Derivations of Kriging Equations

Meredith Franklin

September 30, 2025

1 Lagrange Multipliers

Lagrange multipliers solve constrained extrema for $f(x_1, \ldots, x_n)$ subject to $g(x_1, \ldots, x_n) = 0$ when f, g are continuously differentiable and $\nabla g \neq 0$. At an extremum, $\nabla f = -\lambda \nabla g$.

For $f(x, y) = x^2 + y^2$ and $g(x, y) = x^2 y - 16 = 0$,

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(x^2 y - 16).$$

First-order conditions (FOCs):

$$0 = \frac{\partial L}{\partial x} = 2x + 2\lambda xy = 2x(1 + \lambda y), \tag{1}$$

$$0 = \frac{\partial L}{\partial y} = 2y + \lambda x^2, \tag{2}$$

$$0 = \frac{\partial L}{\partial \lambda} = x^2 y - 16. \tag{3}$$

From (1), $x = 0$ or $\lambda y = -1$. The case $x = 0$ contradicts (3), so $\lambda = -1/y$. Plugging into (2) gives $2y - (x^2)/y = 0 \Rightarrow x^2 = 2y^2$. With (3): $x^2 y = 16 \Rightarrow 2y^3 = 16 \Rightarrow y = 2$ and $x = \pm 2\sqrt{2}$. These two points are minima of f on $x^2 y = 16$; no maximum exists.

2 Ordinary Kriging (OK)

We want to predict $Z(s_0)$ using a linear combination of observations:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} w_i Z(s_i).$$

Unbiasedness condition

Because $E[Z(s_i)] = \mu$ for all i, we have

$$E[\hat{Z}(s_0)] = \sum_{i=1}^{N} w_i E[Z(s_i)] = \mu \sum_{i=1}^{N} w_i.$$

For unbiasedness we require $E[\hat{Z}(s_0)] = E[Z(s_0)] = \mu$, which implies

$$\sum_{i=1}^{N} w_i = 1.$$

Mean squared error (MSE)

The prediction error is

$$\hat{Z}(s_0) - Z(s_0) = \sum_{i=1}^{N} w_i Z(s_i) - Z(s_0).$$

The mean squared error (MSE) is

$$E\big[(\hat{Z}(s_0) - Z(s_0))^2\big] = \mathrm{Var}\big(\hat{Z}(s_0) - Z(s_0)\big).$$

Expanding the variance term:

$$\mathrm{Var}\big(\hat{Z}(s_0) - Z(s_0)\big) = \mathrm{Var}\Big(\sum_{i=1}^{N} w_i Z(s_i)\Big) + \mathrm{Var}[Z(s_0)] - 2\,\mathrm{Cov}\Big(\sum_{i=1}^{N} w_i Z(s_i), Z(s_0)\Big).$$

Each term separately:

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{i=1}^{N} w_i Z(s_i)\Big) &= \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \,\mathrm{Cov}[Z(s_i), Z(s_j)],\\
\mathrm{Var}[Z(s_0)] &= C(0),\\
\mathrm{Cov}\Big(\sum_{i=1}^{N} w_i Z(s_i), Z(s_0)\Big) &= \sum_{i=1}^{N} w_i \,\mathrm{Cov}[Z(s_i), Z(s_0)].
\end{aligned}$$

So the MSE is

$$\mathrm{MSE}(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{C}\boldsymbol{w} + C(0) - 2\boldsymbol{w}^\top \boldsymbol{c},$$

where $\boldsymbol{C} = (C_{ij})$ with $C_{ij} = C(s_i - s_j)$ and $\boldsymbol{c} = (C(s_i - s_0))$.

