Ann. Inst. Statist. Math.
Vol. 43, No. 1, 61–75 (1991)

A KERNEL APPROXIMATION TO THE KRIGING PREDICTOR OF A SPATIAL PROCESS*

MICHAEL L. STEIN

Department of Statistics, The University of Chicago, 5734 University Avenue, Chicago, IL 60637, U.S.A.

(Received September 6, 1988; revised February 13, 1990)

Abstract. Suppose a two-dimensional spatial process $z(x)$ with generalized covariance function $G(x, x') \propto |x - x'|^2 \log |x - x'|$ (Matheron, 1973, *Adv. in Appl. Probab.*, **5**, 439–468) is observed with error at a number of locations. This paper gives a kernel approximation to the optimal linear predictor, or kriging predictor, of $z(x)$ under this model as the observations get increasingly dense. The approximation is in terms of a Kelvin function which itself can be easily approximated by series expansions. This generalized covariance function is of particular interest because the predictions it yields are identical to an order 2 thin plate smoothing spline. For moderate sample sizes, the kernel approximation is seen to work very well when the observations are on a square grid and fairly well when the observations come from a uniform random sample.

Key words and phrases: Thin plate spline, prediction of random fields, Kelvin function, nonparametric regression.

1. Introduction

Universal kriging is the geostatistical term for best linear unbiased prediction under a class of nonstationary spatial processes known as intrinsic random functions (Matheron (1973)). This paper develops a kernel approximation for the universal kriging predictor under a particular intrinsic random function model in two dimensions that is appropriate as the number of observations near the point being predicted increases. This approximation is of interest in nonparametric regression because the universal kriging predictor for this process is equivalent to an order 2 thin plate smoothing spline. Silverman (1984, 1985) has shown that one-dimensional smoothing splines can be approximated by kernel smoothers as

* The author was supported by a National Science Foundation Mathematical Sciences Postdoctoral Research Fellowship.

This manuscript was prepared using computer facilities supported in part by National Science Foundation Grants No. DMS-8601732 and DMS-8404941 to the Department of Statistics at The University of Chicago.

the number of observations near the point being predicted increases. Cox (1983) and Cogburn and Davis (1974) have also obtained connections between spline and kernel smoothing in one dimension under more restrictive conditions. Cox (1984) gives approximations to smoothing splines in two dimensions in terms of the solution of a partial differential equation, but he does not give an explicit solution to this equation; moreover, his results do not apply to the thin plate spline. Unlike the one-dimensional case, where the matrices involved in computing smoothing splines are banded, the corresponding matrix needed to obtain the thin plate smoothing spline has no helpful special structure (Silverman (1984)). Thus, a kernel approximation, if it is sufficiently accurate, is especially valuable in this problem as a computational tool. The approximation is also of value in understanding the behavior of the universal kriging predictor and the equivalent thin plate smoothing spline.

This problem is approached by first deriving the best linear unbiased predictor of a continuous version of this universal kriging problem. The best linear unbiased predictor has a kernel representation, where the kernel is a Kelvin function (Abramowitz and Stegun (1965), p. 379) which is well-tabled and can be easily approximated by series expansions for both small and large values of its argument. This solution to the continuous prediction problem suggests a kernel approximation to the discrete case where the process is observed with noise at a finite set of locations. In Section 2, the order 2 thin plate smoothing spline and its equivalent universal kriging problem are defined. In Section 3, the kernel representation is derived for the continuous problem. Based on this exact solution to the continuous problem, Section 4 gives an approximate kernel solution in terms of a Kelvin function when the process is observed at a large number of locations that are roughly uniformly distributed in a neighborhood of the point at which the process is being predicted. The asymptotic mean square error of this kernel predictor under a class of stochastic models is derived. The Kelvin function kernel predictor is shown to be the asymptotically optimal kernel predictor among kernel predictors whose kernels and their first partial derivatives decay exponentially. In Section 5, the Kelvin function kernel predictor is compared to the optimal predictor under various circumstances. When the observations are on a square grid, the kernel predictor yields weights that are quite close to the thin plate spline weights for moderate sample sizes and a fairly broad range for the smoothing parameter. When the observations are a random sample from a uniform distribution, the optimal weights and the kernel weights are not so close to each other, but the kernel predictor still performs reasonably well in terms of mean square error under a certain class of stochastic models.

2. The thin plate spline and its universal kriging equivalent

Suppose we observe

$$y(x_l) = z(x_l) + e_l \tag{2.1}$$

for $l = 1, \dots, n$, where $x_1, \dots, x_n$ are points in $\mathbb{R}^2$, $z(\cdot)$ is some unknown function, and the e_l's are uncorrelated errors with zero means and equal variances. The order

2 thin plate smoothing spline estimator of $z(\cdot)$ is the function $\hat{z}(\cdot)$ that minimizes

$$n^{-1}\sum_{l=1}^{n}\{y(x_l)-\hat{z}(x_l)\}^2+\lambda\int_{\mathbb{R}^2}\{\hat{z}_{11}(x)^2+2\hat{z}_{12}(x)^2+\hat{z}_{22}(x)^2\}dx,$$

where the subscripts on $\hat{z}(\cdot)$ indicate partial derivatives with respect to the corresponding arguments (Duchon (1976), Meinguet (1979) and Wahba and Wendelberger (1980)). The smoothing parameter λ determines the trade-off between smoothness of the estimated function and its fidelity to the observations; larger values of λ yield smoother estimates that are farther from the observed values. The optimal estimate of $z(x)$ turns out to be linear in the observations; that is, it can be written as $\sum w_l y(x_l)$.

