Semiparametric Bootstrapping For estimating Parameters in Kriging model for Deterministic Simulations

Elmanani Simamora, Susiana and Eri Widyastuti
Matematika FMIPA Universitas Negeri Medan, Jalan Willem Iskandar Pasar V Medan, Indonesia
elmanani_simamora@unimed.ac.id

Abstract. In practice, the parameters in the Kriging model are unknown, but they can be estimated based on the behaviour of the observed data. Parameters in the Kriging model can be estimated based on the consideration of Regression-Kriging models. Regression-Kriging models are Universal Kriging models with polynomials of degree zero (Ordinary Kriging), one (Universal Kriging with degree one), or two (Universal Kriging with degree two). A new method for estimating the parameters in the Kriging model with semiparametric bootstrapping is proposed in this paper. The semiparametric bootstrapping procedure works by combining the bootstrap method and Kriging.

1. Introduction
Classic Kriging in deterministic simulations performs only one input data design that provides simulated output data while ignoring random errors. This means that the number of run for the same input data provides the same output data from the model (Lophaven et al. [1]). Kleijnen [2] states that deterministic simulations only consider a single run for a given scenario using the same combination of input data. Kleijnen and Mehdad [3] affirmed that a deterministic simulation produces a single output for repetitive runs with a given input. Neglecting error in the output data produces exact interpolations of Kriging in deterministic simulations, i.e., the output values match predictions.

The Kriging model parameters consist of correlation parameters, regression parameters, and a process variance parameter. In practice, parameters in the Kriging model are unknown, but they can be estimated from input/output data. Classic Kriging uses the maximum likelihood method to estimate the parameters in the Kriging model. Using estimated Kriging parameters in predictions gives an estimate of the Kriging variance that underestimates the true variance (Den Hertog et al. [4]). For convenience, we refer to this new terminology of parametric Kriging bootstrapping variance as the true Kriging variance. Den Hertog et al. [4] confirmed that there was a mistake in the Classic Kriging formula and provided three algorithms for estimating the Kriging variance using bootstrapped Kriging. The results of the simulation indicate that the estimated variance in Classic Kriging significantly underestimates the estimated variance from parametric Kriging bootstrapping variance. According to [5] the estimated variance in Classic Kriging is biased as it overlooks the randomness of maximum likelihood estimators.

Simamora et al. [6] proposed an estimation of Kriging variance using semiparametric bootstrapping. Next, Simamora et al. [7-8] conducted a comparative study of parametric and
semiparametric bootstrapping based only on simulation results and analytic study. Kleijnen and Mehdad [3] offered a comprehensive future study between [7], [9], [10], and [11].

The terminology of Regression-Kriging can be found in [12], of which a model was adopted from [13]. Regression-Kriging is an interpolation technique that provides the best linear unbiased prediction (BLUP), which combines regression and Kriging in spatially correlated data. Regression-Kriging prediction is a decomposition between deterministic and stochastic parts in a Kriging metamodel. The deterministic part was discussed by [14] as a trend model in which the polynomial degree was given while the stochastic process had zero mean and constant variance. Lophaven et al. [1] provided Regression-Kriging models in deterministic simulations based on the polynomials of degree zero, one, or two. The polynomial of degree zero in geostatistics refers to Ordinary Kriging while polynomials of degree one and two are referred to as Universal Kriging with degrees one and two.

This article, inspired by [6–8], proposes a new method that aims to estimate the parameters of the Kriging model based on semiparametric bootstrapping. The similarity between the Kriging model parameter distribution and normal distribution is determined using bootstrap replication. The composition of this article is as follows. Background, aims, and objectives of the study are presented in Section 1. Section 2 provides a summary of the Classic Kriging theory. A new method for estimating Kriging model parameters based on semiparametric bootstrapping (bootstrapped Kriging) is presented in Section 3. Section 4 presents the result of the deterministic simulation of the test function that were conducted to reveal the behavior of the Kriging model parameter estimation based on semiparametric bootstrapping. Bootstrap replication was used to determine whether or not the sampling distribution of each estimator in the Kriging model is close to a normal distribution. In the last section, we will present conclusions relating to the Classic Kriging estimator and semiparametric bootstrapping based on deterministic simulations.

2. Classic Kriging

In a deterministic simulation model, univariate Kriging assumes an output \( y(x) \in \mathbb{R} \) as crystallization of a stochastic process \( Y(x) \) with \( \mathbb{R} \) (set of real numbers in one dimension). An input \( x = (x_1, x_2, \ldots, x_d) \) is defined as a \( d \)-dimensional location. The stochastic process \( Y(x) \) is stated as the sum of the parts of a linear regression model with random parts. Let \( g(x) = [g_1(x), \ldots, g_p(x)]^T \) and \( \beta = [\beta_1, \ldots, \beta_p]^T \) be function vectors and regression parameters based on linear regression models, respectively, for polynomials of degrees zero, one, and two; the details of each are discussed in Section 2.2. The Regression-Kriging model (metamodel) according to [1] can be expressed as

\[
Y(x) = g^T(x)\beta + Z(x)
\]

Equation (1) was also discussed by Helg [12] and is the sum of deterministic parts \( \sum_{i=1}^p g_i(x)\beta_i \) and a stochastic part \( Z(x) \).

The stochastic process \( Z(x) \) is assumed to have \( \mathbb{E}[Z(x)] = 0 \) and the process variance is \( \mathbb{E}[Z(x)Z(x')] = \sigma^2 \). The covariance of two different inputs \( x \) and \( t \) is

\[
C(\theta, x, t) = \sigma^2 R(\theta, x, t)
\]

where \( R(\theta, x, t) \) is correlation of \( x \) and \( t \), while \( \theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d \) is a vector of correlation parameters. The behavior of \( R(\theta, x, t) \) was determined from the correlation and distance parameters (Lophaven et al. [1]). We used the following Gaussian correlation model:

\[
R(\theta, x, t) = \prod_{i=1}^{d} e^{-|\theta_i|x_i - t_i|^2}.
\]

Suppose the input data \( X = [x_1, \ldots, x_n] \) is an expansion of \( n \) Latin hypercube designs (LHD), as introduced by [15], and symmetric Latin hypercube designs (SLHD), as introduced by [16]. Only a single run of the deterministic simulation is required to obtain the input data. The output data \( Y_X = [y(x_1), \ldots, y(x_n)]^T \) from the test function is the realization of the stochastic vector \( Y = [Y(x_1), \ldots, Y(x_n)]^T \). The \( n \times p \) matrix \( G \) contains \( g^T(x_i) = g_j(x_i) \) for \( i = 1, \ldots, n \) and \( j = 1, \ldots, p \).
The correlation matrix between each \(Z(x_i)\) on the location design is

\[
\mathbf{K} = \begin{bmatrix}
\mathcal{R}(\theta, x_1, x_1) & \cdots & \mathcal{R}(\theta, x_1, x_n) \\
\vdots & \ddots & \vdots \\
\mathcal{R}(\theta, x_n, x_1) & \cdots & \mathcal{R}(\theta, x_n, x_n)
\end{bmatrix}
\]

(5)

\(\mathbf{K}\) matrix in Eq. (5) is a symmetric matrix. From Eq. (5), one obtains the covariance matrix between each \(Z(x_i)\):

\[
\mathbf{C} = \sigma^2 \begin{bmatrix}
\mathcal{R}(\theta, x_1, x_1) & \cdots & \mathcal{R}(\theta, x_1, x_n) \\
\vdots & \ddots & \vdots \\
\mathcal{R}(\theta, x_n, x_1) & \cdots & \mathcal{R}(\theta, x_n, x_n)
\end{bmatrix}
\]

(6)

In general, the parameters \(\theta, \beta,\) and \(\sigma^2\) are unknown but can be estimated based on the behaviour of the input/output data. We used the maximum likelihood method to estimate \(\theta, \beta,\) and \(\sigma^2\).

