THE HAT MATRIX FOR SMOOTHING SPLINES

by

R. L. Eubank

Technical Report No. SMU/DS/TR-179
Department of Statistics ONR Contract

August, 1983

Research sponsored by the Office of Naval Research
Contract N00014-82-k-0207

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, TX 75275

THE HAT MATRIX FOR SMOOTHING SPLINES

by

R. L. Eubank[1]

Department of Statistics, Southern Methodist University, Dallas, Texas

Running head: Hat Matrix for Smoothing Splines

Abstract. The matrix which transforms the data vector to the vector of fitted values for smoothing splines is termed the hat matrix. This matrix is shown to have many of the same properties, and is seen to play the same role in the variances and covariances of the residuals, as its regression analysis counterpart. This fact is utilized to propose several possible diagnostic measures for use with smoothing splines. The extension of these results to include multivariate Laplacian smoothing splines is also indicated.

[1]Research supported in part by Office of Naval Research Contract N00014-82-K-0207.

AMS 1980 subject classification: Primary 62J99, 65D10; Secondary 65D07.

Key Words and Phrases: Diagnostics, Laplacian smoothing splines, leverage values, nonparametric regression, residuals, smoothing splines.

1. Introduction. In regression analysis the matrix which transforms the data vector, $\underline{y}$, to the vector of fitted values, $\hat{\underline{y}}$, is often called the hat matrix. The elements of this matrix appear in the variances and covariances of the residuals and, consequently, play a fundamental role in regression residual diagnostics. The objective of this paper is to show that an analogous matrix which arises in spline smoothing possesses many of the same properties and is, potentially, as important to the development of diagnostic measures as its regression analysis counterpart.

Consider the situation where a vector of responses $\underline{y} = (y(t_1),\ldots$ $\ldots,y(t_n))'$ follows a model of the form

$$\underline{y} = \underline{\eta} + \underline{\varepsilon} \tag{1}$$

for $\underline{\eta} = (\eta(t_1),\ldots,\eta(t_n))'$ a vector of values corresponding to some unknown response function, η, and $\underline{\varepsilon} = (\varepsilon(t_1),\ldots\varepsilon(t_n))'$ a vector of zero mean uncorrelated errors having common variance σ^2. The "time points" $t_1,\ldots t_n$ will be assumed throughout to satisfy $a < t_1 < \cdots$ $\cdots < t_n < b$ for a and b finite constants.

In some instances it is possible to assume that $\underline{\eta}$ satisfies

$$\underline{\eta} = X\underline{\beta} \tag{2}$$

where X is a known $n\times p$ matrix of rank p and $\underline{\beta}$ is a vector of unknown parameters. When (2) holds (1) is a standard linear model and the form of least squares parameter estimates and predicted values for this case are well known. For example, the vector of fitted values $\hat{\underline{y}} = (\hat{y}(t_1),$ $\ldots,\hat{y}(t_n))'$, and residual vector, $\underline{e} = (e(t_1),\ldots,e(t_n))'$, are given by

$$\hat{\underline{y}} = H\underline{y} \tag{3}$$

and

$$\underline{e} = (I-H)\underline{y} \tag{4}$$

where $H = X(X'X)^{-1}X'$. The matrix H in (3) and (4) is a projection operator (i.e., $H^2 = H$) that is frequently called the hat matrix. Using properties of projections it is shown in Hoaglin and Welsch(1978) that the elements of $H = \{h_{ij}\}$ satisfy

$$0 \leq h_{ii} \leq 1, \tag{5}$$

$$-1 \leq h_{ij} \leq 1 \quad \text{for } i \neq j, \tag{6}$$

with

$$h_{ii} = 1 \quad \text{iff } h_{ij} = 0 \text{ for all } i \neq j. \tag{7}$$

The "if" portion of (7) is true provided the model includes a constant term in which case we have $\sum_{j=1}^{n} h_{ij} = 1$.

