AN ALGORITHM OF SURFACE ESTIMATION USING CUBIC B-SPLINE FUNCTION FOR GEOLOGIC MODELING

Susumu NONOGAKI[1], **Kiyoji SHIONO**[1], **Shinji MASUMOTO**[1] **and Masafumi EKAWA**[1]

[1]Department of Geosciences, Graduate School of Science, Osaka City University

3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan

Email: nonogaki@sci.osaka-cu.ac.jp

ABSTRACT

In this paper, we propose a new gridding algorithm applying cubic B-spline function to surface approximation for the purpose of enhancement of the accuracy of geoscientific analysis, especially geologic modeling. This algorithm, coded in Fortran program, is designed to determine the smoothest surface that satisfies a given set of equality data, inequality data and strike-dip data based on the principle of optimization. Examples of surface estimation using a several types of data prove that the algorithm makes it possible to create higher resolution DEM (Digital Elevation Model) than previous algorithms.

1 INTRODUCTION

Gridding is one of the most effective ways to understand irregularly distributed data. There are many algorithms for gridding of geological surface based on observational data (e.g. Davis, 1986; Jones *et al*., 1986). Most of these algorithms are developed for the purpose of determination of reasonable surface $z = f(x, y)$ that satisfies the values z_k at locations (x_k, y_k) (k=1, …, N). In the other hand, in addition to the values z_k such as elevation data, strike-dip data are also important for geological analysis to know the attitude of surfaces. However, there are few algorithms that allow to use strike-dip data in surface estimation.

Horizon2000 (Shiono *et al*., 2001) is one of the programs for optimal determination of geological surface, using the elevation data and the strike-dip data obtained by field survey. In Horizon2000, the surface is discretely approximated by a group of value $f_{ij} = f(x_i, y_j)$ at grid points (x_i, y_j) (i =1, …, Nx ; j =1, …, Ny). An optimal geological surface is determined by solving the simultaneous equation which has $Nx \times Ny$ unknown values based on the exterior penalty function method (Zangwill, 1967). In this algorithm, solution of the problem are the values f_{ij} themselves because of discrete surface approximation, therefore, when the number of grids have to be set very large such as in generating high resolution DEM, the number of unknown values becomes too huge to get solution because a lot of memory are needed to computer.

In order to solve such problem and create higher resolution DEM than in the past, we developed a new algorithm for determination of optimal surface based on Inoue (1985) by applying cubic B-spline function to surface approximation without changing the fundamental algorithm used in Horizon2000.

2 DATA USED FOR SURFACE ESTIMATION

Suppose that an optimal geological surface $z = f(x, y)$ is determined in Cartesian coordinate space with the x-axis pointing towards the east, the y-axis pointing towards the north and z-axis pointing towards the vertical. Although there are many types of data available in field survey, in this case, three main types of data are used for determining an optimal surface as the information at locations (x_p, y_p, z_p).

1) Equality data (Equality constraints): If the surface is exposed at a location (x_p, y_p, z_p), the location provides an equality constrains for the surface;

$$f(x_p, y_p) = z_p \tag{1}$$

2) Inequality data (Inequality constraints): If the surface is located lower than observational data point and only upper geological unit can be seen there, the location provides an inequality constraints for the surface;

$$f(x_p, y_p) < z_p \tag{2a}$$

and if the surface is located upper than observational data point and only lower geological unit can be seen there, the location provides an inequality constraints for the surface;

$$f(x_p, y_p) > z_p \tag{2b}$$

In addition, in order to distinguish each data (constraints) above, optional parameter l_p which is defined by; $l_p = 0$ for the constraint (1), $l_p < 0$ for the constraint (2a) and $l_p > 0$ for the constraint (2b) is added to the information of locations (x_p, y_p, z_p).

3) Strike-dip data: If strike-dip data are measured in the location (x_q, y_q, z_q), the location provides a constraints for the partial derivative of the surface;

$$f_x(x_q, y_q) = -\sin\phi_q \tan\theta_q \tag{3a}$$

$$f_y(x_q, y_q) = -\cos\phi_q \tan\theta_q \tag{3b}$$

where $f_x(x_q, y_q)$ and $f_y(x_q, y_q)$ are respectively the partial derivative with respect to x and y. ϕ_q is the trend of the maximum slope which is represented as degree measured clockwise from north and θ_q is the dip angle.

