Statistica Sinica **23** (2013), 000-000
doi:http://dx.doi.org/10.5705/ss.2010.278

BIVARIATE PENALIZED SPLINES FOR REGRESSION

Ming-Jun Lai and Li Wang

The University of Georgia

Abstract: In this paper, the asymptotic behavior of penalized spline estimators is studied using bivariate splines over triangulations and an energy functional as the penalty. A convergence rate for the penalized spline estimators is derived that achieves the optimal nonparametric convergence rate established by Stone (1982). The asymptotic normality of the proposed estimators is established and shown to hold uniformly over the points where the regression function is estimated. The size of the asymptotic conditional variance is evaluated, and a simple expression for the asymptotic variance is given. Simulation experiments have provided strong evidence that corroborates the asymptotic theory. A comparison with thin-plate splines is provided to illustrate some advantages of this spline smoothing approach.

Key words and phrases: Asymptotic normality, least squares, penalty, spline, triangulation.

1. Introduction

Piecewise polynomial functions, or splines, have proven to be an extremely powerful tool for smoothing. There are numerous applications of nonparametric regression in higher dimensions; see Stone (1994) on the utility of multivariate splines for statistical applications. Theoretical properties of the unpenalized spline estimators have been examined in many directions: for global rates of convergence, see Stone (1985, 1986, 1994), Kooperberg, Stone, and Truong (1995a,b), and Huang (1998); for local asymptotic results, see Zhou, Shen, and Wolfe (1998) and Huang (2003a,b).

Many of the spline-based approaches to multivariate estimation problems involve tensor product spaces. Such spaces are most useful when the data are observed in a rectangular domain. However, the structure of tensor products is undesirable when data are located in domains with complex boundaries and holes. Ramsay (2002) provided examples of such kind of data, and pointed out some serious problems if one uses tensor product methods or wavelet thresholding methods to smooth the data. When the data locations are spread in a general bounded domain of arbitrary shape, a triangulation is the most convenient tool to partition the domain into pieces. Bivariate splines, smooth piecewise polynomial functions over triangulations, are natural extensions of univariate spline functions

over subintervals. They surface in such areas as computer aided geometric design and numerical solution of partial differential equations. Bivariate splines based on triangulations are invariant to affine transformations (Hansen, Kooperberg, and Sardy (1998)), and are more appealing when the given coordinate system of the data is arbitrary, as is the case with special or compositional data.

The theory and computation of bivariate splines become matured and has had continued growth; see the monograph by Lai and Schumaker (2007) for basic theories of bivariate, trivariate, and spherical splines, and see Awanou, Lai, and Wenston (2006) and Baramidze, Lai, and Shum (2006) for some numerical implementations of spherical splines for data fitting and their application to geopotential reconstruction in Lai et al. (2009). We refer to Lai (2008) for a survey of multivariate splines for scattered data fitting and some numerical examples. Guillas and Lai (2010) used these splines for prediction of ozone concentration based on the measurements from all EPA stations scattered around the U.S. continent. Their predictions are consistent for different learning periods (Ettinger, Guillas, and Lai (2012)). Lai and Wenston (2004) applied these splines to numerically solve Navier-Stokes equations and simulated several fluid flows. Huang (2003a,b) has studied the asymptotic behavior of discrete least spline over triangulations for noisy data over random locations.

It is known that penalized regression splines include discrete least squares splines as a special case. When we have regions of sparse data, penalized splines provide a more convenient tool for data fitting than the discrete least squares splines. Penalized bivariate splines over triangulations have been used for data fitting, in particular, in surface generation. Approximation properties of these splines have been studied for noise-free data in approximation theory. For example, von Golitchek and Schumaker (2002a,b) studied the approximation properties of discrete least squares splines and penalized least squares splines for data without noise. However, to the best of our knowledge, statistical aspects of smoothing using bivariate penalized splines over triangulations (BPSOT) have not been discussed for datasets in the presence of random noise, and thus our theoretical and numerical results appear to be new.

