Robust Solution of Systems of Equations in Geostatistics

John G. Manchuk and Clayton V. Deutsch

Many geostatistical modeling algorithms involve solving systems of linear equations. One necessary property is positive definiteness; however, due to various reasons this is not always the guaranteed. In practice, systems may be indefinite or ill conditioned yielding unacceptable results. Two correction schemes are developed, where the first is automatic and uses a standard correction scheme for indefinite systems and regularized approximation for ill conditioned systems. The second is more user intensive allowing a more in-depth manipulation of a problematic matrix. Its correction scheme is based on the eigenvalue decomposition. Regression is used to correct any negative eigenvalues resulting from indefinite systems.

Introduction

Linear systems of equations encountered in geostatistics are derived in two main contexts: 1 – kriging-like systems in spatial settings, and 2 – multivariate systems. The first context involves systems that are encountered during the estimation or simulation of a geologic variable. These processes often involve tens of thousands to millions of systems. For traditional modeling techniques such as simple and ordinary kriging on regular grids, all systems are positive definite and usually yield acceptable results. An unacceptable result is one yielding a negative estimation variance or an extreme estimate beyond the range of a particular variable. Both problems occur from poorly conditioned or indefinite systems. With the advent of more complex modeling algorithms such as those with non-stationary input parameters (Boisvert, Manchuk, and Deutsch, 2007); poorly conditioned and indefinite systems are becoming more prevalent.

The second context involves multivariate matrices relating several geologic variables. They are usually encountered only once over a particular geologic domain. Bayesian updating (Deutsch and Zanon, 2007) is a good example: a correlation matrix is used to merge the information content of several variables. Correlations are calculated from sample data, which do not always lead to a well conditioned matrix. Missing samples, assay error, and redundant variables can cause indefinite systems. By construction, correlation matrices must be positive definite. This and well conditioned matrices ensure the intended results are obtained by the Bayesian updating method.

In both cases, systems must be corrected to be positive definite and well conditioned, thus eliminating the occurrence of negative variances, extreme solutions, and unacceptable estimates. Two corrections schemes are developed: one must be fast and automatic for the case were thousands of systems are encountered and another more sophisticated correction for systems that are only corrected once. Method one involves Householder reduction for eigenvalue calculations (Golub and Van Loan, 1989) and Tikhonov regularization (Boyd and Vandenberghe, 2004), while method two involves QR decomposition (Golub and Van Loan, 1989) to access the whole Eigen system for use in subsequent corrections.

Background

Basic linear systems encountered in geostatistics are calculated from a positive definite variogram function and a pair-wise distance matrix resulting in (1), where C_{ij} is the covariance between two spatial data locations $\mathbf{x}_i$ and $\mathbf{x}_j$, c_i is the covariance between an estimate and $\mathbf{x}_i$, and λ_i is the solution weight applied to location $\mathbf{x}_i$. For the first correction scheme, the system is merged (2). Correlation matrices involved in the second method are typically merged by construction, where all components in (2) are replaced with ρ_{ij} and K with P. For Gaussian variables or standardized systems, the variance in (2), σ^2, is always equal to one. In both contexts, an indefinite system can lead to negative estimation variance. The quadratic term, $\lambda^T C \lambda$,

in the equation for estimation variance (3) is not guaranteed to be positive, leading to the possibility of a negative variance.

$$\left[C_{ij}\right]\left[\lambda_i\right]=\left[c_i\right], i, j=1,...,N \tag{1}$$

$$K=\begin{bmatrix} C_{ij} & c_i \\ c_i^T & \sigma^2 \end{bmatrix} \tag{2}$$

$$\sigma_E^2=\sigma^2+\lambda^T C\lambda-2\lambda^T c \tag{3}$$

Positive definite, but ill conditioned systems pose another problem. The condition number of a matrix is a sensitivity bound for the solution. For large condition numbers, C_{ij} or ρ_{ij} is ill conditioned and the relative error in λ may be much larger than the relative error in c, leading to extreme solution weights and large $\|\lambda\|$. Given these results, the characteristics of a system of equations used to determine if a more robust solution is required are a negative estimation variance or one or more extreme weights. The later is more applicable to method 1, where an extreme weight is defined by (4) and α is a constant.

$$\left|\lambda_i\right|>\alpha\left|c_i\right| \tag{4}$$

Method 1

When thousands of systems are encountered in an estimation or simulation algorithm a fully automatic and fast correction scheme is required. It is impossible to determine a priori when ill conditioned or indefinite systems will occur: they must be detected on the fly. Detection is based on the occurrence of either a negative estimation variance or extreme weights from (3) and (4). For well conditioned systems, the only loss in performance is in checking the results, which is computationally negligible. This is reinforced by Table 1 where two systems are solved 10^5 times using a symmetric Gauss elimination solver. In most cases, the estimation variance is a required quantity.

