University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, Wisconsin 53706

TECHNICAL REPORT NO. 557

CONVERGENCE RATES OF "THIN PLATE" SMOOTHING SPLINES WHEN THE DATA ARE NOISY.

Grace Wahba

This document has been approved for public release and sale; its distribution is unlimited.
CONVERGENCE RATES OF "THIN PLATE" SMOOTHING SPLINES WHEN THE DATA ARE NOISY
(Preliminary report)
Grace Wahba
Department of Statistics
University of Wisconsin
Madison, Wisconsin 53705 U.S.A.

Abstract

We study the use of "thin plate" smoothing splines for smoothing noisy d dimensional data. The model is

\[ z_i = u(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \]

where \( u \) is a real valued function on a closed, bounded subset \( \Omega \) of Euclidean d-space and the \( \epsilon_i \) are random variables satisfying \( E \epsilon_i = 0, E \epsilon_i \epsilon_j = \sigma^2, i \neq j, \epsilon_i \) independent. The \( z_i \) are observed. It is desired to estimate \( u \), given \( z_1, \ldots, z_n \). \( u \) is only assumed to be "smooth", more precisely we assume that \( u \) is in the Sobolev space \( H^m(\Omega) \) of functions with partial derivatives up to order \( m \) in \( L_2(\Omega) \), with \( m > d/2 \). \( u \) is estimated by \( u_{n,m}, \) the restriction to \( \Omega \) of \( \hat{u}_{n,m} \), where \( \hat{u}_{n,m} \) is the solution to:

Find \( \hat{u} \) (in an appropriate space of functions on \( \Omega \)) to minimize

\[ \frac{1}{n} \sum_{i=1}^{n} (u(t_i) - z_i)^2 + \frac{1}{n} \sum_{i=1}^{n} \int_{\Omega} \left( \sum_{k=1}^{m} \frac{x_k^2}{\lambda} \right)^2 dx_1 \ldots dx_d. \]

This minimization problem is known to have a solution for \( \lambda > 0 \), \( \min_{\lambda > 0} n^2 \text{tr}(\Lambda)^{-1} \), provided the \( t_1, \ldots, t_n \) are "unsolvent". We consider the integrated mean square error

\[ \text{RMSE}(\lambda) = \frac{1}{|\Omega|} \int_{\Omega} (u_{n,m}(t) - u(t))^2 dt, \quad |\Omega| = \int_{\Omega} dt, \]

and ER(\lambda), as \( \{t_i\}_{i=1}^{n} \) become dense in \( \Omega \). An estimate of \( \lambda \) which asymptotically minimizes ER(\lambda) can be obtained by the method of generalized cross-validation. In this paper we give plausible arguments and numerical evidence supporting the following conjectures:

Suppose \( u \in H^m(\Omega) \). Then

\[ \min_{\lambda} \text{ER}(\lambda) = O(n^{-2m/(2m+d)}). \]

Research supported by the Office of Naval Research under Grant No. N00014-77-C-0675.
1. Introduction

Consider the model
\[ z_i = u(t_i) + e_i, \quad i = 1, 2, \ldots, n \]  
(1.1)
where \( u \) is some "smooth" function on \( u \), a closed, bounded subset of \( \mathbb{R}^d \), and the \( (e_i) \) are independent, zero mean random variables with common unknown variance \( \sigma^2 \). The \( t_1, \ldots, t_n \) are in \( u \), and \( z = (z_1, \ldots, z_n) \) is observed. It is desired to estimate \( u \) nonparametrically from \( z \).

Our estimate \( \hat{u}_{n,m} \) for \( u \) will be obtained as follows:

Let \( \hat{u}_{n,m} \) be the solution to the following minimization problem: Find \( \hat{u} \in \mathbb{R}^n \) to minimize
\[
\frac{1}{n} \sum_{i=1}^{n} \left[ (\hat{u}(t_i) - z_i)^2 \right] + \frac{m}{n} \int \left[ \frac{1}{2} \sum_{k=1}^{d} \left( \frac{\partial^2 \hat{u}}{\partial t_k^2} \right)^2 \right] dt_1 \ldots dt_d.
\]

For example, when \( d=2, m=2 \), the second or "smoothness penalty" term becomes
\[
\int \left( \hat{u}_{t1}^2 + \hat{u}_{t2}^2 + \hat{u}_{t1}^2 + \hat{u}_{t2}^2 \right) dt_1 dt_2,
\]
which is the bending energy of a thin plate. The space \( \mathbb{R}^n \) is the "Beppo Levi" space \( \mathbb{R}^n = \{ u \in \mathbb{R}^n : \hat{u}_{t1}^2 + \hat{u}_{t2}^2 \in L^2([0,1]) \} \), for \( \forall t_1, \ldots, t_n \in [0,1] \).

where \( D^* \) is the dual of the Schwartz space \( D \) of infinitely differentiable functions with compact support. See Meinert (1978, 1979) for further details. \( u_{n,m} \) is taken as the restriction of \( \hat{u}_{n,m} \) to \( u \).