As pointed out in Ruppert (2002) and Li and Ruppert (2008) for the univariate case, one advantage of using penalized spline is that the number of knots is not essential, provided only that the number is above some minimum depending upon the degree of the spline. For the bivariate case, Xiao, Li, and Ruppert (2010) proposed a penalized spline method for bivariate smoothing using tensor product B-splines and row and column penalties defined in the bivariate P-spline of Marx and Eilers (2005). Here we consider the bivariate penalized spline smoothing over triangulations with energy functional penalties. For the BPSOT estimators, we

observed a similar property in our simulation: the number of triangles is not crucial given that the number is above some minimum depending upon the degree of the smoothness.

The theoretical properties of penalized splines remains less well explored, and most of the work has been confined to the one-dimensional setting. In the univariate case, Wand (1999) derived an asymptotic mean squared error for penalized splines with a fixed number of knots. Hall and Opsomer (2005) derived mean squared error expressions and consistency results for penalized spline estimators using a white-noise model representation. Li and Ruppert (2008) studied the asymptotics of penalized spline estimators using an equivalent kernel representation for B-splines and difference penalties. Depending on the number of knots, sample size, and penalty, Claeskens, Krivobokova, and Opsomer (2008) showed that the theoretical properties of penalized regression spline estimators are either similar to those of regression splines or to those of smoothing splines, with a clear breakpoint distinguishing the cases. The asymptotic behavior of penalized least squares fitting based on multivariate spline is of interest.

We concentrate on bivariate splines mainly because we have good knowledge of the approximation power of bivariate splines for any degree and smoothness over arbitrary triangulation (Lai and Schumaker (2007)). Our results extend these of Huang (2003a,b) on discrete least squares splines over triangulations to the setting of penalized splines, hence are more suitable for sparse data fitting. The proposed BPSOT for regression contains a penalty term that makes the estimation remarkably different from the discrete least squares spline (Huang (2003a,b)). The penalty term involves the smoothness of the fitting spline, but it does not carry information of data locations and observed values; thus, it is difficult to get any direct estimate. Further theoretical development is required to study the properties of the BPSOT estimators.

The rest of the paper is organized as follows. In Section 2, we introduce the notation of triangulation and describe the BPSOT estimator. In Section 3, we formalize the assumptions and present our main asymptotic results. Section 3.1 provides the rate of convergence of the BPSOT estimator to the true mean function. Under some regularity conditions, we show that the rate is optimal. Section 3.2 explains the asymptotic normality of the BPSOT estimator and we give simple expressions for the asymptotic variance. Applications to constructing asymptotic confidence intervals are also discussed. In Section 4, we run numerical experiments to assess the behavior of the BPSOT estimator and to compare the BPSOT estimator with the thin-plate spline estimator. The proofs of the main results are given in the supplementary document online.

2. Triangulations and Bivariate Penalized Spline Estimators

Let Y be the response variable and $\mathbf{X}$ be the predictor variable or design point. Suppose $\mathbf{X}$ ranges over a bounded domain $\Omega \subseteq \mathbb{R}^2$ of arbitrary shape. For any given dataset with design points located in Ω, one can find a polygonal domain (a domain with piecewise linear boundary) to include all the design points. The polygonal domain is allowed have a hole or multiple holes where no design points are located. With this principle, in the following we assume that Ω is a polygonal domain itself.

Suppose that $\{\mathbf{X}_i, Y_i\}_{i=1}^n = \{X_{1i}, X_{2i}, Y_i\}_{i=1}^n$ is an i.i.d sample of size n from the distribution of $(\mathbf{X}, Y)$, and satisfying the following model

$$Y_i = m(\mathbf{X}_i) + \sigma(\mathbf{X}_i)\,\epsilon_i, \quad i = 1, \cdots, n,$$

where $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ is the bivariate conditional mean function and the ϵ_i's are i.i.d random noises with $E(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = 1$, each ϵ_i independent of $\mathbf{X}_i$. Our primary interest is to estimate the unknown function $m(\mathbf{x})$ based on the given observations $\{\mathbf{X}_i, Y_i\}_{i=1}^n$.