Table 1: Time comparison for solving and checking systems

System size	Time(s), without check	Time(s) with weight check	Time(s) with weight and variance checks	% increase
10×10	0.871	0.902	1.137	3.56 / 30.54
30×30	13.59	13.59	15.13	0.00 / 11.33

Correcting systems in this setting is a two stage process: 1 – adjust indefinite systems to be positive definite and 2 – mitigate the occurrence of extreme weights. Both stages are only necessary when a negative variance and extreme weights occur simultaneously. For indefinite systems, the most straightforward correction is inflating the diagonal elements. Eigenvalue decomposition of the system in (2) is used to determine how much the diagonal should be increased to guarantee a positive definite result. The correction is given by (5) where ϖ_k, k=1,…,N+1 denotes the eigenvalues, ε is a small number, I is the identity matrix, and $\tilde{K}$ is re-standardized so the diagonals maintain a value of σ^2.

$$\tilde{K}=\frac{\left[K-(\min(\varpi_k)+\varepsilon)I\right]}{\sigma^2-(\min(\varpi_k)+\varepsilon)} \tag{5}$$

Eigenvalues are calculated using Householder's method, reducing K to a tridiagonal matrix, and QL decomposition. An additional $(N+1)^3/3+30(N+1)^2$ computations are required (Press et al, 1992). For the 10×10 system from Table 1, the initial variance was -0.519 and after applying (5) with ε equal to 10% of the minimum eigenvalue, the variance was 0.037.

When extreme weights are detected, a convex vector optimization technique called regularized approximation can be used (Boyd and Vandenberghe, 2004). An approximate solution vector $\tilde{\lambda}$ is found so that both the residuals $C\tilde{\lambda} - c$ and the magnitude of $\tilde{\lambda}$ are minimized (6).

$$\text{minimize}\left(\left\|C\tilde{\lambda} - c\right\|, \left\|\tilde{\lambda}\right\|\right) \tag{6}$$

Tikhonov regularization is applied to this problem as it has an analytical solution. The system of equations in (7) is solved rather than that in (1) for some regularization term δ. The additional computations in this case are a matrix-matrix product and a matrix-vector product. This technique is applied to the 10×10 system already corrected for the negative variance and for δ ranging from zero to one (Figure 1). This study shows how the method stabilizes the weights by controlling their norm. The estimation variance increases with δ since the solution is suboptimal for (1): i.e. the residual error increases with δ. The sum of the weights also remains nearly constant.

$$\left(C^T C + \delta I\right)\tilde{\lambda} = C^T c \tag{7}$$

Optimization to determine a δ value based on some objective function would be too time consuming for the intended purpose of this solver. Instead a single value is to be chosen and applied to all systems for a particular application. Based on the 10×10 system studied in Figure 1, δ = 0.5 would be adequate. The residual error is small and the weights and weight norm both stabilize quickly. The final estimation variance using this δ is 0.204. In practice, an acceptable δ may be chosen experimentally: slowly increase δ for a modeling application until unacceptable results, such as extreme estimates, are non-existent.

Both of the above cases require the system be solved a second time. Solving the same two systems as in Table 1 with these corrections shows an increase in processing time (Table 2), but not to an unacceptable level. For the values of *N* shown, time increased by 10 fold. In the limiting case for very large *N*, the process is roughly 3.333 times more complex. Note that these increases are only for corrected matrices. For most estimation and simulation applications, the number of poorly conditioned and indefinite systems encountered will likely be small relative to the number of spatial locations being explored.

