Smooth fitting of geophysical data using continuous global surfaces

Stephen D. Billings*, Garry N. Newsam†, and Rick K. Beatson**

ABSTRACT

Continuous global surfaces (CGS) are a general framework for interpolation and smoothing of geophysical data. The first of two smoothing techniques we consider in this paper is generalized cross validation (GCV), which is a bootstrap measure of the predictive error of a surface that requires no prior knowledge of noise levels. The second smoothing technique is to define the CGS surface with fewer centers than data points, and compute the fit by least squares (LSQR); the noise levels are implicitly estimated by the number and placement of the centers relative to the data points. We show that both smoothing methods can be implemented using extensions to the existing fast framework for interpolation, so that it is now possible to construct realistic smooth fits to the very large data sets typically collected in geophysics.

Thin-plate spline and kriging surfaces with GCV smoothing appear to produce realistic fits to noisy radiometric data. The resulting surfaces are similar, yet the thin-plate spline required less parameterization. Given the simplicity and parsimony of GCV, this makes a combination of the two methods a reasonable default choice for the smoothing problem. LSQR smooth fitting with sinc functions defined on a regular grid of centers, effectively low-pass filters the data and produces a reasonable surface, although one not as visually appealing as for splines and kriging.

INTRODUCTION

A very common operation in geophysics is the gridding of irregularly sampled and unevenly spaced noisy data. The gridding is usually applied for visualization of the data and to expedite subsequent processing operations. Current gridding techniques mostly assume that the data is free of noise or that some estimate of the noise level is available. Both these assumptions are usually violated in practice. For example, by its nature radiometric data is a Poisson process and so almost always noisy. In areas of low isotope density or high background radiation levels, the noise can be substantial, and in some cases its variance will be of the same magnitude as the mean signal [especially for uranium; see, e.g., Grasty and Minty (1995)]. Moreover a full analysis and estimation of all contributions to uncertainty in the data is often difficult or impractical.

In this paper, we investigate the application of continuous global surfaces (CGS) to the smoothing of geophysical data. CGS encompasses many different types of surface-fitting methods, including the dual formulation of kriging (e.g. Matheron, 1980; Cressie, 1993), tension splines (Mitasa and Mitas, 1993), radial basis functions (Cheney and Light, 1999), smoothing splines (Wahba, 1990; Hutchinson 1993), sinc interpolation (Shannon, 1949), and the equivalent source method of Cordell (1992). The surfaces defined by each of these methods can be expressed in the following form:

\[ s(x) = p(x) + \sum_{n=1}^{N} \lambda_n \phi(x - x_n) \]  

where \( \phi(x) \) is a fixed, typically radially symmetric, basic function, the \( \lambda_n \) are a set of \( N \) weights corresponding to the \( N \) centers \( \{x_n : n = 1, \ldots, N\} \), and \( p(x) \) is an optional polynomial of degree \( k \) usually included to model global trends in the data. The application of the technique to exact interpolation of geophysical data was investigated in Billings et al. (2002). Here, we extend the technique developed in that paper to smoothed interpolation, or more properly smoothing, of geophysical data by generalized cross validation or least squares.

Manuscript received by the Editor April 11, 2001; revised manuscript received February 6, 2002.
*University of British Columbia, Geophysical Inversion Facility, 2219 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada. E-mail: sbillings@eos.ubc.ca.
†Defence Science and Technology Organisation, Surveillance Systems Division, Post Office Box 1500, Edinburgh, South Australia 5111, Australia. E-mail: garry.newsam@dsto.defence.gov.au.
**University of Canterbury, Department of Mathematics and Statistics, Private Bag 4800, Christchurch, New Zealand. E-mail: r.beatson@math.canterbury.ac.nz.
© 2002 Society of Exploration Geophysicists. All rights reserved.

1823
The main difficulty in applying CGS to large geophysical surveys has been the computational effort to solve the large dense matrix equations that arise, and to subsequently evaluate the CGS on large dense grids. Fortunately, recent developments in iterative algorithms and fast function evaluation techniques can be used to dramatically reduce this cost, making it practical to use CGS to fit most survey data sets. A detailed description of the use of these methods for exact interpolation can be found in Billings et al. (2002) and the references therein. The purpose of this paper is to show that smoothing can be incorporated within this framework. In particular, we show how the equations arising in interpolation by CGS can be extended to include smoothing, and how iterative techniques for their solution can also be appropriately modified.

The first smoothing technique we consider is generalized cross validation (GCV) (see Wahba, 1990). GCV is essentially a bootstrap measure of the predictive error of a surface. It may be calculated by removing each data point in turn, determining the residual at the data point of a surface fitted to the remaining data points, and then summing the squares of the residuals multiplied by appropriate weights. Fortunately calculation and minimization of the GCV measure can be readily accommodated within the linear algebraic techniques used to solve the CGS interpolation equations. The Lanczos procedure is a method for tridiagonalization of a matrix that is used in the conjugate gradient technique (e.g. Golub and van Loan, 1996). By applying the Lanczos technique to the GCV equations, an efficient iterative method (Fast GCV) was developed by Sidje and Williams (1998). We use their methodology for the smoothing of geophysical data.

The second method considered is smoothing by a least squares (LSQR) fit that is applied if fewer centers are included in the CGS surface than there are data points. The level of smoothing is controlled implicitly by the number and placement of the centers. This technique leads to a nonsquare matrix equation which can be solved by the Lanczos-based LSQR method of Paige and Saunders (1982).

The paper is set out as follows. In the first section we review the application of CGS to interpolation of geophysical data. We then describe the Fast GCV method, followed by the LSQR method. The next section describes the application of the method to the smoothing of a radiometric survey. Finally, we review the analysis and performance of the methods and consider their implications for smoothing in practice.

INTERPOLATION WITH CONTINUOUS GLOBAL SURFACES

A CGS surface is parameterized by equation (1). For interpolation, the centers in the CGS definition are usually chosen to be coincident with the data points. The weights, \( \lambda_n \), and the polynomial coefficients of the CGS surface are then determined by the requirement

\[
s(\mathbf{x}_n) = f_n, \quad \text{for } n = 1, \ldots, N,\tag{2}
\]

where the \( \mathbf{x}_n \) are the locations of \( N \) observations, \( f_n \).

