Using Quasi-Newton Methods to Find Optimal Solutions to Problematic Kriging Systems

Steven Lyster

Centre for Computational Geostatistics
Department of Civil & Environmental Engineering
University of Alberta

Solving a kriging system of equations to determine optimal linear estimate weights is straightforward if the left-hand side covariance matrix is strictly positive definite. If the covariance matrix is singular or near-singular, problems are encountered when trying to directly solve the system. Quasi-Newton methods use a recursive algorithm to converge to a local minimum of a given function. In the case of a quadratic function with known first and second derivatives and a positive semidefinite Hessian matrix the quasi-Newton methods are quite efficient and will always converge to a globally optimal solution. Use of these methods could eliminate some types of errors encountered when kriging or simulating a field, particularly when assigning conditioning data to grid nodes.

Introduction

Minimizing the variance of a linear estimate is the basis for most geostatistical estimation. The standard procedure is to express the variance as a function of the estimation weights, then take the derivative with respect to the weights and set the derivative equal to zero. If the second derivative is greater than zero there is a single optimal set of weights at which the variance is a minimum; this is the case if the left hand side covariance matrix in the kriging system of equations is positive definite (Journel and Huijbregts, 1978).

Solving the kriging equations typically involves some technique such as matrix inversion or Gaussian elimination. These methods work well for most kriging problems. However, if the kriging system is ill-behaved problems arise when using conventional analytical methods for directly solving for optimal estimation weights.

Quadratic Form of Estimation Variance

The variance of a linear estimate may be expressed as a quadratic function of the weights:

$$\sigma_E^2 = \frac{1}{2} \cdot X^T \cdot Q \cdot X - X^T \cdot b + \sigma_Z^2 \qquad (1)$$

Where

$$Q = \begin{bmatrix} & & \\ & 2 \cdot E\{Z(u_i) \cdot Z(u_j)\} & \\ & & \end{bmatrix} \qquad b = \begin{bmatrix} \\ 2 \cdot E\{Z(u) \cdot Z(u_i)\} \\ \\ \end{bmatrix} \qquad X = \begin{bmatrix} \\ \lambda_j \\ \\ \end{bmatrix}$$

If the Hessian matrix of the quadratic function, Q, is positive definite then the variance function has a single strict local minimizer. In this case it is possible to find the optimal estimation weights X^* by solving the kriging system of equations $QX = b$.

In cases where Q is not positive definite, most commonly singular or near singular (ie, positive semidefinite), the kriging equations cannot be solved analytically through most methods. Inverting the matrix Q or using a technique such as Gaussian elimination is not possible for singular matrices. There are ways of fixing these problems (Reference??), such as removing one or more data that are causing the singularity, averaging data that occur at the same location, or moving one or more data. These are all ad-hoc solutions; the use of a different optimization algorithm is proposed here to find the optimal weights directly rather than modifying the problem to fit the optimization method.

Quasi-Newton Methods

The standard form of Newton's method recursively uses a quadratic approximation for the objective function to find a minimum (Chong and Zak, 2001). Because the variance is a quadratic function already with a defined Hessian matrix, the standard Newton's algorithm will solve the optimization problem in a single step from any initial point. For a quadratic function, Newton's method simplifies as follows:

$$X^{(k+1)} = X^{(k)} - F\left(X^{(k)}\right)^{-1} \cdot g\left(X^{(k)}\right)$$

Where: $F(X^{(k)})$ is the Hessian of the objective at $X^{(k)}$

$g(X^{(k)})$ is the first derivative of the objective at $X^{(k)}$

This then leads to:

$$X^{(k+1)} = X^{(k)} - Q^{-1} \cdot \left(Q \cdot X^{(k)} - b\right) = Q^{-1} \cdot b \qquad (2)$$

Equation 2 is the same as the solution to the kriging system of equations, and requires the matrix Q to be invertible. In the case of a singular matrix this is not true, and so the same problem arises as before.

Quasi-Newton methods are a family of optimization algorithms that use the same basic principle as Newton's method, but utilize an approximation for the inverse Hessian matrix and therefore do not require matrix inversion or solving systems of equations. The procedure for all of these methods is the same:

1. Select an initial point for the algorithm, $X^{(0)}$, and an initial approximation for the inverse Hessian matrix, H_0.
2. Find the direction of search $d^{(k)} = -H_k \cdot g^{(k)}$

3. Perform a line search to get the minimum objective value in the search direction; for a quadratic function, this simplifies to $\alpha_k = -\dfrac{d^{(k)T} \cdot g^{(k)}}{d^{(k)T} \cdot Q \cdot d^{(k)}}$

4. Get the next estimate for the optimal point, $X^{(k+1)} = X^{(k)} + \alpha_k \cdot d^{(k)}$

5. Compute the next approximation for the inverse Hessian by using

$$\Delta X^{(k)} = X^{(k+1)} - X^{(k)}$$

$$\Delta g^{(k)} = g^{(k+1)} - g^{(k)}$$

H_{k+1} is then found by an equation that depends on the quasi-Newton method used.

