Solution to Kriging Equations

Meredith Franklin

1 Lagrange Multipliers

Lagrange multipliers are used for finding the maxima/minima of multivariate functions that are subject to a constraint. The function $f(x_1, x_2, ..., x_n)$ and constraint $g(x_1, x_2, ..., x_n) = 0$ are continuous, have continuous first partial derivatives and $\nabla g \neq 0$. The two functions meet where their tangent lines are parallel. This is equivalent to saying that f and g meet when their gradients are parallel (since the gradient of a function is perpendicular to the function). Thus, $\nabla f = -\lambda \nabla g$, where the constant λ is called the Lagrange multiplier. The extremum is found by solving the n+1 gradient equations (extremum are when all partials are set equal to 0). For example:

$$f(x, y) = x^2 + y^2 \text{ and } g(x, y) = x^2 y - 16 = 0$$
$$\text{Minimize } L(x, y, \lambda) = x^2 + y^2 + \lambda(x^2 y - 16)$$

$$0 = \frac{\partial L}{\partial x} = 2x + \lambda 2xy \tag{1}$$
$$0 = \frac{\partial L}{\partial y} = 2y + \lambda x^2 \tag{2}$$
$$0 = \frac{\partial L}{\partial \lambda} = x^2 y - 16 \tag{3}$$

Equation (1) gives $2x(1 + \lambda y) = 0$ which requires $x = 0$ or $y = -1/\lambda$. Equation (2) gives $x^2 = -2y/\lambda$. When plugged back into the constraint equation, we find $\lambda = 2$. Thus the minima under the constraint $g = 0$ occurs when $y = 1/2$ and $x = 1/\sqrt{2}$.

2 Ordinary Kriging Equations

Given spatial data $Z(s_i)$ that follows an intrinsically stationary process, i.e. having constant unknown mean μ, known spatial covariance function $C(h)$ for spatial lags $h = s_i - s_j$, and can be written as $Z(s_i) = \mu + \epsilon(s_i)$, we typically want to predict values of the process at unobserved locations, $s_0 \in D$. Kriging is a method that enables prediction of a spatial process based on a weighted average of the observations. In the case of an intrinsically stationary process with constant unknown mean, we use the ordinary kriging (OK) method.

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \omega_i Z(s_i) \tag{4}$$

We want to find the best linear unbiased predictor (BLUP) by minimizing the variance of the interpolation error (i.e. minimize mean square prediction error), $Var(\hat{Z}(s_0) - Z(s_0)) = E[(\hat{Z}(s_0) - Z(s_0))^2]$. For the predictor to be unbiased, $E[\hat{Z}(s_0)] = E[Z(s_0)] = \mu$ is required. Given (4) this means:

$$\begin{aligned}
E[\hat{Z}(s_0)] - E[Z(s_0)] &= E[\sum_{i=1}^{N} \omega_i Z(s_i)] - E[Z(s_0)] \\
&= \sum_{i=1}^{N} \omega_i E[Z(s_i)] - E[Z(s_0)] \\
&= \sum_{i=1}^{N} \omega_i \mu - \mu \\
&= \mu(\sum_{i=1}^{N} \omega_i - 1)
\end{aligned}$$

Thus $\sum_{i=1}^{N} \omega_i = 1$ for unbiasedness to hold and we have a minimization problem with a constraint that can be solved using Lagrange multipliers. We minimize

$$\begin{aligned}
L(\omega_i, \lambda) &= E[(\hat{Z}(s_0) - Z(s_0))^2] + 2\lambda(\sum_{i=1}^{N} \omega_i - 1) \\
&= Var[\sum_{i=1}^{N} \omega_i Z(s_i)] + Var[Z(s_0)] - 2Cov[\sum_{i=1}^{N} \omega_i Z(s_i), Z(s_0)] + 2\lambda(\sum_{i=1}^{N} \omega_i - 1) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j Cov[Z(s_i), Z(s_j)] + Var[Z(s_0)] - 2\sum_{i=1}^{N} \omega_i Cov[Z(s_i), Z(s_0)] + 2\lambda(\sum_{i=1}^{N} \omega_i - 1)
\end{aligned}$$

Recall the variance of a linear combination $Var[\sum_{i=1}^{N} \omega_i Z(s_i)]$ is $\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j Cov[Z(s_i), Z(s_j)]$

Differentiate with respect to ω_i and λ and set equal to 0

$$\frac{\partial L(\omega_i, \lambda)}{\partial \omega_i} = 2\sum_{j=1}^{N} \omega_j Cov[Z(s_i), Z(s_j)] - 2Cov[Z(s_i), Z(s_0)] + 2\lambda = 0$$

Which gives:

$$\sum_{j=1}^{N} \omega_j Cov[Z(s_i), Z(s_j)] + \lambda = Cov[Z(s_i), Z(s_0)]$$

And

$$\frac{\partial L(\omega_i, \lambda)}{\partial \lambda} = 2\sum_{i=1}^{N} \omega_i - 2 = 0$$

