Kriging

Allan Aasbjerg Nielsen
Department of Mathematical Modelling
Technical University of Denmark, Building 321
DK-2800 Lyngby, Denmark
http://www.imm.dtu.dk/∼aa
aa@imm.dtu.dk

14 July 1997

"Kriging" (after the South African mining engineer Danie Krige) is a term used for a family of methods for minimum error variance estimation. Consider a linear estimate $\hat{z}_0 = \hat{z}(\boldsymbol{r}_0)$ at a location $\boldsymbol{r}_0$ based on N measurements $\boldsymbol{z} = [z(\boldsymbol{r}_1), \ldots, z(\boldsymbol{r}_N)]^T = [z_1, \ldots, z_N]^T$

$$\hat{z}_0 = w_0 + \sum_{i=1}^{N} w_i z_i = w_0 + \boldsymbol{w}^T \boldsymbol{z}, \tag{1}$$

where w_i are the weights applied to z_i.

We consider z_i as particular realisations of random variables Z_i, $\boldsymbol{Z} = [Z(\boldsymbol{r}_1), \ldots, Z(\boldsymbol{r}_N)]^T = [Z_1, \ldots, Z_N]^T$. We think of $Z(\boldsymbol{r})$ as consisting of a mean and a residual $Z(\boldsymbol{r}) = \mu(\boldsymbol{r}) + \epsilon(\boldsymbol{r})$ with a constant variance σ^2. $\mathrm{E}\{\epsilon\} = 0$. For the linear estimator we get

$$\hat{Z}_0 = w_0 + \boldsymbol{w}^T \boldsymbol{Z}. \tag{2}$$

The actual error $z_0 - \hat{z}_0$ is generally unknown but for the expectation of the estimation error we get

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \mathrm{E}\{Z_0 - w_0 - \boldsymbol{w}^T \boldsymbol{Z}\} = \mu_0 - w_0 - \boldsymbol{w}^T \boldsymbol{\mu}, \tag{3}$$

where $\mu_0 = \mu(\boldsymbol{r}_0)$ is the expectation value of Z_0 and $\boldsymbol{\mu}$ is a vector of expectation values

$$\boldsymbol{\mu} = \begin{bmatrix} \mu(\boldsymbol{r}_1) \\ \vdots \\ \mu(\boldsymbol{r}_N) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}. \tag{4}$$

We want our estimator to be unbiased, i.e. we request $\mathrm{E}\{Z_0 - \hat{Z}_0\} = 0$ or

$$\mu_0 - w_0 - \boldsymbol{w}^T\boldsymbol{\mu} = 0. \tag{5}$$

The estimation variance is

$$\begin{aligned}\sigma_E^2 = \mathrm{Var}\{Z_0 - \hat{Z}_0\} = \mathrm{Var}\{Z_0\} + \mathrm{Var}\{w_0 + \boldsymbol{w}^T\boldsymbol{Z}\} - 2\mathrm{Cov}\{Z_0, w_0 + \boldsymbol{w}^T\boldsymbol{Z}\} \\ = \sigma^2 + \boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}\boldsymbol{w},\end{aligned} \tag{6}$$

where $\boldsymbol{C}$ is the dispersion of $\boldsymbol{Z}$, $\boldsymbol{C} = \mathrm{D}\{\boldsymbol{Z}\}$. σ_E^2 is a quadratic function in $\boldsymbol{w}$.