3 ALGORITHM

3.1 Representation of surface using cubic B-spline

The N th order B-spline function is the localized piecewise polynomial function of class C^{N-1}, constituted from $N+1$ no-zero sections (de Boor, 1978). In the case of the third order B-spline what is called cubic B-spline, therefore, it has only four non-zero sections defined by five successive knots. Using the B-spline function, the surface $f(x, y)$ is represented with a tensor product form of B-spline functions defined towards x-axis and y-axis. When it is supposed that a domain Ω has a rectangular domain $\Omega_x \times \Omega_y$ in the x-y

plane, and that Ω_x and Ω_y are divided into M_x and M_y sections respectively, that is, Ω consists of $M_x \times M_y$ sections, the surface $f(x, y)$ in the domain Ω can be given by

$$f(x, y) = \Sigma\,\Sigma\, c_{ij}\, N_i(x)\, N_j(y) \qquad (i = 1, \ldots, M_x + 3; j = 1, \ldots, M_y + 3) \tag{4}$$

where $N_i(x)$ and $N_j(y)$ are cubic B-splines, and c_{ij} are coefficients multiplied each tensor product which may be determined in calculation of surface estimation. The surface approximated by cubic B-spline function can represent more complicated shape than by a group of value at grid points discretely.

3.2 Principle of surface estimation

There may be many solutions which satisfy the equality data, inequality data and strike-dip data. In order to determine which surface is an optimal geological surface, assuming that the geological surface must be the smoothest one among the feasible solutions because, we consider the surface determination as an optimization problem:

Find an optimal solution $f(x, y)$ that minimizes

$$J(f) = m_1 J_1(f) + m_2 J_2(f)$$

$$= m_1 \iint_\Omega \{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2\}\,\mathrm{dxdy} / \mathrm{S} + m_2 \iint_\Omega \{(\frac{\partial^2 f}{\partial x^2})^2 + 2(\frac{\partial^2 f}{\partial x \partial y})^2 + (\frac{\partial^2 f}{\partial y^2})^2\}\,\mathrm{dxdy} \tag{5}$$

subject to

$$f(x_p, y_p) - z_p = 0 \qquad ; \text{if } l_p = 0$$

$$f(x_p, y_p) - z_p < 0 \qquad ; \text{if } l_p < 0 \qquad (p = 1, \ldots, N_p)$$

$$f(x_p, y_p) - z_p > 0 \qquad ; \text{if } l_p > 0$$

and

$$f_x(x_q, y_q) + \sin\phi_q \tan\theta_q = 0$$

$$f_y(x_q, y_q) + \cos\phi_q \tan\theta_q = 0 \qquad (q = 1, \ldots, N_q).$$

To solve the problem, we introduce an augmented objective function based on the exterior penalty function method;

$$Q(f; \alpha) = J(f) + \alpha R(f) \tag{6}$$

where $J(f)$ is the functional which evaluates the smoothness of surface, $R(f)$ is the one which evaluates the goodness of fit and α, called penalty, gives the balance between smoothness and goodness of fit. Moreover, $R(f)$ is defined in a form of residual mean of squares as follow;

$$R(f) = R_h(f) + \gamma R_d(f)$$

$$= \Sigma\, \varepsilon_p^{\,2} / N_h$$

$$+ \gamma \Sigma\, [\{f_x(x_q, y_q) + \sin\phi_q \tan\theta_q\}^2 + \{f_y(x_q, y_q) + \cos\phi_q \tan\theta_q\}^2] / N_d \qquad (7)$$

where $R_h(f)$ and $R_d(f)$ respectively evaluates the goodness of fit about equality-inequality data and strike-dip data, and γ gives the balance between them. N_h and N_d are a number of each data. In particular, N_h means a number of data that do not satisfy equality-inequality constraints (1), (2a) and (2b) in calculation. ε_p in $R_h(f)$ is the residual between the surface $f(x_p, y_p)$ and the value z_p at location (x_p, y_p, z_p) given by

$$\varepsilon_p = \begin{cases} f(x_p, y_p) - z_p = 0 & ;\ \text{if } l_p = 0 \\ \max\{f(x_p, y_p) - z_p, 0\} & ;\ \text{if } l_p < 0 \\ \min\{f(x_p, y_p) - z_p, 0\} & ;\ \text{if } l_p > 0. \end{cases} \qquad (8)$$

Thus, applying the surface represented by equation (4) to the augmented objective function (6), a set of coefficients, which give an optimal surface, can be determined by solving the simultaneous equation;

$$\mathrm{A}\, \boldsymbol{c} = \boldsymbol{b} \qquad (9)$$

derived from $\partial Q / \partial c_{ij} = 0$ ($i = 1, \ldots, M_x + 3$; $j = 1, \ldots, M_y + 3$). The optimal surface is finally represented in a form of grid data as well as Horizon2000 by inserting c_{ij} to the equation (4). Separating the number of unknown coefficients in calculation from that of grids enables to create the surface which has any grid size.