We see from (3) that h_{ij} tells us the influence $y(t_j)$ has on the prediction of $y(t_i)$. Because of this the elements of H are frequently examined to aid in the detection of sensitive points among the values of the predictor variables. In particular, one usually examines the diagonal elements h_{ii}, $i = 1,\ldots,n$, for nearness to the bounds 0 and 1. The h_{ii}'s are known as leverage values since they indicate how much influence, or leverage, $y(t_i)$ has in its own prediction. The elements of H also appear in the variances and covariances of the residuals since, from (4),

$$\mathrm{Var}(\underline{e}) = \sigma^2(I-H). \tag{8}$$

As a result of (8), most residual diagnostics involve the leverage values in some fashion.

Of course, in general, $\underline{\eta}$ may not admit a parametric form such as (2). Therefore, it may be necessary to employ some nonparametric

procedure to estimate the function η. If the response function is believed to be smooth, then one choice is to estimate η by the function minimizing the penalized least squares criterion

$$\frac{1}{n}\sum_{j=1}^{n}(y(t_j) - f(t_j))^2 + \lambda\int_a^b\left\{\frac{d^m}{dt^m}f(t)\right\}^2 dt\ ,\ \lambda > 0, \tag{9}$$

over all functions f having m-1 absolutely continuous derivatives and a square integrable mth derivative on [a,b]. For any given λ in $(0,\infty)$ and $n \geq m$ the solution to this problem, $\hat{\eta}_\lambda$ say, is well known (see e.g. Schoenberg(1964)) to be a polynomial spline of order 2m with knots at $t_1,\ldots,t_n$ that is usually called a smoothing spline. The parameter λ governs the amount of smoothing with the extreme values $\lambda = 0$ and ∞ corresponding to a version of spline interpolation and regression on polynomials of order m, respectively. An effective method for the estimation of λ from data can be found in Craven and Wahba(1978). For a discussion of smoothing splines which focuses on their properties as nonparametric regression estimators see Wahba's discussion to Stone(1977).

The smoothing spline is a linear estimator and, hence, the vector of fitted values $\hat{\underline{y}}_\lambda = (\hat{\eta}_\lambda(t_1),\ldots,\hat{\eta}_\lambda(t_n))'$ can be written as

$$\hat{\underline{y}}_\lambda = H(\lambda)\underline{y}. \tag{10}$$

By analogy with the regression case we will call $H(\lambda)$ the <u>hat matrix</u> and its diagonal elements, $h_{ii}(\lambda)$, $i = 1,\ldots,n$, <u>leverage values</u>. The elements of $H(\lambda)$ have the same interpretation as those for H in (3). However, the question arises as to which, if any, of properties (5) - (8) hold for $H(\lambda)$. It should be noted that such conclusions are no longer obvious since $H(\lambda)$ is not a projection operator. However, as shown in

Section 2, properties (5) - (7) still hold as well as several others which closely link $H(\lambda)$ to the hat matrix from polynomial regression. Then, in Section 3 we show that, if properly interpreted, the residuals from a smoothing spline fit satisfy an analog of (8). This allows us to propose some possible smoothing spline residual diagnostics which parallel those utilized in regression analysis. Finally, in Section 4, we sketch the extension of results in Sections 2 and 3 to the multivariate setting.

2. Properties of $H(\lambda)$. In this section several of the basic properties of the hat matrix in (10) are derived. To accomplish this we will require an explicit form for $H(\lambda)$ which can be found, for example, in Wahba (1978). Therefore, let T be the $n \times m$ matrix with ijth element t_i^j, $i = 1,\ldots,n$, $j = 0,\ldots,m-1$, and define

$$H^* = T(T'T)^{-1}T'. \tag{11}$$

The matrix H^* is recognized as the hat matrix for polynomial regression and, consequently, $I-H^*$ is idempotent and can be written as

$$I-H^* = UU' \tag{12}$$

where U is an $n \times (n-m)$ matrix which satisfies

$$U'T = 0_{(n-m)\times m} \tag{13}$$

and

$$U'U = I_{(n-m)\times(n-m)}. \tag{14}$$

Now define the "covariance kernel"

