PRINCIPAL COMPONENT ANALYSIS FOR FIELD SEPARATION

Ton Tich Ai

Department of Physics, College of Science, VNU

Abstract. The article contains different techniques of geophysical data processing by using software Mathematica. Field separation is one of the most important problems in geophysical data interpretation. For potential fields, when there are observational data for the both the profile survey and area survey. The field separation becomes a process of estimation of low frequency component, i.e., the regional anomaly, on the one hand, and high-frequency field component, i.e. the residual or local anomaly on the other hand.

Principal components analysis for field separation provides immediate insight into the structure of field data and is applied for modeling the anomaly field that contains different bodies.

The calculation process is realized by using the computer algebraic system (mathematica). The result of research is used in geophysical field separation for geophysical interpretation.

1. Introduction

Traditional interpretations of the geophysical data have cocentrated on one or two preselected variables or functions of the variables. However, the multivariate structure of the data suggests that statistical techniques of multivariate analysis are appropriate.

Principal components analysis as a multivariate exploratory techniques provides a useful starting point for further investigations. It may also provide insight into the geological processes underlying the data. It is a method for decomposing the total variation of multivariate observations into linearly independent components of decreasing inpotance.

In this article, the principal component analysis is used for field separation and intergrated data processing.

1. Application of the principal component analysic: Field separation for areal survey data

Consider the application of the principal component analysis for field separation when there are areal survey data. Let the set of random values $X_1,...,X_N$ be presented by two-dimensional data file (areal survey data) for the same physical field, in the form of matrix of N rows and n columns. The algorithm of field separation include the following operation:

Calculation of the mean for each profile:

$$\overline{X_i} = \frac{1}{n}\sum_{k=1}^{n} x_{ki}, \qquad i = 1,2,...,N$$

where x_{ki} are the observed field data for the kth point of the ith profile.

Calculation of covariance for each pair of profiles

$$b_{ij} = \frac{1}{n}\sum_{k=1}^{n}(x_{ki} - \overline{x_i})(x_{kj} - \overline{x_j}), \quad i, j = 1, 2, ..., N$$

Construction of covariance matrix B:

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1N} \\ b_{12} & b_{22} & \cdots & b_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ b_{N1} & b_{N2} & \cdots & b_{NN} \end{pmatrix},$$

where *bii* is the observed data variance for the ith profile and $b_{ij} = b_{ji}$

Calculation of the maximum eigenvalue λ_{max} *by solving the matrix equation.*

$$\begin{pmatrix} b_{11} - \lambda & b_{12} & \cdots & b_{1N} \\ b_{21} & b_{22} - \lambda & & b_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ b_{N1} & b_{N2} & & b_{NN} - \lambda \end{pmatrix} = 0.$$

The eigenvalues of this equation $\lambda_1, ..., \lambda_N$ are the roots of the equation, with the determinant of $(B - \lambda I)$ being equal to zero. Father, it is necessary to select the maximum eigenvalue among the obtained roots.

Obtaining the maximum eigenvector of matrix B, which correspond to the λ_{max}, with the aid of the set of equations:

$$\begin{aligned} (b_{11} - \lambda_{max})a_{11} + b_{12}a_{12} + \ldots \quad\quad + b_{1N}a_{1N} &= 0 \\ b_{12}a_{11} + (b_{22} - \lambda_{max})a_{12} + \ldots + b_{2N}a_{1N} &= 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_{1N}a_{11} + b_{2N}a_{12} + \ldots\ldots \quad + (b_{NN} - \lambda_{max})a_{1N} &= 0 \end{aligned}$$

The eigenvector $\overrightarrow{a_1}(a_{11}, a_{12}, ..., a_{1N})$ is determined in terms of normalization $\sum a_{1i}^2 = 1$, i =1, 2, ,...,N. The physical sense of such normalization implies the expression of the transformed data at the same scale as the primary field data.

Finding of the first principal component $Y_1 = \vec{a}X$ or

$$Y_{1K} = (a_{11}, a_{12}, ..., a_{1N})\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{pmatrix} = (Y_{11}, Y_{12}, ..., Y_{1N}).$$

We can regard the values Y_{1K} (k = 1,2, ...,n) as the weight coefficients for each point of the field data. In this connection, the values $a_{1i}(i = 1, 2, ..., N)$ determine the weight coefficient for each profile.

