Limit Kriging

V. Roshan Joseph
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205, USA
roshan@isye.gatech.edu

Abstract

A new kriging predictor is proposed. It gives a better performance over the existing predictor when the constant mean assumption in the kriging model is unreasonable. Moreover, the new predictor seems to be robust to the misspecifications in the correlation parameters. The advantages of the new predictor is demonstrated using some examples from the computer experiments literature.

KEY WORDS: Computer experiments, Geostatistics, Interpolation, Prediction, Spatial statistics.

Submitted for publication
February 14, 2005

1. INTRODUCTION

A kriging model gives an interpolating predictor, which can be used to approximate a function based on its finite number of evaluations. Kriging was originated in geostatistics (Matheron 1963), which completely revolutionized the field. Several empirical studies have proved its superiority over the other interpolating techniques such as splines (see Laslett, 1994). Kriging is now widely used in many fields including computer experiments (see Sacks et al. 1989) and spatial statistics (see Cressie 1991).

There are some problems with the kriging predictor. First we explain those problems using a simple example and then propose a new predictor to overcome them. A few details about the kriging model need to be stated before getting into the problem. Consider a function $y(\boldsymbol{x})$, where $\boldsymbol{x} \in \mathbb{R}^p$. The core idea in kriging is to model this function as a realization from a stochastic process. The kriging model can be stated as follows:

$$Y(\boldsymbol{x}) = \mu + Z(\boldsymbol{x}), \tag{1}$$

where $Z(\boldsymbol{x})$ is a weak stationary stochastic process with mean 0 and covariance function $\sigma^2 R$. In *simple kriging* (see Olea, 1999) μ is assumed to be known. The covariance function is defined as $cov\{Y(\boldsymbol{x}+\boldsymbol{h}), Y(\boldsymbol{x})\} = \sigma^2 R(\boldsymbol{h})$. The correlation function $R(\boldsymbol{h})$ must be a positive semidefinite function with $R(\boldsymbol{0}) = 1$ and $R(-\boldsymbol{h}) = R(\boldsymbol{h})$. The weak stationarity assumption can be relaxed using the intrinsic stationarity assumption (see Wackernagel 2002), but it will not solve the problems with the predictor studied in this article and therefore we will present the results using weak stationary process. Some details about the intrinsic kriging will be given at the end of the next section.

Suppose we evaluated the function at n points $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ and let $\boldsymbol{y} = (y_1, \cdots, y_n)'$ be the corresponding function values. Then the kriging predictor of the function at some $\boldsymbol{x}$ is given by

$$\hat{y}(\boldsymbol{x}) = \mu + \boldsymbol{r}(\boldsymbol{x})' \boldsymbol{R}^{-1} (\boldsymbol{y} - \mu \boldsymbol{1}), \tag{2}$$

where $\boldsymbol{1}$ is a column of 1's having length n, $\boldsymbol{r}(\boldsymbol{x})' = (R(\boldsymbol{x} - \boldsymbol{x}_1), \cdots, R(\boldsymbol{x} - \boldsymbol{x}_n))$, and $\boldsymbol{R}$ is an $n \times n$ matrix with elements $R(\boldsymbol{x}_i - \boldsymbol{x}_j)$. It is the best linear unbiased predictor (BLUP),

which minimizes the mean squared prediction error $E\{\hat{Y}(\boldsymbol{x}) - Y(\boldsymbol{x})\}^2$ under (1). A better understanding of the kriging can be obtained by taking a Bayesian point of view. Assume $Z(\boldsymbol{x})$ to be a Gaussian process. We can consider it as a prior distribution on the function. Then the predictor in (2) is nothing but the posterior mean of the function given the data (see Currin et al. 1991).

Now consider an example from Santner, Williams, and Notz (2003). Suppose the function is $y(x) = \exp(-1.4x)\cos(7\pi x/2)$ over $0 \leq x \leq 1$. We choose seven equally spaced points from $[0, 1]$ for evaluating the function. Thus $x_i = (i - 0.5)/7$ for $i = 1, \cdots, 7$. The function and the seven points are plotted in Figure 1. Suppose we use a Gaussian correlation function $R(h) = \exp(-\theta h^2)$ for $\theta > 0$. Santner, Williams, and Notz suggested using $\theta = 136.1$. Taking $\mu = \int_0^1 y(x)dx \approx -0.01$, the kriging predictor can be obtained from (2). It is also plotted in Figure 1, which shows a reasonably good fit to the true function. Now consider another choice of θ. Figure 2 shows the kriging predictor with $\theta = 300$. The fit is clearly not good in the intervals (x_2, x_3), (x_4, x_5), and (x_6, x_7). The predictor is behaving in the opposite direction of what it should have been. If we take a closer look at Figure 1, we can see the same phenomena but it is not as pronounced as in Figure 2. This has happened because the posterior mean is pulled towards the prior mean ($\mu = -0.01$) in regions where data are not available. It seems that by using the right value of θ we can avoid this problem. But in real practice we do not know the true function and therefore will not be able to choose the best value of θ to get the best fit. There are several methods for estimating the correlation parameters from the data. One popular method is maximum likelihood estimation. In this particular case the maximum likelihood estimate of θ is more than 300, which will make the fit even more worse than the one in Figure 2.

In this article we will propose a modification of the kriging predictor, which seems to mitigate the above problem. The article is organized as follows. In Section 2, we introduce the new predictor. Some of its properties are studied in Section 3. In Section 4, we use two examples from the computer experiments literature to compare the performance of the new predictor with the existing predictor. Some concluding remarks are given in Section 5.

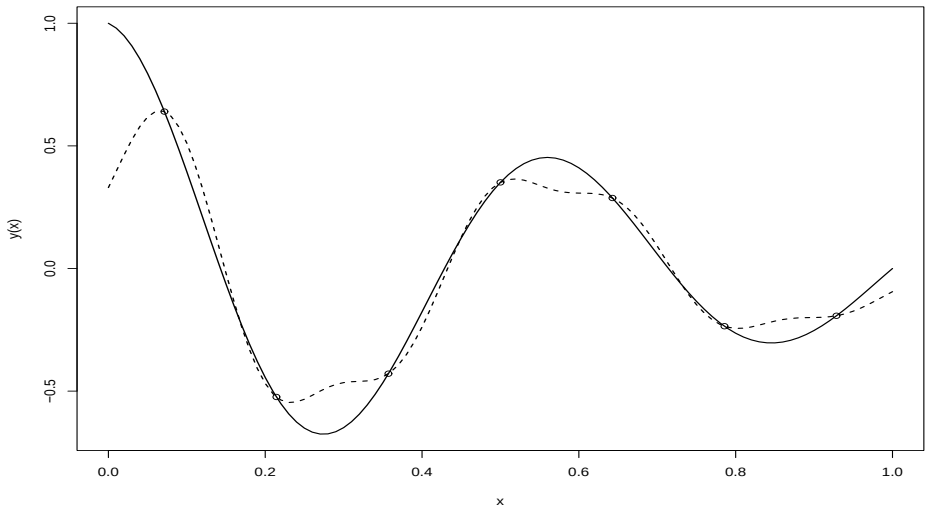

Figure 1: The plot of the function $y(x) = \exp(-1.4x)\cos(7\pi x/2)$ (solid) and the simple kriging predictor (dashed) with $\theta = 136.1$ and $\mu = -0.01$.