6. Set $k = k+1$ and go back to step 2. Repeat until some stopping criterion is met.

The initial guess for $X^{(0)}$ can theoretically be set to anything and the algorithm will still converge to a minimum point. As long as Q is positive semidefinite a minimum point must exist and the algorithm will converge. The simplest possible selection for $X^{(0)}$ is a zero vector, containing all zeros. Because the optimal linear estimate weights are always quite small numbers (typically less than one), a zero vector should be a close enough approximation for the algorithm to converge in a reasonable amount of time.

For the initial approximation of the inverse Hessian, H_0, the only requirements are that it be real, positive definite and symmetric. Within these requirements any matrix may be used and still result in convergence. The easiest matrix to use is an identity matrix; this satisfies all of the conditions and simplifies the math as well.

The intermediate approximations for the inverse Hessian, H_k, are found by using equations that depend on the particular quasi-Newton method being employed. The three method discussed here are the rank one equation algorithm, the DFP algorithm, and the BFGS algorithm (Chong and Zak, 2001). The equations for H_k for each method are:

Rank one:
$$H_{k+1} = H_k + \frac{\left(\Delta x^{(k)} - H_k \cdot \Delta g^{(k)}\right) \cdot \left(\Delta x^{(k)} - H_k \cdot \Delta g^{(k)}\right)^T}{\Delta g^{(k)T} \cdot \left(\Delta x^{(k)} - H_k \cdot \Delta g^{(k)}\right)} \tag{3}$$

DFP:
$$H_{k+1} = H_k + \frac{\Delta x^{(k)} \cdot \Delta x^{(k)T}}{\Delta x^{(k)T} \cdot \Delta g^{(k)}} - \frac{\left(H_k \cdot \Delta g^{(k)}\right) \cdot \left(H_k \cdot \Delta g^{(k)}\right)^T}{\Delta g^{(k)T} \cdot H_k \cdot \Delta g^{(k)}} \tag{4}$$

BFGS:
$$H_{k+1} = H_k + \left(1 + \frac{\Delta g^{(k)T} \cdot H_k \cdot \Delta g^{(k)}}{\Delta g^{(k)T} \cdot \Delta x^{(k)}}\right) \cdot \frac{\Delta x^{(k)} \cdot \Delta x^{(k)T}}{\Delta x^{(k)T} \cdot \Delta g^{(k)}}$$

$$- \frac{\left(H_k \cdot \Delta g^{(k)} \cdot \Delta x^{(k)T}\right) + \left(H_k \cdot \Delta g^{(k)} \cdot \Delta x^{(k)T}\right)^T}{\Delta g^{(k)T} \cdot \Delta x^{(k)}} \tag{5}$$

The difference between the algorithms is the level of sophistication in the update equations for the inverse Hessian. The rank one equation algorithm is the simplest, and finds a symmetric updated approximation. For quadratic objectives with a positive definite Q each iterative

approximation H_k will also be positive definite. Computational problems may be encountered in some cases as the denominator in Equation 3 may approach zero (Chong and Zak, 2001).

The DFP algorithm uses a more robust approach to finding the intermediate matrix approximations. Using this method each matrix H_k is guaranteed to be positive definite. There are no computational divide-by-zero errors when dealing with the DFP algorithm, provided the procedure is stopped once an optimal value has been reached and the delta terms in the denominators begin to approach zero. The main drawback to the DFP algorithm is that it requires accurate line searches to find the search distance for each estimate; this is not a problem as an exact line search of a quadratic function may be performed by using the equation in step 3.

The BFGS algorithm, though complex in appearance, may often be more efficient than the DFP algorithm (Chong and Zak, 2001). The intermediate H_k matrices are guaranteed to be positive definite using BFGS, and there should be no computational problems encountered. The most obvious difference between the DFP and BFGS algorithms is that BFGS does not require very accurate line searches; this is not an issue in this case.

The quasi-Newton methods to employ is up to the user; given the divide-by-zero potential of the rank one algorithm it may be undesirable. The BFGS algorithm may be too robust and not efficient enough for a problem as simple as minimizing an estimation variance.

Example

To demonstrate the use of a quasi-Newton method for solving an unstable kriging system, a simple synthetic problem was created. Figure 1 shows an arrangement of nine data in a 100m by 100m square field. A grid with blocks 10m square was set up for the estimation; the grid is shown in Figure 1. For the purpose of estimation, a linear variogram model with a range of 100m and a nugget effect of 0.1 was used. The point being estimated was arbitrarily set as (35,35).