$$Q(s,t) = \int_a^s \frac{(s-u)^{m-1}(t-u)^{m-1}}{(m-1)!^2}\, du, \quad s \le t, \tag{15}$$

and let Q_n be the $n\times n$ matrix with ijth entry $Q(t_i,t_j)$. Then, if Γ denotes the $(n-m)\times(n-m)$ matrix of eigenvectors and $d_1,\ldots,d_{n-m}$ are the corresponding eigenvalues for the positive definite matrix $U'Q_nU$, the hat matrix $H(\lambda)$ can be expressed as

$$I-H(\lambda) = U\Gamma D(\lambda)\Gamma'U' \tag{16}$$

where $D(\lambda)$ is the diagonal matrix

$$D(\lambda) = \mathrm{diag}((1 + d_1/n\lambda)^{-1},\ldots,(1 + d_{n-m}/n\lambda)^{-1}).$$

With these preliminaries we now establish the following theorem.

<u>Theorem 1</u>. <u>The hat matrix</u> $H(\lambda) = \{h_{ij}(\lambda)\}$ <u>satisfies</u>

$$0 \le h_{ii}(\lambda) \le 1, \tag{17}$$

$$-1 \le h_{ij}(\lambda) \le 1 \quad \underline{\text{for}}\ i \ne j, \tag{18}$$

$$h_{ii}(\lambda) = 1 \quad \underline{\text{iff}}\ h_{ij}(\lambda) = 0\ \underline{\text{for all}}\ i \ne j, \tag{19}$$

<u>and</u> $\sum_{j=1}^{n} h_{ij}(\lambda) = 1$. <u>It is closely related to</u> $H^* = \{h^*_{ij}\}$ <u>in the sense that</u>

$$h_{ii}(\lambda) \downarrow h^*_{ii}\ \underline{\text{as}}\ \lambda \to \infty,\ \underline{\text{if}}\ h^*_{ii} \ne 1. \tag{20}$$

<u>Furthermore</u>, $h_{ij}(\lambda) \to h^*_{ij}$ <u>as</u> $\lambda \to \infty$ <u>and</u>, <u>for</u> λ <u>sufficiently large with</u> $h^*_{ij} \ne 0$, <u>both</u> $h_{ij}(\lambda)$ <u>and</u> h^*_{ij} <u>have the same sign</u>. <u>If instead</u> $\lambda \to 0$ <u>then</u> $h_{ii}(\lambda) \to 0$ <u>and</u>, <u>provided</u> $h^*_{ii} \ne 1$, $h_{ii}(\lambda) \uparrow 1$.

<u>Proof</u>. Set $B = U\Gamma$ and observe that $I-H^* = BB'$. Thus if $h^*_{ii} = 1$ then $b_{ij} = 0$, $j = 1,\ldots,n-m$. Since

$$1-h_{ii}(\lambda) = \sum_{j=1}^{n-m} b_{ij}^2\ (1 + d_j/n\lambda)^{-1} \tag{21}$$

and

$$h_{ij}(\lambda) = -\sum_{r=1}^{n-m} b_{ir}\ b_{jr}(1 + d_r/n\lambda)^{-1}, \tag{22}$$

we see that (17) - (18) hold trivially in this case. If $h^*_{ii} \neq 1$ then $b_{ij} \neq 0$ for some j and it is clear from (21) that $1-h_{ii}(\lambda)$ is a monotonically increasing function with limit $\sum_{j=1}^{n-m} b^*_{ij} = 1-h^*_{ii}$. Thus, $1-(1-h_{ii}(\lambda))$ is monotonically decreasing with limit h^*_{ii} which proves (20). As h^*_{ii} satisfies (5) this establishes (17) as well.

