Another Look at the Kriging Equations

Jason A. McLennan, Oy Leuangthong, and Clayton V. Deutsch

Centre for Computational Geostatisitcs
Department of Civil & Environmental Engineering
University of Alberta

Kriging is an essential element of modern geostatistics. This paper presents a thorough review of the random function, stationarity, and derivation of the kriging equations for simple, ordinary, and universal estimation schemes. Two different trend estimation schemes with kriging are presented. A simple link between unconstrained and constrained kriging is shown.

The Random Function Model

The central paradigm of geostatistics is the description of unsampled locations by random variables representing location-dependant probability distributions about the underlying true value. Figure 1 shows a square domain **D**. Consider within **D** the random function (RF) $Z(\mathbf{u})$ composed of the random variable (RV) set $Z(\mathbf{u}) = \{Z(\mathbf{u}_l)$, for all $\mathbf{u}_l \, \varepsilon \, \mathbf{D}\}$. There are 10 $\mathbf{u}_{l'}$ locations where samples are taken and where the $Z(\mathbf{u}_{l'})$ RV formalism is replaced by the single true or sample value $z(\mathbf{u}_{l'})$. Although it is not required, the RF $Z(\mathbf{u})$ is usually assumed multivariate normal for convenience.

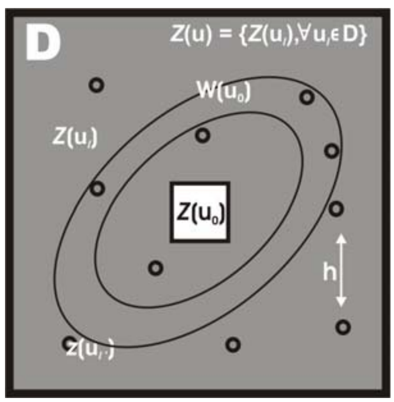


Figure 1: The domain **D**

The probabilistic RF/RV approach is neither motivated nor necessitated by the physical geology of the underlying mineral or hydrocarbon accumulation. The multivariate Gaussian assumption is also unprovoked by any geological interpretation. There may be better approaches. Nonetheless, the Gaussian random function representation is convenient in that each random variable $Z(\mathbf{u}_l)$ can be characterized by its conditional cumulative distribution function (ccdf) $F(\mathbf{u}_l;\, z(\mathbf{u}_l)|(n(\mathbf{u}_l)))$ utilizing the $n(\mathbf{u}_l)$ relevant or close sample data within search neighborhoods or radii centered around each $\mathbf{u}_l$ location. In Figure 1, the value of $n(\mathbf{u}_l)$ is always some subset of the $z(\mathbf{u}_{l'})$, l' = 1,…, 10 vector. Furthermore, the entire RF $Z(\mathbf{u})$ can be fully characterized by the set of all L-

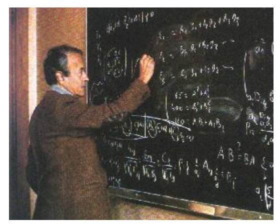
variate cdfs $F(\mathbf{u}_1,\ldots,\mathbf{u}_l;\ z(\mathbf{u}_1)|(n(\mathbf{u}_1),\ldots,\ z(\mathbf{u}_L)|(n(\mathbf{u}_L)))$ for any number L and any choice of the L unsampled locations $\mathbf{u}_l$, $l = 1,\ldots, L$.

When the RF $Z(\mathbf{u})$ model is assumed multi-variate normal meaning all $F(\mathbf{u}_l;\ z(\mathbf{u}_l)|(n(\mathbf{u}_l)))$ cdfs are uni-variate normal and all $F(\mathbf{u}_1,\ldots,\mathbf{u}_L;\ z(\mathbf{u}_1)|(n(\mathbf{u}_1),\ldots,\ z(\mathbf{u}_L)|(n(\mathbf{u}_L)))$ ccdfs are L-variate normal. Remarkably, these ccdfs can be completely parameterized by inferring the single covariance function $\mathrm{COV}\{Z(\mathbf{u}_l), Z(\mathbf{u}_l)+\mathbf{h}\}$ for all separation vectors $\mathbf{h}$ within $\mathbf{D}$. This covariance function is referred to as the *spatial law* of the RF $Z(\mathbf{u})$.

Decomposition

The spatial distribution of a continuous random function variable such as $Z(\mathbf{u})$ is of dual character: partly structured and partly stochastic. The structured component exists from the unique set of depositional events that concentrated the mineral or hydrocarbon and the stochastic component is due to random fluctuations in this depositional process. In fact, Georges Matheron invented the name and field of geostatistics on the basis of this observation:

> … even though mineralization is never so chaotic as to preclude all forms of forecasting, it is never regular enough to allow the use of a deterministic forecasting technique. This is why (at least, simply realistic) estimation must necessarily take into account both features – structure and randomness – inherent in any deposit. Since geologists stress the first of these two aspects, and statisticians stress the second, I proposed, over 15 years ago, the name geostatistics… (Journel, 1978).



Dr. Georges Matheron

This notion of dual character can be represented analytically within the RF/RV formalism. The typical decomposition technique calls for the dissociation of $Z(\mathbf{u})$ into a structured and random component:

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}) \tag{1}$$

where $m(\mathbf{u})$ is the structured component or trend and $R(\mathbf{u})$ is the random component or residual. This decision to split the spatial variability observed into a smoothly varying trend component and a more erratic residual component is often arbitrary [2]. Moreover, the particular additive decomposition in (1) is not really a decision – it is perhaps better described as a necessary implication of the kriging algorithm. That is, all unconstrained and constrained kriging implies the particular dissociation in (1) and no other. This is shown explicitly in later sections.

It is worth reemphasizing that although there are sound geological reasons to consider dissociating a smooth $m(\mathbf{u})$ and more random $R(\mathbf{u})$ component from $Z(\mathbf{u})$, the particular additive decomposition in (1) is arbitrary and not necessarily supported by any geological phenomenon. Better approaches may exist, for example, $m(\mathbf{u})$ and $R(\mathbf{u})$ could be multiplicative. Alternative techniques such as these can be should be investigated.

It is also important to notice that the decomposition in (1) is artificial. Truly there is no $m(\mathbf{u})$ and $R(\mathbf{u})$ samples – only $Z(\mathbf{u})$ is sampled in reality. These variables exist only out of the artificial construct (1) required by kriging. In particular, and as we will see explicitly later, this leads to the requirement of residual covariances, that is, the spatial law of $R(\mathbf{u})$, in the implementation and use of the kriging estimator. Kriging then requires a model of $m(\mathbf{u})$ so that $R(\mathbf{u})$ can be calculated in practice. However, since neither $m(\mathbf{u})$ nor $R(\mathbf{u})$ are available in reality, the spatial variability and model of $m(\mathbf{u})$ is often subjective and difficult to capture correctly. The mean model $m(\mathbf{u})$ almost always robs the residuals of spatial variability needed for accurate spatial interpolation with the kriging equations and uncertainty representation in subsequent simulation. This is the core reason why the theoretical development and practical implementation of kriging are quite different.

Stationarity

It is impossible to rigorously infer the spatial law of $Z(\mathbf{u})$ or $\text{COV}\{Z(\mathbf{u}_l), Z(\mathbf{u}_l)+\mathbf{h}\}$ with the $z(\mathbf{u}_{l'})$ vector. This could only be possible if $z(\mathbf{u}_{l'})$ represented multiple realizations of the true $z(\mathbf{u}_l)$ values at all possible locations $\mathbf{u}_l$ within $\mathbf{D}$ [1]. If this were the case, however, there would be no inference problem left. The sample vector $z(\mathbf{u}_{l'})$ could not possibly contain this amount of information since in reality there is only a single true underlying geologic occurrence and limited sample data. A decision of stationarity must be made in order to substitute the need for repetitive realizations at $\mathbf{u}_l$ locations for scattered (single realizations) sampling at $\mathbf{u}_{l'}$ locations.

