

Kriging Method

1 Introduction

Kriging is a method interpolating values which are modeled by a Gaussian process governed by prior covariances. Least-square linear regression Kriging gives the best linear unbiased prediction of the intermediate values.

1.1 Objetif

Consider the probleme of estimating the value of a continuous attribute at an unsample location using only the data of other location. - not only estimated value but probabilistique distribution - minimum error variance (MVUE) - low-pass filter ? pb: smooth out detail

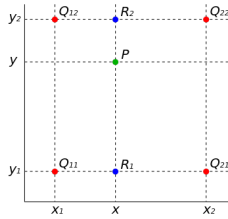
1.2 Other technique

This problem is generalized under the interpolation theory. The classical approach is to postulate a parametric function to fit. Exemple of that include:

- Nearest-neighbor interpolation: Take the value of the nearest data.
- 1D Linear interpolation: Assume a constant slope between each known data. Therefore, for a point x , between x_a and x_b , its value y is:

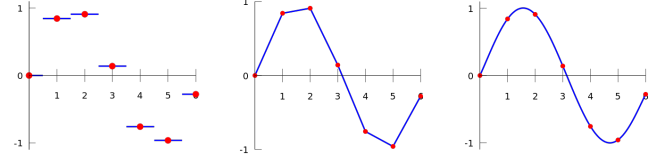
$$y = y_a + (y_b - y_a) \frac{x - x_a}{x_b - x_a} \text{ at the point } (x, y)$$

- Highest dimension linear interpolation. The extension of the linear interpolation is highest dimension does not create a linear function. Rather, these interpolation are a successive 1D interpolation. For instance, in the following figure, P is estimate by first linear interpolation of R_1 and R_2 in the y -direction and then interpolation P , between R_1 and R_2 .



- Polynomial interpolation: Generalisation of linear interpolation with highest degree polynomial equation. For a n -data point problem, there is exactly one polynomial of degree at most $n - 1$ which fit perfectly the data point. Pb: oscillatory effect and computational intensive.
- Spline Interpolation: is a piecewise polynomial.

Exemple of nearest neighbor (left), linear (center) and polynomial (right) . Note that spline look exactly similar than polynomial in this exemple.



1.3 History and Developement

The theoretical basis for the method was developed by the French mathematician Georges Matheron based on the Master's thesis of Danie G. Krige (1951), the pioneering plotter of distance-weighted average gold grades at the Witwatersrand reef complex in South Africa. Krige sought to estimate the most likely distribution of gold based on samples from a few boreholes.

1.4 Variogram

The variogram $2\gamma(\mathbf{u}_1, \mathbf{u}_2)$ is a function describing the degree of spatial dependence of a spatial random field $Z(\mathbf{u})$.

$$2\gamma(\mathbf{u}_1, \mathbf{u}_2) = \text{var} (Z(\mathbf{u}_1) - Z(\mathbf{u}_2)) \\ = E [((Z(\mathbf{u}_1) - m(\mathbf{u}_1)) - (Z(\mathbf{u}_2) - m(\mathbf{u}_1)))^2]$$

For first order stationary $m(\mathbf{u}_1) = m(\mathbf{u}_2)$ and with $\mathbf{h} = \mathbf{u}_2 - \mathbf{u}_1$

$$2\gamma(\mathbf{h}) = E [(Z(\mathbf{u}_0) - Z(\mathbf{u}_0 + \mathbf{h}))^2]$$

- nugget n : The height of the jump of the semi-variogram at the discontinuity at the origin.
- sill s : Limit of the variogram tending to infinity lag distances.
- range r : The distance in which the difference of the variogram from the sill becomes negligible.