The space of polynomials of degree \( k \) in \( d \) variables has dimension \( K = \binom{d+k}{d} \). Hence, if the polynomial \( p \) is present, it introduces additional degrees of freedom that must be removed. Let \( \{ p_i; i = 1, \ldots, K \} \) be a basis for all polynomials in \( d \) dimensions with degree at most \( k \). The theories underpinning most CGS constructions require that the additional degrees of freedom are removed by imposing

\[
\sum_{i=1}^{N} \lambda_n p_i(\mathbf{x}_n) = 0, \quad \text{for all } i = 1, \ldots, K.\tag{3}
\]

The conditions specified in equations (2) and (3) give a linear system of \( N + K \) simultaneous equations in \( N + K \) unknowns. The system may be written in matrix form as

\[
\begin{bmatrix}
A & P \\
P^T & 0
\end{bmatrix}
\begin{bmatrix}
\lambda \\
a
\end{bmatrix}
= \begin{bmatrix}
f \\
0
\end{bmatrix},
\]

where the matrices \( A \) and \( P \) have entries \( A_{mn} = \phi(x_m - x_n) \) and \( P_{mn} = p_i(x_n) \) and \( \lambda, a, \) and \( f \) are vectors of weights, polynomial coefficients, and data, respectively. The matrix in this equation has been shown to be invertible with very mild conditions on the geometry of the centers when the function \( \phi \) is strictly conditionally positive definite of order \( k \) (Micchelli, 1986); that is, \( \lambda^T A \lambda > 0 \) for all \( \lambda \neq 0 \) such that \( P^T \lambda = 0 \). These conditions hold for all the interpolation methods within the CGS framework.

To reduce this matrix to a more amenable form, we compute a QR factorization of the matrix \( P \) of the form \( QR = P \) where \( Q \) is an orthonormal matrix and \( R \) is an upper triangular matrix (see Golub and Van Loan, 1996). The Appendix shows that this can be used to convert equation (4) into the form

\[
C \mu = z,\tag{5}
\]

where \( C = Q^T A Q, z = Q^T f, \lambda = Q^T \mu, \) and \( Q_2 \) is the \( N \times K \) columns of \( Q \). The matrix \( C \) is now strictly positive definite; hence, equation (5) can be solved by the standard conjugate-gradient iterative technique. Moreover, as is shown in the Appendix, matrix-vector multiplications by \( C \) and \( Q_2 \) can be quickly computed in terms of similar multiplications involving \( A \) and \( P \); neither \( C \) or \( Q_2 \) need be explicitly formed or stored.

In Billings et al. (2002), we discussed methods to reduce the \( O(N^3) \) arithmetic operations and \( O(N^2) \) storage required to fit the surface by direct methods, and the \( O(N) \) operations required for a single extra evaluation of the surface. The methodology was based around the following four steps:

1. Replacing direct solution methods by suitable iterative solution methods such as conjugate gradients or generalized minimization of residuals (GMRES) (see, e.g., Saad and Schultz, 1986; Golub and van Loan, 1996). This decouples the \( O(N^3) \) dependence into a product of the number of iterations required for convergence and the cost of the matrix-vector product required at each iteration.
2. Decreasing the number of iterations required to \( O(\log N) \) or \( O(1) \) operations by preconditioning the matrix system (e.g. Beaton et al., 2001).
3. Reducing the cost of a single matrix-vector product to \( O(N \log N) \) or \( O(N) \) operations through the use of fast algorithms for the sums appearing in equation (1) (e.g., Greengard and Rokhlin, 1987; Beaton and Newsam, 1998).
4. Fast algorithms may also be used to decrease the marginal cost of a single extra evaluation to \( O(1) \) operations.

In this paper, we relax the requirement that the surface is approximately fit the data and so make allowances for noise. This
changes the matrix equations that occur and different methods of solution are needed. However, the basic considerations for reducing the operation counts still apply.

SMOOTH FITTING WITH CONTINUOUS GLOBAL SURFACES

In most geophysical applications, the data are contaminated by noise, where the noise shows more rapid spatial variation than the underlying signal. Fitting noisy data by exact interpolation (i.e., constructing a function with the same essential smoothness as the underlying signal that nevertheless still passes exactly through the data points) is likely to produce a wildly oscillating surface. Various techniques exist to construct smooth surfaces that are still reasonable approximate fits to the data. At least three of these can be naturally accommodated within the CGS framework and solution algorithms already developed:

1) Smooth by regularization, using GCV to choose the smoothing parameter.
2) Use fewer centers in the specification of the CGS surface than there are data points, and use LSQR to solve the resulting overdetermined system in a least squares sense.
3) Construct an approximate solution by not iterating conjugate gradients or GMRES to full convergence.

In principle, (1) is the best founded approach, especially when fitting arbitrary functions of unknown variation. It is based on a clear statistical model of the data (signal plus noise) and has been shown to effectively estimate the amount of noise (e.g., Wahba, 1990). In (2), constructing a CGS based on significantly fewer centers than there are data points implicitly restricts its variation to length scales longer than those separating the data points. The resulting least-squares problem is easily formulated, but choosing the centers requires a reasonable prior estimate of the smoothness of the true solution in terms of the length scales of natural variation. These may be available, but are usually unknown and indeed are one of the properties the user would like to determine from the data. The simplest approach to implement is (3): it is based on the observation that the terms in an eigenvector expansion of the solution that are associated with large eigenvalues usually correspond to the smooth or low-frequency components of the surface while the terms associated with the small eigenvalues correspond to the rough or high-frequency components (e.g., Dyn et al., 1988). Iterative methods such as conjugate gradients tend to determine the low-order coefficients in the eigenvector expansion first, which means that they generate a good smooth approximation to the overall surface in the early iterations and only fit the fine details in subsequent iterations. The technique can work well in practice, especially if the convergence rate is monitored and a good stopping point identified. Unfortunately no good, well-founded automatic method has been proposed for this, so we will not consider it further here.

Although (1) is more rigorous, in practice (2) can be more applicable in geophysics if there is a priori information on the spatial bandwidth of the data from which a “natural” nodal spacing may be determined. For example, in the aeromagnetic setting, Rauth and Stromler (1998) showed that there is an effective frequency cutoff imposed by the frequency response to magnetic anomalies observed at height. In the radiometric setting, the smoothing of ground concentrations due to the nondirectional nature of the detector coupled with the high noise levels in the data imposes an upper limit on the spatial frequencies that can be resolved.