The most severe assumption of stationarity entails invariance of the full L-variate joint distribution function under any translation $\mathbf{h}$ within the domain $\mathbf{D}$:

$$\begin{aligned} &F\left(\mathbf{u}_1,\ldots,\mathbf{u}_L; z(\mathbf{u}_1)\,|\,n(\mathbf{u}_1),\ldots,z(\mathbf{u}_L)\,|\,n(\mathbf{u}_L)\right) \\ &\qquad = F\left(\mathbf{u}_1+\mathbf{h},\ldots,\mathbf{u}_L+\mathbf{h}; z(\mathbf{u}_1+\mathbf{h})\,|\,n(\mathbf{u}_1+\mathbf{h}),\ldots,z(\mathbf{u}_L+\mathbf{h})\,|\,n(\mathbf{u}_L+\mathbf{h})\right) \end{aligned} \tag{2}$$

However, virtually all geostatistical techniques can be applied with a much less stringent second-order assumption of stationarity. This entails relation (1) with $L = 2$. All one-variate cdfs $F(\mathbf{u}_l; z(\mathbf{u}_l)|(n(\mathbf{u}_l)))$ are equivalent to the ccdf formed by all available $z(\mathbf{u}_{l'})$ sample values within $\mathbf{D}$. And all joint two-variate ccdfs $F(\mathbf{u}_l, \mathbf{u}_l + \mathbf{h}; z(\mathbf{u}_l)|(n(\mathbf{u}_l), z(\mathbf{u}_l + \mathbf{h})|(n(\mathbf{u}_l + \mathbf{h})))$ are equivalent to the joint distribution of all possible pairs of sample data approximately separated by $\mathbf{h}$ $(z(\mathbf{u}_{l'}), z(\mathbf{u}_{l'} + \mathbf{h} + \mathbf{T}))$. The tolerance tensor $\mathbf{T}$ is required since the $\mathbf{u}_{l'}$ locations are rarely regularly spaced in practice. This assumption or decision of second-order stationarity then implies the following first-order mean and second-order covariance relationships:

1. The mean is independent of location,

$$\mathrm{E}\left\{Z\left(\mathbf{u}_l\right)\right\} = \mathrm{m} \qquad \forall \mathbf{u}_l \in \mathbf{D} \tag{3}$$

2. The covariance is independent of location depending only on the lag vector $\mathbf{h}$,

$$\mathrm{COV}_Z\left(\mathbf{h}\right) = \mathrm{E}\left\{Z\left(\mathbf{u}_l+\mathbf{h}\right)\cdot Z\left(\mathbf{u}_l\right)\right\} - \mathrm{m}^2 \qquad \forall \mathbf{u}_l, \mathbf{u}_l+\mathbf{h} \in \mathbf{D} \tag{4}$$

The assumption of stationarity allows inference. In particular, the spatial law of $Z(\mathbf{u})$ can be assessed by evaluating $\text{COV}\{Z(\mathbf{u}_l), Z(\mathbf{u}_l)+\mathbf{h}\}$ with $\text{COV}_Z(\mathbf{h})$ which is simply the experimental covariance of all pairs of sample data approximately separated by $\mathbf{h}$ $(z(\mathbf{u}_{l'}), z(\mathbf{u}_{l'} + \mathbf{h} + \mathbf{T}))$ within $\mathbf{D}$.

It is worth emphasizing in this work that strangely it is not the spatial law of $Z(\mathbf{u})$ that is required to interpolate $z(\mathbf{u}_{l'})$ data using the unconstrained and constrained kriging equations. It is in fact the spatial law of $R(\mathbf{u})$ that is required for any form of kriging. This is shown explicitly in the next section. The residuals are also assumed second-order stationary (replace Z with R in (3) and (4)). And the $\text{COV}\{R(\mathbf{u}_l), R(\mathbf{u}_l)+\mathbf{h}\}$ spatial law of $R(\mathbf{u})$ is calculated with $\text{COV}_R(\mathbf{h})$, the experimental covariance of all pairs of residual *data* approximately separated by **h** ($r(\mathbf{u}_{l'})$, $r(\mathbf{u}_{l'} + \mathbf{h} + \mathbf{T})$) within **D**. This requires prior modeling of the locally varying mean $m(\mathbf{u})$.

Decisions of stationarity are not manifested from physical phenomenon; rather, they are a necessary consequence of the RF/RV approach. These decisions amount to assume the geology is homogeneous within certain spatial domains, see (3) and (4). Therefore, these decisions are necessarily subjective and can never be refuted, validated, or tested a-priori; however, they can be argued inappropriate a-posteriori. That is, it is always possible to observe large-scale smooth changes or non-stationarities.

A General Kriging Estimator

All forms of kriging including unconstrained (simple) and constrained (ordinary and universal) kriging rely on the artificial construct in (1). Consider in Figure 1 estimation at the central $\mathbf{u}_0$ location using the following linear kriging estimator:

$$z_{\mathrm{K}}^{*}(\mathbf{u}_0) = \mathrm{A} + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \cdot z(\mathbf{u}_{l'}) \tag{5}$$

The number of conditioning data $n(\mathbf{u}_0)$ for estimation depends on the size of the search window $\mathrm{W}(\mathbf{u}_0)$. Two search windows where $n(\mathbf{u}_0) = 2$ and $n(\mathbf{u}_0) = 5$ are shown in Figure 1. A is a constant shift parameter and $\lambda_{\mathrm{K}}(\mathbf{u}_l)$ are the kriging weights assigned to the $z(\mathbf{u}_{l'})$ data. The estimate $z^*_K(\mathbf{u}_0)$ and data $z(\mathbf{u}_{l'})$ can also be represented in probabilistic notation corresponding to the following RV estimator:

$$Z_{\mathrm{K}}^{*}(\mathbf{u}_0) = \mathrm{A} + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \cdot Z(\mathbf{u}_{l'}) \tag{6}$$

The actual error of estimation $x(\mathbf{u}_0)$ is:

$$x(\mathbf{u}_0) = z(\mathbf{u}_0) - z_{\mathrm{K}}^{*}(\mathbf{u}_0) \tag{7}$$

Little can be done about (7) unless its probabilistic version $X(\mathbf{u}_0)$ is considered:

$$X(\mathbf{u}_0) = Z(\mathbf{u}_0) - Z_{\mathrm{K}}^{*}(\mathbf{u}_0) \tag{8}$$

In this case, the expected value and variance of $X(\mathbf{u}_0)$ can be calculated and thus acted upon [3]. In fact, it is these two moments of $X(\mathbf{u}_0)$ that allow the kriging equations to develop. In particular, we require the expected value of $X(\mathbf{u}_0)$ to be zero and the variance of $X(\mathbf{u}_0)$ to be a minimum. This is another reason/advantage of adopting the RF/RV approach. The expected value of the error $X(\mathbf{u}_0)$ is:

$$\begin{aligned} \mathrm{E}\{X(\mathbf{u}_0)\} &= \mathrm{E}\{Z(\mathbf{u}_0)\} - \mathrm{E}\{Z_{\mathrm{K}}^{*}(\mathbf{u}_0)\} \\ &= \mathrm{E}\{Z(\mathbf{u}_0)\} - \mathrm{E}\{\mathrm{A}\} - \mathrm{E}\left\{\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) z_{\mathrm{K}}(\mathbf{u}_{l'})\right\} \\ &= m(\mathbf{u}_0) - \mathrm{A} - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) m(\mathbf{u}_{l'}) \end{aligned} \tag{9}$$

In order for the kriging estimator $Z^*_{\mathrm{K}}(\mathbf{u}_0)$ to be unbiased, this expected error must be zero; therefore, the shift parameter A is set to:

$$\mathrm{A} = m(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) m(\mathbf{u}_{l'}) \tag{10}$$

Indeed, now the $\mathrm{E}\{X(\mathbf{u}_0)\}$ is zero (from (9) and (10)). And the kriging estimator is then:

$$\begin{aligned} Z_{\mathrm{K}}^{*}(\mathbf{u}_0) &= m(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) m(\mathbf{u}_{l'}) + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'}) \\ &= m(\mathbf{u}_0) + \left(\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \left[Z(\mathbf{u}_{l'}) - m(\mathbf{u}_{l'}) \right] \right) \end{aligned} \tag{11}$$

So the unbiased kriging estimator $Z^*_{\mathrm{K}}(\mathbf{u}_0)$ appears as the result of linear estimation of the residual at location $\mathbf{u}_0$ from the residual sample data at the $n(\mathbf{u}_0)$ locations $\mathbf{u}_{l'}$ [3]:

$$\begin{aligned} Z_{\mathrm{K}}^{*}(\mathbf{u}_0) - m(\mathbf{u}_0) &= \left(\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \left[Z(\mathbf{u}_{l'}) - m(\mathbf{u}_{l'}) \right] \right) \\ R^{*}(\mathbf{u}_0) &= \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) R(\mathbf{u}_{l'}) \end{aligned} \tag{12}$$

There are some important observations to make at this point. First notice that the general kriging estimator $Z^*_{\mathrm{K}}(\mathbf{u}_0)$ in formula (12) directly relies on the additive decomposition of the $Z(\mathbf{u})$ RF in formula (1). And recall this decomposition is arbitrary in the sense that there is no undisputable physical evidence for it or actual data for its dissociated components; yet, as (12) shows, it is necessary in order to develop the kriging estimator and subsequent kriging equations. This is unfortunate since there are indeed no $m(\mathbf{u})$ and $R(\mathbf{u})$ *data*. Only $Z(\mathbf{u})$ is sampled in reality and can be used explicitly. This presents some implementation challenges and differentiates the theory and practice of kriging.