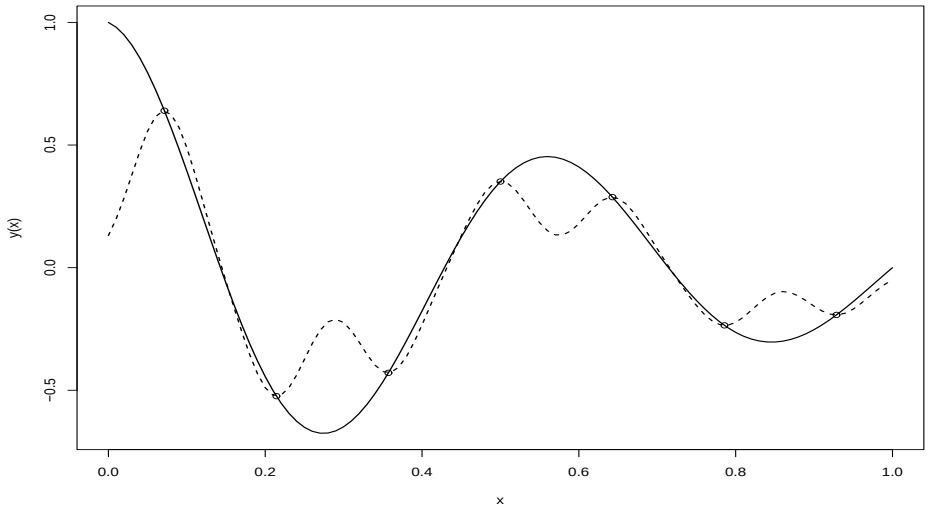

Figure 2: The plot of the function $y(x) = \exp(-1.4x)\cos(7\pi x/2)$ (solid) and the simple kriging predictor (dashed) with $\theta = 300$ and $\mu = -0.01$.

2. THE LIMIT KRIGING

If we can somehow remove μ from (2), then we might be able to overcome the problem exhibited in Figure 2. In order to do this, introduce a new function $\lambda(\boldsymbol{x})$ as follows. Let

$$\hat{y}(\boldsymbol{x}) = \mu + \lambda(\boldsymbol{x})\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \mu\mathbf{1}).$$

Rearranging the terms, $\hat{y}(\boldsymbol{x}) = \mu\{1 - \lambda(\boldsymbol{x})\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}\} + \lambda(\boldsymbol{x})\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{y}$. Assume that $\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1} \neq 0$. Thus if we take $\lambda(\boldsymbol{x}) = 1/\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}$, then μ will no longer be required for prediction. Thus we obtain a new predictor

$$\hat{y}(\boldsymbol{x}) = \frac{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}}. \tag{3}$$

Since $\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}_i) = \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is a unit vector with a 1 at its i^{th} position, we have $\hat{y}(\boldsymbol{x}_i) = \boldsymbol{e}_i'\boldsymbol{y}/\boldsymbol{e}_i'\mathbf{1} = y_i$. Therefore the new predictor is still an interpolator. In addition, it is unbiased under the model in (1), because $E\{\hat{y}(\boldsymbol{x})\} = \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mu\mathbf{1}/\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1} = \mu$.

The earlier work in kriging methodology tries to remove μ from the predictor in a different way. Consider a predictor of the form $\hat{y}(\boldsymbol{x}) = \mu + \boldsymbol{c}(\boldsymbol{x})'(\boldsymbol{y} - \mu\mathbf{1})$. If we minimize the mean squared prediction error (MSPE) with respect to $\boldsymbol{c}(\boldsymbol{x})$, then we will obtain the simple kriging predictor as in (2). Suppose we add a constraint $\boldsymbol{c}(\boldsymbol{x})'\mathbf{1} = 1$ to the above optimization. Then the predictor becomes $\hat{y}(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x})'\boldsymbol{y}$, which does not include μ. By minimizing MSPE under this constraint we obtain

$$\boldsymbol{c}(\boldsymbol{x})' = \{1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}\}\frac{\mathbf{1}'\boldsymbol{R}^{-1}}{\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}} + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}.$$

The corresponding predictor can be written as

$$\hat{y}(\boldsymbol{x}) = \hat{\mu} + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1}), \tag{4}$$

where $\hat{\mu} = \mathbf{1}'\boldsymbol{R}^{-1}\boldsymbol{y}/\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}$. This is known as *ordinary kriging* in geostatistics (see Olea 1999) and is probably the most commonly used form of the kriging predictor in practice. The ordinary kriging predictor is similar to the simple kriging predictor in (2) with μ estimated from the data using generalized least squares. In the Bayesian formulation, $\hat{\mu}$ can be thought

of as an empirical Bayes estimate obtained by maximizing the likelihood. Another Bayesian interpretation can also be given. Suppose we impose an improper prior distribution on μ. Then we can obtain the ordinary kriging predictor as the posterior mean of the unknown function given the data.

Both the ordinary kriging predictor and the new predictor in (3) are linear unbiased predictors. In addition to that, the ordinary kriging predictor is the "best" in the sense of minimizing the MSPE under model (1). In fact, we can obtain the following expressions for the MSPE. Assume the correlation function to be known. Then, for the ordinary kriging

$$MSPE = \sigma^2[1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) + \frac{\{1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}\}^2}{\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}}],$$

and for the new predictor

$$MSPE = \sigma^2[1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) + \frac{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x})}{\{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}\}^2}\{1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}\}^2].$$

By Cauchy-Schwarz inequality $\{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}\}^2 \leq (\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1})\{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x})\}$ and therefore, as expected, the MSPE of the new predictor is always greater than or equal to the MSPE of the ordinary kriging predictor. If this is the case, one would think, why should we even look at the new predictor? The reason is, the ordinary kriging is optimal under the model in (1), but if this model is unreasonable, then other predictors can outperform it. Before evaluating the MSPE under a more reasonable model, let us take a look at the example again given in the introduction. Figure 3 shows the plot of the ordinary kriging predictor in (4) and the new predictor in (3) with $\theta = 136.1$. Note that the new predictor is better than the ordinary kriging predictor in those three intervals (x_2, x_3), (x_4, x_5), and (x_6, x_7), which is more pronounced in Figure 4 with $\theta = 300$. Because $\hat{\mu} = -0.01$, the ordinary kriging predictors behave exactly the same way as the simple kriging predictors in Figures 1 and 2. The most striking feature about the new predictor is that there is no "pulling effect" from the prior mean as it was there for the simple and ordinary kriging predictors. Therefore the technique we applied to remove the effect of the prior mean μ from (2) was successful for the new predictor but not for the ordinary kriging predictor. This is expected because in the ordinary kriging predictor we only replaced the μ in (2) with an estimate of it from the data, which cannot alter the behavior of the simple kriging predictor.

