BIT 24 (1984), 467–472

A NOTE ON THE COMPUTATION OF THE GENERALIZED CROSS-VALIDATION FUNCTION FOR ILL-CONDITIONED LEAST SQUARES PROBLEMS

LARS ELDÉN

Department of Mathematics, Linköping University, S-581 83 Linköping, Sweden

Dedicated to Germund Dahlquist,
on the occasion of his 60th birthday.

1. Introduction.

Extremely ill-conditoned least squares problems arise in many scientific and statistical applications. Let $\|\cdot\|$ denote the Euclidean vector norm, and consider the least squares problem

$$\min_f \|Kf - g\|, \tag{1.1}$$

where the $m \times n$ matrix K, $m \geq n$, is assumed to be very ill-conditoned. To alleviate the ill-conditioning (1.1) is replaced by

$$\min_f \{\|Kf - g\|^2 + \mu^2 \|f\|^2\}. \tag{1.2}$$

This is the *regularization method* of Tikhonov [11] and Phillips [10]; in the statistical literature the solution of (1.2) is called the *ridge estimate* [7].

In some cases it is necessary to choose a regularization of the form

$$\min_f \{\|Kf - g\|^2 + \mu^2 \|Lf\|^2\},$$

where L is equal to a discretization of a differentiation operator. This can be transformed into a problem of the type (1.2) [4].

It is a non-trivial matter to choose a suitable value of the *regularization parameter* μ, which controls the degree of smoothness of the solution. Several methods have been proposed, which explicitly or implicitly correspond to the

Received January 1984. Revised June 1984.

addition of some a priori information into the problem formulation. In contrast, *generalized cross-validation* [7] is a method for choosing a value of the regularization parameter from the data. In this method μ is chosen to minimize

$$V(\mu) = \frac{m^{-1}\|Kf_\mu - g\|^2}{(m^{-1}\mathrm{tr}(I - KM_\mu^{-1}K^T))^2} \tag{1.3a}$$

where f_μ is the solution of (1.2), and

$$M_\mu = K^TK + \mu^2 I. \tag{1.3b}$$

In this procedure it is important to be able to compute function values $V(\mu)$ efficiently. The standard method is to use the *singular value decomposition* (SVD) of K [7].

In this note a modification is described, where instead of computing the SVD of K we use a bidiagonalization

$$K = U\begin{pmatrix} B \\ 0 \end{pmatrix} V^T, \tag{1.4}$$

where U and V are $m \times m$ and $n \times n$ orthogonal matrices respectively, and B is $n \times n$ bidiagonal. In this way we avoid the main part of the work in the SVD algorithm. We first briefly review how to solve (1.2) using the bidiagonalization of K. Then we show that starting from the decomposition (1.4) $V(\mu)$ can be computed in $O(n)$ operations.

This note is similar in spirit to [4]. The algorithm for computing the trace-term in (1.3) was originally given by Grad, Zakrajšek [9] in a different context, see also [3].

2. Generalized cross-validation computations.

The decomposition (1.4) can be computed using a finite number of Householder transformations or Givens rotations [5]. This is the first part of the SVD-algorithm (see e.g. [2], [6]). Using (1.4) we see that (1.2) is equivalent to

$$\min_{\tilde{f}} \{\|B\tilde{f} - \tilde{g}\|^2 + \mu^2\|\tilde{f}\|^2\}, \tag{2.1a}$$

where

$$\begin{aligned} \tilde{f} &= V^Tf, \\ U^Tg &= \begin{pmatrix} \tilde{g} \\ \tilde{\tilde{g}} \end{pmatrix}, \end{aligned} \tag{2.1b}$$

and $\tilde{g}$ is an $n \times 1$ vector. Obviously it is not necessary to compute the matrix U explicitly. It is also seen that (2.1a) is equivalent to

$$\min_{\tilde{f}} \left\| \begin{pmatrix} B \\ \mu I \end{pmatrix} \tilde{f} - \begin{pmatrix} \tilde{g} \\ 0 \end{pmatrix} \right\|,$$

and this least squares problem can be solved efficiently using the algorithm given in [4], where a QR-decomposition is computed

$$Q^T \begin{pmatrix} B & \tilde{g} \\ \mu I & 0 \end{pmatrix} = \begin{pmatrix} B_\mu & g_1 \\ 0 & g_2 \end{pmatrix}. \tag{2.2}$$

Here B_μ is a bidiagonal matrix

$$B_\mu = \begin{pmatrix} \alpha_1 & \beta_1 & & & \\ & \alpha_2 & \beta_2 & & \\ & & \ddots & \ddots & \\ & & & \alpha_{n-1} & \beta_{n-1} \\ & & & & \alpha_n \end{pmatrix}.$$

The algorithm is based on Givens rotations. The solution of (2.1a) (which is denoted $\tilde{f}_\mu$) can then be obtained by solving $B_\mu \tilde{f} = g_1$. The total operation count for the solution of (2.1a) is $O(n)$ (cf. [4]).

It is now easy to compute the norm of the residual vector $\|Kf_\mu - g\|$. Using the decomposition (1.4) we get

$$\|Kf_\mu - g\|^2 = \|B\tilde{f}_\mu - \tilde{g}\|^2 + \|\tilde{\tilde{g}}\|^2. \tag{2.3}$$

Given $\tilde{f}_\mu$ the computation of this quantity can be performed in $O(n)$ operations.