The other moment of $X(\mathbf{u}_0)$ required to develop the kriging equations is its variance:

$$\begin{aligned} \mathrm{VAR}\{X(\mathbf{u}_0)\} &= \mathrm{E}\left\{\left[X(\mathbf{u}_0)\right]^2\right\} - \left[\mathrm{E}\{X(\mathbf{u}_0)\}\right]^2 \\ &= \mathrm{E}\left\{\left[Z(\mathbf{u}_0) - Z_{\mathrm{K}}^{*}(\mathbf{u}_0)\right]^2\right\} - \left[\mathrm{E}\left\{\left[Z(\mathbf{u}_0) - Z_{\mathrm{K}}^{*}(\mathbf{u}_0)\right]\right\}\right]^2 \\ &= \left[Z(\mathbf{u}_0) - m(\mathbf{u}_0)\right] - \mathrm{E}\left\{\left[\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) R(\mathbf{u}_{l'})\right]^2\right\} - \left[\mathrm{E}\left\{\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) R(\mathbf{u}_{l'})\right\}\right]^2 \\ &= \mathrm{E}\left\{\sum_{l'=1}^{n(\mathbf{u}_0)} \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \lambda_{\mathrm{K}}(\mathbf{u}_{i'}) R(\mathbf{u}_{l'}) R(\mathbf{u}_{i'})\right\} - \left[\mathrm{E}\left\{\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) R(\mathbf{u}_{l'})\right\}\right]^2 \\ &= \sum_{l'=1}^{n(\mathbf{u}_0)} \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \lambda_{\mathrm{K}}(\mathbf{u}_{i'}) \mathrm{E}\{R(\mathbf{u}_{l'}) R(\mathbf{u}_{i'})\} - \sum_{l'=1}^{n(\mathbf{u}_0)} \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \lambda_{\mathrm{K}}(\mathbf{u}_{i'}) \mathrm{E}\{R(\mathbf{u}_{l'})\} \mathrm{E}\{R(\mathbf{u}_{i'})\} \\ &= \sum_{l'=1}^{n(\mathbf{u}_0)} \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \lambda_{\mathrm{K}}(\mathbf{u}_{i'}) \left[\mathrm{E}\{R(\mathbf{u}_{l'}) R(\mathbf{u}_{i'})\} - \mathrm{E}\{R(\mathbf{u}_{l'})\} \mathrm{E}\{R(\mathbf{u}_{i'})\}\right] \\ &= \sum_{l'=1}^{n(\mathbf{u}_0)} \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'}) \lambda_{\mathrm{K}}(\mathbf{u}_{i'}) \mathrm{COV}\{R(\mathbf{u}_{l'}) R(\mathbf{u}_{i'})\} \end{aligned} \tag{13}$$

It is already apparent, without yet mentioning any particular type or flavor of kriging, that it is always the spatial law of the unknown and unsampled $R(\mathbf{u})$ RF that is required and not the spatial law of $Z(\mathbf{u})$. The unbiasedness conditions in relation (10), the estimator in (12), and the subsequent error variance in (13) are general results for any type of unconstrained and constrained kriging. The different flavors of kriging simply correspond to different mean models or forms of stationarity in (1) and (12) which require different (unconstrained vs. constrained) forms in the unbiasedness conditions in (9) and minimum error variance in (13).

Simple Kriging Equations

The kriging system results from minimizing $X(\mathbf{u}_0)$ in (13) subject to unbiasedness in (9). Unconstrained and constrained kriging simply consider different models for the mean $m(\mathbf{u})$ and levels of stationarity. Unconstrained kriging is known better as *Simple Kriging*. The simple kriging algorithm assumes the mean is constant over the entire domain **D**:

$$m(\mathbf{u}) = m \qquad \forall \mathbf{u} \in \mathbf{D} \tag{14}$$

This equivalency corresponds to first-order stationarity. The shift parameter A becomes:

$$\mathrm{A} = m - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'})m \tag{15}$$

And the simple kriging estimator $Z^*_{\mathrm{SK}}(\mathbf{u}_0)$ is:

$$\begin{aligned} Z^*_{\mathrm{SK}}(\mathbf{u}_0) &= m - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})m + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})Z(\mathbf{u}_{l'}) \\ &= m + \left(\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) \left[Z(\mathbf{u}_{l'}) - m \right] \right) \end{aligned} \tag{16}$$

Notice the simple kriging estimator is unbiased since the expected error is zero:

$$\begin{aligned} \mathrm{E}\{X(\mathbf{u}_0)\} &= \mathrm{E}\{Z(\mathbf{u}_0)\} - \mathrm{E}\{Z^*_{\mathrm{K}}(\mathbf{u}_0)\} \\ &= m - m - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'})\mathrm{E}\{Z(\mathbf{u}_{l'})\} + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'})m \\ &= m - m - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'})m + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{K}}(\mathbf{u}_{l'})m \\ &= 0 \end{aligned} \tag{17}$$

Simple kriging is referred to as unconstrained since there were no additional constraints that need to be imposed in order to achieve (17). There remains to determine the simple kriging weights $\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})$, $l' = 1,\ldots, n(\mathbf{u}_0)$. These $n(\mathbf{u}_0)$ weights are determined so that the error variance in relation (13) is a minimum. This is done by setting the partial derivatives of (13) with $\lambda_{\mathrm{K}}(\mathbf{u}_{l'})$ replaced with $\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})$ with respect to $\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})$ to zero:

$$\frac{\partial\left(\mathrm{VAR}\{X(\mathbf{u}_0)\}\right)}{2\partial\left(\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\right)} = \mathrm{COV}\{R(\mathbf{u}_{l'})R(\mathbf{u}_0)\} - \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{i'})\mathrm{COV}\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\} = 0 \quad l' = 1,\ldots,n(\mathbf{u}_0) \tag{18}$$

This results in the following system of simple kriging equations:

$$\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{i'})\mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\right\} = \mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_0)\right\} \quad l' = 1,\ldots,n(\mathbf{u}_0) \tag{19}$$

There are $n(\mathbf{u}_0)$ equations with $n(\mathbf{u}_0)$ simple kriging weights $\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})$ to be determined.

The Universal Kriging Equations

Constrained kriging is known as *Universal Kriging* or perhaps better named *Kriging with a Trend* since the $m(\mathbf{u})$ trend component in (1) and (12) is locally varying. The current approach is to assume the $m(\mathbf{u})$ component is a smoothly varying deterministic function of the coordinates vector $\mathbf{u}$ whose unknown parameters are fit from the data within local search windows [4]:

$$m\left(\mathbf{u}^{\mathrm{W}}\right) = \sum_{v=0}^{V} a_v\left(\mathbf{u}^{\mathrm{W}}\right) f_v\left(\mathbf{u}\right) \qquad \forall \mathbf{u}^{\mathrm{W}} \in \mathrm{W}\left(\mathbf{u}\right), \mathbf{u} \in \mathbf{D} \tag{20}$$

The $f_v(\mathbf{u})$'s are known and constant functions of the coordinate vectors over the domain $\mathbf{D}$. The $a_v(\mathbf{u}^{\mathrm{W}})$'s are estimated and constant within local search windows $\mathrm{W}(\mathbf{u})$ centered on the unsampled locations. The actual $m(\mathbf{u}^{\mathrm{W}})$'s trend values are unknown since the $a_v(\mathbf{u}^{\mathrm{W}})$'s are also unknown. From here on, the superscript W will be dropped since it is usual that the conditioning region $\mathbf{u}^{\mathrm{W}}$ is taken as the global domain. For example, it is unrealistic for the parameterization of $m(\mathbf{u}^{\mathrm{W}})$ with the $a_v(\mathbf{u}^{\mathrm{W}})$'s to change for $v > 0$ within a domain where an otherwise stationary random function is assumed.

The shift parameter A becomes:

$$\mathrm{A} = \sum_{v=0}^{V} a_v\left(\mathbf{u}_0\right) f_v\left(\mathbf{u}_0\right) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) \sum_{v=0}^{V} a_v\left(\mathbf{u}_{l'}\right) f_v\left(\mathbf{u}_{l'}\right) \tag{21}$$

And the universal kriging estimator $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ is then:

$$\begin{aligned} Z^*_{\mathrm{UK}}(\mathbf{u}_0) &= \sum_{v=0}^{V} a_v\left(\mathbf{u}_0\right) f_v\left(\mathbf{u}_0\right) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'}) \sum_{v=0}^{V} a_v\left(\mathbf{u}_{l'}\right) f_v\left(\mathbf{u}_{l'}\right) + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'}) \\ &= \sum_{v=0}^{V} a_v\left(\mathbf{u}_0\right) f_v\left(\mathbf{u}_0\right) + \left(\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'}) \left[Z(\mathbf{u}_{l'}) - \sum_{v=0}^{V} a_v\left(\mathbf{u}_{l'}\right) f_v\left(\mathbf{u}_{l'}\right) \right] \right) \end{aligned} \tag{22}$$