The solution to this minimization problem at a point x is identical to the universal kriging predictor of $z(x)$ under a certain intrinsic random function model for $z(\cdot)$. We will only discuss here the specific intrinsic random function that yields the order 2 thin plate smoothing spline as the optimal predictor; more complete expositions of intrinsic random function theory are given by Matheron (1973) and Delfiner (1976). For the model in which we are interested, the mean of $z(x)$ is taken to be linear in $x=(s,\, t)$; that is,

$$Ez(x)=\beta_0+\beta_1 s+\beta_2 t.$$

Other intrinsic random functions can be obtained by allowing $Ez(x)$ to be a polynomial in x of order other than one. In intrinsic random function theory, only the variances of contrasts of $z(\cdot)$ are defined. A contrast is any linear combination of $z(x)$'s that has mean zero for all values of β_0, β_1, β_2. That is, $\sum c_l z(x_l)$ is a contrast if for $x_l=(s_l,\, t_l)$

$$\sum c_l=\sum c_l s_l=\sum c_l t_l=0. \tag{2.2}$$

The key property of an intrinsic random function is that the variance of any contrast can be expressed in terms of a generalized covariance function $G(x,\, x')$ that depends on x and x' only through $x-x'$. In the case of interest here, we will take $z(\cdot)$ to have the generalized covariance function

$$G(x,\, x')=G(x-x')=\theta_1|x-x'|^2\log|x-x'|,$$

where $|\cdot|$ indicates Euclidean distance and θ_1 is a positive parameter. Then the variance of a contrast $\sum c_l z(x_l)$ is given by

$$\sum_{lm} c_l c_m G(x_l,\, x_m)=\theta_1\sum_{lm} c_l c_m|x_l-x_m|^2\log|x_l-x_m|. \tag{2.3}$$

The generalized covariance function has the property of being conditionally positive definite; that is, (2.3) is nonnegative whenever (2.2) is satisfied.

The error of any linear unbiased predictor turns out to be a contrast, so that it is sufficient to specify only the generalized covariance function in order to evaluate the variance of the prediction error. Consider predicting $z(x)$ for $x = (s, t)$ based on $y(x_l)$ for $x_l = (s_l, t_l)$, $l = 1, \ldots, n$. A linear unbiased predictor satisfies $E\{\sum w_l y(x_l)\} = Ez(x)$ for all values of β_0, β_1 and β_2, or

$$\sum w_l = 1, \quad \sum w_l s_l = s \quad \text{and} \quad \sum w_l t_l = t. \tag{2.4}$$

The prediction error, $\sum w_l y(x_l) - z(x)$, is thus a contrast, and

$$\begin{aligned} \operatorname{var}\left(\sum_l w_l y(x_l) - z(x)\right) = \theta_0 \sum_l w_l^2 & \\ + \theta_1 \sum_{lm} w_l w_m |x_l - x_m|^2 \log|x_l - x_m| & \\ - 2\theta_1 \sum_l w_l |x_l - x|^2 \log|x_l - x|, & \end{aligned} \tag{2.5}$$

where $\theta_0 = \operatorname{var}(e_l)$. The universal kriging (best linear unbiased) predictor of $z(x)$ chooses the w_l's to minimize (2.5) subject to (2.4). The correspondence between this minimization problem and the thin plate spline has been noted by Dubrule (1983) and Matheron (1980). Kimeldorf and Wahba (1970, 1971) developed a general theory for correspondences between splines and best linear unbiased predictors, but did not explicitly give the relationship between the order 2 thin plate smoothing spline and its universal kriging equivalent. Another way to characterize the best linear unbiased predictor $\sum w_l y(x_l)$ is by the projection property (Journel (1977))

$$\operatorname{cov}\left(z(x) - \sum w_l y(x_l), \sum u_l y(x_l)\right) = 0$$

for all $u_1, \ldots, u_n$ satisfying $\sum u_l = \sum u_l s_l = \sum u_l t_l = 0$. By comparing (2.8) in Wahba and Wendelberger (1980) to the universal kriging equations (Delfiner (1976)), we see that by setting $\lambda = \theta_0/(8\pi\theta_1 n)$, the optimal estimate of $z(x)$ is the same in each case.

3. The continuous universal kriging problem

We first consider a continuous version of the universal kriging problem. For all $f \in K$, where K is the class of infinitely differentiable functions with bounded support, we "observe"

$$\int f(x)(z(x) + W(x))dx,$$

where $W(\cdot)$ is white noise which is uncorrelated with $z(\cdot)$, and $z(\cdot)$ has generalized covariance function $\theta_1|x|^2 \log|x|$. The integral should be interpreted as a linear random functional on the L^2 closure of the space of functions K; a rigorous treatment of this subject is given by Gel'fand and Vilenkin (1964). Formally, we can compute the covariance of integrals of $W(\cdot)$ by

$$\operatorname{cov}\left(\int f(x)W(x)dx, \int g(x)W(x)dx\right) = \theta_0 \int f(x)g(x)dx.$$

