Geostatistical modeling of large datasets using the incomplete Cholesky decomposition – An application to geological modeling with the potential field method


Developing efficient approaches enabling the determination of geological envelopes is of paramount importance in mining applications. In this context, implicit modeling approaches provide innovative and appealing alternatives to time-consuming classical geological modeling methodologies. The potential field method (PFM) introduced by Lajaunie (1996) has demonstrated its capabilities to this end; it also provides a consistent framework for characterizing, through simulations, the uncertainties related to the estimation of such geological envelopes.

The principle of the potential method is to derive the geometry of the domain under study from a 3D interpolation of a scalar field, known as the potential field. As the underlying co-kriging system is based on the domain intercepts and not on the drillhole information itself, a proper usage of this method imposes the use of additional control points to ensure the respect of information between intercepts or in extrapolation. As the PFM uses a unique neighborhood and the control points can sum up to tens of thousands points, solving the resulting co-kriging poses difficult implementation issues due to the inversion of large matrices.

To overcome this issue, the paper presents an innovative methodology that consists in a hybrid use of both the incomplete Cholesky decomposition (ICD) and principal component analysis (PCA). These combined methods allow building an interpolation which is not necessarily honoring all the input data, due to their large number, but still use their information.

The ICD provides a low rank approximation of any covariance matrix at a reduced computational cost (Fine and Scheinberg, 2002). Its use has been demonstrated for kriging a large data set (Romary, 2013). The idea is to work with a relevant low rank linear combination of the original data set. As a consequence, the exact interpolation property is lost at initial data points. However, this first step can be enhanced to force the interpolation to honor exactly a set of additional samples; the latter can be chosen either manually or using an incremental procedure that selects them based on their estimation error during the first step.

The PCA aims at reflecting a significant part of the data information, measured in terms of variance, to reduce significantly the computational time needed to perform the interpolation. This method is more costly than the ICD because it requires the computation of the eigenvalues and eigenvectors of a “reduced” matrix. However this approach has been considered to be able to reject some eigenvectors depending on their associated eigenvalue, e.g. the variance that they explain.

The paper presents the methodology and demonstrates its efficiency on a real dataset, both in terms of CPU performance and quality of the obtained geological domains.