A unique (continuous) solution is known to exist for any \( \lambda > 0 \) provided
\[
\lambda > d/2,
\]
and the "design" \( t_1, \ldots, t_n \) is "unisolvent", that is, if \( (e_i)_{i=1}^N \) are a basis for the \( N \)-dimensional space of polynomials of total degree \( n-1 \) or less, then \( \sum_{i=1}^{n} e_i \phi_i(t_i) = 0 \), \( i = 1, 2, \ldots, n \), implies that the \( e_i \) are all 0. See Duchon (1976a, 1976b), Meinert (1978, 1979), Painleve (1977, 1978). He henceforth assume these conditions. Duchon has shown that the solution has a representation
\[
\hat{u}_{n,m}(t) = \sum_{j=1}^{n} c_j \{ E_j(\cdot, t) \} + \sum_{j=1}^{m} d_j \phi_j(t),
\]
where
\[
E_j(s,t) = e_j \{ s-t \}^{2m} \log|s-t| \quad \text{m even}
\[
= e_j \{ s-t \}^{2m} \quad \text{m odd}
\]
where, if \( s = (x_1, \ldots, x_d), t = (y_1, \ldots, y_d), \) \( |s-t| = \{ \sum_{j=1}^{d} (x_j-y_j)^2 \}^{1/2}, \)
\[
e_j = (-1)^{d/2+1}/(2^{2m-1} \Gamma(d/2-1)) \quad \text{m even}
\[
= (-1)^{d/2-1}/2^{2m} \Gamma(d/2-1)) \quad \text{m odd}.
\]
The coefficients \( c = (c_1, \ldots, c_m) \) and \( d = (d_1, \ldots, d_m) \) are determined by
\[
\{ \phi_j \} \cdot C \cdot T = z,
\]
\[
T = 0 .
\]
where \( K \) is the \( n \times n \) matrix with \( j^{th} \) entry \( E_j(t_j, t_j) \), \( \phi \) is the \( n \times m \) matrix with \( j^{th} \) entry \( \phi_j(t_j) \) and \( z = (z_1, \ldots, z_n) \). See Duchon (1976, 1977), Painleve (1977, 1978). Wahlba (1979). He discuss the choice of \( \lambda \) shortly.

Let \( u \) be a closed, bounded subset of \( \mathbb{R}^d \). We will suppose that the \( (t_i) \) become dense in \( u \) in such a way that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \rho(t_i) = \frac{1}{|u|} \int \rho(t) dt, \quad |u| = \int_0^1 \rho(t) dt
\]
for any continuous \( \rho \). (However, it will be clear that our results hold under weaker conditions on the distribution of the \( (t_i) \), for example
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \rho(t_i) = \int_0^1 \rho(t) w(t) dt
\]
for some sufficiently nice positive \( w \)). Let \( R(s) \) be the integrated mean square error when \( \lambda \) is used:
\[
R(s) = \frac{1}{n} \sum_{i=1}^{n} \{ \phi_j \}^T \{ \hat{u}_{n,m}(t_i) - u(t_i) \}^2 = \frac{1}{|u|} \int \{ \phi_j \}^T \{ \hat{u}_{n,m}(t) - u(t) \}^2 dt .
\]

The smoothing parameter \( \lambda \) which minimizes \( ER(\lambda) \) can be estimated by the method of generalized cross-validation (GCV), see Craven and Wahba (1979), Golub, Heath and Wahba (1977), Wahba (1979). Pleasing numerical results have been obtained in Monte Carlo studies for \( d=1, m=2 \) (Craven and Wahba (1979)) and\( d=2, m=2 \), Wahba (1979). Convergence rates for \( ER(\lambda) \) have been obtained in the one dimensional case (Wahba (1975)).

