Department of Statistics
University of Wisconsin

1210 W. Dayton St.
Madison, Wisconsin 53706

TECHNICAL REPORT NO. 715

August, 1983

A TRUNCATED SINGULAR VALUE DECOMPOSITION AND OTHER METHODS FOR GENERALIZED CROSS-VALIDATION

by

Douglas M. Bates

Grace Wahba

This research was sponsored by NSF under research grant #MCS-8102732 and by ONR under Contract No. N00014-77-C-0675

A TRUNCATED SINGULAR VALUE DECOMPOSITION AND OTHER METHODS FOR GENERALIZED CROSS-VALIDATION

Douglas M. Bates and Grace Wahba

University of Wisconsin - Madison

Department of Statistics

Abstract

Generalized Cross-Validation (GCV) is an effective method of choosing the smoothing or regularization parameter in data smoothing problems. Unfortunately, the computing cost for the method can be high - particularly with large data sets. One of the most costly parts of this computation is taking the Singular Value Decomposition of a large, generally ill-conditioned, matrix. Since the smaller singular values are not important in further steps, we present a method of determining the larger singular values by a Truncated Singular Value Decomposition (TSVD). Other methods are also presented to use non-iterative matrix decompositions whenever possible and to economize on storage. Applications to the solution of Fredholm integral equations of the first kind and to the computation of thin plate smoothing splines are given.

1. Introduction

The method of generalized cross validation (GCV), described in Craven and Wahba [3], Wahba [20], Golub, Heath, and Wahba [10], and other papers, has proven to be an excellent method of choosing a smoothing parameter or regularization parameter. It has been used in applications from meteorology (see Wahba and Wendelberger [24], Wahba [23]), medicine (see Wahba [23]) and other contexts (see Merz [14], Crump and Seinfeld [4]). However, the method has been criticized recently (Utreras [19], Wecker and Ansley [25]) for its high computational cost since a straightforward application of generalized cross validation to data smoothing is an $O(n^3)$ computation where n is the number of data points. Its use on very large data sets is thus impractical.

We give three methods here for reducing the cost of smoothing and other applications using GCV so the method can be used with large data sets. First, the use of basis functions is outlined in section 2 and the associated GCV calculations are described. The greatest part of the time spent in these calculations is in the determination of the singular value decomposition (SVD) of a large, generally ill-conditioned matrix. We introduce a truncated singular value decomposition (TSVD) in section 3 to provide the information needed in the GCV calculations but at a lower cost. In section 4, we give a means of performing the GCV calculations with the smoothing or penalty component defined by a semi-norm. The method avoids all eigenvalue-eigenvector calculations and uses instead a pivoted Cholesky decomposition thereby lowering the computing burden. A few remarks about generalizations are made in section 5. In section 6 we show how the method can be used to compute cross-validated thin-plate smoothing splines in two or more dimensions with large data sets.

The use of the TSVD is not restricted to GCV calculations. It can be applied in any situation where the larger singular values and associated singular vectors of an ill-conditioned matrix are required. It complements other methods for performing SVD calculations quickly such as those introduced by Golub, Luk, and Overton [11], O'Leary and Simmons [17], Cullum, Willoughby, and Lake [5] and Cuppen [6]. We provide a bound on the error of approximation of the singular values so the impact of the approximation can be assessed and, if necessary, the calculation performed again with a closer approximation.