To verify (18) apply the Cauchy-Schwarz inequality in (22) to obtain $|h_{ij}(\lambda)| \leq \max_{1\leq \ell \leq n}(1-h_{\ell\ell}(\lambda)) \leq 1$. The fact that $h_{ij}(\lambda) \to h^*_{ij}$ as $\lambda \to \infty$ follows immediately from (22) and, by letting $\lambda \to 0$ in (21) - (22), the last statement of the theorem is established. Further, since $h_{ii}(\lambda) = 1$ can occur only when $\lambda = 0$ or $b_{ir} = 0$, $r = 1,\ldots,n-m$, the direct implication in (19) is seen to hold. The converse is obtained by noting that, due to (13), $[I-H(\lambda)]\underline{1} = \underline{0}$, where $\underline{1}$ denotes an $n\times 1$ vector of all unit elements, so that $\sum_{j=1}^{n} h_{ij}(\lambda) = 1$.

Finally, note that $\lim_{\lambda\to\infty} h_{ij}(\lambda) = \lim_{\gamma\to 0} h_{ij}(1/\gamma)$ and that $h_{ij}(1/\gamma)$ is continuous and nonvanishing at $\gamma = 0$ provided $h^*_{ij} \neq 0$. Consequently, well known results regarding the persistance of sign for continuous functions have the implication that $h_{ij}(\lambda)$ and h^*_{ij} must have the same sign for λ sufficiently large.$\square$

As a result of Theorem 1 we can utilize the elements of the hat matrix as a diagnostic tool exactly as in the regression setting. When the value of $h_{ii}(\lambda)$ is near one we see, as a result of (17) - (19) and (21) - (22), that the smoothing spline essentially interpolates $y(t_i)$ and, hence, is pulled toward this data point regardless of its value. Thus points of high leverage should be located and considered as candidates for deletion if the corresponding response value is discrepant. This raises the

question of what is large for a smoothing spline leverage value. One benchmark for comparison is their average value $\Sigma_{i=1}^{n} h_{ii}(\lambda)/n$. In practice, one may also wish to examine a stem-and-leaf plot of the $h_{ii}(\lambda)$, $i = 1,\ldots,n$, to ascertain if some values have strayed from the pack. In the regression setting, Huber(1983) considers points of high leverage as those which exceed a bound he suggests placing somewhere between .2 and .5. This is probably too low for smoothing splines, since $h_{ii}(\lambda)$ majorizes h^*_{ii}, but may still be of some use for comparison purposes.

3. Some Possible Diagnostic Measures. The development of residual diagnostics for regression models such as (1) - (2) is, and has been, an area of considerable research interest. In contrast, diagnostic methods for smoothing splines have, as yet, to receive appreciable attention. The need for such procedures is equally important as in the regression setting since, as might be anticipated from the penalized least squares criterion (9), a smoothing spline fit can be drastically influenced by outlying data values. In this section we suggest some possible diagnostic measures patterned after those employed in regression modeling. Therefore, in order to motivate subsequent developments, let us first discuss some common measures utilized for residual analysis when (2) holds.

If $\underline{\eta}$ satisfies (2) then, from (8), $\mathrm{Var}\,(e(t_i)) = \sigma^2(1-h_{ii})$. Consequently, a standard measure of how well $y(t_i)$ conforms to the fitted model is its studentized residual

$$\tau_i = e(t_i)/\hat{\sigma}(1-h_{ii})^{1/2} \tag{23}$$

where $\hat{\sigma}^2 = \Sigma_{j=1}^{n} e(t_i)^2/(n-p)$. These values are usually compared to those of a Student's t with n-p degrees of freedom. Studentized residuals are not entirely satisfactory as a diagnostic tool, however, and other measures which are often more effective are based on deleted residuals. The ith deleted residual, $e^{[i]}(t_i)$, is obtained as the error from predicting $y(t_i)$ without using its value to fit the model and is known to satisfy

$$e^{[i]}(t_i) = e(t_i)/(1-h_{ii}). \tag{24}$$

A commonly used diagnostic measure obtained using $e^{[i]}(t_i)$ is the studentized deleted residual

$$\tau_i^{[i]} = e(t_i)/\hat{\sigma}^{[i]}(1-h_{ii})^{1/2} \tag{25}$$

where $\hat{\sigma}^{[i]}$ is our previous estimator of σ computed using the "data set" $y(t_1),\ldots,y(t_{i-1}),y(t_{i+1}),\ldots,y(t_n)$. Under the assumption of normally distributed errors $\tau_i^{[i]}$ is known to have a t-distribution with n-p-1 degrees of freedom. For a discussion of how these and other related measures are utilized in regression modeling the reader is referred to Gunst and Mason(1980, Chap. 7).

