Supplementary material to “Lifted Brownian kriging models”: More details on the specification of long range dependence between rays

This section can be considered an expanded version of section 3.2 in the paper “Lifted Brownian kriging models” which includes generalization of the results along with more discussion and examples.

We now show that all radial covariances, which are dominant in analysis of experiments on computer simulations, imply that the surface exhibits peaks and valleys, at least in the long range. To illustrate this, consider the difference along increments of our response as the input is changed,

$$S_1 = Z(\mathbf{h}_1) - Z(\mathbf{0})$$

and

$$S_2 = Z(\mathbf{h}_2 + \mathbf{h}_1) - Z(\mathbf{h}_1),$$

where $\mathbf{h}_1$ and $\mathbf{h}_2$ are any dimension d vectors. Before establishing the mathematics, consider the practical meaning of S_1 and S_2 in a computer simulation. If we increase/decrease inputs by an amount $\mathbf{h}_1$, S_1 is the amount of change in the response of a computer simulation. If we then increase/decrease inputs our model by $\mathbf{h}_2$, S_2 represents the marginal change in the computer simulation response. In general, the relationship between these increments varies based on the size of $\mathbf{h}_1$ and $\mathbf{h}_2$ and the angle between them.

For many computer simulations, similar changes to the inputs should cause similar marginal changes in the response. A way to measure the changes is to look at the correlation

$$\text{corr}(S_1, S_2) = \frac{\text{cov}(S_1, S_2)}{\sqrt{\text{var}(S_1) \text{var}(S_2)}}.$$

In general, if this correlation is negative, if a process rises on a segment from the origin to $\mathbf{h}_1$, it will be expected to decrease in the following segment. Conversely, if this correlation is positive, if a process rises on a segment from the origin to $\mathbf{h}_2$, it will be expected to continue to increase. Thus, this value is directly related to the amount of peaks and valleys in a response surface. The value being negative implies peaks and valleys, the value being positive implies growth. Colloquially, the sign of this correlation can be summarized as “what goes up is expected to come down” versus “what goes up is expected to go up.” As will be discussed in this section, it appears most simulations of physical models have the property of $\text{corr}(S_1, S_2)$ being positive.

Consider the following definition:

Supplemental Definition 1. Let $Z(\cdot)$ be a Gaussian process with $Z(\mathbf{0}) = 0$. For $\mathbf{h}_1$ and $\mathbf{h}_2$ such that $\|\mathbf{h}_1\| = 1$, $\|\mathbf{h}_2\| = 1$ and $\mathbf{h}_1^\top \mathbf{h}_2 > 0$, define the long range correlation between the pair $\mathbf{h}_1$ and

$\mathbf{h}_2$ as

$$T(\mathbf{h}_1, \mathbf{h}_2) = \lim_{\kappa \rightarrow \infty} \text{corr} \{Z(\kappa \mathbf{h}_2 + \kappa \mathbf{h}_1) - Z(\kappa \mathbf{h}_1), Z(\kappa \mathbf{h}_1)\}.$$

This can be considered a measure of the long range structure of the response surface.

We now define strictly stationary covariance functions to investigate their properties. A random function is considered *strictly stationary* if the distribution of $Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)$ is equal to $Z(\mathbf{x}_1 + \mathbf{h}), \dots, Z(\mathbf{x}_n + \mathbf{h})$ for all vectors in $\mathbf{h} \in \mathbb{R}^d$. This directly implies that

$$\text{cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = \phi(\mathbf{x} - \mathbf{x}'), \quad (1)$$

where ϕ is some function. From Christakos (1984), the function ϕ must be such that

$$\phi(\mathbf{h}) < \phi(\mathbf{0}),$$

and thus the covariance function must have a peak when the distance between points is $\mathbf{0}$. Moreover, Christakos (1984) shows that as the distance between points gets large, we have

$$\phi(\mathbf{h}) = \mathcal{O}(\|\mathbf{h}\|^{\frac{1-d}{2}}), \quad \|\mathbf{h}\| \rightarrow \infty.$$

Thus for all problems where the dimension of x is larger than 1, the covariance must decay to $\mathbf{0}$. For radial covariances that are well-defined for all $d > 1$, such as the squared exponential $\phi(\mathbf{h}) = \exp(-\|\mathbf{h}\|^2)$, the rate of decay is faster than any polynomial.