Regularizing the solution with GCV

The standard model underlying methods for automatic smoothing of data is to assume that a datum \( f_n \) can be decomposed into a signal component \( \gamma_n \) and a noise component \( \epsilon_n \), i.e.,

\[
f_n = \gamma_n + \epsilon_n,
\]

where the \( \epsilon_n \) are assumed to be independent and normally distributed but with an unknown variance. Interpolation involves minimizing a penalty function subject to exactly fitting the data; for example, the penalty function for thin-plate splines is

\[
J(s) = \int_{\Omega} \left( \frac{\partial^2 s}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 s}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 s}{\partial y^2} \right)^2 \, \partial x \, \partial y.
\]

(7)

Analogous to this equation occur for tension splines and kriging as discussed in Billings et al. (2002). If the data points are known to be noisy, there is little point in exactly interpolating them; therefore, one instead seeks to trade off minimizing the penalty function against the mean square error in fitting the data. This leads to the following regularized least-squares problem:

\[
\min_{s} \sum_{n=1}^{N} (f_n - s(x_n))^2 + v J(s),
\]

(8)

where the parameter \( v \) governs the tradeoff between goodness of fit and smoothness. The problem thus reduces to deciding on what is a sensible, or optimal, value for \( v \) and then solving the problem given in equation (8).

Taking the second problem first, one obvious way of at least approximately solving equation (8) is to substitute the general CGS expansion of equation (1) into equation (8). If the penalty functional \( J(s) \) is quadratic in \( s \), this gives the regularized discrete least-squares problem,

\[
\min_{\lambda, a} (f - Pa - A\lambda)^T(f - Pa - A\lambda) + v\lambda^T\Omega\lambda,
\]

(9)

where \( \Omega \) is a strictly conditionally positive definite matrix. Indeed, it can be shown (e.g., Wahba, 1990) that if the CGS is the natural expansion associated with the penalty function \( J(s) \), then the solution to equation (9) is also the exact solution to equation (8). Moreover, the matrix \( \Omega \) is identical to the matrix \( A \). This follows since the basic function is the Green’s function of the operator defining the penalty functional (Duchon, 1976; Mitasova and Mitas, 1993). Substituting \( \Omega = A \) and taking derivatives of equation (9) with respect to \( \lambda \) and \( a \) rearranging terms, the solution of equation (9) satisfies the matrix system (e.g., Wahba, 1990)

\[
\begin{bmatrix}
(vI + A) & \mathbf{0} \\
\mathbf{0} & \lambda
\end{bmatrix}
\begin{bmatrix}
\mathbf{P}
\mathbf{P}^T
\end{bmatrix}
\begin{bmatrix}
\lambda \\
\mathbf{a}
\end{bmatrix}
= \begin{bmatrix}
f \\
0
\end{bmatrix}.
\]

(10)

Thus, apart from the addition of the positive constant \( v \) to the diagonal, the matrix system is identical to that for exact interpolation given in equation (4). Therefore, it can be solved efficiently by the same iterative techniques discussed in Billings.
et al. (2002). Note that if, as is for example the case when one uses \( \phi(|x|) = |x|^p \log(|x|) \) for the thin plate spline, the \( \phi \) in the CGS is not the Green’s function itself but rather a multiple of it, then the constant \( v \) in equations (9) and (10) is a multiple of that in equation (8).

Although the derivation and interpretation are different, an exactly analogous statement to equation (10) arises when a nugget effect is included in the dual form of the kriging equations (see, e.g., Matheron, 1980). More specifically, suppose that the kriging semivariogram, \( V(h) \), is taken to have the functional form

\[
V(h) = c_0[1 - \delta(\|h\|)] + V_o(h),
\]

(11)

The function \( V_o(h) \) is continuous and satisfies

\[
V_o(0) = 0 \quad \text{and} \quad \lim_{h \to \infty} V_o(h) = c_1,
\]

(12)

where \( \delta(\|h\|) \) is the Kronecker delta with \( \delta(0) = 1 \) and \( \delta(\|h\|) = 0 \), otherwise; \( c_1 \) is termed the nugget effect, and \( c_0 + c_1 \) is the height to the sill. With these definitions, the correspondence between kriging and spline smoothing can be made explicit by taking the basic function \( \phi(h) \) as \( -V_o(h) \) and the smoothing parameter \( v \) as the nugget effect \( c_0 \).

We now turn to the problem of estimating \( v \). In kriging, the natural interpretation of \( v \) as defining the nugget effect means that it can be determined in the process of estimating the complete semivariogram; this is usually done by restricted maximum likelihood (Zimmerman, 1989). In general, interpolation the usual approach is to adopt the interpretation of equation (8) and to seek a way of estimating the unknown variance of the independent noise process \( e_n \). This is equivalent to estimating an appropriate value of the smoothing parameter at \( v \). In this paper, we propose to do this by generalized cross validation.

GCV is essentially a bootstrap measure of the predictive error of a surface. It is similar to the ordinary cross validation (OCV) measure, which may be calculated by removing each data point in turn, determining the residual at the data point of a surface fitted to the remaining data points, and then summing the squares of the residuals. GCV is a suitably weighted sum of the OCV errors with the weights chosen so that the value of the resulting sum is invariant under orthogonal transformations of the data (e.g., Wahba, 1990). A sensible choice of the smoothing parameter is one that minimizes this predictive error; indeed, it can be shown that under the assumptions above this choice will produce a smooth surface that converges to the true solution with increasing numbers of data points (Wahba, 1990). Fortunately, it can be shown that in the present problem the GCV measure reduces to the following simple expression (Craven and Wahba, 1979):

\[
G(v) = \frac{(N - K)z^T(C + vI)^{-2}z}{[tr((C + vI)^{-1})]^2},
\]

(13)

where \( C = Q^T A Q \) and \( z = Q^T f \) have been transformed by a QR factorization of the polynomial matrix (see the Appendix), and \( tr(C + vI) \) denotes the trace (the sum of the elements on the diagonal). \( G(v) \) can still be difficult to calculate in general, especially the trace term in the denominator. Fortunately, Girard (1987) and Hutchinson (1989) independently developed stochastic approximations of \( G(v) \) that are well suited to large problems. Hutchinson’s estimate has the form

\[
G(v) \approx \frac{(N - K)z^T(C + vI)^{-2}z}{[u^T(C + vI)^{-1}u]^2},
\]

(14)

where \( u \) is a vector with random entries drawn with equal probability from \([-1, 1]\).