2.1. Triangulations

Triangulations have been used for numerical solutions of partial differential equations and computer aided geometry design for many decades. We use τ to denote a triangle that is the convex hull of three points not located in one line. A collection $\triangle = \{\tau_1, \ldots, \tau_N\}$ of triangles is called a triangulation of $\Omega = \cup_{i=1}^N \tau_i$ provided that if a pair of triangles in $\triangle$ intersect, then their intersection is either a common vertex or a common edge. Figure 1 shows some triangulations of a particular polygonal domain. In general, any kind of polygon shapes can be used for the partition of Ω. In this paper we restrict our attention to triangulations of Ω because any polygonal domain of arbitrary shape can be partitioned into finitely many triangles; see Ramsay (2002) for a triangulation of the island of Montreal as an example.

In the rest of the paper, we assume that the domain Ω is partitioned by a set of triangles $\triangle$. In addition, we assume that all $\mathbf{X}_i$'s are inside triangles of $\triangle$. That is, they are not on edges or vertices of triangles in $\triangle$. Otherwise, we can simply count them twice or multiple times if any observation is located inside of an edge or at a vertex of $\triangle$.

Given a triangle $\tau \in \triangle$, we use $|\tau|$ to denote the length of its longest edge, and ρ_τ for the radius of the largest disk that can be inscribed in τ. We call the ratio $\beta_\tau = |\tau|/\rho_\tau$ the shape parameter of τ. When β_τ is small, the triangles are relatively uniform in the sense that all angles of triangles in the triangulation τ are relatively the same. Thus, the triangulation looks more like a uniform

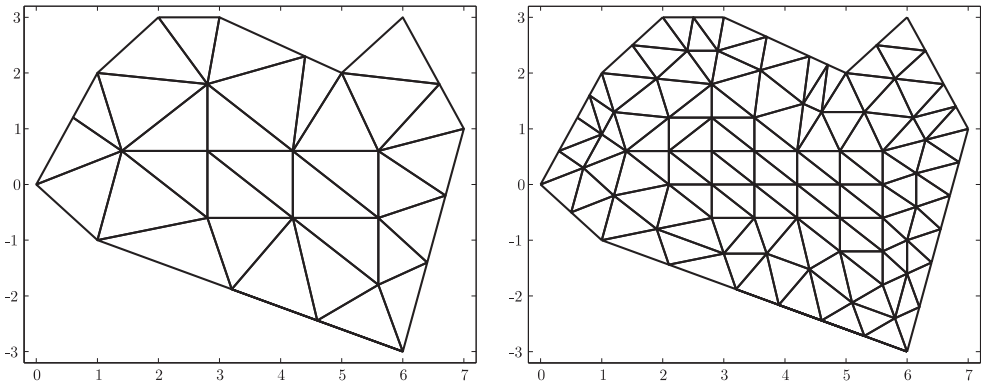


Figure 1. (a) a triangulation; (b) a uniformly refined triangulation.

triangulation and hence, the shape of τ is better. Denote the size of $\triangle$ by $|\triangle| := \max\{|\tau|, \tau \in \triangle\}$, the length of the longest edge of $\triangle$. We say a triangulation $\triangle$ is β-quasi-uniform if there is a positive β such that the triangulation $\triangle$ satisfies

$$\frac{|\triangle|}{\rho_\tau} \leq \beta, \quad \text{for all } \tau \in \triangle. \tag{2.1}$$

This corresponds to Condition A.2 in Huang (2003b). Let N be the number of the triangles in the polygonal domain Ω. From (2.1), we can see that $N \leq (\pi|\Delta|^2)^{-1} A_\Omega \beta^2$, where A_Ω denotes the area of Ω.

It is easy to generate a β quasi-uniform triangulation by starting with an initial triangulation, e.g. by hand drawing and then refining it uniformly and repeatedly. For example, Figure 1 (a) illustrates a triangulation of a polygonal domain with a few triangles and Figure 1 (b) shows the uniformly refined triangulation. We can refine it repeatedly as the approximation power of a bivariate spline space is measured by $|\triangle|$. In addition, as shown in Section 3, the constant in the estimate of spline approximation depends on β. Thus, we take β to be a constant.