The conclusion that a stationary covariance function will approach 0 gives us the ability to investigate the class of all stationary covariances. Returning to the concept of the correlation between directions (proof located at the end of these materials):

Supplemental Theorem 1. Say $Z(\cdot)$ is a Gaussian process with a strictly stationary covariance function defined on $\mathbb{R}^d$, $d \geq 2$. Then conditional on $Z(\mathbf{0}) = 0$, all pairs of directions have a long range correlation of

$$T(\mathbf{h}_1, \mathbf{h}_2) = -\frac{1}{\sqrt{2}}.$$

This demonstrates the lack of flexibility of radial covariances discussed in the introduction of this paper. A correlation coefficient of ≈ -0.7 represents a strong negative relationship. All stationary, and thus radial, covariance functions require some assumption of peaks and valleys in the response surface. Even if this property is appropriate for some models, it is discouraging that there is no room for choice based on the response surface of interest.

To illustrate the restrictive nature of radial covariances, we will study a simple, one dimensional function. Let the response surface be $\exp(x)$, $x \in \mathbb{R}$. Consider the value of $\{y(2h) - y(h)\}\{y(h) - y(0)\}$ with $h > 0$. Any radial covariance implies the expected value of this quantity should be

strictly negative as h gets large.

$$\{y(2h) - y(h)\}\{y(h) - y(0)\} = \exp(h)(\exp(h) - 1)^2.$$

Clearly, this value is strictly positive and gets larger as h gets larger, which means a Gaussian process with a radial covariance would be a poor model of this function.

With this discussion, it appears that new covariances that violate the strictly stationary condition in (1) are needed. In Zhang and Apley (2014, 2015), the authors' attempt to correct this problem by borrowing and extending covariance functions from fractional Brownian fields. In the first paper, the proposal is to treat $Z(\cdot)$ as a fractional Brownian field, with a covariance function

$$c(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1\|_{\mathbf{a}}^{2\beta} + \|\mathbf{x}_2\|_{\mathbf{a}}^{2\beta} - \|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbf{a}}^{2\beta},$$

where $\beta \in (0, 1)$ has an effect on the trends in the response. Zhang and Apley (2014) conjectured that this model is a better representation of computer simulations because of its particular nonstationary nature, by which $c(\mathbf{x}, \mathbf{x}')$ does not necessarily decay to zero as $\|\mathbf{x} - \mathbf{x}'\|$ increases. We revisit this model by looking at the correlation between directions. When $\beta = 1/2$, the field is a standard Brownian field. A well-known property of standard Brownian fields is independent increments, where if $Z(\cdot)$ is a standard Brownian field then

$$\text{cov}\{Z(2\mathbf{h}) - Z(\mathbf{h}), Z(\mathbf{h})\} = 0,$$

independent of the vector $\mathbf{h}$.

The proposed lifted Brownian covariance behaves like a fractional Brownian covariance over long distances. Placing this idea into the definition of correlated directions, we find the following general result (proof located at the end of these materials):

Supplemental Theorem 2. Say $Z(\cdot)$ is a zero mean Gaussian process with a covariance function

$$c(\mathbf{x}, \mathbf{x}') = \psi(\mathbf{x}) + \psi(\mathbf{x}') - \psi(\mathbf{x} - \mathbf{x}') - 1 \quad \text{where } \psi(\mathbf{h}) = \{1 + \|\mathbf{h}\|^{2\gamma}\}^\beta.$$

If θ is the angle between two directions $\mathbf{h}_1$ and $\mathbf{h}_2$, then

$$T(\mathbf{h}_1, \mathbf{h}_2) = \frac{(1 + \cos(\theta))^{\beta\gamma}}{2^{1-\beta\gamma}} - 1.$$

Thus for all pairs of directions with $\cos(\theta) > 0$, there exists a $\beta, \gamma \in (0, 1) \times (0, 1]$ such that they have negative long range correlation and there exists a $\beta, \gamma \in (0, 1) \times (0, 1]$ such that they have positive long range correlation.