The best linear unbiased predictor of $z(0,\, 0)$ is defined by the function f satisfying the unbiasedness conditions

$$\int f(s,\, t)dsdt = 1, \qquad \int sf(s,\, t)dsdt = \int tf(s,\, t)dsdt = 0, \tag{3.1}$$

as well as the projection property

$$\operatorname{cov}\left(z(0) - \int f(x)(z(x) + W(x))dx,\ \int g(x)(z(x) + W(x))dx\right) = 0$$

for all $g \in K$ satisfying

$$\int g(s,\, t)dsdt = \int sg(s,\, t)dsdt = \int tg(s,\, t)dsdt = 0.$$

We can write this covariance as

$$\int g(x)\left[\theta_1 |x|^2 \log|x| - \theta_1 \int f(x')|x - x'|^2 \log|x - x'|dx' - \theta_0 f(x)\right] dx.$$

Thus, we want f to satisfy

$$\theta_1 |x|^2 \log|x| - \theta_1 \int f(x')|x - x'|^2 \log|x - x'|dx' - \theta_0 f(x) = c_0 + c_1 s + c_2 t \tag{3.2}$$

for some constants c_0, c_1, c_2 and all $x = (s,\, t)$. The solution to (3.1) and (3.2) is

$$f(x) = -\frac{1}{2\pi\eta^{1/2}}\operatorname{kei}(\eta^{-1/4}|x|), \tag{3.3}$$

where $\eta = \theta_0/(8\pi\theta_1)$ and $\operatorname{kei}(\cdot)$ is a Kelvin function (Section 9.9, Abramowitz and Stegun (1965)). This f defines the unique best linear unbiased predictor, which follows from the projection property of best linear unbiased predictors and the uniqueness of projections in a Hilbert space (Akhiezer and Glazman (1961), p. 10). This result can be verified using

$$\begin{aligned}
&\int \operatorname{kei}(\eta^{-1/4}|x'|)|x - x'|^2 \log|x - x'|dx' \\
&= \frac{1}{2}\int_0^\infty \int_0^{2\pi} \operatorname{kei}(\eta^{-1/4}r')\{r^2 + r'^2 - 2rr'\cos(\phi - \phi')\} \\
&\quad \times \log\{r^2 + r'^2 - 2rr'\cos(\phi - \phi')\}r'd\phi' dr' \\
&= 2\pi \int_0^\infty r'\operatorname{kei}(\eta^{-1/4}r')\{(r^2 + r'^2)\log(\max(r,\, r')) + \min(r^2,\, r'^2)\}dr' \\
&= 2\pi\eta(-4\operatorname{kei}(\eta^{-1/4}r) - \eta^{-1/2}r^2 \log r),
\end{aligned}$$

where the second equality uses 338.13 from Gröbner and Hofreiter (1950) and the last equality is by lengthy but straightforward calculations using the properties of

Kelvin functions given in Section 9.9 of Abramowitz and Stegun (1965). In fact, we have that the left-hand side of (3.2) is identically zero. It can further be shown that

$$\text{(3.4)}\quad \begin{aligned} &\operatorname{var}\left(z(0,0)+\frac{1}{2\pi\eta^{1/2}}\int \operatorname{kei}(\eta^{-1/4}|x|)(z(x)+W(x))dx\right)\\ &= \frac{\theta_0}{4\pi^2\eta}\int \operatorname{kei}(\eta^{-1/4}|x|)^2dx+\frac{\theta_1}{\pi\eta^{1/2}}\int \operatorname{kei}(\eta^{-1/4}|x|)|x|^2\log|x|dx\\ &\quad+\frac{\theta_1}{4\pi^2\eta}\int \operatorname{kei}(\eta^{-1/4}|x|)\operatorname{kei}(\eta^{-1/4}|x'|)|x-x'|^2\log|x-x'|dxdx'\\ &=(\pi\theta_0\theta_1/8)^{1/2}. \end{aligned}$$

It should be noted that $\operatorname{kei}(\eta^{-1/4}|x|)$ is a solution to the partial differential equation

$$(\eta\Delta^2+1)f=0$$

on $\mathbb{R}^2$, where Δ is the Laplacian, which can be compared to the partial differential equation in Proposition 2.2(ii) of Cox (1984).

An extensive table of $\operatorname{kei}(\cdot)$ is given by Nosova (1961). From 9.9.12 in Abramowitz and Stegun (1965),

$$\begin{aligned} \operatorname{kei}(r) = &-\sum_{j=0}^{\infty}\frac{(-)^j\left(\frac{1}{4}r^2\right)^{2j+1}}{[(2j+1)!]^2}\left\{\log\left(\frac{1}{2}r\right)-\psi(2j+2)\right\}\\ &-\frac{\pi}{4}\sum_{j=0}^{\infty}\frac{(-)^j\left(\frac{1}{4}r^2\right)^{2j}}{[(2j)!]^2}, \end{aligned}$$

where $\psi(\cdot)$ is the digamma function (Abramowitz and Stegun (1965), p. 258). For large r,

$$\operatorname{kei}(r) = -\frac{\pi^{1/2}}{2r}e^{-2^{-1/2}r}\left\{\sin\left(2^{-1/2}r+\frac{\pi}{8}\right)+f(r)\right\},$$

where $f(r)\to 0$ as $r\to\infty$ (Abramowitz and Stegun (1965), 9.10.4).