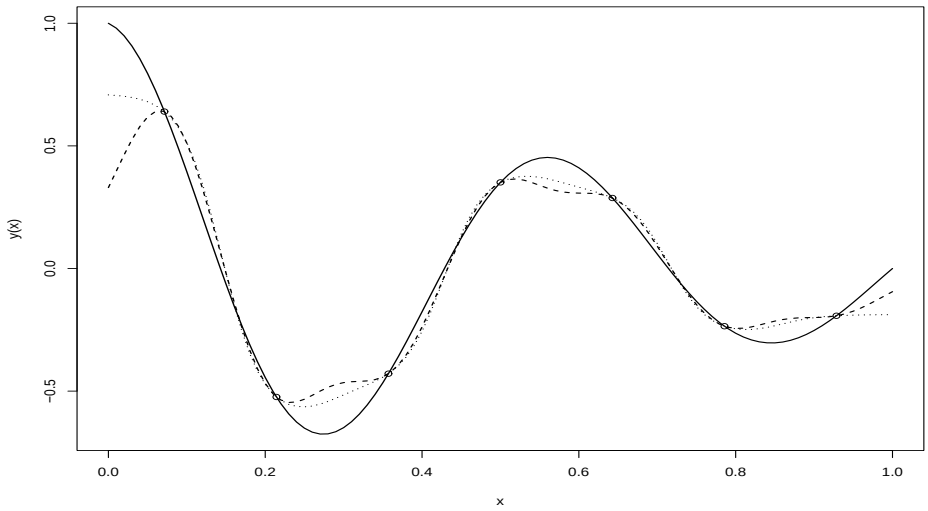

Figure 3: The plot of the function $y(x) = \exp(-1.4x)\cos(7\pi x/2)$ (solid), the ordinary kriging predictor (dashed), and the new predictor (dotted) with $\theta = 136.1$.

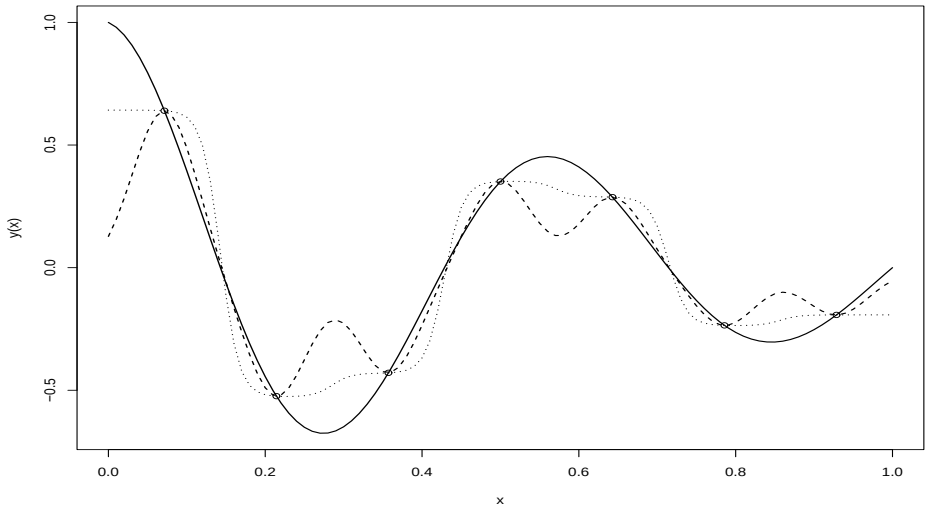

Figure 4: The plot of the function $y(x) = \exp(-1.4x)\cos(7\pi x/2)$ (solid), the ordinary kriging predictor (dashed), and the new predictor (dotted) with $\theta = 300$.

Now we will try to provide more insights into the new predictor. Consider the two plots of the simple kriging predictors in Figure 5. The first one is with $\mu = -0.5$ and the second one is with $\mu = 0.5$ (both are with $\theta = 300$). Note that the first predictor does well in the intervals (x_2, x_3) and (x_6, x_7) and the second predictor does well in the interval (x_4, x_5) compared to the predictors in Figure 2. The reason is, compared to -0.01 the function is closer to -0.5 in the first two intervals and closer to 0.5 in the interval (x_4, x_5). This suggests that if we change μ in the simple kriging predictor depending on $\boldsymbol{x}$, then we might be able to improve the prediction. Therefore, consider a modified predictor of the form

$$\hat{y}(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\{\boldsymbol{y} - \mu(\boldsymbol{x})\boldsymbol{1}\}.$$

If $\mu(\boldsymbol{x})$ is close to the true function $y(\boldsymbol{x})$, then we can expect to get excellent prediction. But unfortunately we do not know the true function. One possible way to circumvent this problem is to use the predicted value of $y(\boldsymbol{x})$ as $\mu(\boldsymbol{x})$. Therefore, consider the following recursive implementation of the above idea. Let

$$\hat{y}_k(\boldsymbol{x}) = \hat{y}_{k-1}(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\{\boldsymbol{y} - \hat{y}_{k-1}(\boldsymbol{x})\boldsymbol{1}\}, \tag{5}$$

where $\hat{y}_k(\boldsymbol{x})$ is the predicted value at the k^{th} stage with $\hat{y}_1(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\{\boldsymbol{y} - \mu(\boldsymbol{x})\boldsymbol{1}\}$. Then it is easy to show that

$$\hat{y}_k(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \frac{[1 - \{1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{1}\}^k]}{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{1}}\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\{\boldsymbol{y} - \mu(\boldsymbol{x})\boldsymbol{1}\}.$$

If $0 < \boldsymbol{r}(\boldsymbol{x})'R^{-1}\boldsymbol{1} < 2$, then as $k \to \infty$ the predictor becomes

$$\lim_{k\to\infty} \hat{y}_k(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \frac{1}{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{1}}\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\{\boldsymbol{y} - \mu(\boldsymbol{x})\boldsymbol{1}\} = \frac{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{1}}, \tag{6}$$

which is the same as the new predictor in (3). Since the new predictor can be obtained as the limit of the simple kriging predictor with varying μ, we will name it as the *limit kriging* predictor. It can also be obtained in a single step as follows. Let $\hat{y}(\boldsymbol{x}) = \lim_{k\to\infty} \hat{y}_k(\boldsymbol{x})$. Then taking limits on both sides of equation (5) we obtain

$$\hat{y}(\boldsymbol{x}) = \hat{y}(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\{\boldsymbol{y} - \hat{y}(\boldsymbol{x})\boldsymbol{1}\}.$$

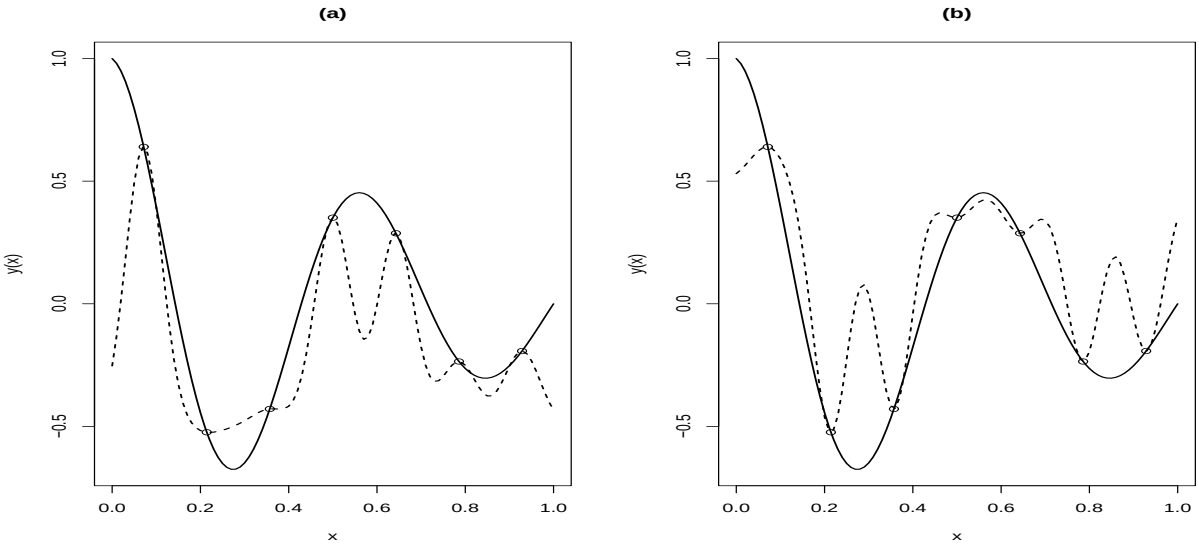

Figure 5: The plot of the function $y(x) = \exp(-1.4x)\cos(7\pi x/2)$ (solid), the simple kriging predictor (dashed) with $\theta = 300$, and (a) $\mu = -0.5$, (b) $\mu = 0.5$.