Simple Kriging

In "simple kriging" we assume that $\mu(\boldsymbol{r})$ is known (often assumed constant). The solution is

$$\hat{Z}_0 - \mu_0 = \boldsymbol{w}^T(\boldsymbol{Z} - \boldsymbol{\mu}) \tag{7}$$

The weights w_i are found by minimising σ_E^2

$$\frac{\partial\sigma_E^2}{\partial\boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\} = \boldsymbol{0}, \tag{8}$$

which gives the simple kriging system

$$\boldsymbol{C}\boldsymbol{w} = \mathrm{Cov}\{Z_0, \boldsymbol{Z}\} \tag{9}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} \\ \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \end{bmatrix}. \tag{10}$$

The minimum squared estimation error is

$$\sigma_{\mathrm{SK}}^2 = \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}) = \sigma^2 - \boldsymbol{w}^T\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}. \tag{11}$$

In SK the mean is known. In practice one must estimate $\mu(\boldsymbol{r})$ prior to the estimation or design an estimation algorithm that requires no prior mean.

Ordinary Kriging

In "ordinary kriging" we assume that the mean is constant for the N points that enter into the estimation of Z_0

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \mu_0(1 - \boldsymbol{w}^T\mathbf{1}) - w_0 = 0 \tag{12}$$

for any μ_0. $\mathbf{1}$ is a vector of ones. This is possible only if $w_0 = 0$ and $\boldsymbol{w}^T\mathbf{1} = 1$.

The weights w_i are found by minimising σ_E^2 with $\boldsymbol{w}^T\mathbf{1} = 1$. Introduce F with Lagrange multiplier -2λ

$$F = \sigma_E^2 + 2\lambda(\boldsymbol{w}^T\mathbf{1} - 1) \tag{13}$$

$$\frac{\partial F}{\partial \boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\} + 2\lambda\mathbf{1} = \mathbf{0} \tag{14}$$

$$\frac{\partial F}{\partial \lambda} = 2(\boldsymbol{w}^T\mathbf{1} - 1) = 0 \tag{15}$$

which gives the ordinary kriging system

$$\boldsymbol{C}\boldsymbol{w} + \lambda\mathbf{1} = \mathrm{Cov}\{Z_0, \boldsymbol{Z}\} \tag{16}$$

$$\mathbf{1}^T\boldsymbol{w} = 1 \tag{17}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{N1} & \cdots & C_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \lambda \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \\ 1 \end{bmatrix}. \tag{18}$$

The minimum squared estimation error is

$$\sigma^2_{\mathrm{OK}} = \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}) = \sigma^2 - \boldsymbol{w}^T\mathrm{Cov}\{Z_0, \boldsymbol{Z}\} - \lambda. \tag{19}$$

OK implies a re-estimation of μ_0 for each new estimation support.

Universal Kriging

In "universal kriging" we assume that the mean can be written as a linear combination of known functions (ideally determined by the physics of the problem dealt with)

$$\mu(\boldsymbol{r}) = \sum_{\ell=0}^{L} a_\ell f_\ell(\boldsymbol{r}) = \boldsymbol{a}^T \boldsymbol{f}(\boldsymbol{r}). \tag{20}$$

If no knowledge exists about the form of $\boldsymbol{f}$ we can use low order (typically second order) polynomials and consider them as local Taylor expansions. By convention $f_0 = 1$.

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \boldsymbol{a}^T \boldsymbol{f}(\boldsymbol{r}_0) - w_0 - \boldsymbol{w}^T \boldsymbol{\mu}. \tag{21}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu(\boldsymbol{r}_1) \\ \vdots \\ \mu(\boldsymbol{r}_N) \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}^T \boldsymbol{f}(\boldsymbol{r}_1) \\ \vdots \\ \boldsymbol{a}^T \boldsymbol{f}(\boldsymbol{r}_N) \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^T(\boldsymbol{r}_1)\boldsymbol{a} \\ \vdots \\ \boldsymbol{f}^T(\boldsymbol{r}_N)\boldsymbol{a} \end{bmatrix} = \boldsymbol{H}\boldsymbol{a} \tag{22}$$

with

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{f}^T(\boldsymbol{r}_1) \\ \vdots \\ \boldsymbol{f}^T(\boldsymbol{r}_N) \end{bmatrix}, \quad \boldsymbol{H}^T = [\boldsymbol{f}(\boldsymbol{r}_1) \ \cdots \ \boldsymbol{f}(\boldsymbol{r}_N)] \tag{23}$$