4. A kernel approximation to the optimal predictor

We now consider using this kernel to predict $z(0,0)$ based on the observations $y(x_1),\ldots,y(x_n)$ under the model defined in Section 2. Specifically, consider a predictor of the form

$$\text{(4.1)}\qquad \frac{c_n^2}{nH_n}\sum_{j=1}^{n}h(c_nx_j)y(x_j),$$

where

$$H_n = n^{-1}c_n^2\sum_{j=1}^{n}h(c_nx_j),$$

$h(x) = h(-x)$ and $h(x)$ is integrable over $\mathbb{R}^2$ with nonzero integral. Without loss of generality, we will take

$$\int_{\mathbb{R}^2} h(x)dx = 1. \tag{4.2}$$

We cannot evaluate the mean square error of this prediction using the generalized covariance function given in Section 2, since the prediction error will not, in general, be a contrast. Instead, we will assume that $z(\cdot)$ is a stationary spatial process with covariance function of the form

$$K(x) = \theta_1\{b_0 - b_1|x|^2 + |x|^2 \log|x| + r(x)\}, \tag{4.3}$$

where for some $\epsilon > 0$, as $|x| \to 0$,

$$r(x) = O(|x|^{2+\epsilon}).$$

A necessary condition for $K(\cdot)$ to be a covariance function is that b_0 and b_1 are positive. However, the local behavior of $z(\cdot)$ is controlled by the $\theta_1|x|^2 \log|x|$ term in (4.3). Thus, a process governed by (4.3) will, over short distances, behave very much like an intrinsic random function with generalized covariance function $\theta_1|x|^2 \log|x|$.

PROPOSITION 4.1. *Suppose $K(x)$ satisfies* (4.3), $c_n \to \infty$ *and* $c_n n^{-1/2} \to 0$ *as* $n \to \infty$, *and that there exists a sequence* $M_1, M_2, \ldots,$ *such that*

$$\underline{\lim} \log M_n / \log n > 0, \tag{4.4}$$

and

$$\sup_{w \in Q_n} \left| \frac{c_n^2}{n} \sum_{j=1}^{n} I_{\{(-M_n, -M_n) \leq w_j \leq w\}} - p_0(u + M_n)(v + M_n) \right| = o((\log n)^{-3/2}) \tag{4.5}$$

where $Q_n = [-M_n, M_n]^2$, $w = (u, v)$, $p_0 > 0$ *and* $w_j = c_n x_j$, *the dependence of* w_j *on* n *being suppressed. Furthermore, assume* $h(x) = h(-x)$, $h(\cdot)$ *satisfies* (4.2) *and both* $h(x)$ *and its first partial derivatives are bounded and decrease exponentially as* $|x| \to \infty$. *Then*

$$\begin{aligned} &\operatorname{var}\left(z(0, 0) - \frac{c_n^2}{nH_n} \sum_{j=1}^{n} h(c_n x_j) y(x_j) \right) \\ &\quad \sim \theta_0 \frac{c_n^2 p_0}{n} \int h(x)^2 dx + \theta_1 \left\{ \frac{-2p_0}{c_n^2} \int h(x)|x|^2 \log|x| dx \right. \\ &\qquad \left. + \frac{p_0^2}{c_n^2} \int h(x)h(x')|x - x'|^2 \log|x - x'| dx dx' \right\}. \end{aligned} \tag{4.6}$$

PROOF. By straightforward algebra,

$$\begin{aligned}
(4.7)\qquad & \operatorname{var}\left(z(0,0)-\frac{c_n^2}{nH_n}\sum_{j=1}^n h(c_nx_j)y(x_j)\right)^2 \\
&= \theta_0\frac{c_n^4}{n^2H_n^2}\sum_{j=1}^n h(c_nx_j)^2+\theta_1 b_1\frac{2c_n^4}{n^2H_n^2}\left|\sum_{j=1}^n h(c_nx_j)x_j\right|^2 \\
&\quad+\theta_1\Bigg[-\frac{2c_n^2}{nH_n}\sum_{j=1}^n h(c_nx_j)\{|x_j|^2\log|x_j|+r(x_j)\} \\
&\qquad+\frac{c_n^4}{n^2H_n^2}\sum_{j,k=1}^n h(c_nx_j)h(c_nx_k)\{|x_j-x_k|^2\log|x_j-x_k| \\
&\qquad+r(x_j-x_k)\}\Bigg] \\
&= \frac{2\theta_1c_n^2}{n^2H_n^2}(b_1+\log c_n)\left|\sum_{j=1}^n h(w_j)w_j\right|^2+\frac{\theta_0c_n^4}{n^2H_n^2}\sum_{j=1}^n h(w_j)^2 \\
&\quad+\theta_1\Bigg\{-\frac{2}{nH_n}\sum_{j=1}^n h(w_j)|w_j|^2\log|w_j| \\
&\qquad+\frac{c_n^2}{n^2H_n^2}\sum_{j,k=1}^n h(w_j)h(w_k)|w_j-w_k|^2\log|w_j-w_k|\Bigg\} \\
&\quad+\theta_1\Bigg\{-\frac{2c_n^2}{nH_n}\sum_{j=1}^n h(w_j)r(c_n^{-1}w_j) \\
&\qquad+\frac{c_n^4}{n^2H_n^2}\sum_{j,k=1}^n h(w_j)h(w_k)r(c_n^{-1}(w_j-w_k))\Bigg\}.
\end{aligned}$$