Solving for $\hat{y}(\boldsymbol{x})$ we get the limit kriging predictor. The above interpretation of the new predictor explains why it performed well compared to the simple/ordinary kriging predictors in Figures 1-4. Because μ was varied depending on the predicted value, the pulling effect from the fixed μ was removed, thereby giving a better prediction.

It is important to elaborate on the last point made in the above discussion. The simple kriging predictor in (2) can be written as

$$\hat{y}(\boldsymbol{x}) = \mu + \sum_{i=1}^{n} c_i(\boldsymbol{x})(y_i - \mu).$$

where $c_i(\boldsymbol{x}) = (\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}))_i$ is the weight given for the i^{th} observation. If we replace μ by $\sum_{i=1}^{n} c_i(\boldsymbol{x})y_i / \sum_{i=1}^{n} c_i(\boldsymbol{x})$, then we get the limit kriging predictor. Typically, the kriging weights are large for observations closer to $\boldsymbol{x}$ and small for observations farther from $\boldsymbol{x}$. In such a case, the limit kriging predictor can be interpreted as a simple kriging predictor with μ replaced by a local average of the y_i's. Therefore the pulling from the global mean as in the case of ordinary kriging is avoided. In this sense limit kriging is similar to the idea of kriging with a moving neighborhood (see Chilès and Delfiner 1999).

As noted in the derivation, the limit in (6) exists only if $0 < \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1} < 2$. Although

this seems to be true under some conditions on the correlation function, we are able to prove only for some special cases given in the Appendix A. Fortunately, the only condition required for the existence of the predictor is $\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1} \neq 0$. A necessary condition for this is $\boldsymbol{r}(\boldsymbol{x}) \neq 0$. For certain correlation functions such as the linear correlation function, cubic correlation function etc. $R(\boldsymbol{h}) = 0$ when the length of $\boldsymbol{h}$ is more than a threshold value. In such cases the limit kriging predictor cannot be computed. The problem can be avoided by defining the predicted value equal to $\hat{\mu}$ when $\boldsymbol{r}(\boldsymbol{x}) = 0$ (so that it gives the same value as the ordinary kriging). In the next section we will show that for certain class of correlation functions the limit kriging predictor exits in the limit as $\boldsymbol{r}(\boldsymbol{x}) \to \mathbf{0}$ and therefore it is defined for all $\boldsymbol{x}$.

Now we will give a justification on why the limit kriging is able to produce better performance over the ordinary kriging in some cases. Note that we cannot compare the performance of two predictors for a given function, because a predictor that is exactly equal to the given function is optimal. We can only compare the average performance of the predictors with respect to a family of functions (such as by using MSPE). Consider a family of functions generated from the model

$$Y(\boldsymbol{x}) = \mu(\boldsymbol{x}) + Z(\boldsymbol{x}), \tag{7}$$

where $Z(\boldsymbol{x})$ is a weak stationary process with mean 0 and covariance function $\tau^2 R$. In particular, we assume the true function to be closer to $\mu(\boldsymbol{x})$ than a constant μ. Let $\boldsymbol{\mu} = (\mu(\boldsymbol{x}_1), \cdots, \mu(\boldsymbol{x}_n))'$. The MSPE of the ordinary kriging under the model in (7) is (for simplicity $\boldsymbol{r}(\boldsymbol{x})$ is denoted as $\boldsymbol{r}$)

$$MSPE_{OK} = \{(1 - \boldsymbol{r}'\boldsymbol{R}^{-1}\mathbf{1})\frac{\mathbf{1}'\boldsymbol{R}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}} + \boldsymbol{r}'\boldsymbol{R}^{-1}\boldsymbol{\mu} - \mu(\boldsymbol{x})\}^2 + \tau^2\{1 - \boldsymbol{r}'\boldsymbol{R}^{-1}\boldsymbol{r} + \frac{(1 - \boldsymbol{r}'\boldsymbol{R}^{-1}\mathbf{1})^2}{\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}}\} \tag{8}$$

and for limit kriging

$$MSPE_{LK} = \{\frac{\boldsymbol{r}'\boldsymbol{R}^{-1}\boldsymbol{\mu}}{\boldsymbol{r}'\boldsymbol{R}^{-1}\mathbf{1}} - \mu(\boldsymbol{x})\}^2 + \tau^2\{1 - \boldsymbol{r}'\boldsymbol{R}^{-1}\boldsymbol{r} + \frac{\boldsymbol{r}'\boldsymbol{R}^{-1}\boldsymbol{r}}{(\boldsymbol{r}'\boldsymbol{R}^{-1}\mathbf{1})^2}(1 - \boldsymbol{r}'\boldsymbol{R}^{-1}\mathbf{1})^2\}. \tag{9}$$

As noted before, when $\mu(\boldsymbol{x})$ is a constant, the MSPE of ordinary kriging will be smaller than that of limit kriging. But when $\mu(\boldsymbol{x})$ is not a constant, the bias term can dominate

the variance term and the MSPE of limit kriging can become smaller. To demonstrate this, suppose there is a linear trend, $\mu(x) = \beta x$ for $x \in \mathbb{R}$. Let the design points be $D = \{1, 2, \cdots, n\}$. For the correlation function $R(h) = \rho^{|h|}$, we obtain $\boldsymbol{R} = (\rho^{|i-j|})_{n\times n}$ and $\boldsymbol{r}(x)' = (\rho^{|x-1|}, \cdots, \rho^{|x-n|})$. Then as shown in the Appendix B, for an $x \in [1, 2]$

$$\begin{aligned} MSPE_{OK} &= \beta^2 \left\{ (1-\phi(x)) \frac{(n+1)}{2} + \frac{(1-2\rho)\rho^{2x-2} + 2\rho - \rho^2}{(1-\rho^2)\rho^{x-1}} - x \right\}^2 \\ &+ \tau^2(1+\rho) \left\{ \frac{1-\phi^2(x)}{1-\rho} + \frac{(1-\phi(x))^2}{n-(n-2)\rho} \right\}, \end{aligned} \tag{10}$$

and

$$MSPE_{LK} = \beta^2 \left\{ 1 + \frac{(1-\rho^{2x-2})}{(1-\rho)(1+\rho^{2x-3})} - x \right\}^2 + 2\tau^2 \frac{(1-\phi(x))}{(1-\rho)} \left\{ 1 + \rho - \frac{\rho(1-\phi(x))}{\phi^2(x)} \right\}, \tag{11}$$

where $\phi(x) = \boldsymbol{r}(x)'\boldsymbol{R}^{-1}\boldsymbol{1} = (\rho^{x-1} + \rho^{2-x})/(1+\rho)$. Since $\phi(x) < 1$ for all $x \in (1, 2)$ and $\rho \in [0, 1]$, $MSPE_{OK} \to \infty$ as $n \to \infty$, whereas the $MSPE_{LK}$ does not increase with n. Thus for n large enough, the MSPE of ordinary kriging can be much larger than that of limit kriging. Therefore, although the ordinary kriging is optimal under the model in (1), the limit kriging can outperform it under a different model. In summary, the ordinary kriging shall be preferred when the function is expected to vary around a constant and limit kriging shall be preferred when the function is expected to have some trends.