2.1. Maximum Likelihood Parameter Estimator in the Kriging Model

\(Z(x)\)in Eq. (1) is a spatially correlated Gaussian process in a deterministic simulation. The maximum likelihood method can be used to determine the estimator from \(\theta, \beta,\) and \(\sigma^2\) (Cressie [17]). The maximum likelihood method is used to maximize the following likelihood function:

\[
L(\theta, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} (y_x - G\beta)^T \mathbf{K}^{-1} (y_x - G\beta) \right\},
\]

(7)

where \(|\mathbf{K}|\) is the determinant of \(\mathbf{K}\) and depends on \(\theta, \beta,\) and \(\sigma^2\). Bain and Engelhardt [18] affirmed that maximizing Eq. (7) also will maximize the logarithm of the likelihood (log-likelihood),

\[
\ln L(\theta, \beta, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2} \ln |\mathbf{K}| - \frac{1}{2\sigma^2} (y_x - G\beta)^T \mathbf{K}^{-1} (y_x - G\beta).
\]

(8)

The necessary condition for maximizing the log-likelihood is to set the first derivative of Eq. (8) with respect to \(\beta\) to a \(1 \times p\) zero vector:

\[
\frac{\partial}{\partial \beta} L(\theta, \beta, \sigma^2) = \frac{1}{\sigma^2} (y_x - G\beta)^T \mathbf{K}^{-1} \mathbf{G} = 0.
\]

(9)

Some matrix operations in Eq. (9) yields Eq. (10) as follows:

\[
y_x^T \mathbf{K}^{-1} \mathbf{G} = \mathbf{G}^T \mathbf{K}^{-1} \mathbf{G} = 0
\]

(10)

Equation (10) is a regression parameter estimator based on linear regression models for polynomials of degree zero, one, and two. In the same way, the log-likelihood can be maximized by taking the first derivative of Eq. (8) with respect to \(\sigma^2\) while taking \(\hat{\beta}\) as defined in Eq. (10):

\[
\frac{\partial}{\partial \sigma^2} \ln L(\theta, \beta, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (y_x - G\beta)^T \mathbf{K}^{-1} (y_x - G\beta) = 0
\]

(11)

The maximize value of \(\hat{\sigma}^2\) in the log-likelihood function is found by taking the second derivative of \(\ln L(\theta, \beta, \sigma^2)\) with respect to \(\sigma^2\) and setting it to zero:

\[
\frac{\partial^2}{\partial \sigma^2} \ln L(\theta, \beta, \sigma^2) = n - \frac{1}{2\sigma^4} (y_x - G\beta)^T \mathbf{K}^{-1} (y_x - G\beta).
\]

(12)

Replacing \(\sigma^2\) with \(\hat{\sigma}^2\) yields \(\hat{\sigma}^2 = \frac{1}{n} (y_x - G\beta)^T \mathbf{K}^{-1} (y_x - G\beta) = \frac{1}{n} (y_x - G\beta)^T \mathbf{K}^{-1} \mathbf{G} \beta \), thus, the value \(\hat{\sigma}^2\) will maximize the log-likelihood function. The second derivative of \(\ln L(\theta, \beta, \sigma^2)\) with respect to \(\beta\) is the matrix

\[
\frac{\partial^2}{\partial \beta^2} \ln L(\theta, \beta, \sigma^2) = -\frac{1}{\sigma^4} (y_x - G\beta)^T \mathbf{K}^{-1} \mathbf{G} \beta.
\]

(13)

Replacing \(\sigma^2\) with \(\hat{\sigma}^2\) yields \(\hat{\beta} = \frac{1}{n} (y_x - G\beta)^T \mathbf{K}^{-1} \mathbf{G} \beta\). As \(\hat{\sigma}^2\) is always positive, \(\mathbf{G}^T \mathbf{K}^{-1} \mathbf{G}\) must be positive definite for a given \(\hat{\beta}\). Suppose any vector \(y = (y_1, \ldots, y_n)^T\) causes \(\mathbf{G}^T \mathbf{K}^{-1} \mathbf{G}\) to be a positive definite matrix. Using a Cholesky decomposition \(\mathbf{K} = \mathbf{L} \mathbf{L}^T\), where \(\mathbf{L}\) is a lower triangular matrix, yields
Suppose \( L^{-1}G = \begin{bmatrix} \mathbf{I}^{-1}_{1} & 0 & \cdots & 0 \\ \mathbf{I}^{-1}_{2} & \mathbf{I}^{-1}_{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}^{-1}_{n} & \mathbf{I}^{-1}_{n} & \cdots & \mathbf{I}^{-1}_{n} \end{bmatrix} \begin{bmatrix} \mathbf{g}^T(x_1) \\ \mathbf{g}^T(x_2) \\ \vdots \\ \mathbf{g}^T(x_n) \end{bmatrix} = \begin{bmatrix} \mathbf{I}^{-1}_{1} \mathbf{g}^T(x_1) \\ \mathbf{I}^{-1}_{2} \mathbf{g}^T(x_2) + \mathbf{I}^{-1}_{2} \mathbf{g}^T(x_2) \\ \vdots \\ \mathbf{I}^{-1}_{n} \mathbf{g}^T(x_n) + \cdots + \mathbf{I}^{-1}_{n} \mathbf{g}^T(x_n) \end{bmatrix} \) with \( \mathbf{I}^{-1}_{k} \) is the element of the matrix \( \mathbf{L}^{-1} \). We define

\[
\begin{align*}
\mathbf{l}^{-1}_{1} \mathbf{g}^T(x_1) &= \mathbf{g}^T(x_1) \\
\mathbf{l}^{-1}_{2} \mathbf{g}^T(x_2) + \mathbf{l}^{-1}_{2} \mathbf{g}^T(x_2) &= \mathbf{g}^T(x_2) \\
\vdots & \\
\mathbf{l}^{-1}_{n} \mathbf{g}^T(x_n) + \cdots + \mathbf{l}^{-1}_{n} \mathbf{g}^T(x_n) &= \mathbf{g}^T(x_n),
\end{align*}
\]