(4.3) can be used to show that all terms in (4.7) containing $r(\cdot)$ are asymptotically negligible, so that

$$\begin{aligned}
(4.8)\qquad & \operatorname{var}\left(z(0,0)-\frac{c_n^2}{nH_n}\sum_{j=1}^n h(c_nx_j)y(x_j)\right)^2 \\
&\sim \frac{\theta_0c_n^4}{n^2H_n^2}\sum_{j=1}^n h(w_j)^2-\frac{2\theta_1}{nH_n}\sum_{j=1}^n h(w_j)|w_j|^2\log|w_j| \\
&\quad+\frac{\theta_1c_n^2}{n^2H_n^2}\sum_{j,k=1}^n h(w_j)h(w_k)|w_j-w_k|^2\log|w_j-w_k|.
\end{aligned}$$

The conditions in Proposition 4.1 allow the sums in (4.8) to be approximated by integrals (see appendix) and the proposition follows. □

An interesting feature of (4.6) is the fact that b_0, b_1, and $r(\cdot)$ do not appear on the right-hand side. This result is not unexpected, since the $\theta_1|x|^2\log|x|$ component of the covariance function in (4.3) controls the local behavior of $z(\cdot)$.

This proposition can be used to obtain a somewhat restrictive optimality property of a Kelvin function kernel predictor. Using the results from Section 3, the function $h(\cdot)$ that minimizes the right-hand side of (4.6) subject to (4.2) is given by (3.3), where $\eta = \theta_0 c_n^4/(8\pi\theta_1 n p_0)$, in which case, the right-hand side of (4.6) equals $\{\pi\theta_0\theta_1/(8np_0)\}^{1/2}$. By setting $c_n = (np_0)^{1/4}$, we get $\eta = \theta_0/(8\pi\theta_1)$ independent of n. Since $-\text{kei}(|x|)/(2\pi)$ satisfies the conditions of the proposition, the kernel predictor of $z(0, 0)$ given by

$$(4.9) \qquad -\left(\frac{p_0}{n\eta}\right)^{1/2} \frac{1}{2\pi H_n} \sum_{j=1}^{n} \text{kei}\left(\left(\frac{np_0}{\eta}\right)^{1/4} |x_j|\right) y(x_j)$$

will, subject to (4.4), (4.5) and $K(\cdot)$ satisfying (4.3), have asymptotic mean square error $\{\pi\theta_0\theta_1/(8np_0)\}^{1/2}$. Furthermore, for any $h(\cdot)$ satisfying the conditions of the proposition and any sequence of c_n's,

$$\underline{\lim}\, n^{1/2}\text{var}\left(z(0,\,0) - \frac{c_n^2}{nH_n}\sum_{j=1}^{n} h(c_n x_j) y(x_j)\right) \geq \left(\frac{\pi\theta_0\theta_1}{8p_0}\right)^{1/2}.$$

This result follows from the proposition when $c_n \to \infty$ and $c_n n^{-1/2} \to 0$. Since the kernel predictor is not mean square consistent along any subsequence of c_n's violating either of these conditions, this bound holds for all c_n's. Thus, the kernel predictor in (4.9) is asymptotically optimal among this class of kernel predictors. A stronger result would be to find conditions under which this kernel predictor is asymptotically optimal among all linear predictors of $z(0, 0)$. Considering the examples on the triangular and spherical covariance functions of Stein and Handcock (1989), it appears that additional conditions on the smoothness of $r(x)$ away from the origin would be needed to obtain such a result. If we could obtain such a result for one particular $K(\cdot)$ and $x_1, x_2, \ldots$ all contained in some bounded region R, then using an argument similar to the proof of (14) by Stein (1988*a*), the kernel predictor would also be asymptotically optimal for the same sequence of x_i's and any covariance function compatible with $K(\cdot)$ on R (see Stein (1988*b*) for the definition of compatibility of covariance functions).

The conditions in the proposition are sufficient to show $H_n \to p_0$ as $n \to \infty$, which accounts for the form of the kernel predictor given in (4.1). For $j = 1, \ldots, n$, we will thus call

$$(4.10) \qquad -\frac{1}{2\pi(p_0 n\eta)^{1/2}}\text{kei}\left(\left(\frac{np_0}{\eta}\right)^{1/4} |x_j|\right)$$

the unadjusted weights, and the weights given in (4.9) the adjusted weights. Both will be used in the next section.

Finally, note that the condition (4.5) is not very strong; it roughly says that $x_1, \ldots, x_n$ need to be approximately uniform in a neighborhood of radius M_n/c_n of the origin. For example, suppose $c_n = (np_0)^{1/4}$, $M_n = n^{\delta}$, $0 < \delta < 1/12$, and $x_1, x_2, \ldots$ are independent identically distributed with density $p(x)$, where

$p(0, 0) = p_0 > 0$ and $p(\cdot)$ has bounded first partial derivatives in a neighborhood of the origin. Then for $w \in Q_n$,

$$E\left(\frac{c_n^2}{n}\sum_{j=1}^{n} I_{\{(-M_n, -M_n) \le w_j \le w\}}\right)$$
$$= c_n^2 \int_{(-M_n/c_n, -M_n/c_n) \le x \le w/c_n} p(x)dx$$
$$= p_0(u + M_n)(v + M_n) + O(n^{3\delta - 1/4}),$$

and

$$\operatorname{var}\left(\frac{c_n^2}{n}\sum_{j=1}^{n} I_{\{(-M_n, -M_n) \le w_j \le w\}}\right) \sim n^{-1/2} p_0^{3/2}(u + M_n)(v + M_n)$$
$$= O(n^{2\delta - 1/2}),$$

so

$$\frac{c_n^2}{n}\sum_{j=1}^{n} I_{\{(-M_n) \le w_j \le w\}} - p_0(u + M_n)(v + M_n) = O(n^{3\delta - 1/4}) + O_p(n^{\delta - 1/4}).$$

Since we only need the supremum over all $w \in Q_n$ to be $o((\log n)^{-3/2})$, we see that (4.5) does not require the x_i's to be very close to uniform in a neighborhood of the origin.