There are many ways to ensure the universal kriging estimator $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ is unbiased. The classic approach is to impose the following $V + 1$ constraints:

$$\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'}) f_v(\mathbf{u}_{l'}) = f_v\left(\mathbf{u}_0\right) \quad v = 0,\ldots,V \tag{23}$$

where $f_v(\mathbf{u})$ are the monomial trend functions evaluated at the unsampled locations $\mathbf{u}_l$ within $\mathbf{D}$ and $f_v(\mathbf{u}_{l'})$ are the monomial trend functions evaluated at the sampled locations $\mathbf{u}_{l'}$. By considering these constraints (23), the resulting universal kriging estimator $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ is then unbiased:

$$
\begin{aligned}
\mathrm{E}\{X(\mathbf{u}_0)\} &= \mathrm{E}\{Z(\mathbf{u}_0)\} - \mathrm{E}\{Z_{\mathrm{K}}^*(\mathbf{u}_0)\} \\
&= \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'})\mathrm{E}\{Z(\mathbf{u}_{l'})\} - \sum_{v=0}^{V} a_v(\mathbf{u}_{l'}) \left[f_v(\mathbf{u}_{l'}) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'})\, f_v(\mathbf{u}_{l'}) \right] \\
&= \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'})\mathrm{E}\{Z(\mathbf{u}_{l'})\} \\
&= \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'}) \sum_{v=0}^{V} a_v(\mathbf{u}_{l'})\, f_v(\mathbf{u}_{l'}) \\
&= \sum_{v=0}^{V} a_v(\mathbf{u}_0) \left[f_v(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'})\, f_v(\mathbf{u}_{l'}) \right] \\
&= 0
\end{aligned}
\tag{24}
$$

Universal kriging can be referred to as constrained kriging since there are $V + 1$ additional constraints (23) that need to be imposed on the system in order to achieve unbiasedness in (24). There are, however, many other possible constraints that can be imposed in order to achieve unbiasedness. These should be investigated. Notice the universal kriging estimator in (22) can be significantly reduced to a linear combination of the $n(\mathbf{u}_0)$ universal kriging weights $\lambda_{\mathrm{UK}}(\mathbf{u}_{l'})$ and RVs $Z(\mathbf{u}_{l'})$, l' = 1,…, $n(\mathbf{u}_0)$, due to the unbiasedness constraints in (23). The form in (22) is emphasized in this work, however, since it is consistent with (12) and (13) and the requirement for the spatial law of $R(\mathbf{u})$.

There remains to determine the universal kriging weights $\lambda_{\mathrm{UK}}(\mathbf{u}_{l'})$. These $n(\mathbf{u}_0)$ weights are determined so that the error variance in relation (13) is a minimum. However, in this case, the minimization must be performed subject to the $V + 1$ constraint equations in (23). These constraints call for the definition of a Lagrangian function $G(\mathbf{u}_0)$ [5] that depend on the $n(\mathbf{u}_0)$ universal kriging weights $\lambda_{\mathrm{UK}}(\mathbf{u}_{l'})$ in addition to the Lagrange parameters $2\mu_v(\mathbf{u}_0)$:

$$
\mathrm{G}(\mathbf{u}_0) = \mathrm{VAR}\{X(\mathbf{u}_0)\} + 2\mu_v(\mathbf{u}_0)\left[\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{l'})\, f_v(\mathbf{u}_{l'}) - f_v(\mathbf{u}_0) \right] \quad v = 0,\ldots,V \tag{25}
$$

The optimal weights $\lambda_{\mathrm{UK}}(\mathbf{u}_{l'})$ are obtained by setting the $n(\mathbf{u}_0)$ partial derivatives of (13) with $\lambda_{\mathrm{K}}(\mathbf{u}_{l'})$ replaced with $\lambda_{\mathrm{UK}}(\mathbf{u}_{l'})$ with respect to $\lambda_{\mathrm{UK}}(\mathbf{u}_{l'})$ to zero and the $(V + 1)$ partial derivatives of (25) with respect to $\mu_v(\mathbf{u}_0)$ to zero:

$$
\begin{aligned}
\frac{\partial\left(\mathrm{VAR}\{X(\mathbf{u}_0)\}\right)}{2\partial\left(\lambda_{\mathrm{UK}}(\mathbf{u}_{l'})\right)} &= \mathrm{COV}\{R(\mathbf{u}_{l'})R(\mathbf{u}_0)\} - \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{i'})\mathrm{COV}\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\} - \sum_{v=0}^{V} \mu_v(\mathbf{u}_0)\, f_v(\mathbf{u}_{l'}) = 0 \\
\frac{\partial\left(\mathrm{G}(\mathbf{u}_0)\right)}{2\partial\left(\mu_v(\mathbf{u}_0)\right)} &= \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{i'})\, f_v(\mathbf{u}_{i'}) - f_v(\mathbf{u}_0) = 0
\end{aligned}
\tag{26}
$$

This results in the following system of universal kriging equations:

$$
\begin{aligned}
&\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{i'})\mathrm{COV}\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\} + \sum_{v=0}^{V} \mu_v(\mathbf{u}_0)\, f_v(\mathbf{u}_{l'}) = \mathrm{COV}\{R(\mathbf{u}_{l'})R(\mathbf{u}_0)\} \\
&\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{UK}}(\mathbf{u}_{i'})\, f_v(\mathbf{u}_{i'}) = f_v(\mathbf{u}_0)
\end{aligned}
\tag{27}
$$

There are $(n(\mathbf{u}_0) + V + 1)$ equations with $n(\mathbf{u}_0)$ universal kriging weights $\lambda_{UK}(\mathbf{u}_{l'})$ and $(V + 1)$ Lagrange parameters $\mu_v(\mathbf{u}_0)$ to be determined.

The Ordinary Kriging Equations

By convention $f_0(\mathbf{u}) = 1$ and $m(\mathbf{u}) = a_0(\mathbf{u})$ in all of relations (20) through (27). This corresponds to the *Ordinary Kriging* case where $m(\mathbf{u})$ is re-estimated to a constant $a_0(\mathbf{u})$ value within local often overlapping search windows $W(\mathbf{u}) \sim \mathbf{u}^W$.

The Spatial Law for Kriging

Recall that the spatial law of the RF $Z(\mathbf{u})$ is $\mathrm{COV}\{Z(\mathbf{u}_l), Z(\mathbf{u}_l)+\mathbf{h}\}$. The goal is to estimate or interpolate $Z(\mathbf{u})$ at unsampled $(\mathbf{u}_l)$ locations from sampled $(\mathbf{u}_{l'})$ locations. However, the unbiased kriging estimator Z^*_K in formula (12) and all universal and constrained kriging estimators, (16) and (22), appear as the result of a linear estimation of the residuals at all the unsampled locations $\mathbf{u}_l$ from the residual at the $n(\mathbf{u}_0)$ sample locations $\mathbf{u}_{l'}$.

It is perhaps unintuitive that the ensuing kriging systems of equations (16) and (22) do not require the spatial law of $Z(\mathbf{u})$. Actually, it is the spatial law of the $R(\mathbf{u})$ RF $\mathrm{COV}\{R(\mathbf{u}_l), R(\mathbf{u}_l)+\mathbf{h}\}$ that is required, see (19) and (27). There are a number of practical implementation challenges associated with inferring the spatial law of $R(\mathbf{u})$ since there are no $r(\mathbf{u}_{l'})$ samples. The only exception is for simple kriging where (14) applies and $\mathrm{COV}\{R(\mathbf{u}_l), R(\mathbf{u}_l)+\mathbf{h}\} = \mathrm{COV}\{Z(\mathbf{u}_l), Z(\mathbf{u}_l)+\mathbf{h}\}$.

Kriging the Trend

System (27) provides the universal kriging weights $\lambda_{UK}(\mathbf{u}_{l'})$ needed to calculate the universal kriging estimate $Z^*_{UK}(\mathbf{u}_0)$. The $a_v(\mathbf{u})$'s and resulting $m(\mathbf{u})$'s are implicitly estimated. However, it is possible to estimate the $m(\mathbf{u})$ values directly. Two approaches are taken in this work. The first approach explicitly estimates $m(\mathbf{u})$ with the $m^*_{UK}(\mathbf{u}_0)$ estimator with the $a_v(\mathbf{u})$ estimates hidden within the algorithm. This is labeled the explicit approach. The other approach estimates $m(\mathbf{u})$ implicitly by first estimating the $a_v(\mathbf{u})$'s with the $a_v^*(\mathbf{u}_0)$ estimator and then setting the final estimate of $m^*_{UK}(\mathbf{u}_0)$ to the weighted linear combination in (20) with $a_v(\mathbf{u})$'s replaced with $a_v^*(\mathbf{u}_0)$. This is labeled the implicit approach. The resulting implicit universal kriging for the mean system reveals a very simple relationship between unconstrained and constrained kriging in the next section.

