

BIHARMONIC SPLINE INTERPOLATION OF GEOS-3 AND SEASAT ALTIMETER DATA

David T. Sandwell

Center for Space Research, University of Texas at Austin

Abstract. Green functions of the biharmonic operator, in one and two dimensions, are used for minimum curvature interpolation of irregularly spaced data points. The interpolating curve (or surface) is a linear combination of Green functions centered at each data point. The amplitudes of the Green functions are found by solving a linear system of equations. In one (or two) dimensions this technique is equivalent to cubic spline (or bicubic spline) interpolation while in three dimension it corresponds to multiquadric interpolation. Although this new technique is relatively slow, it is more flexible than the spline method since both slopes and values can be used to find a surface. Moreover, noisy data can be fit in a least squares sense by reducing the number of model parameters. These properties are well suited for interpolating irregularly spaced satellite altimeter profiles. The long wavelength radial orbit error is suppressed by differentiating each profile. The shorter wavelength noise is reduced by the least squares fit to nearby profiles. Using this technique with 0.5 million GEOS-3 and SEASAT data points, it was found that the marine geoid of the Caribbean area is highly correlated with the sea floor topography. This suggests that similar applications, in more remote, areas may reveal new features of the sea floor.

Introduction

Cubic splines are commonly used to find the smoothest curve [Ahlberg et al., 1967] or surface [DeBoor, 1962; Bhattacharyya, 1969] that passes through a set of irregularly spaced data points. This technique corresponds physically to forcing an elastic beam (or elastic sheet) to match the data points. The interpolating curve (or surface) satisfies the biharmonic equation and therefore has minimum curvature [Briggs, 1974].

Here I present a simpler algorithm for finding the minimum curvature surface that passes through a set of nonuniformly spaced data points. The algorithm is based on the Green function of the biharmonic operator. The interpolating curve (or surface) is a linear combination of Green functions centered at each data point. Their amplitudes are adjusted so that the interpolating surface passes through the points.

Although this new method is relatively inefficient and can be unstable, it is very flexible. For example, the number of Green functions can be made less than the number of data points so that the interpolating curve does not exactly match inaccurate data points. This reduction also stabilizes the calculations. The major advantage of this technique, however, is that slope measurements can be used as data. This feature is important for some remote sensing applications where slopes are measured more accurately than heights. An example of this technique using satellite altimeter data is given below.

Another advantage of this technique is that it is easily applied to interpolation problems in three or more dimensions. In three dimensions, it corresponds to multiquadric interpolation [Hardy, 1971; Hardy and Nelson, 1986]. This method has been used quite successfully by a number of

researchers although it was not clear why it worked so well. The method is not useful in 4 or more dimensions because the Green functions are unbounded at the origin.

Biharmonic Spline Interpolation In One Dimension

The problem is to find a biharmonic function that passes through N data points. Draftsmen in the 19th century solved the problem by attaching weights to an elastic beam or spline and positioning the weights so that the spline passed through the data points. The forces imposed on the spline by each weight kept it bent (see figure 1). For small displacements, the spline has zero fourth derivative except at the weights. The point force Green function for the spline satisfies the biharmonic equation

$$\frac{d^4 \phi}{dx^4} = 6\delta(x) \tag{1}$$

The particular solution to (1) is

$$\phi(x) = |x|^3 \tag{2}$$

When this Green function is used to interpolate N data points, w_i , located at x_i the problem is

$$\frac{d^4 w}{dx^4} = \sum_{j=1}^N 6 \alpha_j \delta(x - x_j) \tag{3}$$

$$w(x_i) = w_i \tag{4}$$

The particular solution to (3) and (4) is a linear combination of point force Green functions centered at each data point. The homogeneous solution is not used.

$$w(x) = \sum_{j=1}^N \alpha_j |x - x_j|^3 \tag{5}$$

The strength of each point force, α_j , is found by solving the linear system

$$w_i = \sum_{j=1}^N \alpha_j |x_i - x_j|^3 \tag{6}$$

If slopes, s_i , are used rather than values, then the α_j 's are determined by solving the following linear system:

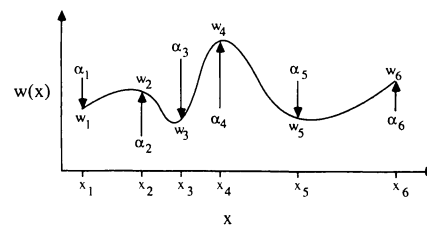


Fig. 1. The biharmonic function $w(x)$ that passes through the data points w_i located at x_i is found by applying point forces α_j to a thin elastic beam or spline.