Which gives:

$$\sum_{i=1}^{N} \omega_i = 1$$

In matrix notation, this can be written as

$$\boldsymbol{C}\boldsymbol{w} = \boldsymbol{D}$$

Where $\boldsymbol{C}$ is the covariance matrix of the observed values and the row and column for the constraint, $\boldsymbol{w}$ is the vector of weights and the Lagrange multiplier, and $\boldsymbol{D}$ is the vector of covariances at the prediction location. Solving for the weights,

$$\boldsymbol{w} = \boldsymbol{C}^{-1}\boldsymbol{D}$$

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \\ \lambda \end{pmatrix} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1N} & 1 \\ C_{21} & C_{22} & \dots & C_{2N} & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ C_{N1} & C_{N2} & \dots & C_{NN} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{N0} \\ 1 \end{bmatrix}$$

We see that $\boldsymbol{C}$ only needs to be calculated (and inverted) once but $\boldsymbol{D}$ is found for every prediction location. The inversion operation can be quite computationally intensive for large N. With the weights we can solve for expected value at the new location $\hat{Z}(s_0) = \sum\limits_{i=1}^{N} \omega_i Z(s_i)$.

The variance of the prediction is found via the MSE:

$$MSE = \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i\omega_j Cov[Z(s_i), Z(s_j)] + Var[Z(s_0)] - 2Cov[\sum_{i=1}^{N} \omega_i Z(s_i), Z(s_0)]$$

Where

$$\begin{aligned} \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i\omega_j Cov[Z(s_i), Z(s_j)] &= \sum_{i=1}^{N} \omega_i \sum_{j=1}^{N} \omega_j Cov[Z(s_i), Z(s_j)] \\ &= \sum_{i=1}^{N} \omega_i (Cov[Z(s_i), Z(s_0)] - \lambda) \end{aligned}$$

So, the MSE gives us the ordinary kriging variance,

$$\sigma^2_{OK} = \sigma^2 - \sum_{i=1}^{N} \omega_i (Cov[Z(s_i), Z(s_0)] - \lambda)$$

Which in matrix form is

$$\sigma^2_{OK} = \sigma^2 - \boldsymbol{w}^T \boldsymbol{D}$$

3 Universal Kriging Equations

When we do not have a constant unknown mean μ in our spatial process, we must expand the above approach to account for a variable mean (i.e. linear or polynomial trend, spatially varying covariates) when making kriging predictions. For example we can define a spatial regression model, $Z(s) \sim \text{MVN}(X\beta, \Sigma)$ where under instrinsic stationarity, Σ is our spatial covariance function $C(h)$ for spatial lags $h = s_i - s_j$. $X\beta$ represents the $k = 1, \ldots, p$ covariates. Another way of expressing the model is $Z(s_i) = \mu(s_i) + \epsilon(s_i)$ where

$$\mu(s_i) = \sum_{k=1}^{p} \beta_k x_k(s_i)$$

As above, we wish to find the BLUP, where (4) becomes

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \omega_i \sum_{k=1}^{p} \beta_k x_k(s_i)$$

$$\begin{aligned} E[\hat{Z}(s_0)] - E[Z(s_0)] &= E[\sum_{i=1}^{N} \omega_i \sum_{k=1}^{p} \beta_k x_k(s_i)] - \sum_{k=1}^{p} \beta_k x_k(s_0) \\ &= \sum_{i=1}^{p} \beta_k \sum_{k=1}^{N} [\omega_i x_k(s_i) - x_k(s_0)] \end{aligned}$$

Where

$$\sum_{i=1}^{N} \omega_i x_k(s_i) = x_k(s_0), k = 1, \ldots, (p-1)$$

We require $\sum_{i=1}^{N} \omega_i = 1$. So there are p constraints in our minimization problem (sum of weights is equal to 1 and the p-1 covariates). We apply the same method as above with Lagrange multipliers λ_k. Minimize:

$$L(\omega_i, \lambda) = E[(\hat{Z}(s_0) - Z(s_0))^2] + 2 \sum_{k=1}^{p} \lambda_k (\sum_{i=1}^{N} \omega_i x_k(s_i) - x_k(s_0))$$

As above, we obtain the result $\boldsymbol{w}^{*} = \boldsymbol{C}^{*-1}\boldsymbol{D}^{*}$ which expands to:

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \\ \lambda_1 \\ \vdots \\ \lambda_p \end{pmatrix} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1N} & x_{11} & \dots & x_{1p} \\ C_{21} & C_{22} & \dots & C_{2N} & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{N1} & C_{N2} & \dots & C_{NN} & x_{N1} & \dots & x_{Np} \\ x_{11} & x_{12} & \dots & x_{1N} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pN} & 0 & \dots & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{N0} \\ x_{10} \\ \vdots \\ x_{p0} \end{bmatrix}$$

And the universal kriging variance is represented by

$$\sigma_{UK}^2 = \sigma^2 - \boldsymbol{w}^{*T}\boldsymbol{D}^{*}$$