Until recently, the most efficient approach for estimating the GCV was to reduce \( C \) to tridiagonal form using Householder reductions (Hutchinson and de Hoog, 1985). This reduction requires \( O(N^3) \) operations, but once carried out it provides a means for rapidly calculating GCV for multiple smoothing parameters. Sidje and Williams (1997), however, developed an efficient iterative method for GCV. Defining \( h(v) = z^T(C + vI)^{-2}z/\|z\|^2 \) and \( g(v) = u^T(C + vI)^{-1}u/\|u\|^2 \) converts equation (14) to

\[
G(v) \approx \frac{(N - K)\|z\|^2h(v)}{[g(v)]^2}.
\]

(15)

The key idea behind the Sidje and Williams algorithm is to approximate the matrix inverses in \( h(v) \) and \( g(v) \) by their tridiagonal Lanczos representation. This requires two separate Lanczos procedures applied to \( C \), one initiated with the vector \( f \) and the other by the vector \( u \). The algorithm is very efficient because, given \( h(k,v) \) and \( g(k,v) \), the estimates at the next iteration, \( h_{k+1}(v) \) and \( g_{k+1}(v) \), can be calculated by short-term recurrences. Furthermore, once good estimates of \( h_k(v) \) and \( g_k(v) \) have been obtained for one \( v \) they can be rapidly generated for an arbitrary \( v \). Generally, this does not require any further Lanczos iterations unless the estimates \( h_k(v) \) and \( g_k(v) \) at the new \( v \) are not accurate enough, in which case the Lanczos procedures resume.

The final part of the algorithm involves calculating the solution \( \mu_{opt} \), and hence \( \lambda_{opt} = Q^T \mu_{opt} \), that corresponds to the optimal smoothing parameter \( v_{opt} \). This can be obtained efficiently by a recurrence formula that involves all the Lanczos basis vectors that defined \( h(v) \). The basis vectors are either stored along the way or can be regenerated when needed. In both cases, there is no need for extra matrix-vector products.

Finally, since GCV is based on the statistical model in equation (6), it also provides an estimate of the standard error in the fitted surface (Wahba, 1990; Hutchinson and Gessler, 1994). This can be compared with that obtainable from fitting a nugget effect in kriging; however, we will not pursue these issues further here.

**Smoothing by using fewer centers than data constraints**

The points at which the data are known and the centers that define the CGS approximation do not have to coincide. In magnetic surveys, the station spacing is of the order of 10 m or less, whereas the line spacing is generally between 50 and 400 m. The substantial over sampling along lines means that not all data constraints need to be used as centers to define a representativet surface. Selection of the defining centers has been discussed in the spline literature by, for example, Hutchinson and De Hoog (1985), where it is referred to as knot selection.

One situation where the number and spacing of centers can be decided on beforehand with some confidence is in sinc interpolation of a signal that is known to be bandlimited (i.e., it has no frequency content above a certain cutoff value \( f_c \)). Any frequencies above this level that are present in the data are noise and should not be allowed to influence the fitted surface.
Using a sinc CGS defined on a regular grid of centers with an internodal spacing of $\Delta r$ gives a cutoff frequency specified by the Nyquist relationship, $f_c = 1/(2\Delta x)$. Least squares fitting in this fashion effectively carries out a low-pass filtering of the available data. Geophysical applications of this form of interpolation occur where convolutions occurring in the measuring process and noise in the data impose effective frequency cutoffs. Such situations occur in airborne magnetic (Rauth and Stohmer, 1998) and radiometric surveys where high frequencies are effectively damped by the measuring process to below the instrument noise level, and so are just not present in the data.

When there are fewer centers than data points, the matrix in equation (4) is not square. The interpolations are overdetermined and cannot be satisfied exactly, so an approximate solution must be chosen. This is usually done by specifying a least-squares solution. With $M$ data points $x_m$ and $N$ centers, $x_\ell$, the constrained least-squares problem is

$$
\min_{\lambda} \| f - A \lambda - P \alpha \|^2 \quad \text{subject to: } P^T \lambda = 0,
$$

(16)

where the terms have the same interpretation as equation (4), except $A_{mn} = \phi(x_m - x_\ell)$, $P$ is $M \times K$ with $P_{mj} = p_j(x_m)$, and $\lambda$ is $N \times K$ with $P_{mj} = p_j(x_\ell)$. QR factorization can again be used to enforce the polynomial constraints, but this time it requires two separate factorizations because $P$ and $\lambda$ are different matrices (see the Appendix for more details). This reduces equation (16) to the form of a standard least-squares problem

$$
\min_\mu \| z - C \mu \|^2,
$$

(17)

where $C = Q_1^T A Q_2$, $z = Q_1^T f$, $\lambda = Q_1 \mu$, and $Q_1$ and $Q_2$ are the relevant parts of the QR factorizations of $P$ and $\lambda$, respectively. Note, that once again the matrix $C$ does not need to be explicitly formed, as matrix-vector multiplication with both $Q_1^T$ and $Q_2$ can be rapidly achieved at each iteration. Once the CGS weights have been found, the polynomial coefficients can be obtained by solving

$$
R \alpha = Q_1^T (f - C \mu),
$$

(18)

where $Q_1$ is the first $K$ columns of $Q$. Solution of equation (18) is trivial as $R$ is a $K \times K$ matrix, typically with $K \leq 3$.