Note that the optimal predictor (that minimizes the MSPE) under model (7) is given by

$$\hat{y}(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}).$$

Since the true function is unknown to the investigator, $\mu(\boldsymbol{x})$ will also be unknown and therefore the above predictor cannot be used in practice. One practical implementation strategy is to let $\mu(\boldsymbol{x}) = \sum_{i=0}^{m} \beta_i f_i(\boldsymbol{x})$, where f_i's are some known functions and β_i's are unknown parameters. This is known as universal kriging in geostatistics. The optimal predictor can be obtained as follows. Usually $f_0(\boldsymbol{x}) = 1$. Let $\boldsymbol{f}(\boldsymbol{x}) = (1, f_1(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x}))'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_m)'$, and $\boldsymbol{F} = (\boldsymbol{f}(\boldsymbol{x}_1), \cdots, \boldsymbol{f}(\boldsymbol{x}_n))'$. Note that $\boldsymbol{F}$ is an $n \times (m+1)$ matrix. Then the universal kriging predictor is given by

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})'\hat{\boldsymbol{\beta}} + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}),$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{y}$. If the f_i's are able to capture some of the global trends in the true function, then this approach could give a better prediction than the ordinary kriging. The limit kriging can also be improved as follows. The idea is to do the limit kriging on $Y(\boldsymbol{x}) - \sum_{i=1}^{m} \beta_i f_i(\boldsymbol{x})$ instead on $Y(\boldsymbol{x})$. This gives us the universal version of the limit kriging as

$$\hat{y}(\boldsymbol{x}) = \tilde{\boldsymbol{f}}(\boldsymbol{x})'\tilde{\boldsymbol{\beta}} + \frac{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \tilde{\boldsymbol{F}}\tilde{\boldsymbol{\beta}})}{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}},$$

where $\tilde{\boldsymbol{f}}(\boldsymbol{x})$ is the $\boldsymbol{f}(\boldsymbol{x})$ without 1, $\tilde{\boldsymbol{F}}$ is the $\boldsymbol{F}$ matrix without the first column of 1's, and $\tilde{\boldsymbol{\beta}}$ is the last m components of $\hat{\boldsymbol{\beta}}$. The above predictor can be interpreted as follows. The universal kriging is used to remove the known trends and then the limit kriging is applied to get a protection against unknown trends.

Another generalization of the ordinary kriging is to use an intrinsic stationary process. The intrinsic stationary model (of order 0) is given by

$$Y(\boldsymbol{x}) = \mu + Z(\boldsymbol{x}),$$

where

$$\begin{aligned} E\{Z(\boldsymbol{x}+\boldsymbol{h}) - Z(\boldsymbol{x})\} &= 0, \\ \frac{1}{2}var\{Z(\boldsymbol{x}+\boldsymbol{h}) - Z(\boldsymbol{x})\} &= \gamma(\boldsymbol{h}). \end{aligned}$$

The $\gamma(\boldsymbol{h})$ is known as a variogram, which is more commonly used in geostatistics than a correlation function (correlogram). The advantage of the intrinsic model is that it can entertain a wider class of processes than the weak stationary processes. Note that a weak stationary process will always be an intrinsic stationary process, but the converse is not true. See Myers (1989) for an interesting discussion on stationarity and examples. Under the intrinsic model, the predictor that minimizes the MSPE is given by (see Cressie 1991)

$$\hat{y}(\boldsymbol{x}) = \hat{\mu}_{IK} + \boldsymbol{\gamma}(\boldsymbol{x})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{y} - \hat{\mu}_{IK}\mathbf{1}), \tag{12}$$

where $\hat{\mu}_{IK} = \mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{y}/\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}$, $\boldsymbol{\gamma}(\boldsymbol{x})' = (\gamma(\boldsymbol{x} - \boldsymbol{x}_1), \cdots, \gamma(\boldsymbol{x} - \boldsymbol{x}_n))$, and $\boldsymbol{\Gamma}$ is an $n \times n$ matrix with elements $\gamma(\boldsymbol{x}_i - \boldsymbol{x}_j)$. In the literature, the above predictor is also called the

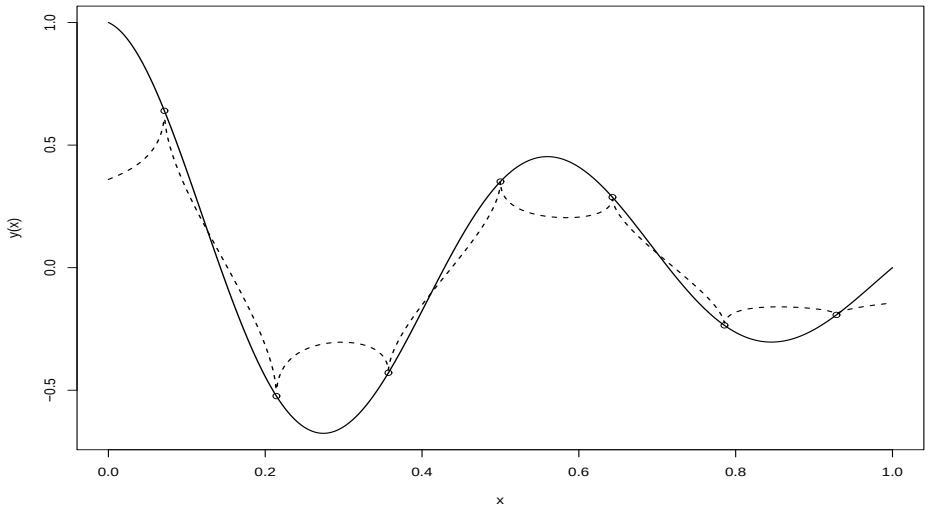

Figure 6: The plot of the function $y(x) = \exp(-1.4x)\cos(7\pi x/2)$ (solid) and the intrinsic kriging predictor (dashed) with $\gamma(h) = |h|^{0.5}$.

ordinary kriging predictor (or intrinsic random function kriging predictor of order 0). The difference between (4) and (12) is that the former uses correlation functions and the latter uses variograms. To avoid confusion, we will refer to (12) as intrinsic kriging predictor. If the process is weak stationary, then there exists a correlation function such that $\gamma(\boldsymbol{h}) = \sigma^2\{1 - R(\boldsymbol{h})\}$. Then, it can be shown that the prediction from the intrinsic kriging is exactly the same as that of the ordinary kriging (Christensen 1990). Thus the limit kriging can outperform the intrinsic kriging under the same situations that we have discussed earlier. It only remains to see whether the intrinsic kriging can perform better when we use variograms for which correlation functions do not exist. To check this consider the example in the introduction. Let $\gamma(h) = |h|^{0.5}$. Note that the power variograms are unbounded and cannot be obtained from correlation functions (see Wackernagel 2002). Figure 6 shows the plot of the predictor. We see that the intrinsic kriging exhibits the same problem as that of the ordinary kriging. Thus extending the weak stationarity to intrinsic stationarity and using unbounded variograms do not solve the problem of "pulling towards the mean".