The Explicit Approach

The form of the mean $m(\mathbf{u})$ as a deterministic function of the coordinates vector $\mathbf{u}$ (20) does not change. The shift parameter A is then also the same. The universal kriging mean estimator $m^*_{UK}(\mathbf{u}_0)$ then also has the same form as (22), except with the universal kriging weights $\lambda_{UK}(\mathbf{u}_{l'})$ replaced by the universal kriging for the mean weights $\lambda^m_{UK}(\mathbf{u}_{l'})$:

$$\begin{aligned} m^*_{UK}(\mathbf{u}_0) &= \sum_{v=0}^{V} a_v(\mathbf{u}_0) f_v(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{UK}(\mathbf{u}_{l'}) \sum_{v=0}^{V} a_v(\mathbf{u}_{l'}) f_v(\mathbf{u}_{l'}) + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{UK}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'}) \\ &= \sum_{v=0}^{V} a_v(\mathbf{u}_0) f_v(\mathbf{u}_0) + \left(\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{UK}(\mathbf{u}_{l'}) \left[Z(\mathbf{u}_{l'}) - \sum_{v=0}^{V} a_v(\mathbf{u}_{l'}) f_v(\mathbf{u}_{l'}) \right] \right) \end{aligned} \tag{28}$$

The error of estimation $Y(\mathbf{u}_0)$ is then:

$$Y(\mathbf{u}_0) = m(\mathbf{u}_0) - m^*_{\mathrm{UK}}(\mathbf{u}_0) \tag{29}$$

The same $(V + 1)$ constraints in (23) ensuring unbiasedness of the $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ estimator, ensures unbiasedness of the $m^*_{\mathrm{UK}}(\mathbf{u}_0)$ estimator:

$$\begin{aligned}
\mathrm{E}\left\{Y(\mathbf{u}_0)\right\} &= \mathrm{E}\left\{m(\mathbf{u}_0)\right\} - \mathrm{E}\left\{m^*_{\mathrm{UK}}(\mathbf{u}_0)\right\} \\
&= \sum_{v=0}^{V} a_v\left(\mathbf{u}_0\right) f_v\left(\mathbf{u}_0\right) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\mathrm{E}\left\{\mathrm{Z}(\mathbf{u}_{l'})\right\} \\
&= \sum_{v=0}^{V} a_v\left(\mathbf{u}_0\right) f_v\left(\mathbf{u}_0\right) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'}) \sum_{v=0}^{V} a_v\left(\mathbf{u}_{l'}\right) f_v\left(\mathbf{u}_{l'}\right) \\
&= \sum_{v=0}^{V} a_v\left(\mathbf{u}_0\right) f_v\left(\mathbf{u}_0\right) - \sum_{v=0}^{V} a_v\left(\mathbf{u}_0^{\mathrm{W}}\right) \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'}) f_v\left(\mathbf{u}_{l'}\right) \\
&= \sum_{v=0}^{V} a_v\left(\mathbf{u}_0\right)\left[f_v\left(\mathbf{u}_0\right) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'}) f_v\left(\mathbf{u}_{l'}\right)\right] \\
&= 0
\end{aligned} \tag{30}$$

Similar to the universal kriging estimator in (22), notice here that the universal kriging estimator for the mean in (28) can be significantly reduced to a linear combination of the $n(\mathbf{u}_0)$ universal kriging for the mean weights $\lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})$ and RVs $Z(\mathbf{u}_{l'})$, $l' = 1,\ldots, n(\mathbf{u}_0)$, due to the unbiasedness constraints in (23). Here, however, (13) is not relevant since the deterministic $m(\mathbf{u}_0)$ term causes the error variance $\mathrm{VAR}\{Y(\mathbf{u}_0)\}$ to take a different form:

$$\begin{aligned}
\mathrm{VAR}\left\{Y(\mathbf{u}_0)\right\} &= \mathrm{E}\left\{\left[Y(\mathbf{u}_0)\right]^2\right\} - \left[\mathrm{E}\left\{Y(\mathbf{u}_0)\right\}\right]^2 \\
&= \mathrm{E}\left\{\left[m(\mathbf{u}_0) - m^*_{\mathrm{UK}}(\mathbf{u}_0)\right]^2\right\} \\
&= \mathrm{E}\left\{\left[m(\mathbf{u}_0)\right]^2 - 2\left[m(\mathbf{u}_0)m^*_{\mathrm{UK}}(\mathbf{u}_0)\right] + \left[m^*_{\mathrm{UK}}(\mathbf{u}_0)\right]^2\right\} \\
&= \mathrm{E}\left\{\left[m(\mathbf{u}_0)\right]^2\right\} - 2\mathrm{E}\left\{m(\mathbf{u}_0)m^*_{\mathrm{UK}}(\mathbf{u}_0)\right\} + \mathrm{E}\left\{\left[m^*_{\mathrm{UK}}(\mathbf{u}_0)\right]^2\right\} \\
&= \mathrm{VAR}\left\{m(\mathbf{u}_0)\right\} - 2\mathrm{COV}\left\{m(\mathbf{u}_0)m^*_{\mathrm{UK}}(\mathbf{u}_0)\right\} + \mathrm{VAR}\left\{m^*_{\mathrm{UK}}(\mathbf{u}_0)\right\} \\
&= \mathrm{VAR}\left\{m^*_{\mathrm{UK}}(\mathbf{u}_0)\right\}
\end{aligned} \tag{31}$$

where the first two terms in the last line are zero since the $m(\mathbf{u}_0)$ component is modeled as a deterministic component. This property ($\mathrm{E}\{m^*_{\mathrm{UK}}(\mathbf{u})m(\mathbf{u})\} = \mathrm{E}\{m^*_{\mathrm{UK}}(\mathbf{u})m(\mathbf{u})\} = 0$) is used in the derivation of the universal kriging for the mean system. The resulting error variance is then:

$$\begin{aligned}\mathrm{VAR}\left\{m^*_{\mathrm{UK}}(\mathbf{u}_0)\right\} &= \mathrm{E}\left\{\left[m^*_{\mathrm{UK}}(\mathbf{u}_0)\right]^2\right\} - \mathrm{E}\left\{\left[m^*_{\mathrm{UK}}(\mathbf{u}_0)\right]\right\}\mathrm{E}\left\{\left[m^*_{\mathrm{UK}}(\mathbf{u}_0)\right]\right\}\\ &= \sum_{l'=1}^{n(\mathbf{u}_0)}\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\lambda^m_{\mathrm{UK}}(\mathbf{u}_{i'})\left[\mathrm{E}\left\{\left[m(\mathbf{u}_{l'})+R(\mathbf{u}_{i'})\right]^2\right\} - \left[\mathrm{E}\left\{\left[m(\mathbf{u}_{l'})+R(\mathbf{u}_{i'})\right]\right\}\right]^2\right]\\ &= \sum_{l'=1}^{n(\mathbf{u}_0)}\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\lambda^m_{\mathrm{UK}}(\mathbf{u}_{i'})\left[\mathrm{E}\left\{\left[R(\mathbf{u}_{i'})\right]^2\right\} - \mathrm{E}\left\{\left[R(\mathbf{u}_{i'})\right]\right\}\mathrm{E}\left\{\left[R(\mathbf{u}_{i'})\right]\right\}\right]\\ &\sum_{l'=1}^{n(\mathbf{u}_0)}\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\lambda^m_{\mathrm{UK}}(\mathbf{u}_{i'})\mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\right\}\end{aligned} \tag{32}$$

The universal kriging weights to the mean $\lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})$, $l' = 1,\ldots, n(\mathbf{u}_0)$ are determined so that the error variance in relation (32) is a minimum subject to the $(V + 1)$ constraint equations in (23). A new Lagrange function for the mean $G_m(\mathbf{u}_0)$ is then defined:

$$\mathrm{G}_m(\mathbf{u}_0) = \mathrm{VAR}\left\{Y(\mathbf{u}_0)\right\} + 2\mu^m_v(\mathbf{u}_0)\left[\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\, f_v(\mathbf{u}_{l'}) - f_v\left(\mathbf{u}_0\right)\right] \quad v = 0,\ldots,V \qquad 33)$$

And the optimal weights $\lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})$ are obtained by setting simultaneously the $n(\mathbf{u}_0)$ partial derivatives of (32) with respect to each of the $\lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})$ weights to zero and the $(V + 1)$ partial derivatives of (31) with respect to each $\mu^m_v(\mathbf{u}_0)$ Lagrange parameter to zero:

$$\begin{aligned}\frac{\partial\left(\mathrm{VAR}\left\{Y(\mathbf{u}_0)\right\}\right)}{2\partial\left(\lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\right)} &= \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{i'})\mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\right\} + \sum_{v=0}^{V}\mu^m_v(\mathbf{u}_0)\, f_v(\mathbf{u}_{l'}) = 0\\ \frac{\partial\left(\mathrm{G}_m(\mathbf{u}_0)\right)}{2\partial\left(\mu^m_v(\mathbf{u}_0)\right)} &= \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{i'})\, f_v(\mathbf{u}_{i'}) - f_v\left(\mathbf{u}_0\right) = 0\end{aligned} \tag{34}$$

Notice that the closeness covariances $\mathrm{COV}\{R(\mathbf{u}_{l'}), R(\mathbf{u}_0)+\mathbf{h}\}$ are all zero. Indeed, this is due to $\mathrm{E}\{m^*_{\mathrm{UK}}(\mathbf{u})m(\mathbf{u})\} = \mathrm{E}\{m(\mathbf{u})m(\mathbf{u})\} = 0$.

This results in the following system of universal kriging for the mean equations:

$$\begin{aligned}&\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{i'})\mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\right\} + \sum_{v=0}^{V}\mu^m_v(\mathbf{u}_0)\, f_v(\mathbf{u}_{l'}) = 0\\ &\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{i'})\, f_v(\mathbf{u}_{i'}) = f_v\left(\mathbf{u}_0\right)\end{aligned} \tag{35}$$

The system in (35) calculates the $\lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})$ and $\mu^m_v(\mathbf{u}_0)$ values that identify the least squares fit of the $m^*_{\mathrm{UK}}(\mathbf{u}_0)$ trend model. Notice that this is equivalent to the universal kriging system in (27) except the closeness (co)variances $\mathrm{COV}\{R(\mathbf{u}_{l'}), R(\mathbf{u}_0)+\mathbf{h}\}$ are zero. There are a total of $(n(\mathbf{u}_0) + V + 1)$ equations with $n(\mathbf{u}_0)$ universal kriging weights $\lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})$ and $(V + 1)$ Lagrange parameters $\mu^m_v(\mathbf{u}_0)$ to be determined.