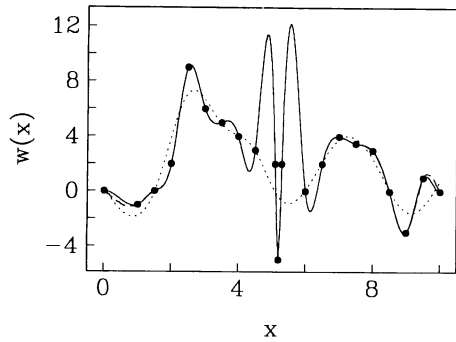


Fig. 2. Cubic spline interpolation (solid curve) of 21 data points (circles) where end slopes are zero. Biharmonic spline interpolation (dashed curve) using the one-dimensional Green function where end slopes are unconstrained. Biharmonic spline interpolation (dotted curve) where every other point force was omitted.

$$s_i = 3 \sum_{j=1}^N \alpha_j |x_i - x_j| (x_i - x_j) \quad (7)$$

Once the α_j 's are determined, the biharmonic function $w(x)$ can be evaluated at any point using equation (5).

A comparison of this technique with the cubic spline method is shown in Figure 2. The solid curve is the cubic spline interpolation of 21 points (circles) while the dashed curve was generated by this new method. The disagreement between the two curves (near the end points) occurs because, in the spline method, the end slopes are constrained to be zero while, in this new method, the end slopes are undefined. When the end slopes are also constrained to be zero for the new method (i.e. by adding two zero slope data to the ends), the two curves agree. Both methods display the dramatic overshoot that is characteristic of cubic spline interpolation.

One of the advantages of this new method is that there can be fewer model parameters (i.e. α_j 's) than data points. In this case, the linear system of equations is overdetermined and can be solved as a least squares problem. The number of parameters can be adjusted so that the standard deviation of the fit is about equal to the standard deviation of the data. An example, where only every other model parameter was used to fit 21 data points, is shown in Figure 2 (dotted curve). This feature is important when large amounts of uncertain data are used or when inaccurate data are combined with accurate data. Note that when fewer α_j 's are used the overshoot problem is less severe.

Biharmonic Spline Interpolation In 2 or More Dimensions

The derivation of the technique in 2 or more dimensions is similar to the 1-dimensional derivation. For N data points in m dimensions the problem is

$$\nabla^4 w(\mathbf{x}) = \sum_{j=1}^N \alpha_j \delta(\mathbf{x} - \mathbf{x}_j) \quad (8)$$

$$w(\mathbf{x}_i) = w_i \quad (9)$$

where ∇^4 is the biharmonic operator and \mathbf{x} is a position in the m -dimensional space. The general solution is

$$w(\mathbf{x}) = \sum_{j=1}^N \alpha_j \phi_m(\mathbf{x} - \mathbf{x}_j) \quad (10)$$

The α_j 's are found by solving the linear system

$$w_i = \sum_{j=1}^N \alpha_j \phi_m(\mathbf{x}_i - \mathbf{x}_j) \quad (11)$$

The biharmonic Green functions, ϕ_m , for each dimension, are given in Table 1. In three dimensions the Green function has discontinuous slope at the origin. To avoid this discontinuity in the multiquadric method, the Green functions are placed on a plane below all of the data points [Hardy and Nelson, 1986]. In four or more dimensions the Green functions are unbounded at the origin so the values of their coefficients are undetermined. I expect this problem can also be overcome by centering the Green functions slightly away from the data points. Because of these difficulties, the biharmonic spline method works best when the number of dimensions is less than three.

As in the one dimensional case, slope data can be used; in higher dimensions each slope is associated with a direction \mathbf{n} . When slope data are used, the following linear system must be solved

$$s_i = (\nabla w \cdot \mathbf{n})_i = \sum_{j=1}^N \alpha_j \nabla \phi_m(\mathbf{x}_i - \mathbf{x}_j) \cdot \mathbf{n}_i \quad (12)$$

where s_i is the slope in the direction \mathbf{n}_i . The gradient of each Green function is also given in Table 1.