Name	$\gamma(h)$	$C(L)$
Exponential	$C_0 [1 - \exp(\frac{-3L}{r})]$	$C_0 \exp(\frac{-3L}{r})$
Spherical	$C_0 [\frac{3L}{2r} - \frac{L^3}{2r^3}]$	$C_0 [1 - \frac{3L}{2r} + \frac{L^3}{2r^3}]$
Gaussian	$C_0 [1 - \exp(\frac{-3L^2}{r^2})]$	$C_0 \exp(\frac{-3L^2}{r^2})$
no-sill (Fractional Gaussian Noise or Brownian Motion)		
Logarithmic	$C_0 \log L$	
Sin	$C_0 [1 - \frac{\sin(rL)}{L}]$	$C_0 \frac{\sin(rL)}{L}$
Cos	$C_0 [1 - \cos(rL)]$	$C_0 \cos(rL)$

2 Kriging Basic Description

2.1 General Definition

All kriging version are elaborations of the regression of the following linear combination:

$$Z^*(\mathbf{u}_0) - m(\mathbf{u}_0) = \sum_{\alpha=1}^{n(\mathbf{u}_0)} \lambda_{\alpha} [Z(\mathbf{u}_{\alpha}) - m(\mathbf{u}_{\alpha})] \quad (2.1)$$

This express the unknown value Z^* at the location \mathbf{u} as a linear combination of value Z at the neighbouring point $\mathbf{u}_{\alpha} \in W(\mathbf{u}_0)$. λ_{α} is the relative importance of each point. And $m(\mathbf{u}) = E[Z(\mathbf{u})]$ is the location-dependant expected value.

As the point \mathbf{u}_0 and \mathbf{u}_α are close, it is assumed that their local means are the same : $m_0 = m(\mathbf{u}_\alpha) = m(\mathbf{u}_0)$. The renaming convention $\tilde{Z}_0^* = Z^*(\mathbf{u}_0) - m_0$ and $\tilde{Z}_\alpha = Z(\mathbf{u}_\alpha) - m_0$ lead to two variation of Equation 2.1

$$Z_0^* = \sum_{\alpha=1}^{n_0} \lambda_\alpha Z_\alpha + \left(1 - \sum_{\alpha=1}^{n_0} \lambda_\alpha\right) m_0 \quad (2.2)$$

$$\tilde{Z}_0^* = \sum_{\alpha=1}^n \lambda_\alpha \tilde{Z}_\alpha \quad (2.3)$$

In this equation, the three unknowns are λ_α , m_0 and \tilde{Z}_0^*

2.2 MVUE

The minimum-variance unbiased estimator (MVUE) is the estimator which is accurate as well as precise.

- Unbiasedness: $B(\theta^*) = E[\theta^*] - \theta = E[\theta^* - \theta] = 0$. The bias is the difference between this estimator's (θ^*) expected value and the true value of the parameter (θ) being estimated.

$$\begin{aligned} B(Z_0^*) &= E[Z_0] - E[Z_0^*] \\ &= \sum_{\alpha=1}^{n_0} \lambda_\alpha E[Z_\alpha] + \left(1 - \sum_{\alpha=1}^{n_0} \lambda_\alpha\right) m_0 - E[Z_0^*] \\ &= \sum_{\alpha=1}^{n_0} \lambda_\alpha m_0 + m_0 - \sum_{\alpha=1}^{n_0} \lambda_\alpha m_0 - m_0 \\ &= 0 \end{aligned}$$

Bias is always achieved for all method.

- Lowest variance: $\sigma_E^{2*} = \text{var}\{\theta^* - \theta\}$. Minimizing this variance try to not only be accurate but also precise. We can rewritten the equation using notation of equation 2.2:

$$\begin{aligned} \sigma_E^{2*} &= \text{var}\{Z_0 - Z_0^*\} \\ &= \text{var}\{(Z_0 - m_0) - (Z_0^* - m_0)\} \\ &= \text{var}\{\tilde{Z}_0 - \tilde{Z}_0^*\} \\ &= \text{var}\{\tilde{Z}_0 - \tilde{Z}_0^*\} \\ &= \text{var}\{\tilde{Z}_0\} + \text{var}\left\{\sum_{\alpha=1}^n \lambda_\alpha \tilde{Z}_\alpha\right\} - 2C\left\{\tilde{Z}_0, \sum_{\alpha=1}^n \lambda_\alpha \tilde{Z}_\alpha\right\} \\ &= \text{var}\{\tilde{Z}_0\} + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C\{\tilde{Z}_\alpha, \tilde{Z}_\beta\} \\ &\quad - 2 \sum_{\alpha=1}^n \lambda_\alpha C\{\tilde{Z}_0, \tilde{Z}_\alpha\} \end{aligned} \quad (2.4)$$

using,

$$\begin{aligned} - \text{var}\{X - Y\} &= \text{var}\{X\} + \text{var}\{Y\} - C\{X, Y\} \\ - \text{var}\left\{\sum_{i=1}^n a_i X_i\right\} &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j C[X_i, X_j] \end{aligned}$$