2. Generalized Cross Validation Calculations

An important setting for GCV involves a data vector $(y_1, y_2, \cdots, y_n)^T$ which is assumed to be of the form

$$y_i = \int_0^1 K(t_i, s) f(s) ds + \varepsilon_i, \quad i=1, \cdots, n \tag{2.1}$$

where K is known and the ε_i are zero mean errors with constant (but unknown) variance. A regularized estimate of the function f, for a given value of the regularization parameter λ, is the minimizer of

$$\frac{1}{n} \sum_{i=1}^{n} [(Kf)(t_i) - y_i]^2 + \lambda \int_0^1 (f^{(m)}(s))^2 ds \quad . \tag{2.2}$$

This minimizer f_λ can be shown to exist and to be linear in the data vector y under suitable, rather general conditions. If $A(\lambda)$ is the n by n influence matrix which satisfies

$$((Kf_\lambda)(t_1), \ldots, (Kf_\lambda)(t_n))^T = A(\lambda) y \tag{2.3}$$

then the GCV estimate $\hat{\lambda}$ of λ is the minimizer of

$$V(\lambda) = \frac{||(I - A(\lambda))y||^2 / n}{[\frac{1}{n} tr(I - A(\lambda))]^2} \tag{2.4}$$

While there are specialized methods of determining $A(\lambda)$ and $tr(I - A(\lambda))$ for certain problems, the determination of $A(\lambda)$ for a general function space and the necessary optimization problems results in a calculation which is $O(n^3)$. When large data sets are being used, however, it was suggested in Wahba [20] that the minimizer of (2.2) in the Sobolev space W_2^m can be approximated to a suitable accuracy by an element in the span of a set of basis functions $\{B_i\}$, i=1,...,p where $p \le n$. In the example of (2.2) a basis of B-splines of degree $2m-1$ is a natural set (see deBoor [7], Nychka, Wahba, Goldfarb, and Pugh [16]). With basis functions, the regularized estimate $f_{\lambda,p}$ is of the form

$$f_{\lambda,p} = \sum_{j=1}^{p} c_{\lambda,j} B_j \tag{2.5}$$

where $c_\lambda = (c_{\lambda,1}, \cdots, c_{\lambda,p})^T$ is the minimizer of

$$\frac{1}{n} || y - Xc ||^2 + \lambda c^T \Sigma c \tag{2.6}$$

and the $n \times p$ matrix X has (i,j)'th entry

$$x_{i,j} = \int_0^1 K(t_i, s) B_j(s) ds \tag{2.7}$$

while the $p \times p$ matrix Σ has (i,j)'th entry

$$\sigma_{i,j} = \int_0^1 B_i^{(m)}(s) B_j^{(m)}(s) ds \ . \tag{2.8}$$

The minimizer could be expressed as

$$c_\lambda = (X^T X + n\lambda\Sigma)^{-1} X^T y \tag{2.9}$$

which gives a form like ridge regression calculations. In fact GCV is shown in Golub et al. [10] to be an excellent method of choosing the ridge parameter in ridge regression.

If we suppose that the symmetric matrix Σ is positive definite, a linear transformation of parameters from c to

$$g = Rc \tag{2.10}$$

with the corresponding transformation from X to

$$Z = XR^{-1} \tag{2.11}$$

where R is the Cholesky factor of Σ so that

$$\Sigma = R^T R \tag{2.12}$$

changes (2.6) to

$$\frac{1}{n}||y-Zg||^2 + \lambda g^T g \tag{2.13}$$

Following Golub et al. [10] we can simplify $V(\lambda)$ by forming the singular value decomposition of Z as

$$Z = UBV^T \tag{2.14}$$

where U is $n \times p$ with $U^T U = I$, B is $p \times p$ and diagonal with diagonal entries $\{b_i\}$, i=1,...,p, and V is $p \times p$ and orthogonal. Using the p dimensional vector

$$z = U^T y \tag{2.15}$$

the predicted y values $(Kf_\lambda(t_1), \ldots, (Kf_\lambda(t_n))$ for a given λ become

$$(Kf_\lambda(t_1), \ldots, Kf_\lambda(t_n)) = y_\lambda = A(\lambda)y = \sum_{j=1}^{p} u_j \frac{b_j^2 z_j}{b_j^2 + n\lambda} \tag{2.16}$$

where u_j is the j'th column of U. This corresponds to an $A(\lambda)$ of

$$A(\lambda) = UB^2(B^2 + n\lambda I)^{-1}U^T \tag{2.17}$$

which, combined with (2.4), gives

$$V(\lambda) = \frac{n[||y||^2 + \sum_{j=1}^{p} \left[\frac{n\lambda}{b_j^2 + n\lambda}\right]^2 z_j^2 - ||z||^2]}{[n + \sum_{j=1}^{p} \frac{n\lambda}{b_j^2 + n\lambda} - p]^2} \tag{2.18}$$