Stone (1978) has recently obtained some rather general results on best achievable pointwise convergence rates for the model (1.1), for any method of estimation of \( u(t) \). Reduced to our case and phrased loosely, his results say that
the rate
\[ E[u(t) - \tilde{u}(t)]^2 = O(u^{-(2m-1)/(2m-1+d)}) \],
where \( \tilde{u}(t) \) is any estimate of \( u(t) \) from the data \( z \), can be achieved for all \( u \in H_m(0) \) but not bettered. In this paper we are concerned with integrated mean square error convergence rates:

\[ E \int_0^1 (u_n(t) - u(t))^2 dt = O(1) \]

of \( u_n \).

It is our goal to give a plausible argument that

1) if \( u \in H^2(0) \),
\[ ER(1) = O(n^{-2m/(4m+d)}) \]

and if

2) if \( u \in H^2(0) \) and some other conditions are satisfied, then
\[ ER(1) = O(n^{-m/(2m+d)}) \]

Our argument follows the arguments given in Wahba (1975, 1977) and Craven and Wahba (1979), and is given in section 2.

2. Plausibility arguments, numerical evidence

Let \( A(x) \) be the \( n \times n \) matrix defined by
\[
\begin{pmatrix}
\tilde{u}_{h,m,n}(t_1) \\
\vdots \\
\tilde{u}_{h,m,n}(t_n)
\end{pmatrix} = A(x) z.
\]

If \( R(x) \) is taken as the middle quantity in (1.5), we have
\[ R(x) = \frac{1}{n} \| A(x)(x\hat{u}) - u \|^2 \]

where \( u = (u(t_1), \ldots, u(t_n))^T \), \( x = (x_1, \ldots, x_n)^T \), and
\[ ER(1) = \frac{1}{n} \| (I - A(x))u \|^2 + \frac{2}{n} \text{Trace } \lambda^2(x) \].

(A(x) is symmetric.)

We call \( \frac{1}{n} \| (I - A(x))u \|^2 \) the "squared bias" and \( \lambda^2(n) \) the variance.

Lemma 1.
\[ \frac{1}{n} \| (I - A(x))u \|^2 \leq \lambda J_u(\hat{u}) \]  

where, for \( \hat{u} \in \mathbb{R}^n \)
\[ J_u(\hat{u}) = \frac{d}{d y} \left[ \int_{y_1}^{y_2} \frac{s_y(\tilde{u}_{h,m,n}(t_1), \ldots, \tilde{u}_{h,m,n}(t_n))}{s_y(\tilde{u}_{h,m,n}(t_1), \ldots, \tilde{u}_{h,m,n}(t_n))} dt_1 \ldots dt_n \right] \]

and \( \hat{u} \) is the element in \( \mathbb{R}^n \) which minimizes \( J_u \) subject to coinciding with \( u \) on \( x \).

Proof.

For each \( i \), \( u(t_i) = \hat{u}(t_i) \). \( A(x)u \) is a vector of values of the function, call it \( \tilde{g}_{h,m,n} \) which is the solution to the problem: Find \( \tilde{v} \in \mathbb{R}^n \) to minimize
\[ \frac{1}{n} \sum_{j=1}^{n} (u(t_j) - \tilde{v}(t_j))^2 + \lambda J_u(\tilde{v}) \]

Therefore
\[ \frac{1}{n} \sum_{j=1}^{n} (u(t_j) - \tilde{g}_{h,m,n}(t_j))^2 + \lambda J_u(\tilde{g}_{h,m,n}) \]
\[ = \frac{1}{n} \| (I - A(x))u \|^2 + \lambda J_u(\tilde{u}) \]
\[ = \frac{1}{n} \| (I - A(x))u \|^2 + \lambda J_u(\hat{u}) \]
We now investigate Trace $A^2(\lambda)$. Let $\mathcal{T}_{n,M}$ be the $n \times M$ dimensional matrix with $j$th entry $a_{j}(t_j)$. Let $R$ be any $n \times n$ matrix of rank $n-M$ satisfying $R \cdot R^T = 0$. Following the results of Anselone and Laurent (1968) it is shown in Wahba (1979) that $c$ and $d$ satisfying $(1.2)$ and $(1.3)$ have the representations

$$c = R^T (R^T R + h^2) R^T$$

and that

$$(1 - A(t)) z = \rho c = n \alpha (R^T R + h^2)^{-1} R^T z , \quad z \in \mathcal{E}_n .$$  \hspace{1cm} (2.3)$$

Hence, if we define $B = R^T R$ and let $b_{m,v} = 1, 2, \ldots, n-M$ be the $m$th eigenvalues of $B$, then

$$\frac{1}{n} \text{Tr} A^2(\lambda) = \frac{1}{n} \sum_{m=1}^{n-M} \frac{b_{m,v}^2}{\sum_{v=1}^{n-M} \frac{1}{\sum_{v=1}^{n-M} (1 + \alpha / b_{m,v})^2}}.$$  \hspace{1cm} (2.4)$$

We remark that $K$ is not, in general, positive definite, however $R^T R$ is, since it is known that $r^T r > 0$ for any non-trivial $r$ satisfying $r^T r = 0$ (See Pefhau (1977), Duchon (1977)).