3 Method

3.1 Simple Kriging

Simple Kriging (SP) assume location-independent mean $m_0 = m$ (first order stationarity) which mean that the mean is the

same throughout the study area (an not only around the simulated point). The mean can be estimated with all known point ($m = \sum Z_i$) and λ_α is the only unknown left. This can be find by minimizing σ_E^{2*} which is found when his derivative is equal to 0 for each λ_α :

$$\frac{\partial \sigma_E^{2*}}{\partial \lambda_\alpha} = 0 = 2 \sum_{\beta=1}^n \lambda_\beta C\{\tilde{Z}_\alpha, \tilde{Z}_\beta\} - 2C\{\tilde{Z}_0, \tilde{Z}_\alpha\} \quad (3.1)$$

Which can be rewritten:

$$\sum_{\beta=1}^n \lambda_\beta C\{\tilde{Z}_\alpha, \tilde{Z}_\beta\} = C\{\tilde{Z}_0, \tilde{Z}_\alpha\} \quad \text{for } \alpha = 1, \dots, n \quad (3.2)$$

A this point, we want to come back to normal notation (Z instead of \tilde{Z}).

$$\begin{aligned} C\{\tilde{X}, \tilde{Y}\} &= E[(X - m_0)(Y - m_0)] - E[(X - m_0)]E[(Y - m_0)] \\ &= E[XY - m_0X - m_0Y + m_0^2] \\ &\quad - (E[X] - m_0)(E[Y] - m_0) \\ &= E[XY] - m_0 E[X] - m_0 E[Y] + m_0^2 \\ &\quad - E[X]E[Y] + m_0 E[X] + m_0 E[Y] - m_0^2 \\ &= E[XY] - E[X]E[Y] \\ &= C\{X, Y\} \end{aligned}$$

This can be rewrite as a matrix notation and can be identify as an linear system:

$$\begin{bmatrix} C\{Z_1, Z_1\} & \dots & C\{Z_1, Z_n\} \\ \vdots & & \vdots \\ C\{Z_n, Z_1\} & \dots & C\{Z_n, Z_n\} \end{bmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{bmatrix} C\{Z_1, Z_0\} \\ \vdots \\ C\{Z_n, Z_0\} \end{bmatrix}$$

The λ can be found with $\mathbf{\Lambda} = \mathbf{C}_{\alpha,\beta}^{-1} \mathbf{C}_{\alpha,0}$, equations 2.1 or 2.3 is used to find the estimator Z_0^* .

We can put back the minimized solution (equation 3.2) into the variance equation 2.4 :

$$\sigma_E^{2*} = \text{var}\{Z_0\} - \sum_{\alpha=1}^n \lambda_\alpha C\{Z_0, Z_\alpha\} \quad (3.3)$$

The error variance in the estimation is equal to the variance of the data $\text{var}\{Z_0\}$ reduced by the information brought by point in the neighbourhood of \mathbf{u}_0 . Note that we used the covariance, but variogram could be interchangeably used

3.2 Ordinary Kriging

Ordinary Kriging doesn't assume first-order stationary (m_0 is unknown). In this case, in order to satisfy equation 2.3, it is useful to remove the second part of the equation by setting

$$\sum_{\alpha=1}^{n_0} \lambda_\alpha = 1 \quad (3.4)$$

This times, minimizing the variance σ_E^{2*} is more trick as it is subject to equation 3.4. Solution is find using the method of the Lagrangian multiplier. It is a strategy for finding the local minima \mathbf{u}_0 of a function $f(\mathbf{u})$ subject to equality constraints

$g(\mathbf{u}) = 0$. The Lagrangian function is define as: $F(\mathbf{u}, 2\mu) = f(\mathbf{u}) + 2\mu g(\mathbf{u})$. This idea behind this method is that if \mathbf{u}_0 is a minimum point, then there is a point μ_0 such that (\mathbf{u}_0, μ_0) is a stationary point for the lagrangian function. Stationary point are points where all partial derivatives of F are zero $\nabla_{\mathbf{u}, \mu} F(\mathbf{u}_0, \mu_0) = 0$