Conjugate gradient methods require the matrix to be symmetric positive definite and cannot be used to solve rectangular systems. One way around this difficulty is to multiply the equation by $C^T$ and then solve the normal equation $C^T C \mu = C^T z$ [algorithm conjugate gradient least squares (CGLS); Hestenes and Stiefel (1952)]. This has the unfortunate consequence that the condition number of the system is squared, which can significantly slow convergence. Additionally, if the matrix $C$ is ill conditioned, CGLS can be unreliable (e.g., Paige and Saunders, 1982). A more reliable Lanczos-based approach can be developed by recognizing that the residual at the least squares solution, $r = z - C \mu$, satisfies $C^T r = 0$. This converts the problem into a system of two equations,

$$
\begin{align*}
\mu + C^T \mu &= z, \\
C^T \mu &= 0,
\end{align*}
$$

(19)

which needs to be solved for the residual $r$ and the weights $\mu$. The LSRQ algorithm of Paige and Saunders (1982) applies the Lanczos procedure to both parts of the above equation. There is a certain relationship between the two Lanczos procedures which the algorithm is able to exploit. LSQR is analytically equivalent to CGLS, but it is more reliable when $C$ is ill conditioned. Both generation of the next Lanczos vector and solving for the iterate $[x, \mu]$ can be accomplished by short-term recurrences. The principle disadvantage of the LSQR algorithm is that it requires the action of $C^T$ as well as $C$, which means that two matrix-vector products are required per iteration. However, one can still reduce the cost of a single iteration by using fast matrix-vector multiplication methods, as calculation of $C^T r$ takes approximately the same computational effort as $C \mu$. Additionally, several operations required in the fast calculation of $C \mu$ do not have to be repeated for the calculation of $C^T r$.

To investigate the conditioning of the matrix that occurs in the least squares formulation of equation (17), we calculated the two-norm condition numbers for the thin-plate spline under various different scenarios (Table 1). We first assumed the centers and data points were coincident and used points distributed within the unit square, both on regular grids and randomly located. From Table 1, we see that randomly distributed points result in worse conditioning for all three matrix formulations. The condition number of $C^T C$ is the square of $C$, which can be substantial, indicating that for large applications some form of preconditioning is essential.

Next, we investigated how the conditioning varies for thin-plate splines and sinc functions under various scenarios. For both, we assumed the centers were distributed on a $10 \times 10$ regular grid in the unit square while the data were available on a $15 \times 15$ regular grid. We then added random perturbations, of increasing size, to the locations of the data points and observed the variation in condition number displayed in Table 2.

<table>
<thead>
<tr>
<th>Perturbations</th>
<th>Thin-plate spline</th>
<th>Sinc function</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>4.9E+06</td>
<td>3.1E+00</td>
</tr>
<tr>
<td>0.1</td>
<td>5.0E+06</td>
<td>3.5E+00</td>
</tr>
<tr>
<td>0.2</td>
<td>5.1E+06</td>
<td>3.7E+00</td>
</tr>
<tr>
<td>0.4</td>
<td>5.2E+06</td>
<td>4.1E+00</td>
</tr>
<tr>
<td>0.8</td>
<td>5.9E+06</td>
<td>7.1E+00</td>
</tr>
<tr>
<td>1.6</td>
<td>2.0E+07</td>
<td>2.1E+01</td>
</tr>
<tr>
<td>Random</td>
<td>1.1E+08</td>
<td>2.5E+02</td>
</tr>
</tbody>
</table>
Last, we investigated how the ratio of the number of weights to the number of data points effects the conditioning of both thin-plate splines and sinc functions (Table 3). The weights are assumed to be distributed on a $10 \times 10$ regular grid in the unit square while the data are randomly located with between 150 and 800 points available. The median condition numbers over 10 different realizations of random numbers are tabulated in Table 3. The condition number of the thin-plate spline is observed to undergo a rapid decrease as the ratio of weights to points decreases. The same behavior is apparent for the sinc function, again with the condition numbers being orders of magnitude smaller than for the thin-plate spline.

APPLICATION TO A RADIOMETRIC SURVEY

We chose to test the smoothing methods on a radiometric survey flown by the Australian Geological Survey Organization in the Jemalong area of New South Wales, Australia, in 1995. The survey was flown in a single day along flight lines approximately 100 m apart at a mean terrain clearance of 60 m. Each measurement involved a 1-s integration time, which corresponds to observations 70 m apart along each flight line. The 256 channel spectra were processed using the Hovgaard and Grasty (1997) noise-adjusted singular-value decomposition (NASVD) technique, which significantly reduced the noise levels in the data (see Minty and McFadden, 1998). The data were then reduced to estimates of potassium, uranium, and thorium concentration using standard processing techniques (e.g., Grasty and Minty, 1995). We then selected a 5 km by 5 km portion of the total survey area (10 km by 12 km) which comprises the 3580 points shown in Figure 1a. The data are very noisy, with a dynamic range of 5.5–13.7 ppm eTh and a noise standard deviation of 0.34 ppm eTh (Minty and McFadden, 1998). We now investigate the application of the different smoothing techniques to fitting of thorium concentration.

Smoothing spline fits to the thorium data

We calculated an interpolating thin-plate spline surface along with two smooth thin-plate spline surfaces: regularization with the parameter chosen by generalized cross validation (the Fast GCV algorithm) and implicit smoothing by using fewer centers than data constraints (the LSQR algorithm).

The interpolating thin-plate spline fit was obtained by the conjugate gradient algorithm and needed 345 matrix-vector products to find reliable estimates of the conjugate gradient algorithm and needed 345 matrix-vector products to reach a tolerance threshold, selected as $||C't||/||C|| \leq \epsilon$, where $\epsilon = 10^{-7}$. The norm $||C't||$ is zero at the least squares solution, while $||C||$ is an estimate of the Frobenius matrix norm (Paige and Saunders, 1982), and $||C||$ is the residual norm.

Contour plots with 0.2-ppm eTh increments are shown in Figure 1 for the interpolating spline (Figure 1b), GCV spline (Figure 1c), and LSQR spline (Figure 1d). The interpolating thin-plate spline has a very complicated surface with numerous bulls-eye–shaped contours and a dynamic range of 5.4–13.8 ppm eTh. Both of the smooth fits have significantly less complicated surfaces and reduced dynamic ranges of 7.0–12.9 ppm eTh for the GCV spline and 6.8–12.7 ppm eTh for the LSQR spline. The standard deviations of the residuals are 0.43 ppm eTh for the GCV spline and 0.50 ppm eTh for the LSQR spline. The GCV spline fit is similar to the LSQR fit, although the latter is slightly smoother as evidenced by its more rounded contours.

The visual improvement in fit of the smooth surfaces is better emphasized by the 3-D color perspective plots shown in Figure 2, which compare the interpolating and GCV splines. The former has substantial jitter and many isolated peaks that occur because the surface is constrained to pass through each of the data points. The GCV spline, on the other hand, appears to be free of noisy jitter and yet faithfully represents the major features in the data.