To justify using diagnostics similar to (23) and (25) for smoothing splines, an analog of (8) must first be shown to hold. This requires that we place an alternative interpretation on $\hat{\eta}_\lambda$ than that which stems from (9). Thus, consider the model

$$y(t) = \Sigma_{j=0}^{m-1} \alpha_j t^j + \sigma_S Z(t) + \varepsilon(t) \ , \ t \in [a,b], \tag{26}$$

where σ_S is a positive scale parameter, Z is a zero mean process with covariance kernel (15) and ε is a white noise process that is uncorrelated with $\{Z(t);t \in [a,b]\}$. It was shown by Kimeldorf and Wahba(1970)

that the best linear unbiased predictor of y(t) based on data sampled from this model is a smoothing spline with $n\lambda = \sigma^2/\sigma^2_S$. The use of $\hat{\eta}_\lambda$ to estimate η can, therefore, be viewed alternatively as approximating η by a polynomial and modeling the remainder, $\eta(t)-\Sigma_{j=0}^{m-1} \alpha_j t^j$, as a stochastic process. As noted by Wecker and Ansley(1983), this approach is commonly utilized in time series analysis where the nature of the true response function is not well understood or adequately described by a low order polynomial. They discuss why it should also be effective for nonparametric regression analysis.

Consider now the residual vector $\underline{e}_\lambda = (e_\lambda(t_1),\ldots,e_\lambda(t_n))'$ which is given by

$$\underline{e}_\lambda = (I-H(\lambda))\underline{y}.$$

If $\underline{y}$ is viewed as a sample from (26) then, using (16) and the fact that $\Gamma'U'(Q_n+n\lambda I)U\Gamma = \text{diag}\ (d_1+n\lambda,\ldots,d_{n-m}+n\lambda)$, we obtain

$$\text{Var}(\underline{e}_\lambda) = \sigma^2(I-H(\lambda)) \tag{27}$$

which parallels (8). This suggests measuring the size of a residual by use of the "<u>studentized residual</u>"

$$\tau_{i,\lambda} = e_\lambda(t_i)/\hat{\sigma}_\lambda(1-h_{ii}(\lambda))^{1/2} \tag{28}$$

where $\hat{\sigma}_\lambda$ is an estimator of σ. Some possible choices for $\hat{\sigma}_\lambda$ can be found in Wahba(1981b) and Wecker and Ansley(1983). In particular, Wahba(1981b) proposes the use of

$$\hat{\sigma}^2_\lambda = \Sigma_{j=1}^n e_\lambda(t_j)^2/\text{tr}(I-H(\lambda)) \tag{29}$$

where tr denotes the matrix trace. Consequently, the value of $\tau_{i,\lambda}$ might be compared to tabulated values for a Student's t distribution with approximately tr(I-H(λ)) degrees of freedom. As

$$\text{tr}(I-H(\lambda)) = \Sigma_{j=1}^n(1-h_{jj}(\lambda)) < \Sigma_{j=1}^n(1-h^*_{jj}) = \text{tr}(I-H^*) = n-p$$

this seems a more conservative choice than the rank of $I-H(\lambda)$ (which is n-p) since one feels, intuitively, that fewer error degrees of freedom should be available from a smoothing spline fit than for polynomial regression.