Figure 1: $T(\mathbf{h}_1, \mathbf{h}_2)$ as a function of the angle between directions for $\gamma\beta = 0.1, 0.2, \dots, 0.9$ (ascending order on the y-axis).

Thus the proposed model allows for flexibility in long range dependence between directions. Figure 1 shows the value of $T(\cdot, \cdot)$ as a function of θ with various values of $\gamma\beta$. Under the proposed covariance, the directions are more correlated if the angle between them, θ , is smaller. This is an intuitively appealing property since one expects similar changes to the input should result in similar marginal changes to the response. If $\gamma\beta$ is less than $1/2$, all directions are negatively correlated, although not as strongly as with stationary covariances. Also, as $\gamma\beta \rightarrow 0$, $T \rightarrow -1/2$. Thus as $\beta\gamma$ gets small, the long range correlation between directions will drop.

We can numerically examine this property by directly computing values from some sample functions and estimate the long range dependency of the function. For the functions described in the simulation study in Section 6, we can estimate the long range dependence by looking at the behavior at the edges of our input space. Via the arguments presented in this section, the long range correlation between directions depends on the angle between them, θ . At the end of these materials, there is an algorithm given to generate an empirical value of $T(\mathbf{h}_1, \mathbf{h}_2)$, called $\hat{T}(\mathbf{h}_1, \mathbf{h}_2)$. With the availability of this empirical value of $T(\cdot, \cdot)$, we can investigate the conjecture made in Section 3 that these types of models tend to have positive long range correlation between directions. Recall that under the radial model of covariance, $T(\cdot, \cdot)$ will always be -0.7 , regardless of the angle between the directions, θ . Figure 2 shows the empirical long range correlation along with the model presented in this paper if $\beta\gamma = 0.95$. The empirical values across the functions do not have a dependency of -0.7 . Instead, it appears that the model proposed in this paper with a large value of $\beta\gamma$ does an adequate job of capturing the dependency between directions. This initial analysis

Figure 2: Estimated values of $T(\mathbf{h}_1, \mathbf{h}_2)$ for the five functions described in Section 6.2, represented in the large dots. The thin line is the model of $T(\mathbf{h}_1, \mathbf{h}_2)$ under the lifted Brownian model if $\beta\gamma = 0.95$.

indicates some promise that lifted Brownian models are better representations of these functions than radial covariances.

Algorithm for the empirical estimate of $\hat{T}(\theta)$

We choose a random unit direction in a d dimensional space, which can be done via simulation of Gaussian random variables, call this vector $\mathbf{h}_1$. Our next objective is to sample from the space of vectors such that $\mathbf{h}_1^\top \mathbf{h}_2 = \cos(\theta)$. To do this, we generate a vector of eight independent Gaussian random variables $\mathbf{z}$ and save $\mathbf{z} - \mathbf{h}_1 \mathbf{h}_1^\top \mathbf{z}$, which is an orthogonal vector to $\mathbf{h}_1$. Normalize this vector and call it $\mathbf{r}_2$, lastly let

$$\mathbf{h}_2 = \cos(\theta)\mathbf{h}_1 + \sin(\theta)\mathbf{r}_2.$$

This procedure creates two random directions with an angle θ between them. For the i th generation at angle θ , let $S_1^i(\theta) = y(\kappa\mathbf{h}_1) - y(\mathbf{0})$ and $S_2^i(\theta) = y(\kappa\mathbf{h}_1 + \kappa\mathbf{h}_2) - y(\kappa\mathbf{h}_1)$. Then set

$$\hat{T}(\theta) = \frac{\sum_i (S_1^i(\theta) - \bar{S}_1(\theta))(S_2^i(\theta) - \bar{S}_2(\theta))}{\sqrt{\sum_i (S_2^i(\theta) - \bar{S}_2(\theta))^2} \sqrt{\sum_i (S_1^i(\theta) - \bar{S}_1(\theta))^2}},$$

where

$$\bar{S}_1(\theta) = \sum_i \frac{1}{N} S_1^i(\theta) \text{ and } \bar{S}_2(\theta) = \sum_i \frac{1}{N} S_2^i(\theta).$$