The Implicit Approach

Estimation of the trend component with $m^*_{\mathrm{UK}}(\mathbf{u}_0)$ in (28) is equivalent to first estimating the $(V + 1)$ trend coefficients $a_v(\mathbf{u})$, then computing the trend estimate simply as a linear combination of the constant trend functionals $f_v(\mathbf{u})$ [5]. This is the essential element of the implicit approach. Consider the following estimator for the mean coefficients:

$$a^*_{v'}(\mathbf{u}_0) = \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'}) \tag{36}$$

The $\lambda_{v'}{}^a(\mathbf{u}_{l'})$'s are the weights assigned to the corresponding $z(\mathbf{u}_{l'})$ sample data for the v'th functional coefficient estimate $a^*_{v'}(\mathbf{u}_0)$. The trend estimator from the explicit approach is then equivalent to:

$$m^*_{\mathrm{UK}}(\mathbf{u}_0) = \sum_{v=0}^{V} a^*_v(\mathbf{u}_0) f_v(\mathbf{u}_0) \tag{37}$$

The error $O(\mathbf{u}_0)$ of estimating each v' functional constant $a^*_{v'}(\mathbf{u}_0)$ is then:

$$O(\mathbf{u}_0) = a_{v'}(\mathbf{u}_0) - a^*_{v'}(\mathbf{u}_0) \tag{38}$$

The expected error can be written as:

$$\begin{aligned}
\mathrm{E}\{O(\mathbf{u}_0)\} &= \mathrm{E}\{a_{v'}(\mathbf{u}_0)\} - \mathrm{E}\{a^*_{v'}(\mathbf{u}_0)\} \\
&= a_{v'}(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) \mathrm{E}\{Z(\mathbf{u}_{l'})\} \\
&= a_{v'}(\mathbf{u}_0) - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) \sum_{v=0}^{V} a_v(\mathbf{u}_0) f_v(\mathbf{u}_0) \\
&= a_{v'}(\mathbf{u}_0)\left[1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) f_{v'}(\mathbf{u}_0)\right] - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) \sum_{\substack{v=0 \\ v\neq v'}}^{V} a_v(\mathbf{u}_0) f_v(\mathbf{u}_0) \\
&= a_{v'}(\mathbf{u}_0)\left[1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) f_{v'}(\mathbf{u}_0)\right] - \sum_{\substack{v=0 \\ v\neq v'}}^{V} a_v(\mathbf{u}_0) \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) f_v(\mathbf{u}_0)
\end{aligned} \tag{39}$$

In order for the expected error $\mathrm{E}\{O(\mathbf{u}_0)\}$ to be zero and $a^*_{v'}(\mathbf{u}_0)$ to be unbiased, the following $(V + 1)$ constraints can be imposed:

$$\begin{aligned}
&\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) f_{v'}(\mathbf{u}_0) = 1 \\
&\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'}) f_v(\mathbf{u}_0) = 0 \quad \forall\, v \neq v'
\end{aligned} \tag{40}$$

Using (40), the $\mathrm{E}\{O(\mathbf{u}_0)\}$ term in (39) is indeed zero. Similar to (31) (since the $a_{v'}(\mathbf{u}_0)$ component is deterministic), the error variance $\mathrm{VAR}\{O(\mathbf{u}_0)\}$ can be significantly reduced to $\mathrm{VAR}\{a^*_{v'}(\mathbf{u}_0)\}$:

$$\begin{aligned}
\mathrm{VAR}\{O(\mathbf{u}_0)\} &= \mathrm{E}\left\{\left[a_{v'}(\mathbf{u}_0) - a^*_{v'}(\mathbf{u}_0)\right]^2\right\} - \mathrm{E}\left\{\left[a_{v'}(\mathbf{u}_0) - a^*_{v'}(\mathbf{u}_0)\right]\right\} \mathrm{E}\left\{\left[a_{v'}(\mathbf{u}_0) - a^*_{v'}(\mathbf{u}_0)\right]\right\} \\
&= \mathrm{VAR}\{a^*_{v'}(\mathbf{u}_0)\}
\end{aligned} \tag{41}$$

And similar to (32), the final error variance is then:

$$\begin{aligned}\mathrm{VAR}\left\{a^*_{v'}(\mathbf{u}_0)\right\} &= \mathrm{E}\left\{\left[a^*_{v'}(\mathbf{u}_0)\right]^2\right\} - \mathrm{E}\left\{\left[a^*_{v'}(\mathbf{u}_0)\right]\right\}\mathrm{E}\left\{\left[a^*_{v'}(\mathbf{u}_0)\right]\right\} \\ &= \sum_{l'=1}^{n(\mathbf{u}_0)}\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{l'})\lambda^a_{v'}(\mathbf{u}_{i'})\mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\right\}\end{aligned} \tag{42}$$

The $\lambda_{v'}{}^{a}(\mathbf{u}_{l'})$ weights are determined so that the error variance in relation (42) is a minimum subject to the (V + 1) constraint equations in (40). A new Lagrange function $G_{v'}{}^{a}(\mathbf{u}_0)$ is then defined for each functional coefficient with the Lagranges $2\mu_{v'}{}^{a}(\mathbf{u}_0)$:

$$\mathrm{G}^a_{v'}(\mathbf{u}_0) = \mathrm{VAR}\left\{O(\mathbf{u}_0)\right\} + 2\mu^a_{v'}(\mathbf{u}_0)\left[\sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^a_{vv'}(\mathbf{u}_{l'})\, f_v(\mathbf{u}_{l'}) - f_v\left(\mathbf{u}_0\right)\right] \quad v = 0,\ldots,V \tag{43}$$

And the optimal weights $\lambda^{m}{}_{\mathrm{UK}}(\mathbf{u}_{l'})$ are obtained by setting simultaneously the $n(\mathbf{u}_0)$ partial derivatives of (32) with respect to each of the $\lambda^{m}{}_{\mathrm{UK}}(\mathbf{u}_{l'})$ weights to zero and the (V + 1) partial derivatives of (31) with respect to each $\mu^{m}{}_{v}(\mathbf{u}_0)$ Lagrange parameter to zero:

$$\begin{aligned}\frac{\partial\left(\mathrm{VAR}\left\{O(\mathbf{u}_0)\right\}\right)}{2\partial\left(\lambda^a_{v'}(\mathbf{u}_{l'})\right)} &= \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{i'})\mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\right\} + \sum_{v=0}^{V}\mu^a_{v'}(\mathbf{u}_0)\, f_v(\mathbf{u}_{l'}) = 0 \\ \frac{\partial\left(\mathrm{G}^a_{v'}(\mathbf{u}_0)\right)}{2\partial\left(\mu^a_{v'}(\mathbf{u}_0)\right)} &= \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{i'})\, f_{v'}(\mathbf{u}_{i'}) = 1 \\ \frac{\partial\left(\mathrm{G}^a_{v'}(\mathbf{u}_0)\right)}{2\partial\left(\mu^a_{v}(\mathbf{u}_0)\right)} &= \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{i'})\, f_{v}(\mathbf{u}_{i'}) = 0 \quad \forall v \neq v'\end{aligned} \tag{44}$$

This results in the following system equations for estimating $a^*{}_{v'}(\mathbf{u}_0)$:

$$\begin{aligned}&\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{i'})\mathrm{COV}\left\{R(\mathbf{u}_{l'})R(\mathbf{u}_{i'})\right\} = \sum_{v=0}^{V}\mu^a_{v'}(\mathbf{u}_0)\, f_v(\mathbf{u}_{l'}) \\ &\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{i'})\, f_{v'}(\mathbf{u}_{i'}) = 1 \\ &\sum_{i'=1}^{n(\mathbf{u}_0)} \lambda^a_{v'}(\mathbf{u}_{i'})\, f_{v}(\mathbf{u}_{i'}) = 0 \quad \forall v \neq v'\end{aligned} \tag{45}$$

This system must be set up and solved v times at each unknown location $\mathbf{u}_0$, once for each functional parameter $f_v(\mathbf{u})$ specified in (37). Also notice that the closeness covariances $\mathrm{COV}\{R(\mathbf{u}_{l'}), R(\mathbf{u}_0)+\mathbf{h}\}$ are all zero. Indeed, this is due to $\mathrm{E}\{a^*{}_{v'}(\mathbf{u}_0)a_{v'}(\mathbf{u}_0)\} = \mathrm{E}\{a_{v'}(\mathbf{u}_0)a_{v'}(\mathbf{u}_0)\} = 0$. The system in (45) then calculates the $\lambda_{v'}{}^{a}(\mathbf{u}_{l'})$ weights and $\mu_{v'}{}^{a}(\mathbf{u}_0)$ values that identify the least squares fit of the $a^*{}_{v'}(\mathbf{u}_0)$ coefficients. Then $m^*{}_{\mathrm{UK}}(\mathbf{u}_0)$ is obtained by combining the $a^*{}_{v'}(\mathbf{u}_0)$ coefficients with the functionals in (37).