yielding the \( n \times p \) matrix \( \mathbf{L}^{-1}G = \begin{bmatrix} \mathbf{g}^T(x_1) \\ \mathbf{g}^T(x_2) \\ \vdots \\ \mathbf{g}^T(x_n) \end{bmatrix} \). Multiplying \( \mathbf{y}^T \) by \( \mathbf{L}^{-1}G \) yields the \( 1 \times n \) matrix \( \mathbf{y}^T(G^T \mathbf{R}^{-1} \mathbf{G}) = \begin{bmatrix} \mathbf{y}^T \mathbf{g}(x_1) \\ \mathbf{y}^T \mathbf{g}(x_2) \\ \vdots \\ \mathbf{y}^T \mathbf{g}(x_n) \end{bmatrix} \), which is used to calculate the transpose matrix and obtain \( \mathbf{y}^T \mathbf{R}^{-1} \mathbf{y} = \mathbf{y}^T \mathbf{g}(x_1) \mathbf{y}, \mathbf{g}(x_2) \mathbf{y}, \mathbf{g}(x_n) \mathbf{y} \). Every element of \( \mathbf{y}^T \mathbf{g}(x_i) = \mathbf{g}^T(x_i) \mathbf{y} \) is a scalar. Note again that

\[
\mathbf{y}^T G^T \mathbf{R}^{-1} G = \mathbf{y}^T (L^{-1}G)^T (L^{-1}G) \mathbf{y} = \sum_{i=1}^{n} (\mathbf{y}^T \mathbf{g}(x_i))^2 > 0. \tag{12}
\]

Based on (12), one can conclude that \( G^T \mathbf{R}^{-1} G \) is positive definite. If matrix \( G^T \mathbf{R}^{-1} G \) is positive definite, then matrix

\[
\mathbf{L}(\mathbf{R}^{-1/2} \mathbf{y}^T \mathbf{R}^{-1/2}) \mathbf{L}(\mathbf{R}^{-1/2} \mathbf{y}^T \mathbf{R}^{-1/2})
\]

is negative definite. The sufficient conditions regarding the estimator of the maximum likelihood with respect to \( \hat{\mathbf{R}} \) have been fulfilled.

Using Eqs. (8), (10) and (11), one obtains

\[
\ln \mathbf{L}(\mathbf{R}, \hat{\mathbf{R}}, \hat{\mathbf{y}}^2) = -\frac{\mathbf{n}}{2} \left( \ln \hat{\mathbf{y}}^2 \mathbf{L}(\mathbf{R})^{-1} + 1 \right). \tag{13}
\]

The estimator for the maximum likelihood for \( \hat{\mathbf{R}} \) is found by maximizing (13) via iteration from an initial value of \( \hat{\mathbf{R}} \). Maximizing (13) means finding \( \hat{\mathbf{R}} = \min_{\mathbf{R}} \left( \hat{\mathbf{y}}^2 \mathbf{L}(\mathbf{R})^{-1} \right) \). Toolbox DACE, available from Lophaven et al. [1], was used to search \( \hat{\mathbf{R}} \). The estimators \( \hat{\mathbf{R}} \) and \( \hat{\mathbf{y}}^2 \) depend on \( \hat{\mathbf{R}} \). Substituting \( \hat{\mathbf{R}} \) into (5) and (6) yields \( \hat{\mathbf{R}} \) and \( \hat{\mathbf{C}} \) as estimators of the correlation and covariance matrices between each \( \mathbf{Z}(\mathbf{x}_i) \) at the planned location.

2.2 Regression-Kriging Models

Lophaven et al. [1] provides linear regression models based on polynomials of degree zero, one, and two. The combination of regression and Kriging with spatially correlated data yields the following Regression-Kriging models (Kriging metamodels):

- Ordinary Kriging

Ordinary Kriging is the most widely used prediction technique in geostatistics, and it assumes constant variance and unknown mean. Ordinary Kriging is a constant trend model (Helg [12]). Lophaven et al. [1] mentioned linear regression models based on degree zero polynomials. Defining \( p = 1 \) yields the deterministic part of the trend model in the form of a constant; Eq. (1) can be written as

\[
Y(\mathbf{x}) = g_1(\mathbf{x}) \beta_1 + Z(\mathbf{x}) = \beta + Z(\mathbf{x}), \tag{14}
\]

with \( g_1(\mathbf{x}) = 1 \). In Eq. (4) becomes a column vector \( \mathbf{1} = [1, \cdots, 1]^T \). Equation (14) provides Ordinary Kriging prediction at an untried location \( \mathbf{x}_0 \) (see the Appendix),

\[
\hat{Y}(\mathbf{x}_0) = \hat{\beta} + \hat{r}(\mathbf{x}_0) \hat{\mathbf{R}}^{-1} (\mathbf{y}_\mathbf{x} - \mathbf{1} \hat{\beta}_1). \tag{15}
\]

- Universal Kriging with Degree One
Universal Kriging with degree one considers the trend model to be a linear function (polynomial of degree one). Lophaven et al. [1] defined $p = d + 1$. Equation (1) for Universal Kriging with degree one is written

$$Y(\mathbf{x}) = \sum_{i=1}^{p} g_i(\mathbf{x}) \beta_i + Z(\mathbf{x})$$

$$Y(\mathbf{x}) = g_1(\mathbf{x}) \beta_1 + g_d(\mathbf{x}) \beta_d + \cdots + g_p(\mathbf{x}) \beta_p + Z(\mathbf{x})$$

(16)

with $g_1(\mathbf{x}) = 1, g_d(\mathbf{x}) = x_1, \cdots, g_p(\mathbf{x}) = x_d$. Given Eq. (4) becomes an $n \times (d + 1)$ matrix. For example, if one takes a two-dimensional input, then $Y(\mathbf{x}) = \beta_1 + x_1 \beta_2 + x_2 \beta_3 + Z(\mathbf{x})$ and

$$G = \begin{bmatrix} 1 & x_1 & x_1 & x_1 \\ 1 & x_2 & x_2 & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n & x_n \end{bmatrix}$$

Equation (16) provides a prediction for Universal Kriging with degree one:

$$\hat{y}_k(\mathbf{x}_0) = \hat{\beta}_1 + x_1 \hat{\beta}_2 + x_2 \hat{\beta}_3 + \mathbf{y}(\mathbf{x}) \mathbf{R}^{-1}(\mathbf{y}_X - G \mathbf{\hat{y}}).$$

(17)

• Universal Kriging with Degree Two

Universal Kriging with polynomial of degree with $p = \frac{1}{2} (d + 1)(d + 2)$ yields a quadratic trend. Equation (1) becomes

$$Y(\mathbf{x}) = \sum_{i=1}^{d+1} g_i(\mathbf{x}) \beta_i + Z(\mathbf{x})$$

$$Y(\mathbf{x}) = g_1(\mathbf{x}) \beta_1 + g_{d+1}(\mathbf{x}) \beta_{d+1} + \cdots + g_p(\mathbf{x}) \beta_p + Z(\mathbf{x}),$$