Figure 1 shows an example of the surface estimated by the program coded in Fortran language based on the algorithm stated above using all types of data. In Figure, black filled circle means the equality data. Minus and plus mark mean inequality data which give constraints (2a) and (2b) respectively. We can find that the surface satisfies all data.

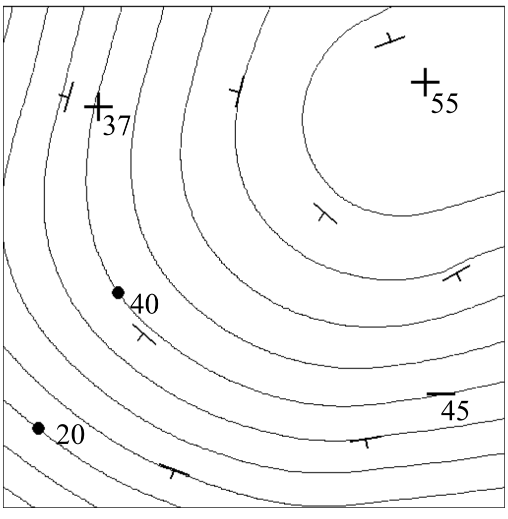

Figure 1. Example map using all types of geologic data.

4 EFFECTIVENESS OF THE ALGORITHM AND APPLICATION

To see the effectiveness of the algorithm, we compared the surfaces estimated by Horizon2000 and new program, using same 40000 equality data. In calculation with both programs, the estimation region is divided into 70 × 70 domains so that the number of unknown values becomes same: each domain includes about 8 constraints. The estimated surfaces were output as grid data which have grid number 71 × 71 in Horizon2000 and 401 × 401 in new program respectively. Difference in the number of grids is for the reason that we have to compare the surfaces directly derived from original observational data. Figure 2 shows the comparison of two surfaces. It clearly shows the difference between two surfaces, especially in the location where the shape of surface changes drastically. We may get more smooth and similar result with Horizon2000 if increasing the number of grid of surface using any interpolation methods such as a bi-cubic B-spline interpolation after first estimation, however, it is considered that such surface do not reflect the information of original data. This indicates that the algorithm enables to create higher resolution DEM than previous algorithm without increasing the quantity of computer memory used in calculation.

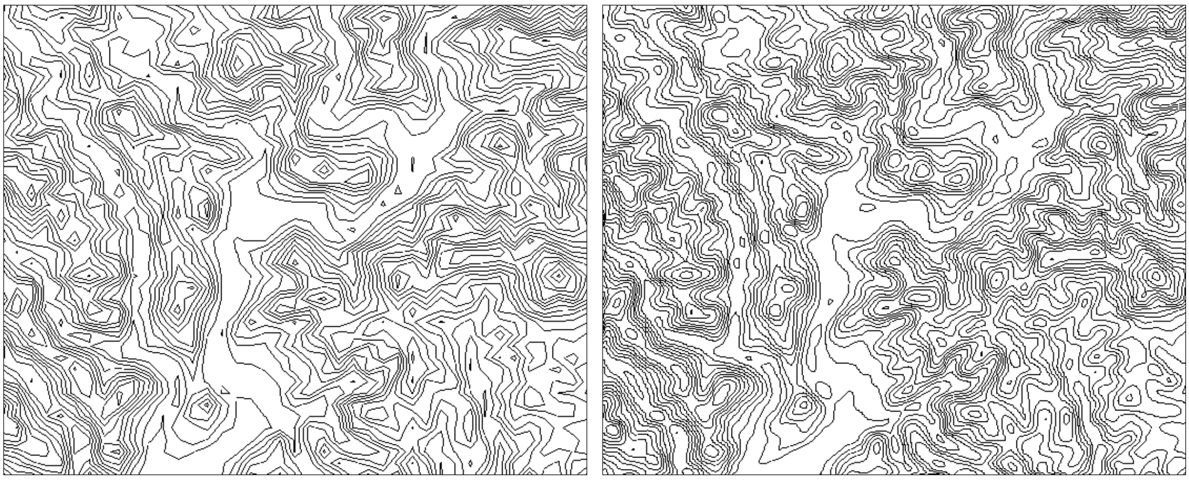

Figure 2. Comparison of surfaces. Left: Horizon2000. Right: New program.