5. Numerical results

In this section, the kernel predictor given in (4.9) is compared to the best linear unbiased predictor in some particular cases. Specifically, for various values of θ_1/θ_0, we consider predicting $z(0, 0)$ based on all observations of the form $y(0.2i, 0.2j)$ within the unit circle. We also consider predicting $z(0, 0)$ based on independent uniformly distributed observations in the unit circle.

From the derivation in the previous section, it is apparent that the accuracy of the approximation depends on how well the various sums in (4.7) can be approximated by integrals. The accuracy of these integral approximations in turn depends on the density and regularity of the points in a neighborhood of the origin. One particular case of the effect of "irregularly" spaced points is the edge effect caused by having no observations beyond a certain boundary, which implies that finite sums are being approximated by integrals over infinite ranges. The exact and approximate weights for predicting $z(0, 0)$ based on all $y(0.2i, 0.2j)$ in the unit circle, where i and j are integers and $\theta_1/\theta_0 = 5$, 1, and 0.2 are given in Table 1. Recall that the corresponding values for the smoothing parameter in the spline formulation of the problem are given by $\lambda = \theta_0/(8\pi\theta_1 n) = 4.91 \times 10^{-4}\theta_0/\theta_1$. The exact weights are based on the universal kriging model described in Section 2. The approximate weights are based on the unadjusted kernel weights given in (4.10) with $n = 81$ and $p_0 = 1/(81 \times 0.2^2)$; kei$(\cdot)$ is evaluated using 9.11.4 and 9.11.9 of Abramowitz and Stegun (1965). The approximate weights are quite close to the optimal weights throughout this wide range of θ_1/θ_0 for those observations near the origin. However, as θ_1/θ_0 decreases, the approximate weights for the observations farther from the origin are increasingly in error. This inaccuracy in the approximation for points near the edge is mainly due to edge effects, which can be seen by predicting $z(0, 0)$ based on all observations $y(0.2i, 0.2j)$ within 2.0 units of the origin and $\theta_1/\theta_0 = 1$. In this case, the approximate weights given in Table 1 are of course unchanged, and the exact weights are (for the points in their given order in Table 1) 0.1245, 0.0785, 0.0567, 0.0321, 0.0245, 0.0108, 0.0081, 0.0059, 0.0017, −0.0004, −0.0008, −0.0011, −0.0015, −0.0018, and −0.0018, which are in excellent agreement with the approximate weights in Table 1. Of course, as noted by Silverman (1985), there are many more points near an edge of the observation region in two dimensions than one. Thus, there is a clear need for some sort of edge correction analogous to the one given by Silverman (1984) for the one-dimensional cubic spline.

Table 1. Exact weights given to $y(x)$ for the thin plate spline estimate of $z(0, 0)$ with various values of $\theta_0 = \theta_1$, together with the approximate weights given by the kernel approximation.[1)]

	θ_1/θ_0					
x	5		1		0.2	
	Exact	Approximate	Exact	Approximate	Exact	Approximate
(0, 0)	.2705	.2802	.1252	.1253	.0584	.0560
(0, 0.2)	.1158	.1185	.0793	.0789	.0460	.0437
(0.2, 0.2)	.0627	.0634	.0575	.0569	.0387	.0367
(0, 0.4)	.0184	.0184	.0329	.0322	.0288	.0272
(0.2, 0.4)	.0089	.0088	.0253	.0246	.0251	.0237
(0.4, 0.4)	−.0023	−.0023	.0114	.0109	.0167	.0162
(0.0, 0.6)	−.0034	−.0034	.0083	.0081	.0145	.0143
(0.2, 0.6)	−.0039	−.0039	.0060	.0059	.0126	.0127
(0.4, 0.6)	−.0037	−.0036	.0014	.0017	.0078	.0088
(0, 0.8)	−.0027	−.0026	−.0019	−.0004	.0040	.0062
(0.2, 0.8)	−.0023	−.0023	−.0026	−.0008	.0029	.0054
(0.6, 0.6)	−.0020	−.0019	−.0026	−.0011	.0020	.0048
(0.4, 0.8)	−.0012	−.0014	−.0041	−.0015	.0000	.0037
(0.6, 0.8)	.0001	−.0005	−.0058	−.0018	−.0039	.0018
(0, 1)	.0001	−.0005	−.0066	−.0018	−.0040	.0018

[1)] Approximate weights given by (4.8). Observation sites consist of all points of the form $(0.2i, 0.2j)$ within distance 1.0 of the origin. Weights are given only for $0 \leq i \leq j$. Other weights can be obtained by symmetry considerations.