The Link between Unconstrained and Constrained Kriging

The unconstrained (simple) and constrained (universal) kriging equations in systems (19) and (27), respectively, depend on the same form of kriging estimator $Z^*{}_{\mathrm{K}}(\mathbf{u}_0)$ in relation (12) and subsequent error variance in relation (13). The spatial law of residuals $R(\mathbf{u})$ is then a common requirement for both systems. However, the reader may be hesitant to accept that both unconstrained and constrained kriging simplify to the same general form of kriging estimator $Z^*_{K}(\mathbf{u}_0)$ in (12) and require the same $R(\mathbf{u})$ spatial law. In fact it can be shown that by setting $m = m^*_{UK}(\mathbf{u}_0)$ in (16) gives back the universal kriging estimator $Z^*_{UK}(\mathbf{u}_0)$, that is:

$$Z^*_{\mathrm{UK}}(\mathbf{u}_0) = m^*_{\mathrm{UK}}(\mathbf{u}_0) + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\left[Z(\mathbf{u}_{l'}) - m^*_{\mathrm{UK}}(\mathbf{u}_0)\right] \tag{46}$$

Formula (46) does not change the general form of (12) or (13). The model assumed for the mean is simply different. This section proves the equality in (46) three different ways. The first approach, adapted from Armstrong [6], explicitly replaces m with $m^*_{UK}(\mathbf{u}_0)$ and checks that system (27) is unchanged. This requires the system of universal kriging for the mean equations in (35). The second approach follows what is referred to as the additivity relationship as described by Chiles and Delfiner [7]. This requires the system of equations for estimating the mean coefficients in (45). The third approach exploits some relatively simple matrix manipulations following a set of advanced geostatistics course notes [8] taught at the University of Alberta. This also requires the system of universal kriging for the mean equations in (35).

The Substitution Approach

The simple kriging estimator $Z^*_{SK}(\mathbf{u}_0)$ in (16) can be rearranged as follows:

$$Z^*_{\mathrm{SK}}(\mathbf{u}_0) = \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'}) + m\left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\right) \tag{47}$$

The stationary mean m is now replaced with the kriging mean estimator $m^*_{UK}(\mathbf{u}_0)$ in (28):

$$\begin{aligned} Z^*_{\mathrm{K}}(\mathbf{u}_0) &= \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'}) + m\left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\right) \\ &= \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'}) + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'}) Z(\mathbf{u}_{l'})\left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\right) \\ &= \sum_{l'=1}^{n(\mathbf{u}_0)} Z(\mathbf{u}_{l'})\left[\lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) + \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\right)\right] \end{aligned} \tag{48}$$

This estimator (48) is in fact equivalent to the universal kriging estimator $Z^*_{UK}(\mathbf{u}_0)$. In order to prove this, it is sufficient to show that the estimator in (48) satisfies the universal kriging system in (27) [6]. The constraints in (23) are checked first. Since the sum of the universal kriging of the mean weights $\lambda^m_{UK}(\mathbf{u}_0)$ is always one for all $v = 0,\ldots, V$:

$$\begin{aligned} \ldots \quad & \sum_{l'=1}^{n(\mathbf{u}_0)} \left[\lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) + \lambda^m_{\mathrm{UK}}(\mathbf{u}_{l'})\left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\right)\right] f_v(\mathbf{u}) \\ &= f_v(\mathbf{u}) \end{aligned} \tag{49}$$

Relation (49) shows that in addition to satisfying the constraints in (23), the second line of the universal kriging system shown in (27) is satisfied. And all that is left to show is that the first line of system (27) is satisfied. The weight expression in (48) is substituted for the universal kriging weights $\lambda_{UK}(\mathbf{u}_0)$ in the first term multiplying the redundancy covariances in (27):

$$
\begin{aligned}
\ldots \quad & \sum_{i'=1}^{n(\mathbf{u}_0)} \left[\lambda_{\mathrm{SK}}(\mathbf{u}_{i'}) + \lambda_{\mathrm{UK}}^{m}(\mathbf{u}_{l'}) \left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) \right) \right] \mathrm{COV}\left\{ R(\mathbf{u}_{l'}) R(\mathbf{u}_{i'}) \right\} \\
&= \sum_{i'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{i'}) \mathrm{COV}\left\{ R(\mathbf{u}_{l'}) R(\mathbf{u}_{i'}) \right\} + \lambda_{\mathrm{UK}}^{m}(\mathbf{u}_{l'}) \left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) \right) \mathrm{COV}\left\{ R(\mathbf{u}_{l'}) R(\mathbf{u}_{i'}) \right\} \\
&= \mathrm{COV}\left\{ R(\mathbf{u}_{l'}) R(\mathbf{u}_0) \right\} - \left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) \right) \sum_{v=0}^{V} \mu_{v}^{m}(\mathbf{u}_0) f_v(\mathbf{u}_{l'})
\end{aligned} \tag{50}
$$

The substitutions in the last line come from the simple kriging system in (19) and the universal kriging of the mean system in (35). Now, since:

$$
\left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) \right) \sum_{v=0}^{V} \mu_{v}^{m}(\mathbf{u}_0) f_v(\mathbf{u}_{l'}) = \sum_{v=0}^{V} \mu_v(\mathbf{u}_0) f_v(\mathbf{u}_{l'}) \tag{51}
$$

Substituting back, the first line of system (27) is indeed preserved:

$$
\begin{aligned}
\ldots \quad & \mathrm{COV}\left\{ R(\mathbf{u}_{l'}) R(\mathbf{u}_0) \right\} - \left(1 - \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'}) \right) \sum_{v=0}^{V} \mu_{v}^{m}(\mathbf{u}_0) f_v(\mathbf{u}_{l'}) + \sum_{v=0}^{V} \mu_v(\mathbf{u}_0) f_v(\mathbf{u}_{l'}) \\
&= \mathrm{COV}\left\{ R(\mathbf{u}_{l'}) R(\mathbf{u}_0) \right\}
\end{aligned} \tag{52}
$$

The Additivity Approach

The additivity approach is more subtle, but is imbedded within an informative additivity theorem [7]. This theorem implies that the universal kriging estimator $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ in (46) can be decomposed into the sum of the simple kriging estimator $Z^*_{\mathrm{SK}}(\mathbf{u}_0)$ and some corrective difference term $Z^*_{\mathrm{D}}(\mathbf{u}_0)$ as follows:

$$
Z^*_{\mathrm{UK}}(\mathbf{u}_0) = Z^*_{\mathrm{SK}}(\mathbf{u}_0) + Z^*_{\mathrm{D}}(\mathbf{u}_0) \tag{53}
$$

The corrective term $Z^*_{\mathrm{D}}(\mathbf{u}_0)$ can be isolated and solved for by manipulating systems (19) and (27) according to (46). Matrix notation will make this process more efficient. The unconstrained kriging system in (19) becomes:

$$
\mathbf{C}\boldsymbol{\lambda}_{\mathrm{SK}} = \mathbf{c} \tag{54}
$$

where $\mathbf{C}$ represents all the redundancy covariances $\mathrm{COV}\{R(\mathbf{u}_{l'}), R(\mathbf{u}_{l'}) + \mathbf{h}\}$, $\mathbf{c}$ is the closeness covariance vector $\mathrm{COV}\{R(\mathbf{u}_{l'}), R(\mathbf{u}_0) + \mathbf{h}\}$, and $\boldsymbol{\lambda}_{\mathrm{SK}}$ represents the $n(\mathbf{u}_0)$ simple kriging weights vector $\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})$. Similarly, system (27) is rewritten:

$$
\begin{aligned}
\mathbf{C}\boldsymbol{\lambda}_{\mathrm{UK}} + \mathbf{F}\boldsymbol{\mu} &= \mathbf{c} \\
\mathbf{F}^{\mathrm{T}}\boldsymbol{\mu} &= \mathbf{f}_0
\end{aligned} \tag{55}
$$

where $\mathbf{F}$ represents the $f_v(\mathbf{u}_{l'})$ coefficients, $\boldsymbol{\mu}$ represents the $\mu_v(\mathbf{u}_0)$ Lagrange parameters, and $\mathbf{f}_0$ represents the $f_v(\mathbf{u}_0)$ values. The solution of $Z^*_{\mathrm{D}}(\mathbf{u}_0)$ is sought after and this requires the $Z^*_{\mathrm{SK}}(\mathbf{u}_0)$ to be subtracted from the $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ estimator, see (46). In making this subtraction, consider now carrying the entire system of equations with the operation. That is, the system in (54) is subtracted first from the first line of system (55) and then from the second line of system (55) to form a new system of equations for $Z^*_{\mathrm{D}}(\mathbf{u}_0)$:

$$
\begin{aligned}
&\mathbf{C}\left(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}}\right)+\mathbf{F}\boldsymbol{\mu}=0\\
&\mathbf{F}^{\mathrm{T}}\left(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}}\right)=\mathbf{f}_0-\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}
\end{aligned}
\tag{56}
$$

