Solution to Ordinary and Universal Kriging Equations

Meredith Franklin
February 6, 2014

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 \( \lambda \) 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:

\[
\begin{align*}
\text{Minimize} \quad & L(x, y, \lambda) = x^2 + y^2 + \lambda (x^2 y - 16) \\
\text{Subject to} \quad & g(x, y) = x^2 y - 16 = 0
\end{align*}
\]

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 \( \mu \), 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.

\[
\hat{Z}(s_0) = \sum_{i=1}^{N} \omega_i Z(s_i) \quad (4)
\]

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), \( \text{Var}(\hat{Z}(s_0) - Z(s_0)) = E[(\hat{Z}(s_0) - Z(s_0))^2] \). For the predictor to be unbiased, \( E[\hat{Z}(s_0)] = E[Z(s_0)] = \mu \) is required. Given (4) this means:

\[
E[\hat{Z}(s_0)] - E[Z(s_0)] = E[\sum_{i=1}^{N} \omega_i Z(s_i)] - E[Z(s_0)]
\]
\[
= \sum_{i=1}^{N} \omega_i E[Z(s_i)] - E[Z(s_0)]
\]
\[
= \sum_{i=1}^{N} \omega_i \mu - \mu
\]
\[
= \mu \left( \sum_{i=1}^{N} \omega_i - 1 \right)
\]

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

\[
L(\omega_i, \lambda) = E[(\hat{Z}(s_0) - Z(s_0))^2] + 2\lambda \left( \sum_{i=1}^{N} \omega_i - 1 \right)
\]
\[
= \text{Var}[\sum_{i=1}^{N} \omega_i Z(s_i)] + \text{Var}[Z(s_0)] - 2\text{Cov}[\sum_{i=1}^{N} \omega_i Z(s_i), Z(s_0)] + 2\lambda \left( \sum_{i=1}^{N} \omega_i - 1 \right)
\]
\[
= \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j \text{Cov}[Z(s_i), Z(s_j)] + \text{Var}[Z(s_0)] - 2 \sum_{i=1}^{N} \omega_i \text{Cov}[Z(s_i), Z(s_0)] + 2\lambda \left( \sum_{i=1}^{N} \omega_i - 1 \right)
\]

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

Differentiate with respect to \( \omega_i \) and \( \lambda \) and set equal to 0

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

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

Which gives:
\[ \sum_{i=1}^{N} \omega_i = 1 \]

In matrix notation, this can be written as
\[ Cw = D \]

Where \( C \) is the covariance matrix of the observed values and the row and column for the constraint, \( w \) is the vector of weights and the Lagrange multiplier, and \( D \) is the vector of covariances at the prediction location. Solving for the weights,
\[ w = C^{-1}D \]

We see that \( C \) only needs to be calculated (and inverted) once but \( D \) 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
\[ \hat{Z}(s_0) = \sum_{i=1}^{N} \omega_i Z(s_i) \]

The variance of the prediction is found via the MSE:
\[ MSE = \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j \text{Cov}[Z(s_i), Z(s_j)] + \text{Var}[Z(s_0)] - 2\text{Cov}\left[ \sum_{i=1}^{N} \omega_i Z(s_i), Z(s_0) \right] \]

Where
\[ \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j \text{Cov}[Z(s_i), Z(s_j)] = \sum_{i=1}^{N} \omega_i \sum_{j=1}^{N} \omega_j \text{Cov}[Z(s_i), Z(s_j)] \]
\[ = \sum_{i=1}^{N} \omega_i (\text{Cov}[Z(s_i), Z(s_0)] - \lambda) \]
So, the MSE gives us the ordinary kriging variance,

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

Which in matrix form is

\[ \sigma_{OK}^2 = \sigma^2 - \mathbf{w}' \mathbf{D} \]

3 Universal Kriging Equations

When we do not have a constant unknown mean \( \mu \) 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, \( \Sigma \) is our spatial covariance function \( C(h) \) for spatial lags \( h = s_i - s_j \). \( X\beta \) represent 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 we (4) becomes

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

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

\[
= \sum_{i=1}^{N} \omega_i \sum_{k=1}^{p} E[\beta_k x_k(s_i)] - \sum_{k=1}^{p} \beta_k x_k(s_0)
\]

We require

\[
\sum_{i=1}^{N} \omega_i = 1
\]

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

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 \( \lambda_k \). Minimize:

\[
L(\omega, \lambda) = E[(\hat{Z}(s_0) - Z(s_0))^2] + 2 \sum_{k=1}^{p} \lambda_k (\sum_{i=1}^{N} \omega_i x_k(s_i) - 1)
\]
As above, we obtain the result \( \mathbf{w}^* = \mathbf{C}^{*-1} \mathbf{D}^* \) which expands to:

\[
\begin{pmatrix}
\omega_1 \\
\omega_2 \\
\vdots \\
\omega_N
\end{pmatrix} = 
\begin{bmatrix}
\begin{array}{cccc}
C_{11} & C_{12} & \cdots & C_{1N} \\
C_{21} & C_{22} & \cdots & C_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
C_{N1} & C_{N2} & \cdots & C_{NN}
\end{array}
\end{bmatrix}^{-1} \begin{bmatrix}
\begin{array}{c}
x_{11} \\
x_{12} \\
\vdots \\
x_{p1}
\end{array}
\end{bmatrix}
\begin{bmatrix}
\begin{array}{c}
x_{11} \\
x_{12} \\
\vdots \\
x_{p1}
\end{array}
\end{bmatrix}
\begin{bmatrix}
\begin{array}{c}
x_{21} \\
x_{22} \\
\vdots \\
x_{p2}
\end{array}
\end{bmatrix}
\begin{bmatrix}
\begin{array}{c}
x_{N1} \\
x_{N2} \\
\vdots \\
x_{pN}
\end{array}
\end{bmatrix}
\begin{bmatrix}
\begin{array}{c}
x_0 \\
x_0 \\
\vdots \\
x_0
\end{array}
\end{bmatrix}
\]

And the universal kriging variance is represented by

\[
\sigma_{UK}^2 = \sigma^2 - \mathbf{w}^*/\mathbf{D}^*
\]