Table 2: Time comparison for solving and correcting systems

System size	Time(s), without corrections	Time(s) with regularization correction	Time(s) with both corrections	% increase
10×10	0.875	3.750	9.250	329 / 957
30×30	13.53	72.64	133.03	437 / 883

Method 2

In multivariate applications, such as Bayesian updating, a system of equations is encountered very few times and a more involved correction scheme can be developed. As an example, the correlation matrix for a set of geologic variables is commonly required. Systems of equations are later derived from blocks of the correlation matrix. Due to the quality and extent of sample data it is possible that these matrices are indefinite, which is invalid by construction.

The dimension of the systems in this context is small, never exceeding the number of variables. Standard eigenvalue decomposition techniques can be applied without being concerned with processing time. QR decomposition will be used to access the eigenvalues and eigenvectors of a correlation matrix, P, resulting in (8), where *Q* is a matrix of eigenvectors and *D* is a diagonal matrix of eigenvalues. (8) contains the simplification that Q^T=Q^{-1} due to the symmetry of P.

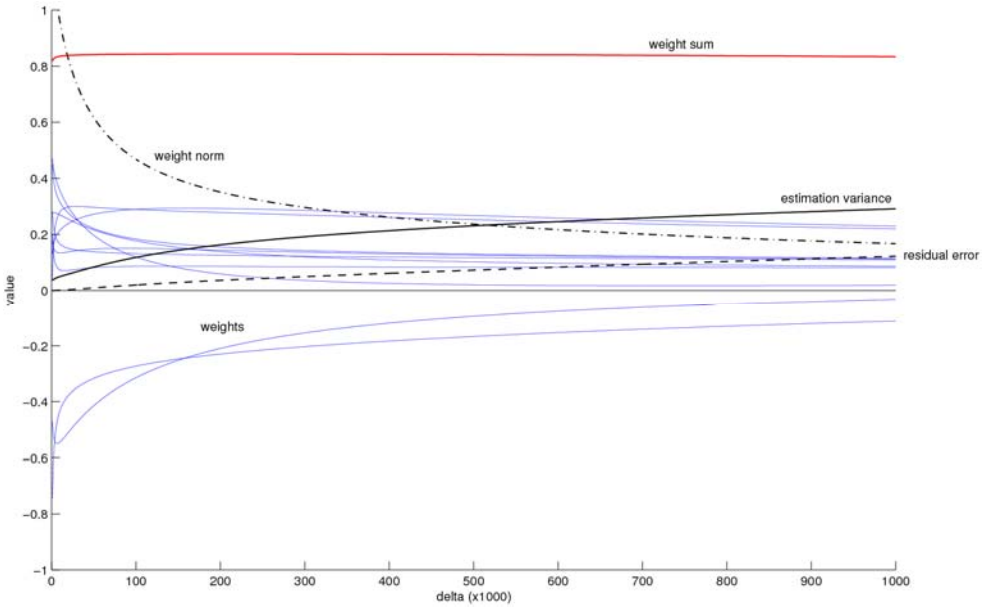

Figure 1: Tikhonov regularization applied to 10×10 system

$$\mathrm{P} = QDQ^T \tag{8}$$

The correction scheme developed here involves the four steps given below. An indefinite correlation matrix involving 37 variables will be used as an example.

1. Eigenvalue decomposition
2. Eigenvalue regression and prediction
3. Reverse calculation of P
4. Standardization of P

Decomposition has already been explained. It is assumed the eigenvalues are sorted from smallest to largest value along the diagonal of D in step 2. Letting q=diag(Q), the regression problem is to fit q as a function of eigenvalue number. Since P is theoretically positive definite, regression is transformed to log space and negative eigenvalues are omitted from the problem. Results using a third degree polynomial fit are given in Figure 2.