A final comparison of the three fits are given in Figure 3a, which shows profiles of each of the surfaces along the flight transect identified by the arrows in Figure 1a. Clearly, the wild oscillations in the interpolating spline are unreasonable given that we estimate the noise has a standard deviation of 0.34 ppm eTh. The GCV and LSQR spline fits are quite different in several parts of the profile, such as between 0.5 and 1.5 km, 2 and 3 km, and 3.5 and 4 km. In particular, the LSQR profile between 2 and 3 km is very smooth while the GCV spline is more oscillatory. Again, this emphasizes the greater smoothness of the LSQR spline fit.

When using GCV the noise levels are estimated directly from the data through the GCV criterion. For LSQR the smoothing occurs implicitly through the number and spacing of the centers in the CGS surface, relative to the data points. Given the difficulties in implicitly estimating the noise in this way, and the much larger number of matrix-vector products required by LSQR, the GCV spline fit is to be preferred.

Spatial prediction by kriging

To obtain a kriged surface the semivariogram first needs to be estimated. A sample semivariogram can be obtained by (e.g., Cressie, 1993)

$$2V(h) = \frac{1}{|N(h)|} \sum_{(m,n) \in N(h)} \left[ f(x_m) - f(x_n) \right]^2, \quad (20)$$

where $N(h) \equiv \{(m, n) : x_m - x_n = h\}$, and $|N(h)|$ is the number of elements in $N(h)$. In principle, $h$ should be a vector to allow...
for possible anisotropy in the data. In practice, we found that the variogram for thorium data was essentially independent of direction, so that we could assume radial symmetry. Also as the survey data is irregularly sampled, there are very few sets of data points separated by exactly the same distance \( h = \| \mathbf{h} \| \). Therefore, the observed distances were binned into a finite collection of intervals with centers \( h_i \) and widths \( 2d_i \) with \( N_i \) pairs in each bin \( \{(m, n); h_i - d_i < \| \mathbf{x}_m - \mathbf{x}_n \| < h_i + d_i \} \).

We calculated a sample semivariogram by equation (20) with 100 equally spaced bins and found that we first needed to remove a quadratic polynomial before the variogram leveled off at large lags. This implies that there is a quadratic drift in the thorium data. We then fitted an exponential semivariogram,

\[
V(\mathbf{h}) = \begin{cases} 
0 & \text{if } \| \mathbf{h} \| = 0 \\
c_o + c_1[1 - \exp(-\| \mathbf{h} \|/a)] & \text{if } \| \mathbf{h} \| \neq 0
\end{cases} \quad (21)
\]

to the first 31 samples in Figure 4a, and found coefficients of \( c_o = 0.24, c_1 = 0.52, \) and \( a = 0.59 \) km. The parameter \( c_o \) is the nugget variance, \( c_o + c_1 \) is the height to the sill, and \( 3a \) is the effective range of the exponential variogram. This method for semivariogram estimation can potentially introduce bias; a more rigorous alternative is to use primary increments (e.g., Cressie, 1993). However, to avoid complicating this illustrative example, the simpler method was used here.

The variogram fit in equation (21) defines a smoothing parameter, equal to the size of the nugget, \( c_o \). The smoothing parameter can then be used when kriging a surface through equation (10). The conjugate gradient method required only 62 matrix-vector products to achieve a fit accurate to \( 10^{-6} \). This approach completely specifies the semivariogram parameters prior to the interpolation and is one method used by kriging practitioners (e.g., Cressie, 1993). However, it is difficult to determine whether the variogram parameters have been successfully estimated. For example, on re-inspection of the semivariogram fit, the curve seems to dip below the data point at the smallest lag, indicating that we may not have made the best estimate of the nugget.

Rather than using the empirical nugget estimate for the smoothing parameter, however, we can instead choose it by applying GCV to the kriging equations and interpreting the result as an estimate of the nugget effect, as advocated by

---

**FIG. 1.** (a) The location of the 3580 data points used for the smooth fits, (b) interpolating thin-plate spline, (c) GCV thin-plate spline, and (d) LSQR thin-plate spline fits to the thorium data. The arrows in (a) show the location of the flight transect used for the profile plots of Figure 3.
Hutchinson (1993). Applying Fast GCV in this way required 148 matrix-vector products and produced the quadratic variation in GCV measure shown in Figure 4b, with a minimum at \( v = 0.19 \). This is slightly smaller than that predicted by directly determining the variogram parameters (\( v = 0.24 \)). In practice, the two kriged surfaces differ little, as emphasized in the contour plots shown in Figure 5. The dynamic range of the fits are 7.2–12.8 ppm eTh for the empirical nugget and 7.0–12.8 ppm eTh for the GCV nugget. In this case, it appears that the empirical and GCV nuggets are comparable; in general, the two estimates may differ and the GCV estimate is preferred.

Comparison of the GCV spline contours (Figure 1c) with the kriging contours (Figure 5), along with the profiles in Figure 3b, shows that the recovered surfaces are very similar. The main difference is that the kriging fits are slightly sharper.

**Smooth fitting using the sinc function**

The sinc function can be an attractive choice as a basic function because it imposes an upper limit on the frequency content of the interpolated surface. When the underlying data is indeed band limited, choosing an appropriately spaced regular grid of centers for a sinc CGS basis will ensure that the frequency content of the data is completely sampled in a parsimonious manner. The spacing of the grid points, \( \Delta x \) and \( \Delta y \), defines the frequency cutoffs by the Nyquist relationships, \( f_x = 1/(2\Delta x) \) and \( f_y = 1/(2\Delta y) \). We fitted a sinc surface with centers defined on the same 100-m spacing regular grid as the thin-plate spline example. This gives a frequency cutoff of 5 km\(^{-1}\) (wavelength 200 m) in both directions, compared to the Nyquist sampling frequencies of 7.1 and 5 km\(^{-1}\) (wavelengths of 140 and 200 m).

![Three-dimensional perspective plots of (a) the interpolating thin-plate spline and (b) the GCV thin-plate spline fits to the thorium data. The same linear color stretch is applied to both plots (7–13 ppm eTh).](image-url)
in the along- and across-line directions, respectively. In the along-line direction, the sinc function is acting as a low-pass filter, whereas across lines the frequency cutoff corresponds with the sampling frequency.