It can be seen that the expression for $V(\lambda)$ in (2.18) is a relatively simple rational function of λ once the $\{b_i\}$ and the $\{z_i\}$ values have been determined.

The major part of the calculation for the method using basis functions is in the singular value decomposition of Z. We discuss in the next section how this can be streamlined. The GCV calculations outlined above apply only when the matrix Σ is positive definite so that the term $c^T \Sigma c$ represents the square of a norm. In many cases, Σ as defined by (2.8) will only be positive semi-definite. In section 4 we will give modifications to these calculations to deal with a positive semi-definite Σ.

3. The Truncated Singular Value Decomposition

We begin by quoting the following theorem on a bound for the error in the singular values when using an approximation to a matrix.

Theorem 1: Let X and Y be $n \times p$ $(n \geq p)$ matrices with singular value decompositions UDV^T and RSW^T respectively. Denote the ordered singular values of X as $\{d_i\}$, i=1,...,p with $d_1 \geq d_2 \geq \cdots \geq d_p$ and the ordered singular values of Y as $\{s_i\}$, i=1,...,p . Then

$$\sum_{i=1}^{p}(d_i - s_i)^2 \leq ||X - Y||_F^2 = tr(X-Y)^T(X-Y)$$

This is quoted in Sun [18] and proven in Mirsky [15].

We will take advantage of this theorem to calculate the SVD of a matrix X_k which is close to X in the sense that $||X - X_k||$ is small but is better conditioned than is X so the iterative portion of the SVD tends to converge faster and the computational burden is reduced. First, we take a pivoted QR decomposition of X using the pivoting scheme from Linpack [8]. That is, we determine Q, $n \times n$ orthogonal, R, $n \times p$ and zero below the main diagonal, and E, a $p \times p$ permutation matrix, such that

$$XE = QR \tag{3.1}$$

and R has the property that

$$r_{k,k}^2 \geq \sum_{i=k}^{j} r_{i,j}^2 \quad (j = k, k+1, \ldots, p). \tag{3.2}$$

If we take the SVD of R_p, the triangular matrix composed of the first p rows of R, as

$$R_p = KDL^T \tag{3.3}$$

we can produce the SVD of X as

$$X = Q_p KDL^T E^T = UDV^T \tag{3.4}$$

where Q_p is the $n \times p$ matrix composed of the first p columns of Q and $U = Q_p K$ is $n \times p$ while $V = EL$ is $p \times p$ and orthogonal. This method would not, however, produce better conditioning for the SVD algorithm since the singular values of R_p are the same as the singular values of X.

To provide better conditioning, we truncate the matrix R_p after the k'th row and take the SVD of the resulting $k \times p$ matrix R_k $(k \leq p)$ as

$$R_k = K_k D_k L_k \tag{3.5}$$

where K_k is $k \times k$ and L_k is $k \times p$. The diagonal elements of D_k are no longer the singular values of X but now represent the singular values of a matrix

$$X_k = Q_p \begin{bmatrix} R_k \\ 0 \end{bmatrix} E^T \tag{3.6}$$

which is different from X. However,

$$||X - X_k||_F = \left[\sum_{i=k+1}^{p} \sum_{j=i}^{p} r_{i,j}^2 \right]^{1/2} \tag{3.7}$$

so we can choose k to be as small as possible subject to the constraint that

$$||X - X_k||_F / ||X||_F \leq \varepsilon \tag{3.8}$$

where ε is a small number, say 10^{-8}. The double sum on the right of (3.7) is easily evaluated a row at a time starting at the p'th row until the constraint (3.8) is violated and the smallest k is determined.