Minimum Curvature Property

In one and two dimensions, it has been shown that a function w has minimum curvature if and only if it satisfies the biharmonic equation [Briggs, 1974]. The total square curvature is

$$C(w) = \int_S (\nabla^2 w)^2 dS \quad (13)$$

where ∇^2 is the Laplacian operator and S is the surface containing w . Note that the Laplacian of w is only approximately equal to the curvature when the gradient of the surface is small. Briggs [1974] first showed that if w has minimum (or maximum) curvature then it satisfies the biharmonic equation. He then showed that if w satisfies the biharmonic equation then it has minimum curvature. Green's theorem was used in both proofs. To use Green's theorem, w and its gradient must be continuous. Since the interpolating curve or surface is a linear combination of Green functions, they must also be continuous and have continuous gradient. Table 1 shows that the 1 and 2-dimensional Green functions satisfy this requirement. However, in 3 dimensions the gradient of the Green function is discontinuous at the origin and in 4 or more dimensions both the Green functions and their gradients are discontinuous. Therefore in 3 or more

TABLE 1. Biharmonic Green Functions

Number of Dimensions m	Green Function $\phi_m(\mathbf{x})$	Gradient of Green Function $\nabla \phi_m(\mathbf{x})$
1	$ x ^3$	$x x $
2	$ x 2(\ln x - 1)$	$x(2\ln x - 1)$
3	$ x $	$x x ^{-1}$
4	$\ln x $	$x x ^{-2}$
5	$ x ^{-1}$	$-x x ^{-3}$
6	$ x ^{-2}$	$-2x x ^{-4}$
m	$ x ^{4-m}$	$(4-m)x x ^{2-m}$

dimensions, biharmonic spline interpolation does not correspond to minimum curvature interpolation.

One of the interesting features of the 2-dimensional Green function is that its curvature behaves as $\ln(r)$ as $r \rightarrow 0$; this seems to violate the minimum curvature principle. Even though the curvature is infinite, the integral of the curvature squared, over a small area surrounding the singularity, is finite.

Practical Considerations

In some cases the system of linear equations (11) and (12) will be singular. As in the cubic spline method, the problem is exactly singular when two data points are located in the same position. Moreover, numerical instabilities occur when the ratio of the greatest distance between any two points to the least distance between two points is large. The linear system is most stable when there are a few equally spaced data points. To determine some of the practical limitations of this new technique, a number of test cases were solved on the computer.

The numerical stability of the linear system depends on the form of the Green function. The 1-dimensional problem is most unstable. This is perhaps due to the rapid growth (r^3) of the Green function resulting in matrix element differences which exceed the machine precision. In practice, the 1-dimensional biharmonic spline method is unstable (32-bit

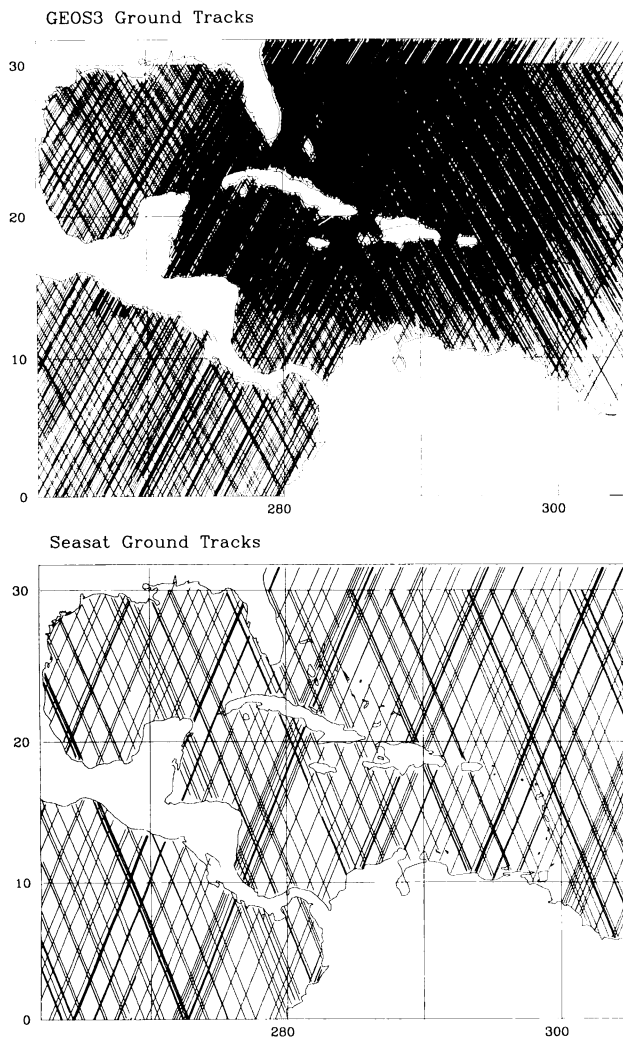


Fig. 3. Data distribution for GEOS-3 and SEASAT altimeters.