Estimation of the field component, characterized by maximum variance, using the matrix expression:

$$x_{ki}^{reg} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \cdots \\ Y_{1n} \end{pmatrix} (a_{11}, a_{12}, ..., a_{1N}) + \overline{x_i} = \begin{pmatrix} y_{11}a_{11} + \overline{x_1} & y_{11}a_{12} + \overline{x_2} & \cdots & y_{11}a_{1N} + \overline{x_N} \\ y_{12}a_{11} + \overline{x_1} & y_{12}a_{12} + \overline{x_2} & \cdots & y_{12}a_{1N} + \overline{x_N} \\ \cdots & \cdots & \cdots & \cdots \\ y_{1n}a_{11} + \overline{x_1} & y_{1n}a_{12} + \overline{x_2} & \cdots & y_{1n}a_{1N} + \overline{x_N} \end{pmatrix}$$

The field component having the maximum variance, ensures the estimation of the regional anomaly when $\lambda_{max} = (70-90\%)\sum \lambda_i$. Since x_{ki}^{reg} is the estimation of the regional anomaly, then the difference $x_{ki}^{loc} = x_{ki} - x_{ki}^{reg}$ will be the estimation of the local one.

On basic of the presented about algorirhm, the program for calculating regional and local anomalies of potential field is made by author in language "Mathematica":

```
<< Statistics`DescriptiveStatistics`
n = Dimensions[data0];
n1 = n[[1]];
n2 = n[[2]];
b = IdentityMatrix[n2];
Do[x[i] = Mean[data0[[i]]], {i, n1}]
Do[Do[b[[i, j]] = Sum[(data0[[k, i]] -
x[i])(data0[[k, j]] - x[j]), {k, n1}]/n1, {i, 1, n1}], {j, 1, n1}]
d = Eigenvectors[b];
{d[[1]].data0};
data1 = Transpose[%].{d[[1]]};
Do[Do[data1[[i, j]] = data1[[i, j]] + x[i], {j, n1}], {i, n2}]
dt0 = ListContourPlot[data0, ContourShading -> False, Contours -> 20,
        FrameLabel -> {"x.100 m", "y.100 m"},
   ContourStyle -> RGBColor[0, 0, 1]];
dt1 = ListContourPlot[
   data1, ContourShading -> False, Contours -> 20, FrameLabel -> {"x.100 m",
"y.100 m"}, ContourStyle -> RGBColor[0, 0, 1]];
dt2 = ListContourPlot[data0 - data1, ContourShading -> False, Contours ->
40, FrameLabel -> {"x.100 m", "y.100 m"}, ContourStyle -> RGBColor[0, 0, 1]];
```

Modeling different field separations. To demonstrate the field separation ability of the method, in this article, the model of three spheres of different prameters is selected.. The results of calculation are presented in figures 1, 2, 3.

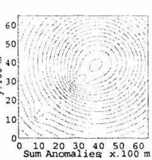

Fig. 1 Total Anomalies

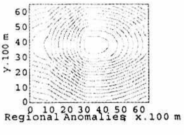

Fig.2 Regional Anomalies

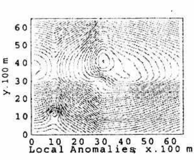

Fig.3 Local Anomalies

2. Conclusions

By using principal components analysis we may emphasize different components from total anomalies in dependence of our interpretation goal. The method was simplified to enable easier and thus possibly geological interpretation of geophysical data in different conditions.

References

1. Ton Tich Ai, *Mathematica for engineer,* National University Publisher, Hanoi 2003.
2. Ton Tich Ai, *Applied Geophysics* ,University Ministry Publisher, Hanoi 1988.
3. Stephen Wolfram, *Mathematica*. Addison-Wesley Publishing Company, Inc. 1988.
4. Tafeev.G.P, Sokolov K.P., *Geological interpretation of magnetic anomalies*. Nedra. Leningrad.1981.
5. Nikitin A.A., *Statistical Processing of Geophysical Data*. Electromagnetic. Research Center. Moscow 1993.
6. J.Trochimczyk and F.Chayes., *Some Properties of Principal Component Scores*. Mathematiacal geology,Vol.10,N0.1,1978