By theorem 1, if $\{d_i\}$, i=1,...,p are the ordered singular values of X and $\{d_{i,k}\}$, i=1,...,p are the ordered singular values of X_k, then

$$\left(\sum_{i=1}^{p} (d_i - d_{i,k})^2\right)^{1/2} \leq \varepsilon ||X||_F = \varepsilon \left(\sum_{i=1}^{p} d_i^2\right)^{1/2} \tag{3.9}$$

The $d_{i,k}$ and the corresponding left singular vectors calculated from $Q_p K_k$ are used to calculate $V(\lambda)$ and to determine an estimate $\hat{\lambda}$. The error bound, $\varepsilon ||X||_F$ is then compared to $n\hat{\lambda}$. If it is much smaller, say by two or more orders of magnitude, the values are accepted. Otherwise, the value of ε is decreased and k is increased to provide a lower error bound.

The end result is that singular values whose squares are comparable to $n\hat{\lambda}$ or greater are well determined. The effect of the other singular values and corresponding singular vectors in calculating $V(\lambda)$ by (2.16) and in calculating $\overline{y}$ is negligible.

This method of calculating the SVD is similar to a method proposed by Chan [2] which also starts with a QR decomposition. However we use a pivoted QR decomposition rather than an unpivoted decomposition and we truncate the resulting R matrix. It is possible to exploit the fact that R_p and R_k are upper triangular when using orthogonal transformations to reduce them to bidiagonal form but Chan [2] points out that this does not result in a substantial computational saving. In fact, we just use the SVD code from Linpack [8] based on the algorithm from Golub and Reinsch [12] for general matrices.

4. Calculations for a Semi-norm

In the case that the matrix Σ is positive semi-definite rather than positive definite, the Cholesky decomposition used in (2.10) cannot usually be computed. We can, however, use a pivoted Cholesky decomposition to form

$$E^T \Sigma E = S^T S \tag{4.1}$$

where E is a $p \times p$ permutation matrix, S is $(p-m) \times p$ with zeroes below the main diagonal, and m is the dimension of the null space of Σ. A QR decomposition of S^T gives

$$S^T = QR \tag{4.2}$$

with Q being $p \times p$ and orthogonal and R being $p \times (p-m)$ and zero below the main diagonal. If we let Q_1 be the first $p-m$ columns of Q and Q_2 be the last m columns of Q, then the columns of Q_2 span the null space of Σ and the columns of Q_1 span the orthogonal complement of the null space. Letting R_1 be the first $p-m$ rows of R, we can now transform to parameters γ and δ where δ lies in the null space of Σ by

$$\begin{pmatrix} \gamma \\ \delta \end{pmatrix} = \begin{pmatrix} R_1^T & 0 \\ 0 & I \end{pmatrix} Q^T E^T c \tag{4.3}$$

with the corresponding transformation of the X matrix to

$$Y = [Y_1 \vdots Y_2] = XEQ\begin{pmatrix}(R_1^T)^{-1} & 0 \\ 0 & I\end{pmatrix} \tag{4.4}$$

The function of γ and δ which is to be optimized is now

$$g_\lambda(\gamma,\delta) = \left|\left|y - Y\begin{pmatrix}\gamma \\ \delta\end{pmatrix}\right|\right|^2 + n\lambda\gamma^T\gamma \tag{4.5}$$

and the second term on the right is in the desired form. However, the first term involves δ as well as γ. We can isolate the dependence of the first term on γ by taking a QR decomposition of Y_2 so