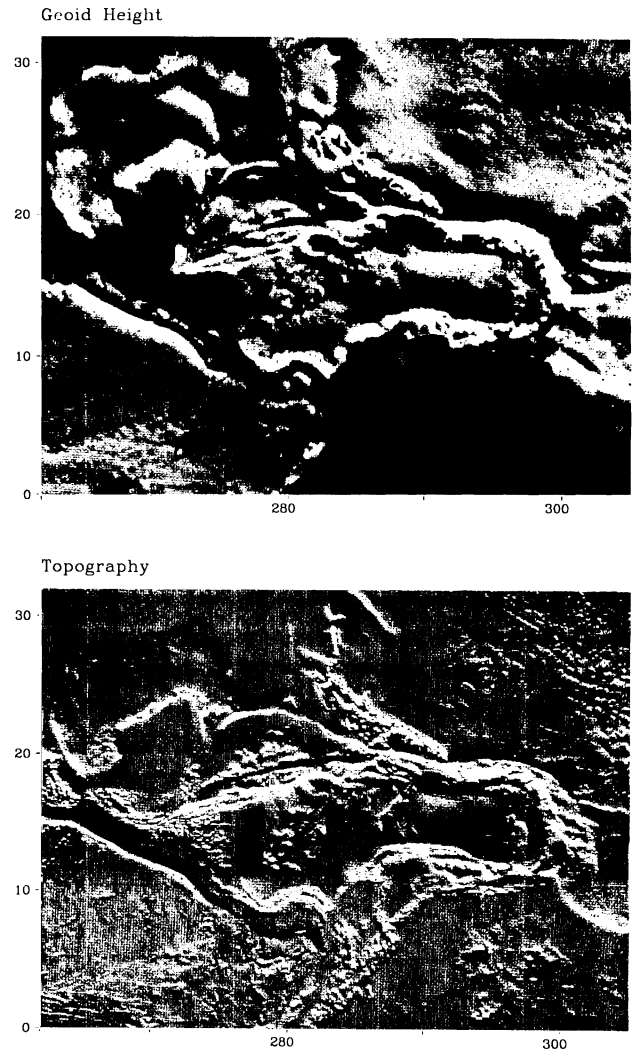


Fig. 4. Geoid height (upper) derived from GEOS-3 and SEASAT profiles (Mercator grid illuminated from the northeast). Topography (lower) (no projection).

word) when more than about 40 irregularly spaced points are used; This is a severe limitation.

In the 2 and 3 dimensional cases, linear systems as large as 400 by 400 are quite stable as long as the data distribution was not too irregular. The practical limitation is computer time which increases as the cube of the number of data points. Both the stability and speed increase when closely spaced α_j 's are omitted from the linear system.

Application To Satellite Altimetry

This new interpolation technique was designed to produce detailed and accurate sea surface topography maps from satellite altimeter profiles. In general, satellite altimeter profiles have two major types of error. Radial orbit error produces an unknown bias in each profile. This causes profiles to disagree where they intersect. Uncorrected profiles introduce stripes in interpolated maps. This type of error can be reduced by adding a bias to each profile such that all of the crossover differences are minimized [Rummel and Rapp, 1977]. However, this crossover adjustment does not work well on the noisy GEOS-3 data. Also it is difficult to find the crossover points of the SEASAT and GEOS-3 profiles since the satellites had different orbit inclinations. Instead of

performing a multisatellite crossover adjustment, the radial orbit error can be suppressed by taking the along-track derivative of each profile [Sandwell, 1984].

The second type of error in satellite altimeter data is white noise; it cannot be eliminated. The interpolation algorithm presented here can accommodate the noise level of the GEOS-3 data which is significantly greater than the noise level of the SEASAT data [Marks and Sailor, 1986].