Another way of judging the quality of the kernel approximation is by comparing the variances of the prediction errors using the exact and kernel weights based on the stochastic model for $z(x)$ described in Section 2. Since the observations are symmetric about the origin, the adjusted kernel predictor given by (4.9) is linear unbiased using the model from Section 2. Table 2 gives the sum of the unadjusted weights, the variance of the errors of the best linear unbiased predictor (the thin plate spline) and its approximation based on the adjusted kernel weights for the same cases as in Table 1. We see that for a very wide range of θ_1/θ_0, the adjusted kernel predictor performs well. Of course, as $\theta_1/\theta_0 \to 0$, the best predictor is just the average of all the observations, and as $\theta_1/\theta_0 \to \infty$, the best predictor is $y(0, 0)$, so it is not surprising that the adjusted kernel performs well for very large and small values of θ_1/θ_0. These results strongly suggest that the adjusted kernel predictor is asymptotically optimal under appropriate conditions on the observations. The approximation to the variance of the error of the best linear unbiased predictor given by $\{\pi\theta_0\theta_1/(8np_0)\}^{1/2}$ is also given in Table 2. Equation (4.8) says that this expression is asymptotically the same as the variance of the adjusted kernel predictor as $n \to \infty$ for fixed θ_0 and θ_1, which in turn, is conjectured to be asymptotically the same as the variance of the optimal predictor. The approximation works quite well for moderate values of θ_1/θ_0 and is not so far off even for very large or small values of θ_1/θ_0.

Table 2. Variances of the prediction error for the thin plate spline and its adjusted kernel approximation.[1)]

θ_1/θ_0	Adjustment to kernel[2)]	Variance of prediction error		Approximate variance[3)]
		Spline	Adjusted kernel	
25	1.0957	0.02129	0.02145	0.02507
5	1.0153	0.05411	0.05416	0.05605
1	1.0853	0.1252	0.1312	0.1253
0.2	1.1077	0.2922	0.3048	0.2802
0.04	0.8866	0.6674	0.6680	0.6267

[1)] Based on same observations as in Table 1 with $\theta_1 = 1$.

[2)] Sum of the unadjusted kernel weights, used to rescale them so that they satisfy the unbiasedness condition (2.4).

[3)] Given by $\{\pi\theta_0\theta_1/(8np_0)\}^{1/2}$.

As a final example, take 81 points (the same as in the previous example) chosen independently and uniformly on the unit disk, and again consider predicting $z(0, 0)$. In this case, the adjusted kernel predictor will not be linear unbiased under the model described in Section 2, so we will assume that $z(\cdot)$ is a stationary process with unknown constant mean and covariance function $2\theta_1 x K_1(x)$, where $K_1(\cdot)$ is a modified Bessel function (Abramowitz and Stegun, (1965), p. 374). This covariance function is of the form given in (4.3) with $b_0 = 2$, $b_1 = (2\gamma - 1)/4$

where γ is Euler's constant, and $r(x) = O(|x|^4 \log |x|)$ as $x \to 0$. The optimal predictors are based on this model, and the adjusted kernel predictor uses (4.9) with $n = 81$ and $p_0 = 1/\pi$. Fifteen sets of observations were generated, and in each case, the optimal and adjusted kernel weights were computed. Using these, the relative efficiency of the adjusted kernel predictor (the mean square error of the adjusted kernel predictor divided by the mean square error of the optimal predictor) and the sum of the absolute differences of the two sets of weights were calculated. The results are summarized in Table 3. The smaller the value of θ_1/θ_0, the closer the adjusted kernel predictor comes to the optimal predictor by either measure. Since smaller values of θ_1/θ_0 correspond to more smoothing, we would expect the unevenness in the observations to matter less. We also see that the stochastic measure of closeness of the two predictors, the relative efficiency, suggests a smaller discrepancy between the predictors than the sum of the absolute differences in the weights. That is, it is possible to get a very good predictor in terms of mean square error under the stochastic model without the weights of the predictor being all that close to the weights of the optimal predictor. The optimal weights change very little when they are generated using the nonstationary stochastic model corresponding to the thin plate spline. Thus, while we cannot evaluate the mean square error of the adjusted kernel predictor under this model, we do have that the last two columns of Table 3 remain practically unchanged when the optimal weights are calculated using the nonstationary model.

Table 3. A comparison of optimal to adjusted kernel prediction with 81 independent uniformly distributed observations on the unit disk.[1)]

θ_1/θ_0	Efficiency[2)]		Sum of absolute differences in weights	
	Median	Minimum	Median	Maximum
5	0.8138	0.3905	0.3872	1.5255
1	0.9381	0.8771	0.2628	0.4288
0.2	0.9910	0.9420	0.1119	0.2271

[1)] Optimal predictor based on model with constant unknown mean and covariance function of continuous part of process given by $2\theta_1|x|K_1(|x|)$. Adjusted kernel predictor given by (4.9). Results based on 15 simulations.

[2)] Mean square error of optimal predictor divided by mean square error of adjusted kernel predictor.

Acknowledgements

The author is grateful to Professor Wing Wong and the referee for several helpful suggestions.

Appendix

Using an argument similar to the one given below, it can be shown that (4.4) and (4.5) together with the conditions on $h(\cdot)$ given in the proposition imply that the right-hand side of (4.8) is asymptotically equivalent to the right-hand side of (4.6) and that $H_n \to p_0$. Thus, (4.6) follows if we can show

$$\text{(A.1)} \qquad \frac{c_n^2}{n^2}\log n \left|\sum_{j=1}^{n} h(w_j)w_j\right|^2 = o(c_n^{-2}).$$