$$Y_2 = FG = [F_1 \vdots F_2]\begin{bmatrix}G_1 \\ 0\end{bmatrix} \tag{4.6}$$

and pre-multiplying everything in the left hand term by F^T. This produces

$$g_\lambda(\gamma,\delta) = ||w_1 - T_1\gamma - G_1\delta||^2 + ||w_2 - T_2\gamma||^2 + n\lambda\gamma^T\gamma \tag{4.7}$$

where

$$w = \begin{pmatrix}w_1 \\ w_2\end{pmatrix} = F^T y \tag{4.8}$$

and

$$T = \begin{pmatrix}T_1 \\ T_2\end{pmatrix} = F^T Y_1 = \begin{pmatrix}F_1^T Y_1 \\ F_2^T Y_1\end{pmatrix} \tag{4.9}$$

The leading term on the right hand side of (4.7) can be made zero for any choice of γ by solving

$$G_1\delta = w_1 - T_1\gamma \tag{4.10}$$

so we can proceed as in the case of the positive definite Σ as described in section 2 with g replaced by γ. That is, we form

$$T_2 = UBV^T \tag{4.11}$$

and set

$$z = U^T w_2 \tag{4.12}$$

to get

$$V(\lambda) = \frac{n[\,||w_2||^2 + \sum_{j=1}^{p-m} \left[\frac{n\lambda}{b_j^2+n\lambda}\right]^2 z_j^2 - ||z||^2]}{[n + \sum_{j=1}^{p-m} \frac{n\lambda}{b_j^2+n\lambda} - (p-m)]^2} \tag{4.13}$$

Once the optimal λ is chosen, γ can be calculated as

$$\gamma = V(B^2 + n\lambda I)^{-1} Bz \tag{4.14}$$

then δ from (4.10) and finally c from solving (4.3).

5. Generalizations

The methods above are broadly applicable to smoothing problems and general ill-posed linear operator equations involving functions defined on the plane, the sphere, and in d dimensions. See, for example Wahba [21] and the references cited there. Let H be a Hilbert space of real valued functions on some compact index set τ, let J(.) be a seminorm on H with M dimensional null space, and suppose the data are of the form

$$y_i = L_i f + \varepsilon_i, \qquad i=1,\dots,n$$

where the $L_1, \cdots, L_n$ are n bounded linear functionals on H. Let $B_1, \cdots, B_p$ be p suitably chosen basis functions in H, then f_λ is obtained as the minimizer in span$\{ B_1, \cdots, B_p\}$ of

$$\frac{1}{n}\sum_{i=1}^{n}(L_i f - y_i)^2 + \lambda J(f).$$

The matrix Σ of (2.8) has i,j'th entry $<B_i, B_j>$ where $<B_i, B_j>$ is the semi inner product associated with J, and the matrix X of (2.7) has i,j'th entry $L_i B_j$.

6. Thin plate smoothing splines

With a trivial extension, the methods above can be used to obtain an efficient approximation to a thin plate smoothing spline (TPSS) in two or more dimensions, with a very large data set, by using the basis functions described in Wahba [22].

The model for the TPSS in two dimensions is

$$y_i = f(x_{1i}, x_{2i}) + \varepsilon_i, \quad i=1,\ldots,n \;,$$

and the TPSS estimate f_λ for f is the minimizer, in an appropriate space H of functions possessing square integrable derivatives of total order m, of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_{1i}, x_{2i}))^2 + \lambda J_m(f) \tag{6.1}$$

where

$$J_m(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{\nu=0}^{m}\binom{m}{\nu}\left[\frac{\partial^m}{\partial x_1^{\nu}\partial x_2^{m-\nu}} f(x_1, x_2)\right]^2 dx_1 dx_2 \,. \tag{6.2}$$

In particular, for $m=2$

$$J_2(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(f_{x_1x_1}^2 + 2f_{x_1x_2}^2 + f_{x_2x_2}^2)dx_1 dx_2 \,. \tag{6.3}$$