The LSQR algorithm applied to the sinc data converged much faster than the thin-plate spline fit, taking only 24 matrix-vector products to reach the same tolerance threshold ($10^{-7}$). The reasons for this were explored in the discussion just preceding the results section (see Table 2). A contour plot of the surface is shown in Figure 6a and a profile in Figure 3c. The surface is very complicated and has a dynamic range of 5.6–15.0 ppm eTh, which is even larger than the dynamic range of the data (5.5–13.7 ppm eTh). Inspection of the profile shows that the surface is close to an interpolating fit, indicating that we have not allowed enough smoothing. Notice that there are numerous peaks where the sinc profile overshoots the data (e.g., at ~0.4 km and 1.9 km). This exacerbates an already noisy data set by creating taller and narrower peaks about outlying data points, and explains the large dynamic range.

Clearly, we have not incorporated enough smoothing. As a remedy, we calculated two new surfaces, one defined on a 150-m grid and the other on a 200-m grid. The frequency cutoffs of these two surfaces are 0.33 and 0.25 km$^{-1}$, respectively (wavelengths of 300 and 400 m). The resulting contour plots (Figures 6b and 6c) show that the new surfaces do a much better job of smoothing the data, especially the 200-m nodal spacing.

**Fig. 3.** Profiles of the different fits along the transect marked by arrows in Figure 1(a): (a) interpolating, GCV, and LSQR thin-plate spline fits; (b) normal kriging, GCV kriging, and GCV thin-plate spline fits; and (c) sinc fits with 100-, 150-, and 200-m nodal spacing.

**Fig. 4.** (a) Sample semivariogram binned into 100 equally spaced increments of distance and the best fitting exponential model ($c_0=0.240$, $c_1=0.523$, and $a=0.59$) fit to the first 31 values. (b) Variation in the GCV measure with smoothing parameter.
spacing surface. This point is further emphasized by the profile plots in Figure 3c.

Compared with the kriging and thin-plate spline fits, the sinc surfaces are more oscillatory or peaked. This results from the fundamental paradigm underlying sinc methods of bandlimited signals and the resulting oscillatory basic function. The paradigms underlying splines and kriging are quite different and result in smoother, more pleasing fits to the data.

**DISCUSSION**

In this paper, we considered two methods for smoothing noisy geophysical data: GCV and LSQR. We have shown that both approaches can be easily accommodated within the linear algebraic framework already used to construct exact interpolants. Consequently, they can be coupled with direct solution methods to fit smooth surfaces to data sets with 10,000 samples or less. For larger problems, they can be coupled with the preconditioned iterative solvers and fast matrix-vector product techniques mentioned here in the interpolation section and described in more detail in Billings et al. (2002). In other settings, these techniques have been shown to be capable of fitting surfaces to millions of data points on standard workstations. Moreover, Billings et al. (2002) noted that fitting a smooth continuous surface to data had several important advantages compared with the output of existing discrete techniques involving combinations of gridding and smoothing by linear filters. In particular, the surfaces produced are visually more pleasing, they provide better interpolations of gaps in the data, they are better able to handle variations in data density, and they support more accurate postprocessing of the interpolated data. Consequently, we believe there is a strong case for the methods described here becoming the approach of choice for fitting smooth CGSs to large geophysical data sets.

![Contour plots for the sinc surfaces calculated for (a) 100-m nodal spacing, (b) 150-m nodal spacing, and (c) 200-m nodal spacing.](image)

![Contour plots for the kriged surface with the nugget chosen (a) from the sample semivariogram and (b) from GCV.](image)
The most rigorous of the smoothing methods considered in this paper is GCV, which minimizes an intuitively reasonable and statistically well-founded performance measure. It was successfully applied here both to the selection of the smoothing parameter in a thin-plate spline fit and to the selection of the nugget effect in kriging. In both cases, the result was visually reasonable and required no intervention by the user.

There has been considerable debate in the kriging literature as to whether splines or kriging give a superior fit (e.g., Cressie, 1993; Hutchinson, 1993). In principle, if the data is indeed well modeled as a spatial random field with an exponential semivariogram and if the parameters in this semivariogram have been successfully estimated, then the kriged surface is “more accurate” than a spline fit; it is the best linear predictor that can be obtained from the data. In practice, (and as we observed here) there is usually little difference between the kriged surface and a thin-plate spline fit. Moreover, Hutchinson (1993) has shown that the same is true of fits to simulated data, even when they are simulated using the same assumptions underlying the variogram that is used to krig the surface. Therefore, as Hutchinson (1993) observed, since it is difficult to decide whether the kriging assumptions are indeed met, and since fitting the spline requires significantly less calculation and fitting (it does not require estimating the semivariogram), there is much to be said for using thin-plate smoothing splines in place of kriging.

The GCV constructions of thin-plate spline smoothing surfaces are preferable to those obtained with the LSQR method. In the latter, the noise levels have to be implicitly estimated by the user through the choice of the number of supporting centers for the surface. As far as we are aware, there are no good rules of thumb for making this choice. Additionally, even after appropriate rescaling, the conditioning of the LSQR system can be quite poor, resulting in slow convergence of iterative algorithms for its solution. Therefore, compared to GCV, LSQR seems a generally less efficient and effective approach. Nevertheless, it is attractive in certain situations, in particular when fitting data with a known finite bandwidth. In this case, LSQR with regularly spaced centers is equivalent to a low-pass filtering of the data. However, in the example here, the fitted surface was still not as pleasing to the eye as the spline or kriging surfaces.

Finally, we mentioned earlier a third method of smoothing (premature termination of iterative solution algorithms), although we did not further analyze it here due to the lack of good methods of selecting a stopping point, either automatically or through the use of prior knowledge. Nevertheless, the simplicity of this approach means that it could be immediately coupled with existing fast algorithms for exact interpolation to smooth very large surveys.

ACKNOWLEDGMENTS

We thank AGSO–Geoscience Australia for supplying the radiometric data used in this paper. Matlab code suitable for small- to medium-size geophysical surveys will be made available by the authors. See http://www.geop.ubc.ca/~sbilling/cgs.html for details.

REFERENCES


Giraud, D., 1987, Un algorithme simple et rapide pour la validation croisée généralisée sur des problèmes de grande taille: Inform. et Mathematiques de Grenoble RR 669-M.


Sidje, R. B., and Williams, A. B., 1997, Fast generalized cross validation: Mathematics Department, University of Queensland, research report.


