Journal of Computational and Applied Mathematics 235 (2010) 102–107

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Efficient algorithms for robust generalized cross-validation spline smoothing

Mark A. Lukas[a,*], Frank R. de Hoog[b], Robert S. Anderssen[b]

[a] Mathematics and Statistics, Murdoch University, South Street, Murdoch WA 6150, Australia
[b] CSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra ACT 2601, Australia

ARTICLE INFO

Article history:
Received 23 July 2009
Received in revised form 23 April 2010

MSC:
65F30
65D10
62G08

Keywords:
Cholesky decomposition
Generalized cross-validation
Smoothing matrix
Smoothing parameter
Spline
Trace

ABSTRACT

Generalized cross-validation (GCV) is a widely used parameter selection criterion for spline smoothing, but it can give poor results if the sample size n is not sufficiently large. An effective way to overcome this is to use the more stable criterion called robust GCV (RGCV). The main computational effort for the evaluation of the GCV score is the trace of the smoothing matrix, $\operatorname{tr} A$, while the RGCV score requires both $\operatorname{tr} A$ and $\operatorname{tr} A^2$. Since 1985, there has been an efficient $O(n)$ algorithm to compute $\operatorname{tr} A$. This paper develops two pairs of new $O(n)$ algorithms to compute $\operatorname{tr} A$ and $\operatorname{tr} A^2$, which allow the RGCV score to be calculated efficiently. The algorithms involve the differentiation of certain matrix functionals using banded Cholesky decomposition.

1. Introduction

A common task in data analysis is to fit a smooth curve to noisy data

$$y_i = f(x_i) + \varepsilon_i, \qquad a \le x_1 < x_2 < \cdots < x_n \le b, \quad i = 1, \ldots, n, \tag{1}$$

where it is assumed that the random errors ε_i are uncorrelated, with zero mean and equal variance σ^2. Smoothing splines are widely used for this purpose [1,2]. The natural polynomial smoothing spline of degree $2m-1$ can be defined as the minimizer f_λ of the functional

$$n^{-1}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int_a^b (f^{(m)}(x))^2 \mathrm{d}x \tag{2}$$

over all functions f for which $f^{(m)}$ is square integrable. The most frequently used smoothing spline is the cubic spline, for which $m = 2$.

It is well known that the quality of the fit depends critically on the choice of the smoothing parameter $\lambda > 0$. If λ is too small, then f_λ is too rough, and, if λ is too large, then f_λ is overly smooth and is not faithful to the data. In fact, as $\lambda \to \infty, f_\lambda$ approaches the least squares polynomial of degree $m-1$. A popular and practical parameter selection criterion is

* Corresponding author. Tel.: +61 893602423; fax: +61 893606332.
E-mail addresses: M.Lukas@murdoch.edu.au (M.A. Lukas), Frank.deHoog@csiro.au (F.R. de Hoog), Bob.Anderssen@csiro.au (R.S. Anderssen).

0377-0427/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2010.05.016

generalized cross-validation (GCV) [3,2]. This criterion usually performs well for problems with large sample size n. However, it can be unreliable for smaller n, and, even for large n, it occasionally gives a parameter value that is far too small. For this reason, two more stable extensions of GCV, called robust GCV (RGCV) [4–6] and modified GCV [7], were developed. The RGCV and modified GCV criteria have favourable small-sample and large-sample properties, and they perform very well in simulations [8,9].

Denote $\boldsymbol{f}_\lambda = (f_\lambda(x_1), \ldots, f_\lambda(x_n))^T$ and let $A(\lambda)$ be the smoothing matrix defined by $\boldsymbol{f}_\lambda = A(\lambda)\boldsymbol{y}$. The main computational effort in using GCV is the calculation of the trace $\operatorname{tr} A(\lambda)$, which is often referred to as the degrees of freedom for the spline. For the representation of f_λ in [10,11], the smoothing matrix can be written in terms of the inverse of a certain banded matrix of bandwidth $2m + 1$. Using the band structure and Cholesky decomposition, Hutchinson and de Hoog [12] developed an efficient $O(m^2n)$ algorithm to compute the diagonal elements of $A(\lambda)$, and hence to find $\operatorname{tr} A(\lambda)$. The diagonal elements of $A(\lambda)$, called the leverage values, are also used to obtain confidence intervals for the spline estimate [13]. With the local support basis for f_λ in [14], the method in [15] yields another $O(m^2n)$ algorithm for the calculation of $\operatorname{tr} A(\lambda)$ and the leverage values (see also [16, Sect. 3.8.1]). There are also efficient $O(m^2n)$ algorithms for the GCV criterion based on QR factorization [17,18]. The modified GCV criterion requires the same calculations as GCV, and so the same $O(m^2n)$ algorithms can be used.