This is the Pearson correlation coefficient between S_1 and S_2 . This can be considered an analogue to the empirical covariance or the empirical variogram. If κ is chosen large enough, we arrive at an empirical estimate of the quantity discussed in Section 3. For our analysis in Section 6, we rescaled all input domains to the unit cube and choose $\kappa = 1/3$. While this can result in some values outside of the input domain, it did not cause any problems for these particular functions as they are all defined outside of the input domain.

Proof of Supplemental Theorem 1

Proof. Without loss of generality, assume that $\phi(\mathbf{0}) = 1$. First we show the covariance between points conditioned on the observation that $Z(\mathbf{0}) = 0$. Following standard computations, this is given by

$$\begin{aligned}\phi_{Z(\mathbf{0})=0}(x, x') &= \mathbb{E}(Z(\mathbf{x})Z(\mathbf{x}')) - \mathbb{E}(Z(\mathbf{x})Z(\mathbf{0}))\mathbb{E}(Z(\mathbf{0})Z(\mathbf{0}))^{-1}\mathbb{E}(Z(\mathbf{0})Z(\mathbf{x}')), \\ &= \phi(\mathbf{x} - \mathbf{x}') - \phi(\mathbf{x})\phi(\mathbf{x}').\end{aligned}$$

Let $\mathbf{h}_3 = \mathbf{h}_1 + \mathbf{h}_2$. The variance of $Z(\kappa\mathbf{h}_1) - Z(\mathbf{0})$ is $1 - \phi(\kappa\mathbf{h}_1)^2$ and the variance of $Z(\kappa\mathbf{h}_3) - Z(\kappa\mathbf{h}_1)$ is

$$2 - \phi(\kappa\mathbf{h}_1)^2 - \phi(\kappa\mathbf{h}_3)^2 - 2(\phi(\kappa\mathbf{h}_2) - \phi(\kappa\mathbf{h}_1)\phi(\kappa\mathbf{h}_3)),$$

and thus their limits approach 1 and 2 respectively as $\kappa \rightarrow \infty$. The covariance between $Z(\kappa\mathbf{h}_3) - Z(\kappa\mathbf{h}_1)$ and $Z(\kappa\mathbf{h}_1) - Z(\mathbf{0})$ is the same as the covariance between $Z(\kappa\mathbf{h}_3) - Z(\kappa\mathbf{h}_1)$ and $Z(\kappa\mathbf{h}_1)$, which is

$$\phi(\kappa\mathbf{h}_2) - \phi(\kappa\mathbf{h}_1)\phi(\kappa\mathbf{h}_3) - 1 + \phi(\kappa\mathbf{h}_1)^2.$$

This has a limit as $\kappa \rightarrow \infty$ of -1 . Thus the correlation has limit

$$-\frac{1}{\sqrt{1}\sqrt{2}} = -\frac{1}{\sqrt{2}}.$$

□

Proof of Supplemental Theorem 2

Proof. Let $\mathbf{h}_3 = \mathbf{h}_1 + \mathbf{h}_2$. Under the traditional fractional Brownian model with parameter β , the variance of $Z(\kappa\mathbf{h}_1) - Z(\mathbf{0})$ is $2\kappa^{2\beta}$ and the variance of $Z(\kappa\mathbf{h}_3) - Z(\kappa\mathbf{h}_1)$ is

$$2\kappa^{2\beta}\|\mathbf{h}_1\|^{2\beta} + 2\kappa^{2\beta}\|\mathbf{h}_3\|^{2\beta} - 2\kappa^{2\beta}(\|\mathbf{h}_3\|^{2\beta} + \|\mathbf{h}_2\|^{2\beta} - \|\mathbf{h}_1\|^{2\beta}) = 2\kappa^{2\beta}.$$