Craven and Wahba(1978) showed that deleted residuals for smoothing splines satisfy

$$e_{\lambda}^{[i]}(t_i) = e_{\lambda}(t_i)/(1-h_{ii}(\lambda)) \tag{30}$$

and, hence, a diagnostic measure which may be more sensitive than (28) is the "<u>studentized deleted residual</u>"

$$\tau_{i,\lambda}^{[i]} = e_{\lambda}^{[i]}(t_i)/\hat{\sigma}_{\lambda}^{[i]}(1-h_{ii}(\lambda))^{1/2} \tag{31}$$

where $\hat{\sigma}_{\lambda}^{[i]}$ is the estimator of σ obtained without using $y(t_i)$ to fit the spline. If $H^{[i]}(\lambda)$ denotes the hat matrix constructed from $t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_n$ then the value of $\tau_{i,\lambda}^{[i]}$ might be compared to values from a t-distribution with $tr(I-H^{[i]}(\lambda))$ " degrees of freedom". The computation of studentized deleted residuals is aided by the fact that

$$tr(H^{[i]}(\lambda)) = \left\{ \sum_{\substack{j=1 \\ j\neq i}}^{n} h_{jj}(\lambda) + \frac{h_{ji}(\lambda)^2}{1-h_{ii}(\lambda)} \right\} \tag{32}$$

and

$$\hat{\sigma}^{[i]} = \left\{ \sum_{\substack{j=1 \\ j\neq i}}^{n} (e_{\lambda}(t_j) + h_{ji}(\lambda)\ e_{\lambda}^{[i]}(t_i))^2/tr(I-H^{[i]}(\lambda)) \right\}^{\frac{1}{2}} \tag{33}$$

which can be established from (30) and Lemma 3.1 of Craven and Wahba(1979). Equations (32) - (33) have the implication that $\tau_{i,\lambda}^{[i]}$ may be computed without actually deleting the ith observation and refitting the smoothing spline which parallels results which hold in the regression setting (c.f. Hoaglin and Welsch (1978)).

The routine examination of leverage values, studentized residuals and studentized deleted residuals appears to be an effective method of guarding against data points that are overly influential in smoothing spline fits. Furthermore, by using the framework set out in this section various other diagnostic measures, which may provide additional information for this purpose, can be developed. For example, one possible definition of a "Cook's distance measure" is $D_i = h_{ii}(\lambda)\tau^2_{i,\lambda}/[(1-h_{ii}(\lambda))\mathrm{tr}H(\lambda)]$ (c.f. Gunst and Mason(1980)).

4. Multivariate Smoothing Splines. In this section we sketch the extension of results in Sections 2 and 3 to the case of more than one independent variable. Our attention will be restricted to multivariate Laplacian smoothing splines.

Suppose now that we have data $(y_1,\underline{t}_1),\ldots,(y_n,\underline{t}_n)$ where $\underline{t}_j = (x_{1j},\ldots,x_{dj})'$ is a $d\times 1$ vector of values for d independent variables $x_1,\ldots,x_d$. Then, if the data follows a model of the form (1) with η now representing a smooth function on Ω, some closed bounded subset of $\mathbb{R}^d$, and $\underline{t}_j \in \Omega$, $j = 1,\ldots,n$, a reasonable smoothing criterion is to estimate η by the minimizer of

$$\frac{1}{n}\sum_{j=1}^{n}(y_j-f(\underline{t}_j))^2 + \lambda \sum_{i_1,\ldots,i_m=1}^{d} \int_{\mathbb{R}^d}\left\{\frac{\partial^m f(\underline{t})}{\partial x_{i_1}\cdots\partial x_{i_m}}\right\}^2 dx_1\ldots dx_d.$$

Provided $m > d/2$, $n > \binom{m+d-1}{d}$ and under certain restrictions on the $\underline{t}_j$'s the solution to this problem, when formulated in the appropriate function space, is known as a Laplacian or "Thin Plate" smoothing spline (see Wahba (1979), Wahba and Wendelberger(1980) or Wendelberger(1981) for details). It is known (c.f. equation (2.3) of Wahba(1979)) that the hat matrix in this setting

admits a representation similar to (16) where H* now corresponds to regression analysis on polynomials in $x_1,\ldots,x_d$ of total order at most m. Consequently, Theorem 1 is readily seen to hold in this setting and, hence, the examination of leverage values should also prove to be a useful diagnostic tool for Laplacian smoothing splines. This may be particularly true for $d \geq 3$ since, in this case, outliers in the independent variables are no longer easily detected. In addition, it follows from the theory in Kimeldorf and Wahba(1970) that a Laplacian smoothing spline can be regarded as a best linear unbiased predictor for a model which parallels (26). As a result, the diagnostic measures discussed in Section 3 can be utilized for this type of multivariate spline as well. We point out that such conclusions can also be drawn for spline smoothing on the sphere as discussed in Wahba(1981a).