The aim of this paper is to develop and investigate efficient exact algorithms for the RGCV criterion. The RGCV score function requires the calculation of both $\operatorname{tr} A(\lambda)$ and $\operatorname{tr} A^2(\lambda)$. We develop two pairs of new $O(m^2n)$ algorithms to calculate these quantities. The algorithms use an approach involving the differentiation of certain matrix functionals [19], and are based on the Cholesky decomposition of a banded matrix. One of the algorithms for $\operatorname{tr} A(\lambda)$ is similar to the algorithm in [12].

In addition to exact methods, there are other methods that approximate $\operatorname{tr} A(\lambda)$ and $\operatorname{tr} A^2(\lambda)$. In particular, using the known asymptotic behaviour of the eigenvalues τ_i in the Demmler–Reinsch diagonalization $A(\lambda) = Q\operatorname{diag}(1 + \lambda\tau_i)^{-1}Q^T$, asymptotic estimates of both $\operatorname{tr} A(\lambda)$ and $\operatorname{tr} A^2(\lambda)$ can be obtained [20]. The estimate of $\operatorname{tr} A(\lambda)$ was used in [21] to derive an asymptotic GCV selection criterion. By using the asymptotic estimate of $\operatorname{tr} A^2(\lambda)$, one can also derive an asymptotic RGCV criterion. A different approach is to use a stochastic estimator of $\operatorname{tr} A(\lambda)$ and so approximate the GCV score [22,23]. This can be extended easily to estimate $\operatorname{tr} A^2(\lambda)$ with little extra effort (since, having estimated $\operatorname{tr} A(\lambda)$ by $\boldsymbol{u}^T A\boldsymbol{u}$ for a pseudo-random vector $\boldsymbol{u}$, then $\operatorname{tr} A^2(\lambda)$ can be estimated as $\|A\boldsymbol{u}\|^2$).

Besides its use in the RGCV criterion, the function $\operatorname{tr} A^2(\lambda)$ also arises in the variance estimate [24]

$$\hat{\sigma}^2 = \|(I - A(\lambda))\boldsymbol{y}\|^2 / \operatorname{tr}((I - A(\lambda))^2),$$

where $A(\lambda)$ is the smoothing matrix above. Therefore, the algorithms developed here also apply to the calculation of $\hat{\sigma}^2$. The three quantities $\operatorname{tr} A$, $\operatorname{tr} A^2$ and $\operatorname{tr}(2A - A^2)$, all have useful interpretations as degrees of freedom [25]. Note that these quantities coincide for parametric linear regression (since $A^2 = A$ in this situation).

The paper is organized as follows. After some preliminaries in Section 2, the first pair of algorithms for $\operatorname{tr} A(\lambda)$ and $\operatorname{tr} A^2(\lambda)$ is developed in Section 3. These algorithms use an approach that yields the diagonals of $A(\lambda)$ and $A^2(\lambda)$. The second pair of algorithms, based on log determinant relationships for $\operatorname{tr} A(\lambda)$ and $\operatorname{tr} A^2(\lambda)$, is developed in Section 4. We compare the efficiencies of the algorithms in Section 5.

2. Preliminaries

The smoothing spline f_λ can be computed efficiently using a local support spline basis for $f_\lambda^{(m)}$ [26,10,11], that is, with the representation

$$f_\lambda^{(m)} = \sum_{i=1}^{n-m} c_i M_i,$$

where the M_i are B-splines. Because of the continuity conditions at the knots, the spline f_λ is uniquely determined by the coefficients $\boldsymbol{c} = (c_1, \ldots, c_{n-m})^T$ and values $\boldsymbol{a} = (f_\lambda(x_1), \ldots, f_\lambda(x_n))^T$. These coefficients and values can be computed by solving

$$(H + n\lambda G^T G)\boldsymbol{c} = G^T\boldsymbol{y}, \tag{3}$$

$$\boldsymbol{a} = \boldsymbol{y} - n\lambda G\boldsymbol{c}, \tag{4}$$

where H and G^TG are symmetric, positive definite band matrices of bandwidth $2m - 1$ and $2m + 1$, respectively. The $(n - m) \times (n - m)$ matrix H has elements

$$h_{ij} = \int_a^b M_i(x)M_j(x)\mathrm{d}x,$$

and the $n \times (n - m)$ matrix G is an $(m + 1)$-banded lower triangular matrix with the non-zero elements of column i equal to the coefficients of the mth divided difference based on $x_i, \ldots, x_{i+m}$.

Let $p = \lambda^{-1}$, and define $B = B(p) = nG^TG + pH$. Then, from (3) and (4), the smoothing matrix $A(\lambda)$, defined by $\boldsymbol{f}_\lambda = A(\lambda)\boldsymbol{y}$, satisfies

$$I - A(\lambda) = nGB^{-1}(p)G^T. \tag{5}$$

Note that $A(\lambda)$ is symmetric and $I - A(\lambda)$ is non-negative definite.