The minimizer f of (6.1) can be obtained from the work of Duchon [9] (where H is also described), and has the representation:

$$f_\lambda(t) = \sum_{i=1}^{n} c_i E_m(t - t_i) + \sum_{\nu=1}^{M} d_\nu \varphi_\nu(t), \tag{6.4}$$

where $t = (x_1, x_2)^T$, $E_m(t) = [2^{2m-1}\pi\{(m-1)!\}^2]^{-1}||t||^{2m-2}\ell n||t||$,

$$\varphi_1(t), \ldots, \varphi_M(t) = 1, x_1, x_2, x_1^2, x_1x_2, x_2^2, \cdots, x_2^{m-1},$$

where $M=\binom{m+1}{2}$ is the number of polynomials of total degree $\le m-1$, and the vectors $c=(c_1, \cdots, c_n)^T$ and $d=(d_1, \cdots, d_M)^T$ are solutions to

$$\begin{aligned}(K + n\lambda I)c + Td &= y \\ T^T c &= 0\,.\end{aligned} \tag{6.5}$$

Here K is the $n \times n$ matrix with (i,j)'th entry $E_m(t_i - t_j)$, and T is the $n \times M$ matrix with j,ν entry $\varphi_\nu(t_j)$. The $\{\varphi_\nu\}$ span the null space of J_m and T of rank M is sufficient for a unique minimizer. The use of GCV to choose λ in this problem, and a computational approach good for n up to several hundred is given in Wahba and Wendelberger [24] see also Wendelberger [26], Wahba [21]. Generalizations to three and higher dimensions are discussed in Duchon [9] as well as the above references and the results below also generalize in a straightforward manner, but we restrict the discussion to two dimensions to avoid excessive notation.

There are several generalizations of B splines to more than one dimension. The generalization suggested in Wahba [22] appears to be a natural one for use with the penalty functional J_m of (6.2) and goes as follows: Choose p "knots" $s_1, s_2, \ldots, s_p$ in Euclidean 2 space, distributed throughout the same region as the data points $t_1, t_2, \ldots, t_n$. Letting S be the $p \times M$ matrix with i,ν'th entry $\varphi_\nu(s_i)$, the $s_1, \ldots, s_p$ must be chosen so that S is of rank M. Let B_p be

$$span\left\{\sum_{i=1}^{p} u_{ji} E_m(\cdot - s_i), j=1,\ldots,p-M\right\} \cup \left\{\varphi_\nu, \nu=1,\ldots,M\right\}$$

where the $u_j = (u_{j1},\ldots,u_{jp})^T$, $j=1,\ldots,p-M$ are linearly independent and satisfy $S^T u_j = 0$. B_p is a p dimensional subspace of H. Then the minimizer f_λ of (6.1) in B_p has a representation

$$f_\lambda(t) = \sum_{i=1}^{p} \alpha_i E_m(t - s_i) + \sum_{\nu=1}^{M} \vartheta_\nu \varphi_\nu(t), \tag{6.6}$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)^T$ satisfies

$$S^T \alpha = 0, \tag{6.7}$$

and it can be shown that α and $\vartheta = (\vartheta_1, \ldots, \vartheta_M)^T$ are the minimizers of

$$\frac{1}{n} \| y - (L\alpha + T\vartheta) \|^2 + \lambda \alpha^T J \alpha \tag{6.8}$$

subject to (6.7), where L is the $n \times p$ matrix with (i,j)'th entry $E_m(t_i - s_j)$ and J is the $p \times p$ matrix with (i,j)'th entry $E_m(s_i - s_j)$. It is known that J is strictly positive definite over the $p-M$ dimensional subspace of Euclidean p space perpendicular to the M columns of S.