In addition to the errors, the distribution of satellite altimeter data is very irregular. Most detailed geoid maps or gravity maps have been constructed using only SEASAT data [Haxby, 1985]. The along-track resolution of SEASAT altimeter profiles is about 30 km while the cross-track resolution is typically greater than 100 km (i.e. profile spacing > 50 km). Using a sophisticated interpolation method Haxby [1985] has retained some of the along-track resolution in his detailed gravity maps. However, these SEASAT-derived gravity maps do not accurately resolve features between profiles. The only way to improve the resolution is to fill the data gaps. Here I use the less accurate GEOS-3 data to interpolate between SEASAT profiles.

The biharmonic spline algorithm was used to combine the along-track slopes of the GEOS-3 and SEASAT data into a consistent geoid height map of the Caribbean area. In this area there are many more GEOS-3 data points (460,000) than SEASAT data (83,000) as shown in Figure 3. Before interpolating, each profile was first differentiated to remove the bias and then low-pass filtered to remove the shortest wavelength noise (< 30 km). After filtering only every other point was kept and each slope value was assigned a direction as well as an uncertainty (2 μ rad for SEASAT, 10 μ rad for GEOS-3).

The interpolation algorithm cannot accommodate all of the data points simultaneously so, the area was broken up into sub-areas each containing less than 400 data points. To avoid discontinuities, sub-areas were made to overlap by 50% on each edge. Each sub-area was then interpolated using the 2-dimensional biharmonic spline algorithm. Finally, adjacent sub-areas were blended together in the central half of the overlap region by using a distance weighting algorithm. Maximum detail and resolution was attained if model coefficients (i.e. α_j 's) were used at every 4th slope point and closely spaced α_j 's were omitted. After finding the α_j 's, the surface was calculated on a Mercator grid at 5 minute longitude intervals. The derivatives of the surface were also calculated analytically by differentiating (10) before summing the series.

A gray-tone image of the gridded surface is shown in Figure 4. (northeast illumination) along with the gridded topography [Van Wykhouse, 1973]. The geoid appears as a low-pass filtered version of the topography. They are highly correlated in the intermediate wavelength band (60 km - 600 km).

Even though these ocean areas are well surveyed, the geoid map reveals some new features. First, while the

topography of the Cayman trough is symmetric about the center of the trough, the geoid is asymmetric. Slopes in the geoid are much higher along the active transform faults than they are along the passive fracture zones. Second, the geoid map shows a previously undiscovered lineation that is parallel to the Beata Ridge and extends from Costa Rica to the eastern edge of Jamaica. It is more apparent when the geoid is illuminated from the north. Finally, the Atlantic fracture zones are evident in the geoid even though some of the fracture zone topography is buried by sediment. These results suggest that similar applications, in more remote areas, may reveal new features of the sea floor.

Acknowledgements. I thank Jan Garmany and the reviewer for their critical comments and suggestions. This work was supported by NASA under contract number NAG 5-787. University of Texas, Institute for Geophysics contribution # 680.

References

- Ahlberg, J. H., E. N. Nilson and J. L. Walsh, The Theory of Splines and Their Applications, Academic Press, New York, 1967.
- Bhattacharyya, B. K., Bicubic spline interpolation as a method for treatment of potential field data, Geophysics, **34**, 402-423, 1969.
- Briggs, I. C., Machine contouring using minimum curvature, Geophysics, **39**, 39-48, 1974.
- De Boor, C., Bicubic spline interpolation, J. Math. and Phys., **41**, 212-218, 1962.
- Hardy, R. L., Multiquadric equations of topography and other irregular surfaces, J. Geophys. Res., **76**, 1905-1915, 1971.
- Hardy, R. L. and S. A. Nelson, A multiquadric-biharmonic representation and approximation of the disturbing potential, Geophys. Res. Lett., **13**, 18-21, 1986.
- Haxby, W. F., Gravity Field of the World's Oceans, Lamont-Doherty Geol. Obs., Palisades, New York, 1985.
- Marks, K. M., and R. V. Sailor, Comparison of GEOS-3 and SEASAT altimeter resolution capabilities, Geophys. Res. Lett., **13**, 697-700, 1986.
- Rummel, R., and R. H. Rapp, Undulation and anomaly estimation using GEOS-3 altimeter data without precise satellite orbits, Bull. Geod., **51**, 73-88, 1977.
- Van Wykhouse, R., SYNAPS (Synthetic Bathymetry Profiling Systems), Technical Report TR-233, Naval Oceanographic Office, Washington, D. C., 1973.

D. Sandwell, Center for Space Research,
University of Texas at Austin, Austin, TX 78712

(Received October 10, 1986;
accepted November 24, 1986.)