3. SOME PROPERTIES

Consider an exponential correlation function of the form $R(\boldsymbol{h}) = \exp\{-\theta d(\boldsymbol{h})\}$ with $\theta > 0$, where $d(\boldsymbol{h}) = \sum_{i=1}^{p} h_i^{\alpha}$ for some $\alpha \in (0, 2]$. Let $D = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ be the design. Consider any point $\boldsymbol{x} \in \mathbb{R}^p$ and let $\boldsymbol{x}_N$ be the nearest point in D, i.e. $\min_i d(\boldsymbol{x} - \boldsymbol{x}_i) = d(\boldsymbol{x} - \boldsymbol{x}_N)$. Now $\boldsymbol{r}(\boldsymbol{x}) = \exp\{-\theta d(\boldsymbol{x} - \boldsymbol{x}_N)\}\boldsymbol{\delta}_N$, where $\boldsymbol{\delta}_N$ is an n-dimensional vector with elements $\exp[-\theta\{d(\boldsymbol{x} - \boldsymbol{x}_i) - d(\boldsymbol{x} - \boldsymbol{x}_N)\}]$. In the limit as $\theta \to \infty$, we have that $\boldsymbol{\delta}_N \to \boldsymbol{e}_N$. In addition, $\boldsymbol{R} \to \boldsymbol{I}_n$, the n-dimensional identity matrix. Then for the limit kriging

$$\lim_{\theta\to\infty} \hat{y}(\boldsymbol{x}) = \lim_{\theta\to\infty} \frac{\boldsymbol{\delta}_N' \boldsymbol{R}^{-1} \boldsymbol{y}}{\boldsymbol{\delta}_N' \boldsymbol{R}^{-1} \boldsymbol{1}} = \frac{\boldsymbol{e}_N' \boldsymbol{y}}{\boldsymbol{e}_N' \boldsymbol{1}} = y_N, \tag{13}$$

whereas for the ordinary kriging

$$\lim_{\theta\to\infty} \hat{y}(\boldsymbol{x}) = \begin{cases} \hat{\mu} & , \boldsymbol{x} \notin D \\ y(\boldsymbol{x}) & , \boldsymbol{x} \in D \end{cases}.$$

Thus, for the limit kriging the predicted value is the closest function value in the design, but for the ordinary kriging it is the mean value. This shows that the prediction from the limit kriging will be better if the function does not vary around a constant. This statement can be made more rigorous by comparing the MSPE of the two predictors under the model in (7). Taking the limit as $\theta \to \infty$ in (8) and (9), we obtain (for $\boldsymbol{x} \notin D$)

$$MSPE_{OK} = \{\bar{\mu} - \mu(\boldsymbol{x})\}^2 + \tau^2(1 + \frac{1}{n})$$

and

$$MSPE_{LK} = \{\mu(\boldsymbol{x}_N) - \mu(\boldsymbol{x})\}^2 + 2\tau^2,$$

where $\bar{\mu} = \sum_{i=1}^{n} \mu(\boldsymbol{x}_i)/n$. Clearly, although the second term is larger for the limit kriging compared to the ordinary kriging, the first term can be much smaller. Thus roughly speaking, the limit kriging performs better than the ordinary kriging when the variation in the function with respect to $\bar{\mu}$ is much larger than τ^2. As an example, consider a function $y(x) = \sin(2x)$ over $x \in [0, 1]$. Let $R(h) = \exp(-\theta h^2)$. Suppose again we use a 7-point design as in the example given in the introduction. The true function and the limit kriging predictors are

plotted in Figure 7 for various values of θ. We can see that the limit kriging predictor becomes like a step function as θ increases. The ordinary kriging predictors are also plotted in Figure 7. Clearly the limit kriging predictor is uniformly better than the ordinary kriging predictor for all values of θ. It is clear from Figure 7 that the variation in the limit kriging predictor with respect to the correlation parameter θ is smaller compared to that of the ordinary kriging predictor. Thus the limit kriging predictor is not greatly affected by a misspecification of θ.

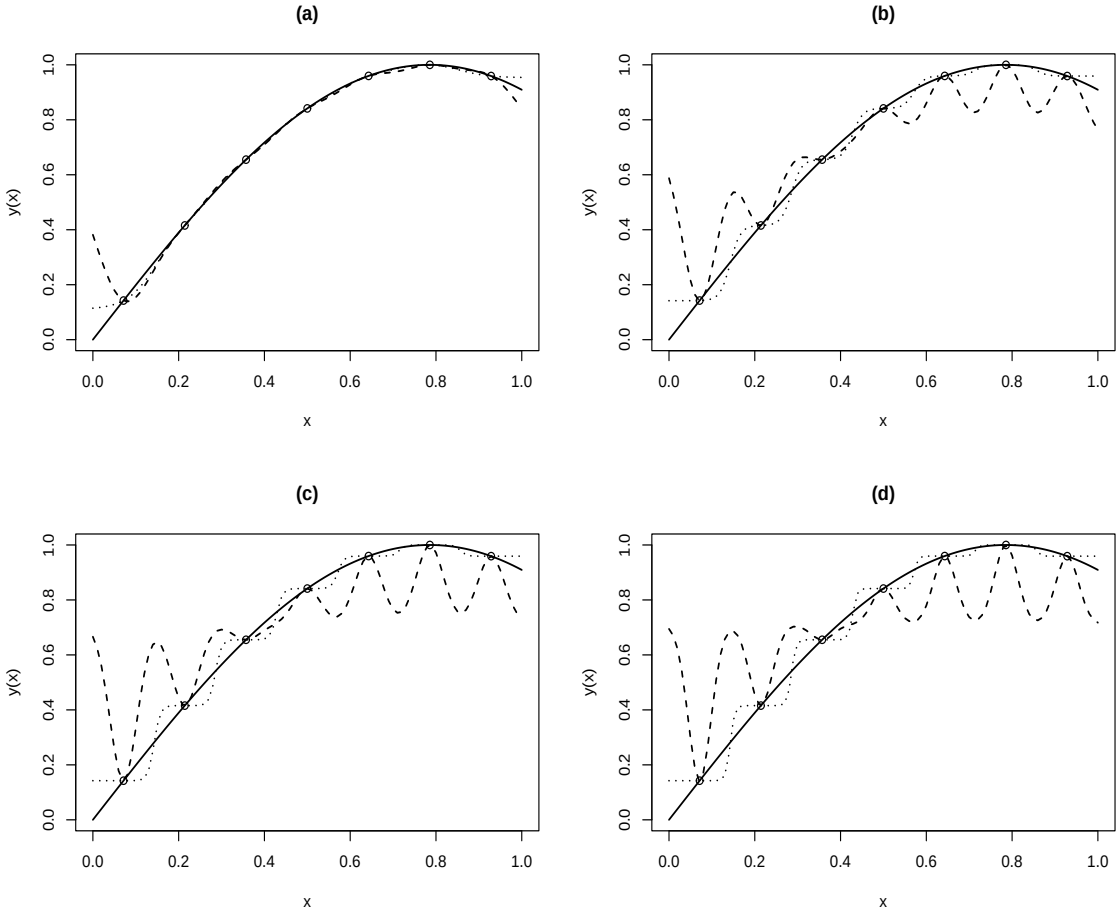

Figure 7: The plot of the function $y(x) = \sin(2x)$ (solid), the ordinary kriging predictor (dashed), and limit kriging predictor (dotted) with (a) $\theta = 100$, (b) $\theta = 300$, (c) $\theta = 500$, (d) $\theta = 700$.