In our case the Lagrangian function is:

$$F(\lambda_\alpha, 2\mu) = \sigma_E^2 + 2\mu \left(\sum_{\alpha=1}^{n_0} \lambda_\alpha - 1 \right) \quad (3.5)$$

Leading up to finding the stationary point

$$\begin{bmatrix} \frac{\partial F}{\partial \lambda_\alpha} = \sum_{\beta=1}^n \lambda_\beta C\{Z_\alpha, Z_\beta\} - C\{Z_0, Z_\alpha\} + \mu \\ \frac{\partial F}{\partial \mu} = \sum_{\alpha=1}^{n_0} \lambda_\alpha - 1 \end{bmatrix} = 0 \quad (3.6)$$

The procedure is similar to simple kriging but with a new constraint in our system.

$$\begin{bmatrix} C\{Z_1, Z_1\} & \cdots & C\{Z_1, Z_n\} & 1 \\ \vdots & & & \vdots \\ C\{Z_1, Z_n\} & \cdots & C\{Z_n, Z_n\} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{bmatrix} C\{Z_1, Z_0\} \\ \vdots \\ C\{Z_n, Z_0\} \\ 1 \end{bmatrix}$$

Once the λ have been found, equation 2.3 can be simplified and help us compute the value at the unsample location.

$$Z_0^* = \sum_{\alpha=1}^{n_0} \lambda_\alpha Z_\alpha \quad (3.7)$$

In a similar matter, variance of the computed value can be found with

$$\sigma_E^2 = \text{var}\{Z_0\} - \sum_{\alpha=1}^n \lambda_\alpha C\{Z_0, Z_\alpha\} - \mu \quad (3.8)$$

3.3 Kriging with a Trend

The assumption made in the general description that the mean is constant in the search neighbour $m(\mathbf{u}_0) = m(\mathbf{u}_\alpha), \alpha \in W(\mathbf{u}_0)$ might not hold in the presence of a trend. In the Universal Kriging (with a trend), the local mean in a search neighbor is modelled with a smoothly varying function:

$$m(\mathbf{u}_\alpha) = \sum_{k=0}^K a_k f_k(\mathbf{u}_\alpha)$$

Where f_k are known function while the coefficient a_k are unknown. Note that if $k=0$, we end up with Ordinary Kriging. Equation 2.3 became:

$$Z_0^* = \sum_{\alpha=1}^{n_0} \lambda_\alpha Z_\alpha + \left(f_{k,0} - \sum_{\alpha=1}^{n_0} \lambda_\alpha f_{k,\alpha} \right) \sum_{k=0}^K a_k = 0^K a_k \quad (3.9)$$

We follow the same procedure as Ordinary kriging and find a similar equation to equation 3.6

$$\begin{cases} C\{Z_0, Z_\alpha\} = \sum_{\beta=1}^n \lambda_\beta C\{Z_\alpha, Z_\beta\} + \sum_{k=0}^K \mu_k f_{k,\alpha} & \text{for } \alpha = 1, \dots, n_0 \\ f_{k,0} = \sum_{\alpha=1}^{n_0} \lambda_\alpha f_{k,\alpha} & \text{for } k = 0, \dots, K \end{cases} \quad (3.10)$$

$$\begin{bmatrix} C\{Z_\alpha, Z_\beta\} & [f_{k,\alpha}]^T \\ [f_{k,\beta}] & [0] \end{bmatrix} \begin{pmatrix} [\lambda_\beta]^T \\ [\mu_k]^T \end{pmatrix} = \begin{bmatrix} C\{Z_\alpha, Z_0\}^T \\ [f_{k,0}]^T \end{bmatrix} \quad (3.11)$$

Once the λ_β and μ_k are found, kriging estimate and variance can be computed.