**Lemma 2.**

Suppose there exist $p_1$, and $\lambda_2 > 0$ such that

$$\lim_{\lambda \to \infty} \frac{b_{m,v}}{\lambda} \frac{b_{m,v} - \lambda_2}{\lambda} = 0$$

then, for some constant $\lambda_2$,

$$\frac{1}{n} \text{Tr} A^2(\lambda) \geq \frac{\lambda_2}{n \alpha^{1/2}} (1 + o(1)) .$$  \hspace{1cm} (2.4)$$

**Outline of Proof.**

$$\frac{1}{n} \text{Tr} A^2(\lambda) = \frac{1}{n} \lambda \sum_{m=1}^{n-M} \left(1 + \frac{h}{b_{m,v}}\right)^2 + \frac{1}{n} \lambda \sum_{m=1}^{n-M} \left(1 + \frac{h}{b_{m,v}}\right)^2$$

$$= \frac{1}{n} \sum_{m=1}^{n-M} \frac{\lambda}{(1 + \alpha / b_{m,v})^2} + \frac{\lambda}{n \alpha^{1/2}} (1 + o(1)) .$$

($A$ more rigorous argument can be found in Craven and Wahba 1979)).
are \( e_n \) and \( \beta_n \), so that the eigenvalues \( \lambda_n \) satisfy
\[
\sin \frac{\beta_n}{\beta_0} = \left( \frac{\beta_n}{\beta_0} \right) \frac{\sin \beta_0}{\beta_0}
\]
and this result is independent of the shape of \( D \). Going to \( d = 3 \) dimensions, the eigenvalues for \( \Delta = \omega_0 \) on a rectangle with sides \( a_1, a_2 \) and \( a_3 \) and suitable boundary conditions are
\[
\lambda_{n,m,p} = \left( \frac{\pi}{a_1} \right)^2 \left( \frac{\pi}{a_2} \right)^2 \left( \frac{\pi}{a_3} \right)^2 e_n = 1, 2, \ldots
\]
and, by counting the number of triplets \((n,m,p)\) in the ellipse
\[
v^2 \left( \frac{x^2}{a_1^2} + \frac{y^2}{a_2^2} + \frac{z^2}{a_3^2} \right) \leq c
\]
one obtains that
\[
\sin \frac{\beta_n}{\beta_0} = \left( \frac{\beta_n}{\beta_0} \right) \frac{\sin \beta_0}{\beta_0}
\]
or
\[
\frac{\beta_n}{\beta_0} = \sqrt{1 - \frac{c}{a_1^2 a_2^2 a_3^2}} \, (1 + o(1))
\]
See Courant and Hilbert (1953). Similarly the eigenvalues for \( \Delta^d \) satisfy
\[
\frac{\lambda_{n}}{n^2/3} = \left( \frac{\pi}{a_1} \right)^2 \left( \frac{\pi}{a_2} \right)^2 \left( \frac{\pi}{a_3} \right)^2 (1 + o(1))
\]
and, extending the argument to \( d \) dimensions gives
\[
\frac{\lambda_n}{n^{2d/3}} = \left( \frac{\pi}{a_1} \right)^2 \left( \frac{\pi}{a_2} \right)^2 \left( \frac{\pi}{a_3} \right)^2 (1 + o(1))
\]
where \( V_d \) is the volume of the sphere of radius 1 in \( d \) dimensions. Therefore, we conjecture that the rate of decrease of the eigenvalues \( \lambda_n \) of \( K \) is \( \sim n^{-d/2} \).