For small values of μ cancellation occurs in the computation of the norm of the residual vector (2.3), and there may be large errors. Cancellation can be avoided by basing the computations on the identity

$$B(B^TB + \mu^2 I)^{-1}B^T - I = -\mu^2(BB^T + \mu^2 I)^{-1};$$

(the identity is easily proved, e.g. using the SVD of B). Note that a decomposition of $BB^T + \mu^2 I$ can be obtained using a variant of the algorithm in [4].

It is somewhat more complicated to compute the trace-term in (1.3). Using the relation

$$\operatorname{tr}(BA) = \operatorname{tr}(AB), \tag{2.4}$$

we can simplify

$$\operatorname{tr}(I-KM_\mu^{-1}K^T) = m-\operatorname{tr}(KM_\mu^{-1}K^T)$$

$$= m-\operatorname{tr}(I-\mu^2M_\mu^{-1}) = m-n+\mu^2\operatorname{tr}M_\mu^{-1}.$$

Then from (1.4) and (2.4) we get

$$\operatorname{tr}(M_\mu^{-1}) = \operatorname{tr}(VB^TBV^T+\mu^2I)^{-1} = \operatorname{tr}(B^TB+\mu^2I)^{-1}.$$

A recursive procedure for computing the diagonal elements of the inverse of a band matrix can be obtained easily when the Cholesky decomposition of the matrix is known [8]. For completeness we here derive the procedure using a slightly different approach than that in [8].

From (2.2) it is seen that

$$B^TB+\mu^2I = B_\mu^TB_\mu,$$

and letting b_i^T denote the ith row of B_μ^{-1}, we obtain

$$\operatorname{tr}(M_\mu^{-1}) = \operatorname{tr}((B_\mu^TB_\mu)^{-1}) = \sum_{i=1}^{n}\|b_i\|^2.$$

Using the identity $B_\mu B_\mu^{-1} = I$ we get the relation

$$\alpha_n b_n = e_n,$$

$$\alpha_i b_i = e_i-\beta_i b_{i+1}, \qquad i = n-1, n-2, \ldots, 1,$$

where e_i denotes the ith unit vector. Since B_μ^{-1} is upper triangular b_{i+1} is orthogonal to e_i, and it follows that

$$\|b_n\|^2 = 1/\alpha_n^2,$$

$$\|b_i\|^2 = (1+\beta_i^2\|b_{i+1}\|^2)/\alpha_i^2, \qquad i = n-1, n-2, \ldots, 1. \tag{2.5}$$

Using the recursion (2.5) we can compute the trace term in (1.3) and thus also function values $V(\mu)$ in $O(n)$ operations.

It is also possible to compute the first derivative of $V(\mu)$ in $O(n)$ operations [3], [9].

In Table 1 we give the results of some timing experiments, where we compare our method to that based on the SVD. The test problem is an integral equation of the first kind given in [10], Example 1. The integral equation is discretized

using the trapezoidal rule giving an $n \times n$ matrix K. The algorithms were programmed in Fortran 77, and run on a DEC 20 computer under the TOPS 20 operating system. Single precision is used. The SVD subroutine is from the NAG library, and it computes only one of the orthogonal matrices. The bidiagonalization algorithm is a modification of the first part of the SVD subroutine in LINPACK [2]. Neither of the orthogonal matrices in (1.4) is computed explicitly.

Table 1 *Execution time in seconds for the two algorithms. The cross-validation function is evaluated for* 20 *different values of* μ.

Dimension	Execution time in seconds	
	SVD	Bidiagonalization
25	1.6	0.95
50	6.3	2.4
100	37	13

The SVD algorithm required approximately 0.95 iterations per singular value. This is in agreement with [1], where it is observed that for certain ill-conditioned matrices the SVD algorithm converges faster than for well-conditioned matrices.

Our tests indicate that for large problems the method based on the bidiagonalization is more efficient than that based on the SVD.

It should be remarked, however, that the modifications of the SVD algorithm described in [1] give a considerable speed-up, and therefore our method should be compared to that in [1]. So far we have not been able to pursue this.

Acknowledgement.

Helpful advice from Prof. Gene H. Golub is gratefully acknowledged.

REFERENCES

1. J. J. M. Cuppen, *The singular value decomposition in product form*, SIAM J. Sci. Stat. Comp. 4 (1983), 216–222.
2. J. J. Dongarra, J. R. Bunch, C. B. Moler, G. W. Stewart, *Linpack users' guide*, Soc. for Industrial and Appl. Math., Philadelphia, 1979.
3. L. Eldén, *Numerical methods for the regularization of Fredholm integral equations of the first kind*, Department of Mathematics, Linköping University, Report LiH-MAT-R-1974-7, 1974.
4. L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT 17 (1977), pp. 134–145.
5. G. H. Golub and W. Kahan, *Calculating the singular values and pseudoinverse of a matrix*, SIAM J. Numer. Anal. 2 (1965), 205–224.
6. G. H. Golub and C. Reinsch, *Singular value decomposition and least squares solutions*, Numer. Math. 14 (1970), pp. 403–420.

7. G. H. Golub, M. Heath and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics 21 (1979), pp. 215–223.
8. G. H. Golub and R. J. Plemmons, *Large-scale geodetic least-squares adjustment by dissection and orthogonal decomposition*, Lin. Alg. Appl. 34 (1980), 3–27.
9. J. Grad and E. Zakrajšek, *LR algorithm with Laguerre shifts for symmetric tridiagonal matrices*, Comput. J. 15 (1972), pp. 268–270.
10. D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. Assoc. Comput. Mach. 9 (1962), pp. 84–97.
11. A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Dokl. Akad. Nauk SSSR 151 (1963), pp. 501–504 = Soviet Math. Dokl. 4 (1963), pp. 1035–1038.