First, (4.4) and the requirement that $h(\cdot)$ decay exponentially imply that only those $w_j \in Q_n$ need be considered. Partition Q_n into d_n squares with sides of length $M_n d_n^{-1/2}$. (4.5) allows us to choose d_n such that $M_n d_n^{-1/2} = o((\log n)^{-1/2})$ and the left-hand side of (4.5) is $o(M_n^2 d_n^{-1}(\log n)^{-1/2})$. Let $w_j = (u_j,\, v_j)$, $S_1, \ldots, S_{d_n}$ be the d_n squares partitioning Q_n and $W_i = (U_i,\, V_i)$ be the center of S_i. Then,

$$\begin{aligned}
\text{(A.2)} \qquad \left|\frac{c_n^2}{n}\sum_{w_j\in Q_n} h(w_j)u_j\right| \le{}& \frac{c_n^2}{n}\sum_{k=1}^{d_n}\sum_{w_j\in S_k}|h(w_j)u_j - h(W_k)U_k| \\
&+\frac{c_n^2}{n}\sum_{k=1}^{d_n}|h(W_k)U_k|\left|\sum_{w_j\in S_k} 1 - \frac{n}{c_n^2}p_0\frac{M_n^2}{d_n}\right| \\
&+\sum_{k=1}^{d_n}\left|h(W_k)U_k p_0 M_n^2 d_n^{-1} - p_0\int_{S_k} h(w)u\,dw\right| \\
&+p_0\left|\int_{Q_n} h(w)u\,dw\right|.
\end{aligned}$$

Using the assumption that the first partial derivatives of $h(\cdot)$ are bounded and decay exponentially, there exist positive constants C and α, independent of n, such that

$$\begin{aligned}
\frac{c_n^2}{n}\sum_{k=1}^{d_n}\sum_{w_j\in S_k}|h(w_j)u_j - h(w_k)U_k| &\le \frac{CM_n^2}{d_n}\sum_{k=1}^{d_n} e^{-\alpha|w_k|} \\
&= o((\log n)^{-1/2}).
\end{aligned}$$

The third term on the right-hand side of (A.2) is handled similarly. The fourth term is easily shown to be negligible using (4.4), the exponential decay of $h(\cdot)$ and $\int h(w)u\,dw = 0$. Finally,

$$\begin{aligned}
\frac{c_n^2}{n}\sum_{k=1}^{d_n}|h(w_k)U_k|\left|\sum_{w_j\in S_k} 1 - \frac{n}{c_n^2}p_0\frac{M_n^2}{d_n}\right| &= \frac{c_n^2}{n}\sum_{k=1}^{d_n}|h(w_k)U_k| o(M_n^2 d_n^{-1}(\log n)^{-1/2}) \\
&= \frac{c_n^2}{n} o((\log n)^{-1/2}).
\end{aligned}$$

Combining these results, (A.2) is $o((\log n)^{-1/2})$ and (A.1) follows.

REFERENCES

Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Functions*, Dover, New York.

Akhiezer, N. I. and Glazman, I. M. (1961). *Theory of Linear Operators in Hilbert Space* (translated by M. Nestell), Frederick Ungar, New York.

Cogburn, R. and Davis, H. T. (1974). Periodic splines and spectral estimation, *Ann. Statist.*, **2**, 1108–1126.

Cox, D. D. (1983). Asymptotics of *M*-type smoothing splines, *Ann. Statist.*, **11**, 530–551.

Cox, D. D. (1984). Multivariate smoothing spline functions, *SIAM J. Numer. Anal.*, **21**, 789–813.

Delfiner, P. (1976). Linear estimation of non-stationary spatial phenomena, *Advanced Geostatistics in the Mining Industry, Proceedings of NATO A.S.I.* (eds. E. M. Gurascio, M. David and Ch. Huijbregts), 49–68, Reidel, Dordecht.

Dubrule, O. (1983). Two methods with different objectives: splines and kriging, *Math. Geol.*, **15**, 245–257.

Duchon, J. (1976). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces, *RAIRO Anal. Numér.*, **10**, 5–12.

Gel'fand, I. M. and Vilenkin, N. Ya. (1964). *Generalized Functions, Vol. 4, Applications of harmonic analysis* (translated by A. Feinstein), Academic Press, New York.

Gröbner, W. and Hofreiter, N. (1950). *Integraltafel Zweiter Teil Bestimmte Integrale*, Springer, Vienna.

Journel, A. G. (1977). Kriging in terms of projections, *Math. Geol.*, **9**, 563–586.

Kimeldorf, G. and Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Ann. Math. Statist.*, **41**, 495–502.

Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian Splines, *J. Math. Anal. Appl.*, **33**, 82–95.

Matheron, G. (1973). The intrinsic random functions and their applications, *Adv. in Appl. Probab.*, **5**, 439–468.

Matheron, G. (1980). Splines and kriging: their formal equivalence, Internal Report, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.

Meinguet, J. (1979). Multivariate interpolation at arbitrary points made simple, *Z. Angew. Math. Phys.*, **30**, 292–304.

Nosova, L. N. (1961). *Tables of Thomson Functions and Their First Derivatives* (translated by P. Basu), Pergamon Press, New York.

Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method, *Ann. Statist.*, **12**, 898–916.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting, *J. Roy. Statist. Soc. Ser. B*, **47**, 1–52.

Stein, M. L. (1988*a*). An application of the theory of equivalence of Gaussian measures to a prediction problem, *IEEE Trans. Inform. Theory*, **34**, 580–582.

Stein, M. L. (1988*b*). Asymptotically efficient prediction of a random field with a misspecified covariance function, *Ann. Statist.*, **16**, 55–63.

Stein, M. L. and Handcock, M. S. (1989). Some asymptotic properties of kriging when the covariance function is misspecified, *Math. Geol.*, **21**, 171–190.

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