Let the QR decomposition of S be

$$S = [Q_1 : Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \tag{6.9}$$

where Q_1 is $p \times M$, Q_2 is $p \times (p-M)$ and R_1 is $M \times M$. Then by (6.7), α has a representation

$$\alpha = Q_2\psi$$

for some $\psi=(\psi_1,\ldots,\psi_{p-M})^T$ and (6.8) becomes

$$\frac{1}{n}||y-LQ_2\psi-T\vartheta||^2+\lambda\psi^T Q_2^T J Q_2\psi \ . \tag{6.10}$$

Now let $W^T W$ be the Cholesky decomposition of $Q_2^T J Q_2$ and let $\gamma=W\psi$. Then (6.10) becomes

$$\frac{1}{n}||y-LQ_2W^{-1}\gamma-T\vartheta||^2+\lambda\gamma^T\gamma$$

which is of the form (4.5). Thus γ and $\hat{\lambda}$ may be found exactly as in Section 4, and thence α and ϑ in the representation (6.6) .

The space B_p contains the "bellshaped functions" of Dyn and Levin [1], and a spanning set for B_p has been successfully used as basis functions by Hutchinson et al. [13] in the context of thin plate smoothing splines to estimate Australian solar radiation as a function of latitude and longitude. The procedure here can also be extended to the solution of integral equations involving functions of two or more variables.

7. Discussion

Three methods are proposed here for making the use of generalized cross-validation on large data sets easier. The use of basis functions with either a positive definite or positive semi-definite penalty functional can be simplified by the methods of sections 2 and 3. The time-consuming step of calculating the SVD can be speeded by using the TSVD of section 3 and the particular applications to TPSS are aided by the methods of section 6.

8. Acknowledgement

We would like to thank Gene Golub for pointing out Theorem 1 to us, and for other helpful discussions.

References

[1] **N. Dyn and D. Levin**, A procedure for generating diagonal dominance in ill-conditioned systems originating from integral equations and surface interpolation, TR#81-18, School of Mathematical Sciences, Tel Aviv University, June 1981 .

[2] **T. F. Chan**, An improved algorithm for computing the singular value decomposition, ACM Trans. Math. Software, 8 (1982). pp.72-83.

[3] **P. Craven and G. Wahba**, Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation, Numer. Math., 31 (1979). pp.377-403.

[4] **J. G. Crump and J. H. Seinfeld**, A new algorithm for inversion of aerosol size distribution data, Aerosol Sci. and Tech., 1 (1982). pp.15-34.

[5] **J. Cullum, R. A. Willoughby, and M. Lake**, A Lanczos algorithm for computing singular values and vectors of large matrices, SIAM J. Sci. Statist. Comput., 4 (1983). pp.197-215.

[6] **J. J. M. Cuppen**, The singular value decomposition in product form, SIAM J. Sci. Statist. Comput., 4 (1983). pp.216-222.

[7] **C. DeBoor**, A Practical Guide to Splines, Springer-Verlag, New York, 1978. (Applied Mathematical Sciences No. 27)

[8] **J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart**, Linpack User's Guide, SIAM, Philadelphia, 1979.

[9] **J. Duchon**, Splines minimizing rotation invariant seminorms in Sobolev spaces, Constructive Theory of Functions of Several Variables, W. Schempp and K. Zeller, ed., Springer-Verlag, New York, 1976, pp.85-100. (Lecture Notes in Mathematics No. 571)

[10] **G. H. Golub, M. Heath, and G. Wahba**, Generalized cross validation as a method for choosing a good ridge parameter, Technometrics, 21 (1979). pp.215-224.

[11] **G. H. Golub, F. T. Luk, and M. L. Overton**, A block Lancos method for computing the singular values and corresponding singular vectors of a matrix, ACM Trans. Math. Software, 7 (1981). pp.149-169.

[12] **G. H. Golub and C. Reinsch**, Singular value decompositions and least squares solutions, Numer. Math., 14 (1970). pp.403-420.