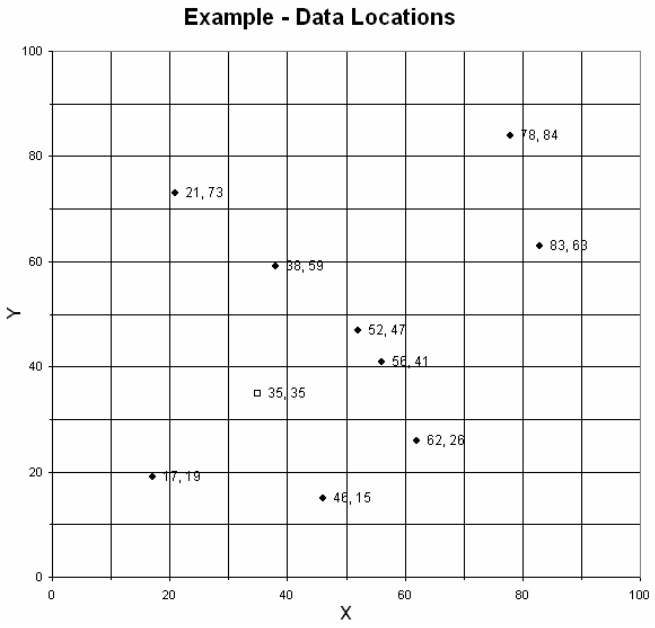

Figure 1: Arrangement of data for the example.

Two cases were considered: first, solving the kriging system of equations for the grid node at (35,35) using both matrix inversion and the DFP algorithm; then, the data were assigned to the closest grid nodes and the system was solved again. It may be easily seen from Figure 1 that two of the data were assigned to the same grid node; this created a singular matrix that could not be inverted. Because the data were only moved slightly the results for the two cases were very similar; this allowed a comparison of the exact solution from the first case to the quasi-Newton solution from the second case.

Table 1 shows the results of solving for the minimum estimation variance in both cases, with both the exact and DFP solutions for case one.

	Case 1		Case 2
Value	Exact	DFP	DFP
λ_1	0.2726	0.2726	0.2258
λ_2	0.0802	0.0802	0.0616
λ_3	0.1802	0.1802	0.2856
λ_4	0.1972	0.1972	0.2292
λ_5	0.1579	0.1579	0.0953
λ_6	0.1261	0.1261	0.0953
λ_7	0.0564	0.0564	0.0624
λ_8	-0.0116	-0.0116	-0.0055
λ_9	-0.024	-0.024	-0.012
Variance	0.2865	0.2865	0.2905

Table 1: Solutions found for minimizing the estimation variance in the example.

For case one, the DFP algorithm produced exactly the same results as solving the kriging system of equations through matrix inversion. This shows that quasi-Newton methods are a valid way of solving the system. When the data were relocated to the grid node locations and the Hessian matrix became singular, the exact solution could no longer be found. The DFP algorithm in this case returned results quite similar to the first case, with the optimal variance being very close to the same value. The variance should be close to the same for both cases as moving the data a few meters should not completely change the kriging matrices; that this is indeed the result found shows that even when the covariance matrix is singular a quasi-Newton method will return valid results.

Another notable result is that the weights given in case 2 to the collocated conditioning data are identical. This is probably the most ideal way to handle the problem of multiple solutions; with a singular covariance matrix there are many possible solutions. Each of the solutions must have the same sum of weights for the collocated data; beyond this constraint, any combination of weights is possible without affecting the optimality of the estimation variance. That the quasi-Newton solution resulted in the collocated data receiving equal weighting with no explicit control suggests the method will always return this result; this is as unbiased as is possible in a case with many solutions.

Nine iterations were performed for the DFP algorithm; with nine data, this is the maximum number that should be required to converge to a solution. Table 2 shows the progress of the algorithm for each iteration. From this information it may be seen that the algorithm actually

converged to within four decimal places at iteration five in both cases. It is clear that performing the maximum number of iterations is not necessary to find a solution.

Iteration	Variance	
	Case 1	Case 2
0	1.0000	1.0000
1	0.3325	0.3409
2	0.2892	0.2953
3	0.2871	0.2919
4	0.2866	0.2906
5	0.2865	0.2905
6	0.2865	0.2905
7	0.2865	0.2905
8	0.2865	0.2905
9	0.2865	0.2905

Table 2: Progression of the DFP algorithm in the example through nine iterations when minimizing the estimation variance.

Conclusions

Quasi-Newton methods may be used to find optimal solutions to the minimization of estimation variances without directly solving the kriging system of equations. Optimal solutions are found even when the covariance matrix is not strictly positive definite and analytical solutions are not attainable directly. The stability and efficiency of quasi-Newton methods for very large problems has not been proved; if these techniques are employed to avoid issues with singular matrices and problematic kriging systems extra attention must be paid to ensure the results are reasonable.

References

E.K.P. Chong and S.H. Zak. *An Introduction to Optimization*. Second Edition. John Wiley and Sons, Inc, New York, 2001.

A.G. Journel, Ch.J. Huijbregts. *Mining Geostatistics*. The Blackburn Press, Caldwell, New Jersey, 2003; Original printing 1978.