The covariance between $Z(\kappa\mathbf{h}_3) - Z(\kappa\mathbf{h}_1)$ and $Z(\kappa\mathbf{h}_1) - Z(\mathbf{0})$ is the same as the covariance between $Z(\kappa\mathbf{h}_3) - Z(\kappa\mathbf{h}_1)$ and $Z(\kappa\mathbf{h}_1)$, which is

$$\kappa^{2\beta}\|\mathbf{h}_3\|^{2\beta} + \kappa^{2\beta} - \kappa^{2\beta} - 2\kappa^{2\beta} = \kappa^{2\beta}(2 + 2\cos(\theta))^\beta - 2\kappa^{2\beta}.$$

Thus

$$\text{corr}\{Z(\kappa\mathbf{h}_1 + \kappa\mathbf{h}_2) - Z(\kappa\mathbf{h}_1), Z(\kappa\mathbf{h}_1) - Z(\mathbf{0})\} = \frac{(2 + 2\cos(\theta))^\beta - 2}{2}.$$

We now generalize to the proposed covariance with $\psi(\mathbf{h}) = (1 + \|\mathbf{h}\|^{2\gamma})^\beta$. The variance of

$Z(\kappa \mathbf{h}_1) - Z(\mathbf{0})$ is $2\psi(\kappa \mathbf{h}_1) - 2$ and the variance of $Z(\kappa \mathbf{h}_3) - Z(\kappa \mathbf{h}_1)$ is

$$2\psi(\kappa \mathbf{h}_3) - 2 + 2\psi(\kappa \mathbf{h}_1) - 2 - 2(\psi(\kappa \mathbf{h}_3) + \psi(\kappa \mathbf{h}_1) - \psi(\kappa \mathbf{h}_2) - 1),$$

or

$$2\psi(\kappa \mathbf{h}_2) - 2.$$

The covariance between $Z(\kappa \mathbf{h}_3) - Z(\kappa \mathbf{h}_1)$ and $Z(\kappa \mathbf{h}_1) - Z(\mathbf{0})$ is the same as the covariance between $Z(\kappa \mathbf{h}_3) - Z(\kappa \mathbf{h}_1)$ and $Z(\kappa \mathbf{h}_1)$, which is

$$\psi(\kappa \mathbf{h}_3) + \psi(\kappa \mathbf{h}_1) - \psi(\kappa \mathbf{h}_2) - 1 - 2\psi(\kappa \mathbf{h}_1) + 2,$$

or

$$\psi(\kappa \mathbf{h}_3) - \psi(\kappa \mathbf{h}_2) - \psi(\kappa \mathbf{h}_1) + 1.$$

We now use the fact that the function $\psi(\|\mathbf{h}\|)$ behaves like $\|\mathbf{h}\|^{2\beta\gamma}$ for large $\|\mathbf{h}\|$ and thus we have that as κ gets large,

$$\text{corr} \{Z(\kappa \mathbf{h}_1 + \kappa \mathbf{h}_2) - Z(\kappa \mathbf{h}_1), Z(\kappa \mathbf{h}_1) - Z(\mathbf{0})\} \rightarrow \frac{\|\mathbf{h}_3\|^{2\beta\gamma} - \|\mathbf{h}_1\|^{2\beta\gamma} - \|\mathbf{h}_2\|^{2\beta\gamma}}{2\|\mathbf{h}_2\|^{\beta\gamma}\|\mathbf{h}_1\|^{\beta\gamma}}.$$

□

Supplemental References

- Christakos, G. (1984), “On the problem of permissible covariance and variogram models,” *Water Resources Research*, 20, 251–265.
- Zhang, N. and Apley, D. W. (2014), “Fractional Brownian fields for response surface metamodeling,” *Journal of Quality Technology*, 46, 285–301.
- (2015), “Brownian Integrated Covariance Functions for Gaussian Process Modeling: Sigmoidal Versus Localized Basis Functions,” *Journal of the American Statistical Association*, 00–00.