3.4 Block Kriging

Block Estimation is used when the estimated location is not a point but has an area. In this case, the estimator is assumed to be the arithmetic average of all point inside the volume $(\mathbf{u}_{0,i})$:

$$Z_V^*(\mathbf{u}_0) = \frac{1}{N} \sum_{i=0}^N Z^*(\mathbf{u}_{0,i})$$

This is a arithmetic average and transformation might be necessary. The procedure is similar to ordinary (or simple) kriging except that the covariance between the known point and unknown volume is computed by averaging the covariance between known point and point inside the volume

$$C\{Z_\alpha, Z_{V,0}\} = \frac{1}{N} \sum_{i=0}^N C\{Z_\alpha, Z_{0,i}\}$$

$$\begin{bmatrix} C\{Z_\alpha, Z_\beta\} & [1]^T \\ [1] & [0] \end{bmatrix} \begin{pmatrix} [\lambda_\beta]^T \\ [\mu] \end{pmatrix} = \begin{bmatrix} C\{Z_\alpha, Z_{V,0}\}^T \\ [1] \end{bmatrix} \quad (3.12)$$

3.5 Dual Kriging

Dual Kriging is just a way of presenting kriging where the estimate is express as a linear combination of covariance instead of values. Solution for λ ($\Lambda = \mathbf{C}_{\alpha,\beta}^{-1} \mathbf{C}_{\alpha,0}$) can be incorporated in the general definition of kriging (equation 2.1) and lead up to:

$$Z^*(\mathbf{u}_0) - m(\mathbf{u}_0) = \sum_{\alpha=1}^{n(\mathbf{u}_0)} d_\alpha C\{Z(\mathbf{u}_\alpha), Z(\mathbf{u}_0)\} \quad (3.13)$$

The weight d_α are obtain by identification of data

$$\begin{bmatrix} C\{Z_\alpha, Z_\beta\} \\ [1] \end{bmatrix} [d_\beta]^T = \begin{bmatrix} [Z_\alpha]^T - m_0 \\ [0] \end{bmatrix} \quad (3.14)$$

With dual kriging we are not minimizing an error variance but "calibrate" the weight on the known values. What's really interesting here is that the weight d_α are only depending on \mathbf{u}_α and \mathbf{u}_β but not on \mathbf{u}_0 . This made possible to compute the weight once and use them for several point.

3.6 Filtering

One exemple of filtering is with a nugget effect: we don't want the direct neighbouring point ($Z(\mathbf{u}_{min})$ with $\|\mathbf{u}_{min} - \mathbf{u}_0\| < d_{min}$) to influence the estimated value. Equation 3.13 helps us to see how filtering can be apply to kriging by simply putting covariance at certain distance equal to zero.

3.7 Non-linear Kriging

Log-Normal Kriging is a non-linear kriging procedure which (1) reduces the variability of sample point and (2) transform log-normal distributed variables in normale distribution. The back-transforme depend on error variance. (not very used) *Multi-gaussian* convert with a normal z-score the data.

3.8 Indicator Kriging

final result represent the uncertain distribution at unsampled location (instead of single value) and has not simple back transform which therefore require a conditional simulation technique. Indicator variable express the confidence in the sample values. It therefore allows incomplete or soft information (value less than 1). There are not distribution assumed in the data.

$$I(u_j, x_t) = \begin{cases} 1, & \text{if } X(u_j) < x_t \\ 0, & \text{if } X(u_j) > x_t \end{cases}$$

4 Multivariate Kriging

4.1 Cokriging

Cokriging is used to estimate one variable (primary Z) value knowing other sampled variables (covariables Y). It use the assumption that the variables are spatially related with a linear relationship. The relationship must be strong: (1) high correlation, (2) physical link and (3) good history matching. This procedure is very computer expensive and numerical unstable.