The above discussion throws some light on the optimal sampling of points to observe

$y(\boldsymbol{x})$. Let $\mathcal{X}$ be the region of interest and $D = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ denotes the sampling design, where $\boldsymbol{x}_i$'s are distinct points in $\mathcal{X}$. Since for limit kriging as $\theta \to \infty$, $\hat{y}(\boldsymbol{x}) \to y_N$, the prediction will be good if $\boldsymbol{x}$ is close to $\boldsymbol{x}_N$. Therefore a good design criterion is to

$$\min_{D} \max_{\boldsymbol{x} \in \mathcal{X}} d(\boldsymbol{x} - \boldsymbol{x}_N),$$

which is the minimax distance criterion proposed by Johnson, Moore, and Ylvisaker (1990). Thus, the minimax distance design seems to be the most natural design to be used in conjunction with limit kriging. Note that for ordinary kriging $\hat{y}(\boldsymbol{x}) \to \hat{\mu}$ as $\theta \to \infty$ (for $\boldsymbol{x} \notin D$) and therefore it does not possess the above connection with the minimax distance design. Johnson, Moore, and Ylvisaker (1990) had shown that as the correlation becomes weak (as $\theta \to \infty$), a minimax distance design is asymptotically G-optimum (i.e. it minimizes the maximum prediction variance). Interestingly, as we have shown here that when $\theta \to \infty$ limit kriging is more meaningful to use than ordinary kriging with a minimax distance design. This could be considered as another justification for preferring limit kriging over ordinary kriging, because if one strongly believes in the usefulness of minimax distance design, then it is the limit kriging that produces the expected prediction behavior and not the ordinary kriging.

Although kriging is not recommended for extrapolation, it is interesting to see how the predictors behave as we try to extrapolate. For the ordinary kriging, as any of the components of $\boldsymbol{x}$ tends to $\pm\infty$, $\boldsymbol{r}(\boldsymbol{x}) \to 0$ and therefore $\hat{y}(\boldsymbol{x}) \to \hat{\mu}$. The behavior of the limit kriging is not that easy to see, because both numerator and denominator in (3) goes to 0. But by using (13), we can see that for θ large enough the extrapolated value is y_N. For small values of θ, the extrapolation depends on the correlation function. For $R(h) = \exp(-\theta h^2)$, $(x - x_i)^2 - (x - x_N)^2 \to \infty$ as $x \to \pm\infty$ and therefore, $\hat{y}(x) \to y_N$. This can be observed in Figures 3, 4, and 7.

4. EXAMPLES

Example 1: Consider the following test function used by Currin et al. (1991):

$$y(\boldsymbol{x}) = \{1 - \exp(-0.5/x_2)\}\frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20},$$

over $\boldsymbol{x} \in [0,1]^2$. We use a 4^2 full factorial design with levels $(0.125, 0.375, 0.625, 0.875)$, which is the minimax distance design in 16 points. Let the correlation function be

$$R(\boldsymbol{h}) = \exp(-\sum_{i=1}^{2} \theta_i h_i^2).$$

The maximum likelihood estimate of $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ can be obtained by minimizing the function (see Santner, Williams, and Notz 2003, page 66)

$$n \log\{\sigma^2(\boldsymbol{\theta})\} + \log det\{\boldsymbol{R}(\boldsymbol{\theta})\},$$

where $\sigma^2(\boldsymbol{\theta}) = \{\boldsymbol{y} - \hat{\mu}(\boldsymbol{\theta})\mathbf{1}\}'\boldsymbol{R}^{-1}(\boldsymbol{\theta})\{\boldsymbol{y} - \hat{\mu}(\boldsymbol{\theta})\mathbf{1}\}/n$ and $\hat{\mu}(\boldsymbol{\theta}) = \mathbf{1}'\boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{y}/\mathbf{1}'\boldsymbol{R}^{-1}(\boldsymbol{\theta})\mathbf{1}$. We obtain $\hat{\boldsymbol{\theta}} = (1.9046, 0.1725)$. Now the limit kriging and ordinary kriging predictors can be computed from (3) and (4). To compare their performance, 400 points were randomly sampled from $[0,1]^2$. Because the true function is known, the actual prediction errors can be computed for both the methods. Define

$$RMSPE = \left[\frac{1}{400}\sum_{i=1}^{400}\{\hat{y}(\boldsymbol{x}_i) - y(\boldsymbol{x}_i)\}^2\right]^{1/2}.$$

The RMSPE of the ordinary kriging is 1.094 and that of the limit kriging is 1.029. Thus the limit kriging gives a better prediction in this example. Now consider the case with $\theta_1 = \theta_2 = \theta_{eq}$. The RMSPE is calculated for different values of θ_{eq} (using the same set of 400 design points) and is given in Table 1. We can see that the limit kriging shows a remarkable improvement over the ordinary kriging at high values of θ_{eq}.

Example 2: Morris, Mitchell, and Ylvisaker (1993) used the following borehole function for the comparison of several methods in computer experiments. The function is given by

$$y = \frac{2\pi T_u(H_u - H_l)}{ln(r/r_w)\left[1 + \frac{2LT_u}{ln(r/r_w)r_w^2K_w} + \frac{T_u}{T_l}\right]},$$

Table 1: RMSPE values for the ordinary and limit kriging predictors.

	Example 1		Example 2	
θ_{eq}	Ordinary	Limit	Ordinary	Limit
1	0.998	0.997	25.49	22.66
10	1.097	1.094	42.82	32.37
100	1.830	1.180	43.50	36.66
1000	2.624	1.252	43.50	37.75

where the ranges of interest for the eight variables are $r_w : (0.05, 0.15)$, $r = (100, 50000)$, $T_u = (63070, 115600)$, $H_u = (990, 1110)$, $T_l = (63.1, 116)$, $H_l = (700, 820)$, $L = (1120, 1680)$, and $K_w = (9855, 12045)$. We scale these eight variables to $(0, 1)$ and denote them as $x_1, x_2, \cdots, x_8$. Morris, Mitchell, and Ylvisaker gave a 10-run design using the maximin distance criterion. In that design each variable takes only two levels 0 or 1. We use the same design except that the two levels are taken as 0.25 and 0.75. Like in their paper we use the Gaussian product correlation function $R(\boldsymbol{h}) = \exp(-\sum_{i=1}^{8} \theta_i h_i^2)$. By constraining each θ_i in $[0.01, 4.6]$, they obtained the maximum likelihood estimates of $\boldsymbol{\theta}$ from their design as $(0.71, 0.31, 0.01, 0.03, 0.01, 0.04, 0.21, 0.01)$. We use these estimates to compute the ordinary and limit kriging predictors. A random sample of 400 points were obtained from $[0, 1]^8$ and the RMSPE for both the predictors were computed. The RMSPE for the ordinary kriging was 11.97 and that of the limit kriging was 11.30. Thus in this example also the limit kriging does a better job than the ordinary kriging. Again consider the case with equal θ's. Table 1 also tabulates the RMSPE of both the predictors for different values of θ_{eq}. We see that the limit kriging performs better than the ordinary kriging.