Acknowledgement. The author would like to thank Dick Gunst for several helpful conversations during the preparation of this manuscript.

REFERENCES

Craven, Peter and Wahba, Grace(1979). Smoothing noisy data with spline functions. Numer. Math. 31, 377-403.

Gunst, Richard F. and Mason, Robert L.(1980). Regression Analysis and Its Applications: a Data-Oriented Approach. Marcel Dekker, New York.

Hoaglin, David C. and Welsch, Roy E.(1978). The hat matrix in regression and ANOVA. Amer. Statist. 32, 17-22. Corrigenda, 32, 146.

Huber, Peter J.(1983). Minimax aspects of bounded-influence regression. J. Amer. Statist. Assoc. 78, 66-80.

Kimeldorf, George S. and Wahba, Grace(1970). Spline functions and stochastic processes. Sankhyā Ser. A 32, 173-180.

Schoenberg, I.J.(1964). Spline functions and the problem of graduation. Proc. Nat. Acad. Sci. USA 52, 947-950.

Stone, Charles J.(1977). Consistent nonparametric regression. Ann. Statist. 5, 595-645.

Wahba, Grace(1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. J. Roy. Statist. Soc. Ser. B 40, 364-372.

Wahba, Grace(1979). Convergence rates of "Thin Plate" splines when the data are noise. In Smoothing Techniques for Curve Estimation, T. Gasser and M. Rosenblatt, eds. Lecture Notes in Mathematics No. 757, 233-245, Springer-Verlag.

Wahba, Grace(1981a). Spline interpolation and smoothing on the sphere. SIAM J. Sci. Statist. Comput. 2, 5-16. Erratum, 3, 385-386.

Wahba, Grace(1981b). Bayesian confidence intervals for the cross validated smoothing spline. To appear J. Roy. Statist. Soc. Ser. B.

Wahba, Grace and Wendelberger, James(1980). Some new mathematical methods for variational objective analysis using splines and cross validation. Monthly Weather Review 108, 1122-1143.

Wecker, William E. and Ansley, Craig F.(1983). The signal extraction approach to nonlinear regression and spline smoothing. J. Amer. Statist. Assoc. 78, 81-89.

Wendelberger, James G.(1981). The computation of Laplacian smoothing splines with examples. Tech. Rep. No. 648, Dept. of Statist., Univ. of Wisconsin-Madison.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
SMU/DS/TR-179

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
THE HAT MATRIX FOR SMOOTHING SPLINES

5. TYPE OF REPORT & PERIOD COVERED
Technical Report

6. PERFORMING ORG. REPORT NUMBER
SMU/DS/TR-179

7. AUTHOR(s)
R. L. Eubank

8. CONTRACT OR GRANT NUMBER(s)
N00014-82-k-0207

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Southern Methodist University
Dallas, TX 75275

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 042-479

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Arlington, VA 22217

12. REPORT DATE
August, 1983

13. NUMBER OF PAGES
14

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any putpose of The United States Government.

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)
Diagnostics, Laplacian smoothing splines, leverage values, nonparametric regression, residuals, smoothing splines.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
The matrix which transforms the data vector to the vector of fitted values for smoothing splines is termed the hat matrix. This matrix is shown to have many of the same properties, and is seen to play the same role in the variances and covariances of the residuals, as its regression analysis counterpart. This fact is utilized to propose several possible diagnostic measures for use with smoothing splines. The extension of these results to include multivariate Laplacian smoothing splines is also indicated.

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)