(18)

where $g_1(\mathbf{x}) = 1, g_{d+1}(\mathbf{x}) = x_1, \cdots, g_{d+2}(\mathbf{x}) = x_{d+1}, \cdots, g_{d+1}(\mathbf{x}) = x_{d+2}, \cdots, g_p(\mathbf{x}) = x_d$. Given Eq. (4) becomes an $n \times (\frac{1}{2} (d + 1)(d + 2))$ matrix. Suppose, the input point is two dimensions with $d = 2$, i.e., $p = 6$. Therefore, Eq. (18) becomes $Y(\mathbf{x}) = \beta_1 + x_1 \beta_2 + x_2 \beta_3 + x_1^2 \beta_4 + x_2^2 \beta_5 + x_1x_2 \beta_6 + Z(\mathbf{x})$ and

$$G = \begin{bmatrix} 1 & x_1 & x_1 & x_1^2 & x_1 & x_1^2 \\ 1 & x_2 & x_2 & x_2^2 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n & x_n^2 & x_n & x_n^2 \end{bmatrix}$$

Equation (18) yields a prediction for the Universal Kriging with degree two:

$$\hat{y}_k(\mathbf{x}_0) = \hat{\beta}_1 + x_1 \hat{\beta}_2 + x_2 \hat{\beta}_3 + x_1^2 \hat{\beta}_4 + x_2^2 \hat{\beta}_5 + x_1x_2 \hat{\beta}_6 + \mathbf{y}(\mathbf{x}) \mathbf{R}^{-1}(\mathbf{y}_X - G \mathbf{\hat{y}}).$$

(19)

3. Bootstrapped Kriging

Kleijnen and Mehdad [5] affirmed that the variance estimation of Classic Kriging is biased. This occurs because classical kriging predictions become nonlinear predictions such that bootstrapping can be used to estimate correlation parameters, Regression-Kriging parameters, and the process variance parameter. The bootstrap method applies randomness to the output data. Den Hertog et al. [4], as a pioneer, introduced parametric bootstrapping for estimating the actual Kriging variance. Parametric bootstrapping works by generating new output data (bootstrap samples) that mimics the original output data. The new output data is used in the Classic Kriging prediction, referred to as bootstrapped kriging, thus it provides the maximum likelihood estimation. The prediction from Classic Kriging overlooks randomness of the maximum likelihood estimator.

Simamora et al. [6] introduced semiparametric bootstrapping to obtain bootstrap samples. The study of Simamora et al. [6] related the estimated Kriging variance as an alternative to determine the actual kriging variance. Semiparametric bootstrapping considers a sample using the original output data. Simamora et al. [7] conducted a comparative study of the estimated variance from semiparametric and parametric Kriging bootstrapping. The simulation results indicate that the estimation from semiparametric bootstrapping is smaller than that from parametric bootstrapping. Simamora et al. [8] then examined the asymptotic properties of estimation in semiparametric bootstrapping.

This study is an extension to the work of Simamora et al. [6–8], which offers a new method for estimating parameters in the Kriging model, namely correlation parameters, Regression-Kriging...
parameters, and the process variance parameter. This study focuses on the behaviour of parameter estimators in the Regression-Kriging model. A new method for estimating parameters in the Kriging model is contained in the following algorithm.

3.1. Estimator Algorithm in Semiparametric Bootstrapping
The algorithm in the semiparametric bootstrap method for estimating Kriging model parameters in deterministic simulations is as follows:

1. Specify input/output data based on:
   - Input data is generated from the computer experimental design.
   - Output data is generated from multimodal test functions.
2. The correlation parameter, Regression-Kriging parameter, and process variance parameter are estimated with maximum likelihood methods based on input/output data as shown in Section 2.1.
3. Data are centralized by separating each output from the mean of the output data as follows:
   \[
   \overline{y}_X = y_X - \overline{y}_X, \quad \text{with} \quad \overline{y}_X = \frac{\sum_{i=1}^{n} y(x_i)}{n}
   \]
4. The estimator of the covariance matrix \( \overline{C} \) is determined from the input/output data, as shown in the final part of Section 2.1.
5. The lower triangular matrix \( L \) is determined using a Cholesky decomposition \( \overline{C} \).
6. \( U_n = L^{-1} \overline{y}_X = (u_1, u_2, \ldots, u_n)^T \), which is an uncorrelated sample, is calculated. Theorem 3.1 will prove that the vector \( U_n \) has uncorrelated (or statistically independent) components.
7. The uncorrelated bootstrap samples \( U_n^{(b)} \) for \( b = 1, 2, \ldots, B \) are retrieved from step 6.
8. The correlated bootstrap sample \( \mathbf{y}_X^{(b)} = \mathbf{\hat{\theta}} + \mathbf{L} \mathbf{U}_n^{(b)} \) is determined with \( \mathbf{\hat{\theta}} = \mathbf{G} \mathbf{\hat{\beta}} + \mathbf{\hat{R}}^T(\mathbf{x}) \mathbf{R}^{-1}(\mathbf{y}_X - \mathbf{G} \mathbf{\hat{\beta}}) \), which is then referred to as a quasi-bootstrap sample from the observation data.
9. The correlation parameters, Regression-Kriging, and process variance parameters are determined from each correlated bootstrap sample \( \mathbf{\hat{\theta}}^{(b)}, \mathbf{\hat{\beta}}^{(b)}, \mathbf{\hat{\sigma}}^{(b)} \) from step 8. This is a process of matching the Kriging model based on correlated bootstrap samples.
10. The bootstrap estimators for the parameters \( \mathbf{\theta}, \mathbf{\beta}, \) and \( \sigma^2 \) are specified, which are the averages of the bootstrap estimators:
   \[
   \overline{\mathbf{\theta}} = \frac{\sum_{b=1}^{B} \mathbf{\hat{\theta}}^{(b)}}{B}, \quad \overline{\mathbf{\beta}} = \frac{\sum_{b=1}^{B} \mathbf{\hat{\beta}}^{(b)}}{B}, \quad \overline{\mathbf{\sigma}^2} = \frac{\sum_{b=1}^{B} \mathbf{\hat{\sigma}}^{(b)}}{B}.
   \]

A new theorem regarding certain properties of \( U_n \) in step 6 is presented in Theorem 3.1 as follows.

**Theorem 3.1** If \( L \) is a lower (upper) triangular matrix based on a Cholesky decomposition \( \overline{C} = \mathbf{L} \mathbf{L}^T \) and \( U_n = L^{-1} \overline{y}_X \), then the vector of \( U_n \) satisfies

1. \( \mathbb{E}(U_n) = \mathbf{L}^{-1} \overline{\mathbf{G}} \overline{\mathbf{\beta}} \).
2. \( \text{Var}(U_n) = \mathbf{I} \), where \( \mathbf{I} \) is an \( n \times n \) identity matrix.