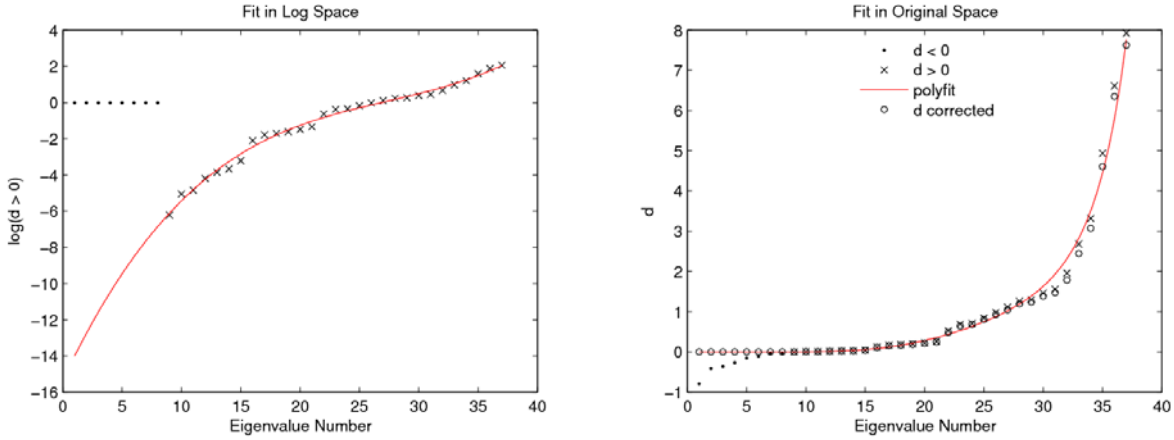

Figure 2: Initial eigenvalues, regression curve, and corrected eigenvalues

In this example, there are M = 8 negative eigenvalues. These M values are replaced by their predicted values given by the polynomial fit. Updated values are stored in a diagonal matrix $\tilde{D}$, which replaces D in (8) for calculating the corrected correlation matrix $\tilde{\mathrm{P}}$. The advantage to this method, over setting the first M eigenvalues to a small number, is rank preservation of all eigenvalues.

After calculating $\tilde{\mathrm{P}}$ from the adjusted eigenvalues, there is no guarantee the diagonal entries will equal 1. A standardization procedure is used in step 4 to accommodate this: for the k^{th} entry, $\tilde{\mathrm{P}}_{kk}$, all the elements of row k and column k are scaled by $\tilde{\mathrm{P}}_{kk}^{\frac{1}{2}}$. The standardized correlation matrix is denoted $\overline{\mathrm{P}}$ and is calculated using (9). Eigenvalues of P are shown in Figure 2, right. All eigenvalues are positive with a minimum of 7.32×10^{-7} and a maximum of 7.63. The initial and final correlation matrices and their absolute difference are shown in Figure 3.

$$\begin{aligned} \overline{\mathrm{P}}_{ik} &= \tilde{\mathrm{P}}_{ik} / \tilde{\mathrm{P}}_{kk}^{\frac{1}{2}} \quad i = 1,...,N \\ \overline{\mathrm{P}}_{kj} &= \tilde{\mathrm{P}}_{kj} / \tilde{\mathrm{P}}_{kk}^{\frac{1}{2}} \quad j = 1,...,N \end{aligned} \tag{9}$$

Very few correlation coefficients undergo a large change through the correction process. Correlations with the largest change are most influential in causing the indefinite matrix. An influence curve can be analyzed to choose a significant absolute change for further analysis (Figure 3, bottom right). For a series of change values, the number of variables above can be calculated. In this example, a value of 0.15 might be selected due to the rapid change in slope of the curve. Those variables with a change greater than 0.15 can be highlighted and further analyzed (Figure 4)

Conclusion

Two main categories of systems of linear equations encountered in geostatistics may require adjustments to ensure acceptable solutions. In certain applications, thousands of systems are encountered and a fully automatic correction scheme has been developed to handle indefinite and ill conditioned systems. The main advantage of the proposed correction is its non-iterative solution. Indefiniteness or negative variance is corrected in one analytic step through calculating the minimum eigenvalue and extreme weights are corrected by regularization with a fixed parameter. Some fine tuning of the regularization parameter may be the only necessary adjustment. Increase in execution time is the price to ensure well conditioned solutions, although the increase is not detrimental. In many cases, systems will not need any correction. Calculation of eigenvalues is the most expensive component of the solver and methods for approximating only the minimum eigenvalue (Watkins, 2007) would improve performance.

Other applications may encounter very few ill conditioned matrices and a more elaborate user intensive correction process can be implemented. In these cases, users fit the eigenvalue curve to ensure positive definiteness, then recalculate the corrected matrix. Problematic variables can be identified through the correction and further analysis may be carried out, for example taking a closer look at cross plots between the problematic variables for outliers. It is the author's opinion that in nearly all cases, few variables will be responsible for an indefinite (correlation) matrix.