The solution of system (56) must take into account the general least squares estimate of the mean general formula in (46). For this, the system in (45) is solved. Setting **a** to the $a^*_{v'}(\mathbf{u})$ estimates and **z** to the data vector, the general least squares solution is [7, 9]:

$$
\mathbf{a}^*=\left(\mathbf{F}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{F}\right)\mathbf{F}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{z}
\tag{57}
$$

The system (56) can be resolved for $(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}})$ from either the first or second line in (56). Using the second line:

$$
\begin{aligned}
&\mathbf{F}^{\mathrm{T}}\left(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}}\right)=\mathbf{f}_0-\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}\\
&\mathbf{F}\mathbf{C}^{-1}\mathbf{F}^{\mathrm{T}}\left(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}}\right)=\mathbf{F}\mathbf{C}^{-1}\left[\mathbf{f}_0-\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}\right]\\
&\left(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}}\right)=\left(\mathbf{F}\mathbf{C}^{-1}\mathbf{F}^{\mathrm{T}}\right)^{-1}\mathbf{F}\mathbf{C}^{-1}\left[\mathbf{f}_0-\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}\right]\\
&\left(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}}\right)=\mathbf{a}^*\left[\mathbf{f}_0-\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}\right]
\end{aligned}
\tag{58}
$$

Returning to summation notation:

$$
\begin{aligned}
\left(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}}\right)&=\mathbf{a}^*\left[\mathbf{f}_0-\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}\right]\\
&=\sum_{v=0}^{V}a_v^*(\mathbf{u}_0)\left[f_v(\mathbf{u}_0)-\sum_{l'=1}^{n(\mathbf{u}_0)}\left[\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})f_v(\mathbf{u}_{l'})\right]\right]\\
&=\sum_{v=0}^{V}a_v^*(\mathbf{u}_0)f_v(\mathbf{u}_0)-\sum_{l'=1}^{n(\mathbf{u}_0)}\left[\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\sum_{v=0}^{V}a_v^*(\mathbf{u}_{l'})f_v(\mathbf{u}_{l'})\right]\\
&=m^*_{\mathrm{UK}}(\mathbf{u}_0)-\sum_{l'=1}^{n(\mathbf{u}_0)}\left[\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})m^*_{\mathrm{UK}}(\mathbf{u}_{l'})\right]
\end{aligned}
\tag{59}
$$

The $(\boldsymbol{\lambda}_{\mathrm{UK}}-\boldsymbol{\lambda}_{\mathrm{SK}})$ quantity is, therefore, a trend adjustment involving the universal kriging least squares mean estimate $m^*_{\mathrm{UK}}(\mathbf{u})$. Now, recombining the result of (59) with (53):

$$
\begin{aligned}
Z^*_{\mathrm{UK}}(\mathbf{u}_0)&=Z^*_{\mathrm{SK}}(\mathbf{u}_0)+Z^*_{\mathrm{D}}(\mathbf{u}_0)\\
&=\sum_{l'=1}^{n(\mathbf{u}_0)}\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\mathrm{Z}(\mathbf{u}_{l'})+m^*_{\mathrm{UK}}(\mathbf{u}_0)-\sum_{l'=1}^{n(\mathbf{u}_0)}\left[\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})m^*_{\mathrm{UK}}(\mathbf{u}_{l'})\right]\\
&=m^*_{\mathrm{UK}}(\mathbf{u}_0)+\sum_{l'=1}^{n(\mathbf{u}_0)}\lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\left[\mathrm{Z}(\mathbf{u}_{l'})-m^*_{\mathrm{UK}}(\mathbf{u}_{l'})\right]
\end{aligned}
\tag{60}
$$

The result in (46) is achieved.

The Matrix Manipulation Approach

This is perhaps the most direct and comprehensible of all 3 approaches. The left hand side matrix of the universal kriging system in (55) can be written as:

$$
\mathbf{T}=\begin{bmatrix}\mathbf{C} & \mathbf{F}\\ \mathbf{F}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}
\tag{61}
$$

with $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_{\mathrm{UK}}, \boldsymbol{\mu}]^{\mathrm{T}}$ and $\mathbf{t} = [\mathbf{c}, \mathbf{f}_0]^{\mathrm{T}}$; therefore, $\boldsymbol{\lambda} = \mathbf{T}^{-1}\mathbf{t}$. The universal kriging system of equation is then written in terms of the simple kriging system and weights. From (55):

$$\boldsymbol{\lambda}_{\mathrm{UK}} = \mathbf{C}^{-1}\mathbf{c} - \mathbf{C}^{-1}\boldsymbol{\mu}\mathbf{F} = \boldsymbol{\lambda}_{\mathrm{SK}} - \boldsymbol{\mu}\boldsymbol{\lambda}\mathbf{F} \tag{62}$$

since $\mathbf{C}\boldsymbol{\lambda}\mathbf{F} = \mathbf{F}$. Solving for $\boldsymbol{\mu}$ from (62) as:

$$-\boldsymbol{\mu} = \frac{\boldsymbol{\lambda}_{\mathrm{UK}} - \boldsymbol{\lambda}_{\mathrm{SK}}}{\boldsymbol{\lambda}\mathbf{F}} = \frac{1 - \mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}}{\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}} \tag{63}$$

And substituting (63) into (62) results in the following universal kriging weights:

$$\boldsymbol{\lambda}_{\mathrm{UK}} = \boldsymbol{\lambda}_{\mathrm{SK}} + \left(\frac{1 - \mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}}{\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}}\right)\boldsymbol{\lambda}\mathbf{F} \tag{64}$$

And universal kriging estimator:

$$Z^*_{\mathrm{UK}}(\mathbf{u}_0) = \mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{UK}} = \mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}} + \left(\frac{1 - \mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}}{\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}}\right)\mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F} \tag{65}$$

The universal kriging for the mean weights from (35) can be expressed as:

$$\boldsymbol{\lambda}^{\mathrm{m}}_{\mathrm{UK}} = -\boldsymbol{\mu}^{\mathrm{m}}\mathbf{C}^{-1}\mathbf{F} = -\boldsymbol{\mu}^{\mathrm{m}}\boldsymbol{\lambda}\mathbf{F} \tag{66}$$

And the universal kriging for the mean estimator becomes:

$$m^*_{\mathrm{UK}}(\mathbf{u}_0) = \mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{m}}_{\mathrm{UK}} = -\boldsymbol{\mu}^{\mathrm{m}}\mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F} = \frac{\mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}}{\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}} \tag{67}$$

Since $-\boldsymbol{\mu}^{\mathrm{m}} = 1/\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}$ (comes from (66)). Now the universal kriging estimator is written using (65) and (67):

$$\begin{aligned} Z^*_{\mathrm{UK}}(\mathbf{u}_0) &= \mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}} + \frac{\mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}}{\mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{F}}\left(1 - \mathbf{F}^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{SK}}\right) \\ &= m^*_{\mathrm{UK}}(\mathbf{u}_0) + \sum_{l'=1}^{n(\mathbf{u}_0)} \lambda_{\mathrm{SK}}(\mathbf{u}_{l'})\left[Z(\mathbf{u}_{l'}) - m^*_{\mathrm{UK}}(\mathbf{u}_0)\right] \end{aligned} \tag{68}$$

which matches (46).

All three of these proofs show that constrained kriging can be performed in two steps:

1. Least squares estimation of the local mean $m^*_{\mathrm{UK}}(\mathbf{u}_0)$ using universal kriging;

2. Applying the simple kriging estimator in (16) with the stationary mean m replaced with the estimated mean $m^*_{\mathrm{UK}}(\mathbf{u}_0)$.

Furthermore, this proves the kriging estimator $Z^*_{\mathrm{K}}(\mathbf{u}_0)$ construct in relation (12) is the same for unconstrained and constrained kriging. This same construct leads to the error variance in relation (13) and the common need for the $R(\mathbf{u})$ covariance or spatial law.

References

[1] Journel, A.G. and Huijbregts, Ch.J. *Mining Geostatistics*. The Blackburn Press, New Jersey, 1978.

[2] Journel, A.G. and Rossi, M.E. *When do we Need a Trend Model in Kriging?*, Mathematical Geology, vol. 21, no. 7, 1989.

[3] Journel, A.G. *Fundamentals of Geostatistics in Five Lessons*. American Geophysical Union 1989.

[4] Deutsch, C.V. and Journel, A.G. *GSLIB: Geostatistical Software and User's Guide (Second Edition)*, Oxford University Press, New York, 1998.

[5] Goovaerts, P. *Geostatistics for Natural Resource Evaluation*. Oxford University Press, New York, 1997.

[6] Armstrong, M. *Basic Linear Geostatistics*, Springer-Verlag Berlin Heidelberg, New York, 1998.

[7] Chiles, J. and Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York, 1999.

[8] Deutsch, C. V. *MINE 614: Advanced Topics in Geostatistics*. University of Alberta, 2000.

[9] Rao, C.R. *Linear Statistical Inference and its Applications*. Wiley, New York, 1973.