**Proof**

The output data \( \mathbf{y}_X = [y(x_1), \ldots, y(x_n)]^T \) from Section 2 can be written as a matrix equation
\[
\mathbf{y}_X = \begin{bmatrix} y(x_1) \\ \vdots \\ y(x_n) \end{bmatrix} = \begin{bmatrix} \mathbf{g}^T(x_1) \mathbf{\beta} + \mathbf{Z}(\mathbf{x}_1) \\ \vdots \\ \mathbf{g}^T(x_n) \mathbf{\beta} + \mathbf{Z}(\mathbf{x}_1) \end{bmatrix} = \mathbf{G} \mathbf{\beta} + \mathbf{Z} \mathbf{(x)},
\]
where \( \mathbb{E}(\mathbf{Z} \mathbf{(x)}) = [\mathbb{E}(\mathbf{Z}(x_1)), \ldots, \mathbb{E}(\mathbf{Z}(x_n))]^T = \mathbf{0} \) yields the following properties
with degree one provide

\[
E(U_n) = \mathbf{L}^{-1}(\mathbf{y}_X) = \mathbf{L}^{-1}E(\mathbf{y}_X) = \mathbf{L}^{-1}E(\mathbf{b} + \mathbf{Z}(\mathbf{x})) = \mathbf{L}^{-1}\mathbf{b}.
\]

(2) \[
\text{Var}(U_n) = \text{Var}(U_n - E(U_n))(U_n - E(U_n))^T = \mathbf{L}^{-1}E(\mathbf{y}_X - \mathbf{L}^{-1}\mathbf{b})(\mathbf{y}_X - \mathbf{L}^{-1}\mathbf{b})^T \mathbf{L}^{-T} = \mathbf{L}^{-1}E(\mathbf{Z}(\mathbf{x})\mathbf{Z}(\mathbf{x})^T) \mathbf{L}^{-T}
\]

In view of \( E(\mathbf{Z}(\mathbf{x})\mathbf{Z}(\mathbf{x})^T) = \mathbf{C} = \mathbf{L}\mathbf{L}^T \), one obtains \( \text{Var}(U_n) = (\mathbf{L}^{-1}\mathbf{L})(\mathbf{L}^T \mathbf{L}^{-T}) = \mathbf{L}\text{If} \text{Var}(U_n) = \mathbf{I} \), where \( U_n \) is the sample, then the samples are independent and identically distributed.

4. Result of the Deterministic Simulations

The deterministic simulation takes one run for the planned location or input data. Kenny et al. [16] proposed using SLHD in place of LHD. The aim of SLHD is to offer more compromising design in terms of computational capabilities and design optimization. Using SLHD generates twenty-five point locations as input data with dimension \( d = 2 \), where each \( x_i = (x_{i1}, x_{i2}) \in [-5, 10] \times [0, 15] \subset \mathbb{R}^2 \). Twenty-one location points were substituted into the Branin test function,

\[
y(x_{i1}, x_{i2}) = \frac{x_{i2}^2}{4\pi^2} - \frac{5}{\pi} \cos x_{i1} - 6 \pm 10 \left(1 \mp \frac{1}{\pi}\right) \cos x_{i1} + 10, \quad i = 1, \ldots, 21,
\]

yielding the output data \( \mathbf{y}_X \).

Figure 1 shows a Classic Kriging prediction simulation with initial value of the correlation parameter \( \mathbf{\theta} = [2 \ 2] \) and \( \mathbf{\theta} = (\beta_1, \beta_2) \in (0, 0) \times [0, 15] \). The correlation model follows Eq. (3), which is the Gaussian. Figure 1 (a) shows a plot of the Branin test function. Figure 1 (b), (c), and (d) show predictions from Ordinary Kriging and Universal Kriging with polynomial of degrees one and two, respectively. All three predictions yield the same curve surfaces that are close to the true function. Deterministic simulations provide Kriging predictions at input points that give an estimate that is equal to the true function values \( \mathbf{\bar{y}}_h(x_i) = \mathbf{y}(x_i) \). This shows that predictions at input points have no errors [1].

![Figure 1](image1.png)

Figure 1. Simulation of Classic Kriging uses input data from SLHD generation and output data from Branin test function, (a) Function True, (b) Prediction of Ordinary Kriging, (c) Prediction of Universal Kriging with degree One, (d) Prediction of Universal Kriging with degree Two.

Table 1 shows the Classic Kriging method for estimating parameters in the Kriging model. The correlation parameters estimator in Ordinary Kriging and Universal Kriging with degree one provide
the same estimates but they do not assure that this always happens. In general, the three prediction models provide different parameter estimators in Kriging models. The Universal Kriging model with degree two produces an estimate of process variance that is relatively smaller than the Ordinary Kriging model and Universal Kriging model with degree one. The estimated process variance in Universal Kriging with degree one is smaller than that from Ordinary Kriging. This situation does not always ensure that Universal Kriging with degree two is always better than Ordinary Kriging and Universal Kriging with degree one.

Table 1. Parameter Estimation of Kriging Model with Maximum Likelihood Method

| Prediction               | Parameter Estimators | Regress-Kriging (β) | Process Variance (σ²) |
|-------------------------|----------------------|---------------------|-----------------------|
| Ordinary Kriging        | [0.8409 0.4585]      | 0.6688              | 3.441 × 10³           |
| Universal Kriging with  |                      |                     |                       |
| Degree One              |                      |                     |                       |
| Universal Kriging with  | [0.8409 0.4585]      | 0.6820              | 3.329 × 10³           |
| Degree Two              |                      | -0.2673             |                       |
|                         |                      | 0.2053              |                       |
| Universal Kriging with  | [1.0000 0.2500]      | -0.7374             | 1.807 × 10³           |
| Degree Two              |                      | -0.2689             |                       |
|                         |                      | 0.3682              |                       |
|                         |                      | 0.4157              |                       |
|                         |                      | 0.7539              |                       |
|                         |                      | 0.4230              |                       |

One way of looking at the quality of measure of the accuracy of the Regression-Kriging Models is based on the variance of Kriging predictions or often also called Kriging variance (see[4]). The process variance is not as a measure of the accuracy of the model although the process variance is directly proportional to the Kriging variance (see the equation for Kriging variance in the Appendix). We assume that the correlation parameter greatly influences the accuracy of the model. For the simplicity of observing the variance of kriging predictions of the Regression-Kriging Models, we take the one-dimensional inputs in [4] i.e function of expected waiting time in the steady-state of M/M/1 queuing model with the design of six input points spaced equally.

Figure 2. (a) Plot of function of expected waiting time in the steady-state of M/M/1 queuing model. (b) Kriging prediction based on the Regression-Kriging models. (c) Variance of Kriging prediction in untried based on the Regression-Kriging models.
Figure 2 (b) shows that prediction errors are relatively smaller than on based Universal Kriging with degree one and two so that the Kriging Variance is smaller than the two another models as shown by Figure 2 (c).