APPENDIX

SPD FORM OF MATRIX EQUATIONS

In this Appendix, we give a method for converting the equations for interpolation or smoothing to a strictly positive definite form. The method is a generalization of one developed in Powell (1994) for the thin-plate spline interpolation equations, which in turn was based on the work of Sibson and Stone (1991).

Our method uses a QR factorization based on Householder reflections (e.g., Golub and Van Loan, 1996). A Householder matrix \( H \) is symmetric, orthogonal (\( H^T H = I \)), and is specified by a single vector \( v \) through the equation \( H = I - 2vv^T/v^Tv \). Householder reflections are usually used to zero selected elements within a given column of a matrix. Consider first the matrix equations that occur in interpolation [equation (4)]. Applied to the \( N \times K \) polynomial matrix \( P \) of equation (4), the QR factorization method generates \( K \) Householder matrices, \( H_1 \cdots H_K \), such that

\[
H_K \cdots H_1 P = \begin{bmatrix} R \\ 0 \end{bmatrix} \Rightarrow P = H_1 \cdots H_K \begin{bmatrix} R \\ 0 \end{bmatrix},
\]

(A-1)

where \( R \) is a \( K \times K \) upper triangular matrix. Let \( Q = H_1 \cdots H_K \). Then, the first \( K \) columns, \( Q_1, \ldots, Q_K \), span the column space of \( P \), so that \( P = Q_1 R \), while the remaining \( N-K \) columns, \( Q_2 \), span the null space of \( P^T \). Therefore, the polynomial constraints can be made automatic by requiring that \( \lambda = Q_2 \mu \) for some \( \mu \in \mathbb{R}^{N-K} \).

The original system given in equation (4) may then be written as

\[
AQ_2 \mu + Pa = f.
\]

(A-2)

Premultiplying both sides of the equation by \( Q_1^T \) and noting that \( Q_1^T P = 0 \), we find

\[
(Q_1^T AQ_2) \mu = Q_1^T f,
\]

(A-3)

or alternatively,

\[
C \mu = z,
\]

(A-4)

where \( C = Q_1^T AQ_2 \), and \( z = Q_1^T f \). Since \( A \) is positive-definite over the space of \( \lambda \), the identity \( \mu^T (Q_1^T AQ_2) \mu = \lambda^T A \lambda \) shows that the matrix \( C \) is strictly symmetric positive (negative-) definite. Thus, equation (A-4) is suitable for solution by the conjugate gradient method.

It remains to find the polynomial coefficients. From equation (A-2), \( Pa = f - A \lambda \), and the identity \( P = Q_1 R \) plus the orthogonality of \( Q_1 \) imply that

\[
Ra = Q_1^T (f - A \lambda).
\]

(A-5)

This is an upper triangular \( K \times K \) matrix equation that is trivial to solve for the unknown polynomial coefficients. Alternatively, we can evaluate \( A \lambda \) onto \( K \) or more data points and obtain the polynomial coefficients by a least squares fit.

When there are fewer defining centers than data points, a constrained least-squares problem arises that involves the rectangular \( M \times N \) matrix \( A \), the \( M \times K \) matrix \( P \), and the \( N \times K \) matrix \( P \):

\[
\min_{\lambda,a} ||f - A \lambda - Pa||^2 \quad \text{subject to:} \quad \hat{P}^T \lambda = 0. \quad (A-6)
\]

Construct QR factorizations of \( P = QR \) and \( \hat{P} = QR \) in the same way as before. The constraints on the weights imply that \( \lambda = Q_2 \mu \) for some \( \mu \in \mathbb{R}^{N-K} \). Because the matrix \( [Q_1, Q_2]^T \) is orthonormal,

\[
||f - A \lambda - Pa||^2 = \left\| (Q_1, Q_2)^T (f - A \lambda - Pa) \right\|^2. \quad (A-7)
\]

The equation can be separated into two components:

\[
\min_{\mu,a} \left( \left|\left| Q_1^T f - Q_1^T AQ_2 \mu - Q_1^T Pa \right|\right|^2 + \left|\left| Q_2^T f - Q_2^T AQ_2 \mu - Q_2^T Pa \right|\right|^2 \right). \quad (A-8)
\]

Observing that \( Q_1^T P = 0 \) and letting \( C = Q_1^T AQ_2, \; z = Q_1^T f, \) and \( Q_1 P = R \) converts expression (A-8) to

\[
\min_{\mu, a} \left( \left|\left| Q_1^T f - Q_1^T A \lambda \right|\right|^2 + \left|\left| z - \hat{C} \mu \right|\right|^2 \right). \quad (A-9)
\]

Now the second norm may be minimized by choice of \( \mu \), and then the first norm made zero by choice of \( a \). Thus, the minimum of the quantity (A-8) over all choices of \( \mu \) and \( a \) is achieved when \( \mu \) solves the unconstrained least-squares problem

\[
\min_{\mu} \left|\left| z - \hat{C} \mu \right|\right|^2. \quad (A-10)
\]

and \( a \) is then the solution of the triangular system

\[
Ra = Q_1^T (f - A \lambda). \quad (A-11)
\]

where \( \lambda = Q_2 \mu \). The least squares problem (A-10) is solved by the LSQR algorithm.

The advantage of the Householder approach for QR factorization is that neither the matrix \( Q_1 \) nor the product \( C = Q_1^T AQ_2 \) have to be explicitly formed and stored. The matrix \( Q_2 \) is the last \( N-K \) columns of the matrix \( Q \), which is constructed from a sequence of Householder matrices. The action of \( Q \) on a vector can be evaluated by calculating a matrix-vector product for each Householder matrix. Due to their special structure, each Householder matrix-vector multiply can be calculated very quickly in about 4NK flops (e.g. Golub and Van Loan, 1996). As there are \( K \) Householder matrices, the matrix-vector product of \( Q_1 \), or of \( Q_2 \), can be calculated in approximately \( 4NK \) flops. The calculation of \( Q_2 \mu \) can be achieved by padding \( \mu \) with \( K \) leading zeros to form \( \mu' \) and then calculating \( Q_2 \mu' \). Similarly, matrix-vector multiplication with \( Q_1 \) can be achieved by multiplying by \( Q_1^T \) and then dropping the first \( K \) rows of the resultant vector. Therefore, the QR factorization parts of the matrix-vector multiplication \( C \mu \) can be applied rapidly at each iteration by first calculating \( v = Q_2 \mu \), followed by \( k = Av \), and finally \( Q_1 \hat{v} \).