5. CONCLUSIONS

In this article we have proposed a modified version of the kriging, known as limit kriging, and discussed some of its properties. The ordinary kriging predictor minimizes the mean squared prediction error under the model in (1), whereas the limit kriging predictor is not "optimal" in this sense. This could be the reason why it was never studied before, besides

having a long history for the kriging methodology. The optimality of the ordinary kriging holds under the weak stationarity assumption of the model which requires a constant mean. The true function can exhibit some global trends and therefore the stationarity assumption may not be realistic in practice. Universal kriging could be an option if the trends are known, but is rarely the case. In most practical situations ordinary kriging is used (see the comments in Currin et al. 1991 and Welch et al. 1992). The limit kriging is found to produce a superior performance over the ordinary kriging when the true function does not vary around a constant. This is because, the limit kriging uses a varying μ and therefore gives a better local fit than the ordinary kriging.

Global trends are quite common in engineering systems. For examples, in electroplating the plated thickness increases with plating time and current, in turning operation the surface roughness of the machined parts increases with feed rate, and so on. Whereas "regression effect" towards mean may be present in geological problems. Therefore, ordinary kriging may be more suitable for some geostatistical applications, but limit kriging may perform better in engineering applications.

Limit kriging also seems to be robust against the misspecifications of some correlation parameters. This is a great advantage because the investigator usually does not have precise information about the correlation parameters.

ACKNOWLEDGEMENTS

The author would like to thank the Associate Editor and two referees for their valuable comments and suggestions. The research was supported by the US National Science Foundation CAREER grant DMI-0448774.

APPENDIX A: BOUNDS FOR $\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}$

We claim that $0 < \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1} < 2$ for all $\boldsymbol{x} \in \mathcal{X}$ under some conditions on the correlation function. This is proved for some special cases given below.

Assume that $R(\boldsymbol{h}) \in (0, 1]$ for all $\boldsymbol{h}$. It is easy to verify the claim for $n = 1$ and $n = 2$, because for $n = 1$,

$$\boldsymbol{r}_1(\boldsymbol{x})'\boldsymbol{R}_1^{-1}\mathbf{1}_1 = R(\boldsymbol{x} - \boldsymbol{x}_1)$$

and for $n = 2$,

$$\boldsymbol{r}_2(\boldsymbol{x})'\boldsymbol{R}_2^{-1}\mathbf{1}_2 = \frac{R(\boldsymbol{x}-\boldsymbol{x}_1)+R(\boldsymbol{x}-\boldsymbol{x}_2)}{1+R(\boldsymbol{x}_1-\boldsymbol{x}_2)}.$$

Note that here we use subscripts for $\boldsymbol{r}(\boldsymbol{x})$, $\boldsymbol{R}$, and $\mathbf{1}$ to explicitly show their relation with the number of design points. The claim is difficult to prove analytically for $n \geq 3$. We prove the following case for sufficiently large n.

PROPOSITION 1. *If $R(\boldsymbol{h})$ is a continuous function in $\boldsymbol{h}$, then there exists an N and a design $D_N = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ such that $0 < \boldsymbol{r}_n(\boldsymbol{x})'\boldsymbol{R}_n^{-1}\mathbf{1}_n < 2$ for all $n > N$, $D_n \supset D_N$, and for all $\boldsymbol{x} \in \mathcal{X}$.*

PROOF: We have $\boldsymbol{r}_n(\boldsymbol{x}_i)'\boldsymbol{R}_n^{-1}\mathbf{1}_n = 1$ for all $i = 1, \cdots, n$. Since $R(\boldsymbol{h})$ is a continuous function in $\boldsymbol{h}$, $\boldsymbol{r}_n(\boldsymbol{x})'\boldsymbol{R}_n^{-1}\mathbf{1}_n$ is a continuous function in $\boldsymbol{x}$. Therefore, there exists an $\epsilon_i > 0$ such that

$$|\boldsymbol{r}_n(\boldsymbol{x})'\boldsymbol{R}_n^{-1}\mathbf{1}_n - 1| < 1 \text{ if } |\boldsymbol{x}-\boldsymbol{x}_i| < \epsilon_i,$$

for all $i = 1, \cdots, n$. Define the ball around $\boldsymbol{x}_i$ as $\{\boldsymbol{x} : |\boldsymbol{x}-\boldsymbol{x}_i| < \epsilon_i\}$. The proposition follows, because $\mathcal{X}$ can be covered with a finite number of balls. ♢

APPENDIX B: PROOF OF EQUATIONS (10) and (11)

Using a matrix inversion result (Cressie, 1991, page 133), we obtain

$$\boldsymbol{R}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & -\rho & 1 \end{pmatrix}.$$

For $\mu(x) = \beta x$, $\boldsymbol{\mu} = \beta(1, 2, \cdots, n)'$ and for $x \in [1, 2]$, $\boldsymbol{r}(x) = \rho^{1-x}(\rho^{2x-2}, \rho, \rho^2, \cdots, \rho^{n-1})'$. Thus we obtain

$$\begin{aligned} \mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1} &= \frac{n-(n-2)\rho}{1+\rho}, \\ \mathbf{1}'\boldsymbol{R}^{-1}\boldsymbol{\mu} &= \beta\frac{(n+1)}{2}\frac{(n-(n-2)\rho)}{1+\rho}, \end{aligned}$$

$$
\begin{aligned}
\boldsymbol{r}(x)'\boldsymbol{R}^{-1}\mathbf{1} &= \frac{\rho^{x-1}+\rho^{2-x}}{1+\rho}, \\
\boldsymbol{r}(x)'\boldsymbol{R}^{-1}\boldsymbol{\mu} &= \beta\rho^{1-x}\frac{(1-2\rho)\rho^{2x-2}+2\rho-\rho^2}{1-\rho^2}, \\
\boldsymbol{r}(x)'\boldsymbol{R}^{-1}\boldsymbol{r}(x) &= \frac{\rho^{2x-2}-2\rho^2+\rho^{4-2x}}{1-\rho^2}.
\end{aligned}
$$

Substituting the above expressions in (8) & (9) and simplifying, we obtain (10) & (11). $\diamondsuit$

REFERENCES

Chilès, J-P. and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, New York: Wiley.

Christensen, R. (1990), "The Equivalence of Predictions from Universal Kriging and Intrinsic Random-Function Kriging," *Mathematical Geology*, 22, 655-664.

Cressie, N. A. (1993), *Statistics for Spatial Data.* New York: Wiley.

Currin, C., Mitchell, T. J., Morris, M. D. and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments," *Journal of American Statistical Association*, 86, 953-963.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, 26, 131-148.

Laslett, G. M. (1994), "Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications," *Journal of the American Statistical Association* 89, 391-400.

Matheron, G. (1963), "Principles of Geostatistics," *Economic Geology*, 58, 1246-1266.

Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993), "Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction," *Technometrics*, 35, 243-255.

Myers, D. E. (1989), "To Be or Not to Be ... Stationary? That Is the Question," *Mathematical Geology*, 21, 347-362.

Olea, R. A. (1999), *Geostatistics for Engineers and Earth Scientists*, Boston: Kluwer Academic Publishers.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409-423.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer.

Wackernagel, H. (2002), *Multivariate Geostatistics*, New York: Springer.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992), "Screening, Predicting, and Computer Experiments," *Technometrics*, 34, 15-25.