The simulation results in Classic Kriging for the Branin test function were used to execute the estimation algorithm for the maximum likelihood in semiparametric bootstrapping. Simulations were conducted using the Regression-Kriging model.

- Ordinary Kriging

Table 2 shows a bootstrapping estimate for the Kriging model parameters, where the number of bootstrap samples is \( B = 1000, B = 2000, \) and \( B = 3000 \). The Classic Kriging and bootstrap estimates provide different results for each model parameter, while the bootstrapping estimates for \( B = 1000, B = 2000, \) and \( B = 3000 \) are similar. The correlation parameters estimated with bootstrapping become closer to those from Classic Kriging, while the estimated parameters with Regression-Kriging and the process variances differ from Classic Kriging. The bootstrap estimation for the process variance parameter decreases and approaches the classical Kriging estimate as \( B \) increases.

Table 2. The Bootstrap Estimation for Parameters of the Kriging Model with \( B = 1000, B = 2000, \) and \( B = 3000 \) Using Prediction of Ordinary Kriging

| Estimation | Number of Bootstrap Samples (B) |
|------------|---------------------------------|
|            | 1000 | 2000 | 3000 |
| \( \hat{\theta}^2 \) | [0.90692, 0.51429] | [0.90118, 0.51478] | [0.90367, 0.51474] |
| \( \hat{\beta}^2 \) | -0.036582 | -0.033026 | -0.037966 |
| \( \hat{\sigma}^2 \) | 4511.223 | 4438.812 | 4421.823 |

Figure 3 (a) and (b) shows that the sampling distribution for correlation parameters based on bootstrapping is very different from a Gaussian distribution. Figure 3 (c) shows that sampling distribution for the Regression-Kriging parameters approaches a Gaussian distribution. Figure 3 (d) shows that the sampling distribution of the process variance does not approach a Gaussian distribution.

Figure 3. Sampling Distribution of Parameters of the Kriging Model Based on \( B = 3000 \) Bootstrap Replication with Prediction of Ordinary Kriging.
Universal Kriging with Degree One

Table 3 shows that the correlation parameters estimated with bootstrapping are exactly the same as those estimated with Classic Kriging. The model parameters estimated with bootstrapping do not change significantly as \( B \) increases. Bootstrap estimates for Regression-Kriging parameters yield results that are different from those estimated with Classic Kriging. The process variance parameters estimated with bootstrapping increase and approach the estimate from Classic Kriging as \( B \) increases.

**Table 3.** The Bootstrap Estimation for Parameters of the Kriging Model with \( B = 1000 \), \( B = 2000 \) and \( B = 3000 \) Using Prediction of Universal Kriging with Degree One.

| Estimation | \( \tilde{\theta} \) | Number of Bootstrap Samples (\( B \)) |
|------------|---------------------|-------------------------------------|
|            | 1000                | 2000                                | 3000                                |
| \( \hat{\theta}^* \) | \([1.03609, 0.56645]\) | \([1.0145, 0.56226]\)                | \([1.016640, 56260]\)               |
| \( \hat{\beta}_1 \)  | 0.009116            | 0.005335                            | 0.009424                            |
| \( \hat{\beta}_2 \)  | -0.02523            | -0.03018                            | -0.02583                            |
| \( \hat{\beta}_3 \)  | 0.008226            | 0.011766                            | 0.015061                            |
| \( \hat{\theta}_4 \)  | 2447.775            | 2521.622                            | 2538.040                            |

Figure 4 shows that the sampling distribution from each parameter based on bootstrapping allows one to come to the same conclusion as that from the Ordinary Kriging method. (1) The sampling distribution of correlation parameters is very different from a Gaussian distribution. (2) The sampling distribution for Regression-Kriging parameters approaches a Gaussian distribution. (3) The sampling distribution of the process variance is different from a Gaussian distribution.

**Figure 4.** Sampling Distribution of Parameters of the Kriging Model Based on \( B = 3000 \) Bootstrap Replication with Prediction of Universal Kriging with Degree One.
Universal Kriging with Degree Two

Table 4 shows the same conclusions regarding the Ordinary Kriging and Universal Kriging methods with degree one. In particular, (1) the bootstrapping estimate for the correlation parameter is similar to estimate of Classic Kriging. (2) Bootstrap estimation for each model parameter, where the number of bootstrap samples increases (\( B = 1000 \), \( B = 2000 \), and \( B = 3000 \)) does not significantly affect estimation. (3) Bootstrap estimates for Regression-Kriging parameters yield different estimates compared to Classic Kriging. (4) As B increases, the bootstrap estimation for the process variance parameters approaches the estimate from Classic Kriging.

Table 4. The Bootstrap Estimation for Parameters of the Kriging Model with \( B = 1000 \), \( B = 2000 \) and \( B = 3000 \) Using Prediction of Universal Kriging with Degree Two.

| Estimation | Number of Bootstrap Samples (\( B \)) |
|------------|--------------------------------------|
|            | 1000  | 2000  | 3000  |
| \( \hat{\theta} \) | [1.66261, 0.51795] | [1.67383, 0.52266] | [1.667440, 0.51674] |
| \( \hat{\beta}_1 \) | -0.132044 | -0.118087 | -0.118563 |
| \( \hat{\beta}_2 \) | -0.045378 | -0.036540 | -0.040892 |
| \( \hat{\beta}_3 \) | -0.004389 | 0.002197 | 0.004115 |
| \( \hat{\beta}_4 \) | 0.048623 | 0.039941 | 0.043354 |
| \( \hat{\beta}_5 \) | -0.053150 | -0.042412 | -0.041848 |
| \( \hat{\beta}_6 \) | 0.083487 | 0.078411 | 0.075171 |
| \( \hat{\sigma}^2 \) | 1711.891 | 1737.013 | 1750.782 |

Figure 5 shows that the sampling distribution of each parameter based on bootstrapping gives the same conclusion as that from Ordinary Kriging and Universal Kriging with degree one. (1) The sampling distribution of correlation parameters is very different from a Gaussian distribution. (2) The sampling distribution for Regression-Kriging parameters approximates a Gaussian distribution. (3) The sampling distribution of process variance does not approach a Gaussian distribution.
5. Conclusions and Suggestions
The accuracy of the Kriging model is not quantified by the process variance but by the Kriging variance. The estimated Kriging variance does not depend directly on the estimated process variance, even though they are directly proportional but influenced by the correlation parameter. General in the field of Geostatistic that the correlation parameter often ignored (see [17]).

We provide a conjecture that the estimated Kriging variance depends on the estimated correlation parameter. The estimated correlation parameters depend on the configuration and amount of input/output data. Verification must be performed through further analytic studies and simulations.

Den Hertog et al. [4] shows that the difference between the estimated variance with bootstrapping and can be greater than that estimated from Classic Kriging. One cannot claim that the variance of the classical Kriging prediction is smaller than the predicted Kriging with bootstrapping. Some researchers such as [3–8] revealed that the estimated variance in Classic Kriging underestimates the true Kriging variance.

