2)$. For example, the MLE for the order of the GIBF on the ith experiment for a certain noise level might have been $(1, 1, 1)$, whereas the MLE for the order of the GIBF on the jth experiment, $i \neq j$, for the same noise level might have been $(1, 1, 2)$. Thus, the results in Tables 3 and 4 are for fixed orders of GIBF, namely $(1, 1, 1)$-GIBF and $(1, 1, 2)$-GIBF, and are not chosen using maximum likelihood estimation. Overall, $(1, 1, 2)$-GIBF gave better predictions than $(1, 1, 1)$-GIBF. Similar to the credit risk example, finding the MLEs for the parameters of $(1, 1, 1)$-GIBF and $(1, 1, 2)$-GIBF required more effort than the GRFs corresponding to the Gaussian and Matérn covariance functions, but was still relatively fast and efficient.

7 Conclusion

In this paper, we introduced a new class of GRFs called GIBFs, focusing on their use with stochastic kriging for deterministic and stochastic simulation metamodeling. We gave a probabilistic representation of GIBFs and discussed several of their properties, including the Markov property and the absence of mean reversion, as well as the differentiability of the resulting metamodel. Using stochastic kriging, we showed how to implement GIBFs and used several examples to assess the prediction ability of stochastic kriging with GIBFs. These examples exhibited the benefit gained using stochastic kriging with GIBFs instead of the GRFs corresponding to the Gaussian and Matérn covariance functions. The examples also showed the improvement in performance when gradient estimates were included in the prediction. As the dimension increases, the number of parameters in our parameterization of GIBFs can quickly become difficult to handle. However, other parameterizations exist. Since the computational effort required to find the MLEs of the parameters of GIBFs of any order increases substantially as the dimension increases, other parameterizations, especially those with fewer parameters, should be investigated.

Acknowledgments

This article is based upon work supported by the National Science Foundation under Grant No. CMMI-0900354. Portions of this paper were published in Salemi et al. [2013].

References

B. Ankenman, B. L. Nelson, and J. Staum. Stochastic kriging for simulation metamodeling. *Operations Research*, 58(2):371–382, 2010.

A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, New York, NY, 2004.

J. J. Borkowski and G. F. Piepel. Uniform designs for highly constrained mixture experimentes. *Journal of Quality Technology*, 41(1):35–47, 2009.

X. Chen, B. E. Ankenman, and B. L. Nelson. The effects of common random numbers of stochastic kriging metamodels. *ACM Transactions of Modeling and Computer Simulation*, 2012.

X. Chen, B. Ankenman, and B. L. Nelson. Enhancing stochastic kriging metamodels with gradient estimators. *Operations Research*, 61(2):512–528, 2013.

J. A. Fill and F. Torcaso. Asymptotic analysis via Mellin transforms for small deviations in $l2$-norm of integrated Brownian sheets. *Probability Theory and Related Fields*, 130(2):259–288, 2004.

P. Glasserman, W. Kang, and P. Shahabuddin. Fast simulation of multifactor portfolio credit risk. *Operations Research*, 56(5):1200–1217, 2008.

Chong Gu and Grace Wahba. Smoothing spline anova with component-wise bayesian "confidence intervals". *Journal of Computational and Graphical Statistics*, 2:97–117, 1992.

H. Holden, B. Øksendal, J. Ubøe, and T. Zhang. *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach*. Springer Science and Business Media (Universitext), New York, NY, second edition, 2010.

V. R. Joseph. Limit kriging. *Technometrics*, 48(4):458–466, 2006.

V. R. Joseph, Y. Hung, and A. Sudjianto. Blind kriging. *Journal of Mechanical Design*, 130(3), 2008.

J. Kleijnen and W. van Beers. Robustness of kriging when interpolating in random simulation with heterogeneous variance: Some experiments. *European Journal of Operational Research*, 165:826–834, 2005.

D. G. Krige. A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis, University of Witwatersrand, 1951.

H. Künsch. Gaussian Markov random fields. *Journal of the Faculty of Science, University of Tokyo, Mathematics*, 26:53–73, 1979.

R. Li and A. Sudjianto. Analysis of computer experiments using penalized likelihood in Gaussian kriging models. *Technometrics*, 47(2):111–120, 2005.

M. Liu and J. Staum. Stochastic kriging for efficient nested simulation of expected shortfall. *Journal of Risk*, 12(3):3–27, 2010.

T. J. Mitchell and M. D. Morris. The spatial correlation function approach to response surface estimation. In *Proceedings of the 1992 Winter Simulation Conference*, pages 565–571, Piscataway, NJ, 1992. IEEE.

C. J. Paciorek and M. J. Schervish. Nonstationary covariance functions for Gaussian process regression. In *Proceedings of the Conference on Neural Information Processing Systems*. MIT Press, 2004.

L. D. Pitt. A Markov property for Gaussian processes with a multidimensional parameter. *Archive for Rational Mechanics and Analysis*, 43(5):367–391, 1971.

J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989.

P. Salemi, J. Staum, and B. L. Nelson. Generalized integrated Brownian fields for simulation metamodeling. In *Proceedings of the 2013 Winter Simulation Conference*, Piscataway, NJ, 2013. IEEE.

T. J. Santner, B. J. Williams, and W. I. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag New York, LLC, 2010.

J. Shao. *Mathematical Statistics*. Springer Texts in Statistics, New York, NY, 2010.

W. van Beers and J. Kleijnen. Kriging for interpolation in random simulation. *The Journal of the Operational Research Society*, 54(3):255–262, 2003.

G. Wahba. *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.

W. Xie, B. L. Nelson, and J. Staum. The influence of correlation functions on stochastic kriging metamodels. In *Proceedings of the 2010 Winter Simulation Conference*, Piscataway, NJ, 2010. IEEE.

J. Yin, S. H. Ng, and K. M. Ng. Kriging metamodel with modified nugget-effect: The heteroscedastic variance case. *Computers and Industrial Engineering*, 61(3):760–777, 2011.