Constrained minimization

We minimize MSE subject to $\mathbf{1}^\top \boldsymbol{w} = 1$ using the Lagrangian

$$L(\boldsymbol{w}, \lambda) = \boldsymbol{w}^\top \boldsymbol{C}\boldsymbol{w} + C(0) - 2\boldsymbol{w}^\top \boldsymbol{c} + 2\lambda(\mathbf{1}^\top \boldsymbol{w} - 1).$$

First-order conditions (FOCs)

Differentiate with respect to $\boldsymbol{w}$ and λ:

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{w}} &= 2\boldsymbol{C}\boldsymbol{w} - 2\boldsymbol{c} + 2\lambda\mathbf{1} = \mathbf{0},\\
\frac{\partial L}{\partial \lambda} &= 2(\mathbf{1}^\top \boldsymbol{w} - 1) = 0.
\end{aligned}$$

Kriging system

Thus we obtain

$$\boldsymbol{C}\boldsymbol{w} + \lambda\mathbf{1} = \boldsymbol{c}, \qquad \mathbf{1}^\top \boldsymbol{w} = 1.$$

Block matrix form:

$$\begin{bmatrix} \boldsymbol{C} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ 1 \end{bmatrix}.$$

Ordinary kriging variance

Plugging the optimal $\boldsymbol{w}$ and λ back into the MSE yields

$$\sigma^2_{\mathrm{OK}} = C(0) - \boldsymbol{w}^\top \boldsymbol{c} - \lambda.$$

3 Universal Kriging (UK)

Suppose the mean is no longer constant but a linear combination of known functions:

$$Z(s) = \mu(s) + \varepsilon(s), \qquad \mu(s) = \sum_{k=1}^{p} \beta_k f_k(s).$$

Predictor

We still predict as

$$\hat{Z}(s_0) = \sum_{i=1}^{N} w_i Z(s_i).$$

Unbiasedness constraints

We require $E[\hat{Z}(s_0)] = E[Z(s_0)]$. That is,

$$\sum_{i=1}^{N} w_i \mu(s_i) = \mu(s_0).$$

Since $\mu(s) = \sum_k \beta_k f_k(s)$, this implies

$$\sum_{i=1}^{N} w_i f_k(s_i) = f_k(s_0), \qquad k = 1, \ldots, p.$$

Matrix form: with $F \in \mathbb{R}^{N\times p}$, $F_{ik} = f_k(s_i)$ and $\boldsymbol{f}_0 = (f_1(s_0), \ldots, f_p(s_0))^\top$, the constraints are

$$F^\top \boldsymbol{w} = \boldsymbol{f}_0.$$

Mean squared error (MSE)

As before, the MSE is

$$\mathrm{MSE}(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{C} \boldsymbol{w} + C(0) - 2\boldsymbol{w}^\top \boldsymbol{c}.$$

Constrained minimization

Now we minimize MSE subject to $F^\top \boldsymbol{w} = \boldsymbol{f}_0$. Introduce multipliers $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^\top$ and write

$$L(\boldsymbol{w}, \boldsymbol{\lambda}) = \boldsymbol{w}^\top \boldsymbol{C} \boldsymbol{w} + C(0) - 2\boldsymbol{w}^\top \boldsymbol{c} + 2\boldsymbol{\lambda}^\top (F^\top \boldsymbol{w} - \boldsymbol{f}_0).$$

First-order conditions (FOCs)

Differentiate:

$$\frac{\partial L}{\partial \boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\boldsymbol{c} + 2F\boldsymbol{\lambda} = \boldsymbol{0},$$
$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = 2(F^\top \boldsymbol{w} - \boldsymbol{f}_0) = \boldsymbol{0}.$$

Kriging system

This gives the UK system:

$$\begin{bmatrix} \boldsymbol{C} & F \\ F^\top & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{f}_0 \end{bmatrix}.$$

Universal kriging variance

The minimized MSE is

$$\sigma^2_{\mathrm{UK}} = C(0) - \boldsymbol{w}^\top \boldsymbol{c} - \boldsymbol{\lambda}^\top \boldsymbol{f}_0.$$