The Kriging variance with bootstrapping is the true Kriging variance considering the randomness of maximum likelihood estimators. The underestimate in the classical Kriging variance causes the parameter estimate in the Kriging model based on semiparametric bootstrapping to be different from that in Classic Kriging. The increasing size of the bootstrap sample leads to the conclusion that the bootstrap estimator will approach the Classic Kriging estimation. Further studies consider the increasing size of the input/output data, illustrating the consistency of the estimators.
The estimated sampling distribution of the Kriging model parameters based on bootstrapping yields the following conclusions. First, the sampling distribution of the correlation parameters is very different from a Gaussian distribution. Second, sampling distributions of Regression-Kriging parameters approximate a Gaussian distribution. Third, sampling distributions of the process variance parameter do not approach a Gaussian distribution. Further studies can consider the increased size of the input/output data to determine the asymptotic distribution of the sampling parameters in the Kriging model.

**Acknowledgements:** I would like to thank the Ministry of Research, Technology and Higher Education of The Republic of Indonesia for funding this research. This research is included in the research scheme "Penelitian Dasar Unggulan Perguruan Tinggi" and honourable Prof. Dr. J.P.C. (Jack) Kleijnen, Prof. Subanar, and Prof. Sri Haryatmi have provided useful suggestions for this research.

**References**

[1] Lophaven S.N, Nielsen, H.B and Søndergaard J 2002DACE: A matlab Kriging toolbox (Vol. 2). IMM Informatics and Mathematical Modelling The Technical University of Denmark.

[2] Kleijnen J.P 2014Simulation-optimization via Kriging and bootstrapping: A survey Journal of Simulation;8(4):241–250.

[3] Kleijnen J.P and Mehdad E 2016Estimating the variance of the predictor in stochastic Kriging. Simulation Modelling Practice and Theory;66:166–173.

[4] Den Hertog D, Kleijnen J.P and Siem A.Y 2006. The correct Kriging variance estimated by bootstrapping. Journal of the Operational Research Society;57(4):400–409.

[5] Kleijnen J.P and Mehdad E 2013Conditional simulation for efficient global optimization. In 2013 Winter Simulations Conference (WSC);969–979. IEEE.

[6] Simamora E, Subanar and Kartiko S.H 2014The procedure of Kriging variance estimation based on semiparametric bootstrapping in deterministic simulation. International Journal of Applied Mathematics and Statistics™;52(7):99–110.

[7] Simamora E, Subanar and Kartiko S.H 2015Asymptotic property of semiparametric bootstrapping Kriging variance in deterministic simulation. Applied Mathematical Sciences; 9(50):2477–2491.

[8] Simamora E, Subanar, Kartiko S.H 2015 A comparison study of parametric and semiparametric bootstrapping in deterministic simulation. International Journal of Applied Mathematics and Statistics™;53(5):172–181.

[9] Abt M 1999 Estimating the prediction mean squared error in Gaussian stochastic processes with exponential correlation structure. Scandinavian Journal of Statistics;26(4):563–578.

[10] Harville D. A 2008 Accounting for the estimation of variances and covariances in prediction under a general linear model: an Overview. Tatra Mountains Mathematical Publications; 39(1):1–15.

[11] C. Thiart, M. Ngwenya and L. Haines 2014 Investigating ‘optimal’ Kriging variance estimation using an analytic and a bootstrap approach. J. South. Afr. Inst. Min. Metall;114:613–619.

[12] Hengl T. A 2009 Practical Guide to Geostatistical Mapping. Amsterdam: Hengl.

[13] Christensen R 2001 Linear Models for Multivariate, Time Series, 49 and Spatial Data, 2nd Edition. Springer Verlag. New York.

[14] Acar E 2013Effects of the Correlation Model, the Trend Model, and the Number of Training Points on the Accuracy of Kriging Metamodels. Expert Systems; 30(5):418–428.

[15] McKay M.D, Beckman R J and Conover W.J 2000 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics;42(1):55–61.

[16] Kenny Q.Y, Li W and Sudjianto A 2000 Algorithmic Construction of Optimal Symmetric Latin Hypercube Designs, Journal of Statistical Planning and Inference 90(1), 145–159.

[17] Cressie N 1993 Statistics for spatial data. Wiley. New York;1993.
[18] Bain L.J and Engelhardt M 1987 Introduction to probability and mathematical statistics. Brooks/Cole.
[19] Santner T.J, Williams B.J and Notz W.I 2003 The Design and Analysis of Computer Experiments. Springer Science and Business Media.
[20] Nocedal J and Wright S.J 1999 Numerical Optimization. Springer Science & Business Media.
Appendix

Kriging predictions on $x_0$ are stated as $\hat{y}(x_0) = \hat{\mathbf{a}}^T \mathbf{y}_x$, where $\hat{\mathbf{a}} = [\lambda_1, \cdots, \lambda_n]^T$ is a weight vector and $\mathbf{Z} = [Z(x_1), \cdots, Z(x_n)]^T$ is an error vector at $n$ design sites. The maximum likelihood estimators in Section 2.1 can be used to write the Kriging predictions as $\hat{y}(x_0) = \hat{\mathbf{a}}^T (\mathbf{G} \mathbf{b} + \mathbf{Z})$. Santner et al. [19] stated that Kriging is the best linear unbiased predictor (BLUP) if it meets the requirements of unbiased Kriging prediction and minimizes the Kriging variance. The error in the Kriging prediction at point $x_0$ is defined as

$$\mathbf{y}(x_0) - y(x_0) = \hat{\mathbf{a}}^T (\mathbf{G} \mathbf{b} + \mathbf{Z}) - (\mathbf{G}^T(x_0) \mathbf{b} + Z(x_0)) = \hat{\mathbf{a}}^T \mathbf{Z} - Z(x_0) + [\mathbf{G}^T \hat{\mathbf{a}} - y(x_0)]^T \mathbf{b}.$$ 

In order for the Kriging predictor to be unbiased, it must fulfill $E[\mathbf{y}(x_0) - y(x_0)] = 0$ that $\mathbf{G}^T \hat{\mathbf{a}} = y(x_0)$. The Kriging variance is the mean squared prediction error (MSPE) at an untried location $x_0$,

$$\text{MSPE}(\mathbf{y}(x_0)) = E[(\hat{\mathbf{a}}^T \mathbf{Z} - Z(x_0))^2] = E[Z^2(x_0)] + \hat{\mathbf{a}}^T \mathbf{Z} - 2 \hat{\mathbf{a}}^T ZZ(x_0) | E[Z^2(x_0)] = \kappa^2$$

and $E[Z] = \mathbf{C} = \sigma^2 \mathbf{R}$, where $\mathbf{R}$ and $\mathbf{C}$ are defined in Eqs. (2.5) and (2.6), respectively.