[13] **M. F. Hutchinson, T. H. Booth, J. P. McMahon, and M. A. Nix**, Estimating monthly mean values of daily total solar radiation for Australia, CSIRO Division of Water and Land Resources, 1983. (to appear, Solar Energy)

[14] **P. Merz**, Determination of adsorption energy distribution by regularizarion and a characterization of certain adsorption isotherms, J. Comput. Physics, 38 (1980). pp.64-85.

[15] **L. Mirsky**, Symmmetric gauge functions and unitarily invariant norms, Quart. J. Math. Oxford, 11 (1960). pp.50-59.

[16] **D. Nychka, G. Wahba, D. Goldfarb, and T. Pugh**, Cross validated spline methods for the estimation of three dimensional tumor size distributions from observations on two dimensional cross sections. , TR#711,

Dept. of Statistics, U. of Wisconsin, 1983 .

[17] **D. P. O'Leary and J. A. Simmons**, A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems, SIAM J. Sci. Statist. Comp., 2 (1981). pp.474.

[18] **J-G. Sun**, Perturbation analysis for the generalized singular value problem, SIAM J. Numer. Anal., 20 (1983). pp.611-625.

[19] **F. Utreras**, Optimal smoothing of noisy data using spline functions, SIAM J. Sci. Statist. Comput., 2 (1981). pp.349-362.

[20] **G. Wahba**, Ill posed problems: Numerical and statistical methods for mildly, moderately, and severely ill posed problems with noisy data, TR#595, Dept. of Statistics, U. of Wisconsin, 1980 (To appear in the Proceedings of the International Conference on Ill Posed Problems, M.Z. Nashed, ed.)

[21] **G. Wahba**, Surface fitting with scattered, noisy data on Euclidean d-spaces and on the sphere, Rocky Mountain Journal of Mathematics (1983). (To appear)

[22] **G. Wahba**, Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data, Approximation Theory III, W. Cheney, ed., Academic Press, 1980, pp.905-912.

[23] **G. Wahba**, Constrained regularization for ill posed linear operator equations, with applications in meteorology and medicine, Statistical Decision Theory and Related Topics III, Vol. 2, S.S. Gupta and J.O. Berger, ed., Academic Press, 1982, pp.383-418.

[24] **G. Wahba and J. Wendelberger**, Some new mathematical methods for variational objective analysis using splines and cross-validation, Monthly Weather Review, 108 (1980). pp.36-57.

[25] **W. E. Wecker and C. F. Ansley**, The signal extraction approach to non-linear regression and spline smoothing, J. Amer. Statist. Ass., 78 (1983). pp.81-89.

[26] **J. Wendelberger**, The computation of Laplacian smoothing splines with examples, TR#648, Dept. of Statistics, U. of Wisconsin, 1981 .

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. 715
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): A TRUNCATED SINGULAR VALUE DECOMPOSITION AND OTHER METHODS FOR GENERALIZED CROSS-VALIDATION
5. TYPE OF REPORT & PERIOD COVERED: Scientific Interim
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Douglas M. Bates, Grace Wahba
8. CONTRACT OR GRANT NUMBER(s): ONR N00014-77-C-0675
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, University of Wisconsin, 1210 W. Dayton St., Madison, WI 53706
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, 800 N. Quincy St., Arlington, VA
12. REPORT DATE: August 1983
13. NUMBER OF PAGES: 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Generalized Cross-Validation (GCV) is an effective method of choosing the smoothing or regularization parameter in data smoothing problems. Unfortunately, the computing cost for the method can be high-particularly with large data sets. One of the most costly parts of this computation is taking the Singular Value Decomposition of a large, generally ill-conditioned, matrix. Since the smaller singular values are not important in further steps, we present a method of determining the larger singular values by a Truncated Singular Value Decomposition (TSVD). Other methods are also presented to use non-iterative matrix decompositions whenever possible and to economize on storage. Applications to the solution of Fredholm integral

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

equations of the first kind and to the computation of thin plate smoothing splines are given.