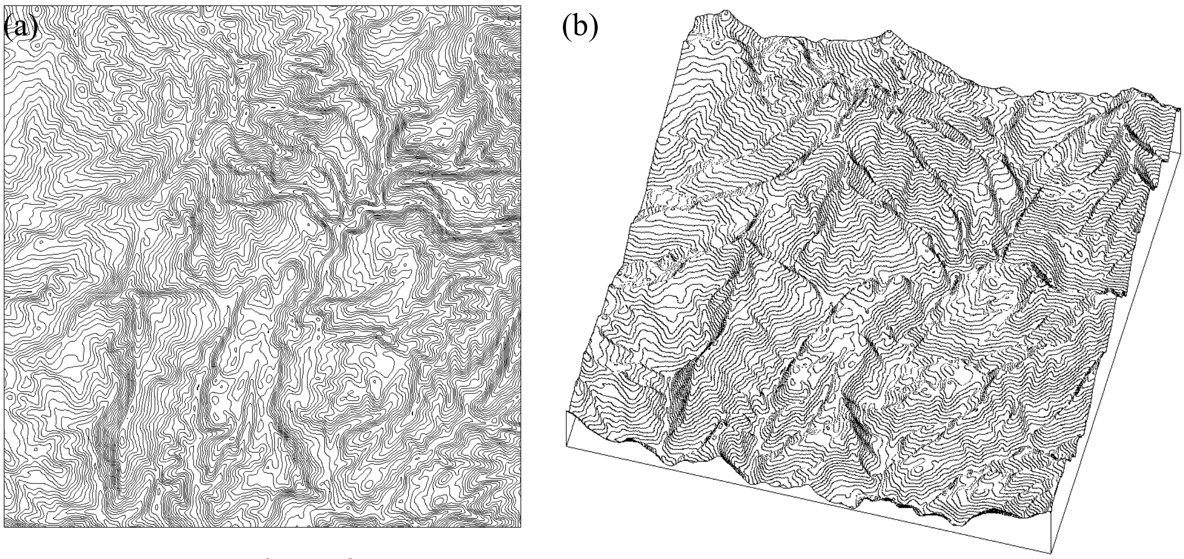

Figure 3. DEM generated from a topographic map.
(a) 2D visualization, (b) 3D visualization.

Furthermore, we tried to apply the algorithm to STRIPE method (Noumi, 2003), which is one of the most efficient DEM generation methods. This method generates DEM from a topographic map based on the idea that an elevation $f\,(x_p, y_p)$ at a point (x_p, y_p) in the space between two successive contour lines h_1 and h_2 must be

$$h_1 < f\,(x_p, y_p) < h_2. \quad (10)$$

Figure 3 shows an example of DEM generation using STRIPE method.

5 CONCLUSION

In this research, we developed an algorithm of surface estimation for geologic modeling applying the cubic B-spline function to approximate the surface $f\,(x, y)$ in calculation. The algorithm assumes that an optimal geological surface is the smoothest among surfaces that satisfy the observational data, and determine it by finding a set of unknown coefficients, which minimizes augmented objective function Q, based on the exterior penalty method. The examples of surface estimated using this algorithm indicate the effectiveness of the algorithm, that is, this algorithm allows to create higher resolution DEM than previous algorithms.

However, the relationship between the parameters used in estimation and the accuracy of the result have not explored in detail yet. In particular, little is known about the effectiveness of parameter m_1, m_2 in functional $J\,(\,f\,)$ although they may be very important parameters for geology. In order to use this algorithm more effectively, further tests using a variety of data will be needed.

REFERENCES

de Boor, C., 1978. *A Practical Guide to Splines*. Springer-Verlag, New York, 392p.

Davis, J. C., 1986. *Statistics and Data Analysis in Geology* (2nd Edition). John Wiley & Sons, New York, 646p.

Inoue, H., 1985. A smooth fitting based on a generalized inverse theory. *Geological Data Processing* (*Joho Chishitu*), 10, 105-134.

Jones, T. A., Hamilton, D. E., and Johnson, C. R., 1986. *Contouring Geologic Surfaces with the Computer*. Van Nostrand Reinhold Company, 314p.

Noumi, Y. 2003. Generation of DEM Using Inter-Contour Height Information on Topographic Map. *Journal of Geosciences Osaka City University*, 46, 217-230.

Shiono, K., Noumi, Y., Masumoto, S., and Sakamoto, M., 2001. Horizon2000: Revised Fortran program for optimal determination of geologic surfaces based on field observation including equality-inequality constrains and slope information. *Geoinformatics*, 12, 229-249.

Zangwill, W. I., 1967. Non-linear programming via penalty function. *Management Science*, 13, 344-358.