Let \( K(s,t) \) be a kernel with a Mercer-Hilbert Schmidt expansion on \( D \),
\[
K(s,t) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(s) \varphi_n(t)
\]
where the eigenvalues \( \lambda_n \) are absolutely summable and the eigenfunctions \( \varphi_n \) are an orthonormal set on \( L_2(D) \). Then, for large \( n \),
\[
K(t_j, t_k) \geq \frac{1}{n} \max \| \varphi_n(t_j) \| \| \varphi_n(t_k) \| ,
\]
and provided
\[
\int_0^1 [ \phi(t_j) \phi(t_k) ] \varphi_n(t) dt = 1, \quad x \neq y
\]
we see that the eigenvalues \( \lambda_{n,m} = 1, 2, \ldots, n \), say, of the matrix \( K \) with \( jk \)th entry
\[ y = \frac{1}{\lambda} \int_0^1 k_0 e^{\alpha_0} d^2 \]

where \( \rho = (\varrho(t_1), \ldots, \varrho(t_M)) \), \( \eta = (\eta_1, \ldots, \eta_M) \), and \( \delta^2 = (\delta^2_1, \ldots, \delta^2_M) \) is a vector of quadrature errors which we must assume are negligible in the limit. Similarly

\[ T^T \delta^2 = \delta^2, \]

where \( \delta^2 \) is a vector of quadrature errors which we must assume are negligible in the limit. Assuming \( T^T \delta = 0 \), then \( \delta = R_0 \eta \), for some \( n \times M \) vector \( \eta \).

Then

\[ R^T u = \frac{1}{\lambda} \int_0^1 R^T R \eta - R^T \eta + \text{negligible terms} \]

and

\[ n \left( \frac{1}{\lambda} \int_0^1 R(R^T + \alpha I)^{-1} R^T \right)^2 R^T u \]

\[ \leq n \left( \frac{1}{\lambda} \int_0^1 R(R^T)^{-1} R^T \right)^2 R^T u \]

\[ = \frac{1}{\lambda} \left( \| \eta_0 \| \right)^2 = \frac{1}{\lambda} \left( \| \eta \| \right)^2 = \frac{1}{\lambda} \left( \| \int_0^1 f^2(t) dt (1+o(1)) \right) . \]

**Theorem 2.**

Suppose

\[ u(t) = \int_0^1 \eta_0(s) \nu(s) ds + \sum_{\nu=1}^N \delta_\nu \eta(t) \]

for some \( \rho \) piecewise continuous with \( \int_0^1 \eta_0(s) \nu(s) ds = 0, \nu = 1, 2, \ldots, M \). Then (assuming the conclusions of lemmas 3 and 4),

\[ \min_\lambda R(x) = O(n^{-2m}/(4m \lambda)) . \]

**Proof.**

Using (2.4), (2.5) and (2.7) gives

\[ R(x) \leq k_3 \frac{1}{\lambda} \left( \| \eta_0 \| \right)^2 \]

where \( k_3 \) and \( k_4 \) are constants. Setting \( \lambda = O(n^{-2m}/(4m \lambda)) \) gives the result.
space \( H^m(\Omega) \) of functions with partial derivatives up to order \( m \) in \( L^2(\Omega) \), with \( m > d/2 \). \( u \) is estimated by \( u_{n,m,\lambda} \), the restriction to \( \Omega \) of \( \tilde{u}_{n,m,\lambda} \), where \( \tilde{u}_{n,m,\lambda} \) is the solution to: Find \( \tilde{u} \) (in an appropriate space of functions on \( \mathbb{R}^d \)) to minimize
\[
\frac{1}{n} \sum_{i=1}^{n} (u(t_i) - z_i)^2 + \lambda \sum_{i_1, \ldots, i_m=1}^{d} \frac{\partial^{m} u}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_m}} \left( \frac{2}{dx_{i_1} dx_{i_2} \ldots dx_{i_m}} \right)
\]
This minimization problem is known to have a solution for \( \lambda > 0, m > d/2, n > M(m+d-1), \) provided the \( t_1, \ldots, t_n \) are "unisolvent". We consider the integrated mean square error
\[
R(\lambda) = \frac{1}{|\Omega|} \int_{\Omega} \left( u_{n,m,\lambda}(t) - u(t) \right)^2 dt, \quad |\Omega| = \int_{\Omega} dt
\]
and \( ER(\lambda) \), as \( \{t_i\}_{i=1}^{n} \) become dense in \( \Omega \). An estimate of \( \lambda \) which asymptotically minimizes \( ER(\lambda) \) can be obtained by the method of generalized cross-validation. In this paper we give plausible arguments and numerical evidence supporting the following conjectures:
Suppose \( u \in H^m(\Omega) \). Then
\[
\min_{\lambda} ER(\lambda) = O(n^{-2m/(2m+d)}).
\]
Suppose \( u \in H^m(\Omega) \) and certain other conditions are satisfied. Then
\[
\min_{\lambda} ER(\lambda) = O(n^{-4m/(4m+d)}).
\]