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \boldsymbol{a}^T \boldsymbol{f}(\boldsymbol{r}_0) - w_0 - \boldsymbol{w}^T \boldsymbol{H}\boldsymbol{a} = \boldsymbol{a}^T(\boldsymbol{f}(\boldsymbol{r}_0) - \boldsymbol{H}^T\boldsymbol{w}) - w_0 = 0 \tag{24}$$

for any $\boldsymbol{a}$. This is possible only if $w_0 = 0$ and $\boldsymbol{H}^T\boldsymbol{w} = \boldsymbol{f}(\boldsymbol{r}_0)$.

The weights w_i are found by minimising σ_E^2 with $\boldsymbol{H}^T\boldsymbol{w} = \boldsymbol{f}_0$. Introduce F with $L+1$ Lagrange multipliers $-2\boldsymbol{\lambda}$ where $\boldsymbol{\lambda} = [\lambda_0, \ldots, \lambda_L]^T$

$$F = \sigma_E^2 + 2\boldsymbol{\lambda}^T(\boldsymbol{H}^T\boldsymbol{w} - \boldsymbol{f}(\boldsymbol{r}_0)) \tag{25}$$

$$\frac{\partial F}{\partial \boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\} + 2\boldsymbol{H}\boldsymbol{\lambda} = \boldsymbol{0} \tag{26}$$

$$\frac{\partial F}{\partial \boldsymbol{\lambda}} = 2(\boldsymbol{H}^T\boldsymbol{w} - \boldsymbol{f}(\boldsymbol{r}_0)) = \boldsymbol{0} \tag{27}$$

which gives the universal kriging system

$$\boldsymbol{C}\boldsymbol{w} + \boldsymbol{H}\boldsymbol{\lambda} = \mathrm{Cov}\{Z_0, \boldsymbol{Z}\} \tag{28}$$

$$\boldsymbol{H}^T\boldsymbol{w} = \boldsymbol{f}(\boldsymbol{r}_0) \tag{29}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & f_0(\boldsymbol{r}_1) & \cdots & f_L(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} & f_0(\boldsymbol{r}_N) & \cdots & f_L(\boldsymbol{r}_N) \\ f_0(\boldsymbol{r}_1) & \cdots & f_0(\boldsymbol{r}_N) & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_L(\boldsymbol{r}_1) & \cdots & f_L(\boldsymbol{r}_N) & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \lambda_0 \\ \vdots \\ \lambda_L \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \\ f_0(\boldsymbol{r}_0) \\ \vdots \\ f_L(\boldsymbol{r}_0) \end{bmatrix}. \tag{30}$$

The minimum squared estimation error is

$$\sigma^2_{\mathrm{UK}} = \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}) = \sigma^2 - \boldsymbol{w}^T\mathrm{Cov}\{Z_0, \boldsymbol{Z}\} - \boldsymbol{\lambda}^T\boldsymbol{f}(\boldsymbol{r}_0). \tag{31}$$

When applying UK $\boldsymbol{C}$ and $\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}$ must be estimated after removal of the trend that we are estimating along with the stochastic part. This is a problem that limits the use of UK (one could argue that this characteristic doesn't go too well with the adjective "universal").

We can use UK to apply an external drift for example by setting $L = 1$

$$\boldsymbol{a}^T\boldsymbol{f}(\boldsymbol{r}) = \sum_{\ell=0}^{L} a_\ell f_\ell(\boldsymbol{r}) = a_0 + a_1 f_1(\boldsymbol{r}). \tag{32}$$

An external drift is easily allowed for before estimating $\boldsymbol{C}$ and $\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}$.

References

A.G. Journel and Ch.J. Huijbregts (1978). *Mining Geostatistics.* Academic Press.

E.H. Isaaks and R.M. Srivastava (1989). *Applied Geostatistics*. Oxford University Press.