The procedure is similar, this time the estimate is a weighted linear combination of the surrounding value of the primary and those of the secondary variable. In order to find the weight, in addition to the covariance structure of the primary variable (CZ_α, Z_α), we also need the one of the secondary variable (CZ_α, Y_β) and the cross-covariance of both (CY_α, Y_β).

$$\tilde{Z}_0^* = \sum_{\alpha^Z=1}^n \lambda_{\alpha^Z}^Z Z_{\alpha^Z} + \sum_{\alpha^Y=1}^m \lambda_{\alpha^Y}^Y Y_{\alpha^Y} \quad (4.1)$$

In ordinary cokriging, we need $m_{Z,0} - \sum_{\alpha}^Z \lambda_{\alpha^Z}^Z m_{Z,0} - \sum_{\alpha}^Y \lambda_{\alpha^Y}^Y m_{Y,0} = 0$, which is achives for both $\sum_{\alpha} \lambda_{\alpha} = 1$ (as usual) and $\sum_{\beta} \lambda_{\beta} = 0$. This additional constraint can be put in the matrix format :

$$\begin{bmatrix} C\{Z, Z\} & C\{Z, Y\} & [1]_{n \times 1} & [0]_{n \times 1} \\ C\{Y, Z\} & C\{Y, Y\} & [1]_{n \times 1} & [0]_{n \times 1} \\ [1]_{1 \times n} & [0]_{1 \times n} & 0 & 0 \\ [0]_{1 \times n} & [1]_{1 \times n} & 0 & 0 \end{bmatrix} \begin{pmatrix} \lambda_Z \\ \lambda_Y \\ \mu_Z \\ \mu_Y \end{pmatrix} = \begin{bmatrix} C\{Z_0, Z\} \\ C\{Z_0, Y\} \\ 1 \\ 0 \end{bmatrix}$$

4.2 Kriging with external Drift

4.3 Collocated cokriging

4.4 Cosimulation

4.5 ...

5 Search Strategie

Search strategie are used in computing covariance matrices. Three reasons encourages algorithm to only a limited number of neraby conditioning data :

1. CPU time to solve a kriging system is $\mathcal{O}(n^3)$ and memory storage is $\mathcal{O}(n^2)$,

2. Knowlegde of covariance for very distance over half or third of domain is poorly known
3. local search allow local rescaling of the mean for ordinary kriging.

Three technique will be explained:

- The simplest implementation is to check for all available point and retaining only the one inside the ellipsoid (Exhaustive search). The previous work was sorting the point and selecting the n-closest (?)
- Composite sampling presented in (?) selects the n-closest points and then split the remaining point into quadrant (or octant). When building the covariance matrix, points on each quadrant are being treated as block points. This results in a similar weight for all points inside the same quadrant.
- The super Block Search (??) consist of creating a coarse grid network surimposing the data and associated each data point to its corresponding block. When searching the neighbourhood, distance is computed on super blocks center, retaining those blocks following inside the ellipse, and all data points contained in the selected blocks are used for kriging. This method is particularly useful for scatter data with an unchanging number of point. The number (and size) of the blocks is a critical parameter to avoid boundary artefact.
- Spiral Search (?) is very efficient for regular gridded data as it looks sequentially at all grid nodes starting by the closer one and going further away. It can stop whenever a certain number of point is reached.
- Quadrant Search is especially interesting for clustered data as it splits the search window in 4 (or 8 in 3D) zones and can limit the number of points in each quadrant (or octant) separately. This force an equal directional sampling.
- For gridded sequential simulation with scattered conditioning hard data, a two-part search combining both super block search for hard data and spiral search for previously simulated point is suggested by ?. One advantages is that hard data can be selected in a different manner than previously simulated point and therefore it is possible to give a special effort to reproduce the hard data effect.
- Search Tree Structure (?) successively divide the space in quadrant (or octant for 3D) indexing each region. In this way, different area of the space are sub-divided at different level. The nearest search become a simple query on the tree using the index of the region to find the corresponding nearest point. The memory usage is particularly low because each node only store its parent nodes up to the root. This method is useful for unstructured grids.

Cross validation of subspace can help to select the search criteria.

6 Limitation

- 1st order stationary (mean cst) required

- Only applicable for Gaussian-related stochastic model
- if search long enough : global minima is achieved.
- In theory, spacial structure is preserve. But in practise, because each simulation is not perfect (exact mean, variance and covariance) spatial statistical structure is not preserve. This become a serious problem when the number of realisation is higher than the number of grid cells (see ?)
- Preserve only variogram structure (covariance) not multi-point
- for linear problems, GDM's optimizations converge exponentially to the global minimum but not for non linear.
- Something about getting accurate estimate... using a prior constraint ?

7 To go futher

- Cross validation
-