References

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, 716

Deutsch CV, Zanon SD (2007) Direct prediction of reservoir performance with Bayesian updating. JCPT, Vol. 46, No. 2

Boisvert JB, Manchuk JG, Deutsch CV (2007) Calculating distance in the presence of locally varying anisotropy. In Zhao P, Agterberg F, Cheng Q (Eds.) Proc. IAMG'07 Geomethematics and GIS Analysis of Resources, Environment and Hazards, 82-85

Golub GH, and Van Loan CF (1989) Matrix computations. The John Hopkins University Press, 728

Watkins SD (2007) The matrix eigenvalue problem: GR and Krylov subspace methods. Society for Industrial and Applied Mathematics, 452

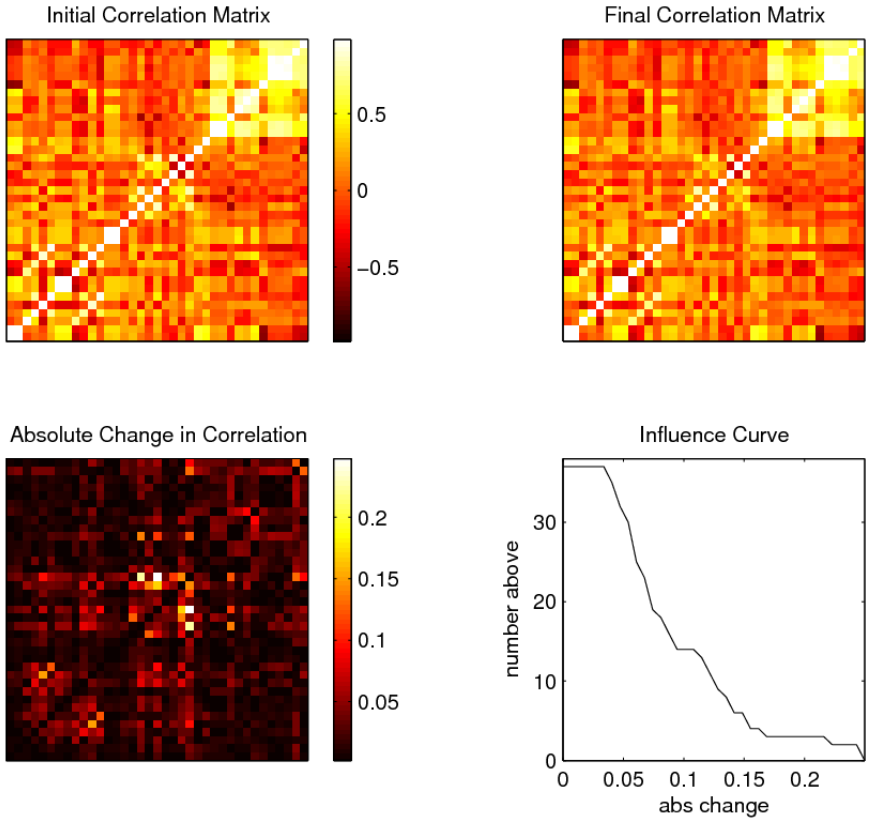

Figure 3: Initial and final correlation matrices and their absolute difference

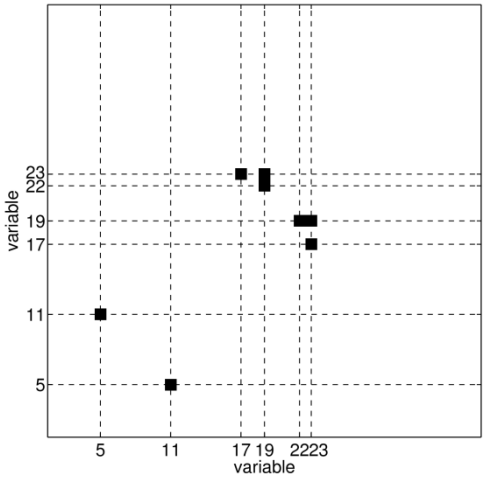

Figure 4: Most influential variables causing indefiniteness