The covariance of $Z(x_0)$ for every element in $\mathbf{Z}$ is defined as

$$E[ZZ(x_0)] = \sigma^2 [\mathbf{R}(\theta, x_1, x_1), \cdots, \mathbf{R}(\theta, x_n, x_0)]^T = \sigma^2 \mathbf{R}(x_0).$$

The Kriging variance can be rewritten as $\text{MSPE}(\mathbf{y}(x_0)) = \sigma^2 (1 + \hat{\mathbf{a}}^T \mathbf{R} \mathbf{a} - 2 \hat{\mathbf{a}}^T \mathbf{R}(x_0)).$ The Kriging variance can be minimized with respect to the weight vector $\hat{\mathbf{a}}$ with the constraint $\mathbf{G}^T \hat{\mathbf{a}} = y(x_0)$ and introduced into a Lagrangian: $L(\hat{\mathbf{a}}, \mathbf{K}) = \sigma^2 (1 + \hat{\mathbf{a}}^T \mathbf{R} \mathbf{a} - 2 \hat{\mathbf{a}}^T \mathbf{R}(x_0)) - \kappa^2 (\mathbf{G}^T \mathbf{a} - y(x_0))^2$, where $\mathbf{K}$ is a vector of Lagrange multipliers. The solution to $L(\hat{\mathbf{a}}, \mathbf{K})$ with the constraint $\mathbf{G}^T \mathbf{a} = y(x_0)$ is based on first order necessary conditions of an optimization problem [20] is

$$\begin{bmatrix} \nabla_{\hat{\mathbf{a}}} L(\hat{\mathbf{a}}, \mathbf{K}) & \kappa^2 \mathbf{R} \mathbf{a} - \mathbf{r}(x_0) \end{bmatrix} = 0$$

and

$$\begin{bmatrix} \nabla_{\hat{\mathbf{a}}} L(\hat{\mathbf{a}}, \mathbf{K}) & \kappa^2 \mathbf{R} \mathbf{a} - \mathbf{r}(x_0) \end{bmatrix} = 0.$$

which is a system of Kriging equations (Cressie, 1993). If rewritten as a matrix equation, one obtains

$$\begin{bmatrix} 0 & \mathbf{G}^T \mathbf{a} \end{bmatrix} \begin{bmatrix} \mathbf{r}(x_0) \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{e}(x_0) \mathbf{R} \end{bmatrix}$$

so that

$$\begin{bmatrix} \mathbf{r}(x_0) \mathbf{R} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{G}^T \mathbf{a} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{e}(x_0) \mathbf{R} \end{bmatrix}.$$

The last equation can be solved by modifying the inverse theorem for partitioned matrices.

Theorem (Appendix) Suppose A is an $(n + p) \times (n + p)$ matrix that is partitioned into four sub-matrix blocks $A = \begin{bmatrix} 0 & \mathbf{G}^T \mathbf{a} \\ \mathbf{G} \mathbf{R} \end{bmatrix}$. If $\mathbf{R}$ is nonsingular, then

$$A^{-1} = \begin{bmatrix} (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})^{-1} & 0 \\ \mathbf{R}^{-1} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

where $\mathbf{I}$ is an $n \times n$ identity matrix.

Proof

The Schur complement $A$ with respect to $\mathbf{R}$ is defined $A|\mathbf{R} = 0 - \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} = -\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}$. The shape of the diagonalized blocks is

$$\begin{bmatrix} \mathbf{I} & -\mathbf{G} \mathbf{R}^{-1} \mathbf{G} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{G} \mathbf{R} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{G}^T \mathbf{a} \\ \mathbf{G} \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{I} \mathbf{R}^{-1} \mathbf{G} \mathbf{R} \\ 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{e}(x_0) \mathbf{R} \end{bmatrix}$$

Using the definition of the Schur complement $A$ with respect to $\mathbf{R}$ and performing matrix multiplication on the right side from the last equation yields

$$\begin{bmatrix} 0 & \mathbf{G}^T \mathbf{a} \\ \mathbf{G} \mathbf{R} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})^{-1} & 0 \\ \mathbf{R}^{-1} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix}.$$

We write again
\[ \begin{bmatrix} -\frac{\kappa}{2\sigma^2} \end{bmatrix} = \begin{bmatrix} -(G^T\mathcal{K}^{-1}G)^{-1} & (G^T\mathcal{K}^{-1}G)^{-1}G^T \mathcal{K}^{-1} - \mathcal{K}^{-1}G(x_u) \mathcal{K}^{-1} \end{bmatrix} \begin{bmatrix} \mathcal{K}(x_u) \end{bmatrix} \]

Simplifying yields

\[ -\frac{\kappa}{2\sigma^2} = \kappa' = (G^T\mathcal{K}^{-1}G)^{-1} \left( (G^T\mathcal{K}^{-1}G)^{-1}G^T \mathcal{K}^{-1} \right) \begin{bmatrix} \mathcal{K}(x_u) \end{bmatrix} \]

Substituting \( \lambda = \mathcal{K}^{-1}(r(x_u) - \mathcal{K}^*) \) to \( \mathcal{K}(x_u) = \lambda^T \mathcal{K} \) yield \( \mathcal{K}(x_u) = (r(x_u) - \mathcal{K}^*)^T \mathcal{K}^{-1} \mathcal{K} \).

Remember that \( \hat{\beta} = (G^T\mathcal{K}^{-1}G)^{-1}G^T \mathcal{K}^{-1} \mathcal{K} \) in Section 2.1 yields the Kriging prediction \( \mathcal{K}(x_u) = G^T(x_u) \hat{\beta} + r^T(x_u) \mathcal{K}^{-1}(\mathcal{K} - \mathcal{K}) \), while the Kriging variance is

\[ \text{MSPE} \left( \mathcal{K}(x_u) \right) = \sigma^2 \left( 1 + \lambda^T \mathcal{K} \lambda - 2 \lambda^T r(x_u) \right) = \sigma^2 \left( 1 + \lambda^T (\mathcal{K} \mathcal{K} - 2 r(x_u)) \right) \]

\[ = \sigma^2 \left( 1 + (G\mathcal{K}^* - r(x_u))^T \mathcal{K}^{-1} (G\mathcal{K}^* + r(x_u)) \right) \]

\[ = \sigma^2 \left( 1 + \lambda^T (G^T \mathcal{K}^{-1} G)^{-1} A - r^T(x_u) \mathcal{K}^{-1} r(x_u) \right) \]

with \( A = G^T \mathcal{K}^{-1} r(x_u) - \mathcal{K}(x_u) \). Substituting the estimated model parameters into the Kriging prediction yields the Classic Kriging prediction, \( \hat{\mathcal{K}}(x_u) = G^T(x_u) \hat{\beta} + r^T(x_u) \mathcal{K}^{-1}(\mathcal{K} - \mathcal{K}) \).

Similarly, substituting the estimated model parameters into the Kriging variance yields the Classic Kriging variance, \( \text{MSPE}_K \left( \hat{\mathcal{K}}(x_u) \right) = \sigma^2 \left( 1 + \lambda^T (G^T \mathcal{K}^{-1} G)^{-1} A - r^T(x_u) \mathcal{K}^{-1} r(x_u) \right) \).