In Figure 2, we show a triangulation that is not β quasi-uniform when the pattern of triangles at the low-right corner continues to a desired level so that the condition in (2.1) is not satisfiable in this case. Indeed, $|\triangle|$ is a fixed number while the smallest ρ_τ can be as small as one wishes so that no β can be found to have (2.1).

In practice, we can find a triangulation for a given dataset in several ways. For example, we can find a polygon Ω containing all the design points of the data and triangulate Ω by hand or on the computer to have a triangulation

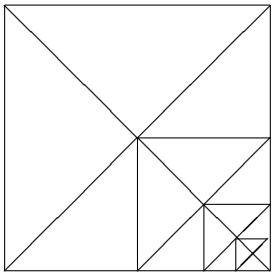


Figure 2. An illustration of non-β quasi-uniform triangulation.

$\triangle_0$. Then, we can uniformly refine $\triangle_0$ several times to have a desired triangulation. The Delaunay algorithm is a good way to triangulate the convex hull of an arbitrary dataset; see MATLAB program `delaunay.m`. Additionally, we can generate a triangular mesh through the DistMesh Generator; see `http://persson.berkeley.edu/distmesh/`. DistMesh is a MATLAB code for generation of unstructured triangular and tetrahedral meshes. It is short and simple, and appears to produce high quality triangular meshes. A detailed description of the program is provided in Persson and Strang (2004).

2.2. Penalized spline estimators

For an integer $r \geq 0$, let $C^r(\Omega)$ be the space of all r-times continuously differentiable functions over Ω. Let $S_d^r(\triangle) = \{s \in C^r(\Omega) : s|_\tau \in \mathbf{P}_d, \tau \in \triangle\}$ be a spline space of degree d and smoothness r over triangulation $\triangle$, where $s|_\tau$ is the polynomial piece of spline s restricted to triangle τ, and $\mathbf{P}_d$ is the space of all polynomials of degree less than or equal to d. For notation simplicity, let $\mathcal{S} := S_{3r+2}^r(\triangle)$ for a fixed smoothness $r \geq 1$; we know that such a spline space has the optimal approximation order (rate of convergence) for noise-free datasets; see Lai and Schumaker (1998, 2007).

To discuss the asymptotics of the BPSOT estimator, we need some notation. For any function f over the closure of domain Ω, let $E_n(f) = n^{-1}\sum_{i=1}^n f(\mathbf{X}_i)$ and $E(f) = E[f(\mathbf{X})]$. Write the empirical inner product and norm as $\langle f_1, f_2\rangle_{n,\Omega} = E_n(f_1 f_2)$ and $\|f_1\|_{n,\Omega}^2 = \langle f_1, f_1\rangle_{n,\Omega}$ for measurable functions f_1 and f_2 on Ω. The theoretical L^2 inner product and the induced norm are given by $\langle f_1, f_2\rangle_{L^2(\Omega)} = E(f_1 f_2)$ and $\|f_1\|_{L^2(\Omega)}^2 = \langle f_1, f_1\rangle_{L^2(\Omega)}$. Let $\|f\|_{\infty,\Omega} = \sup_{\mathbf{x}\in\Omega}|f(\mathbf{x})|$, and let $|f|_{\upsilon,\infty,\Omega} = \max_{i+j=\upsilon}\|D_{x_1}^i D_{x_2}^j f(\mathbf{x})\|_{\infty,\Omega}$ be the largest value of the maximum norms of all the υth order derivatives of f over the closure of Ω, where D is derivative operator.

To fix the penalized spline method, let

$$\mathcal{E}_\upsilon(f) = \sum_{\tau\in\triangle}\int_\tau \sum_{i+j=\upsilon}\binom{\upsilon}{i}(D^i_{x_1}D^j_{x_2}f)^2dx_1dx_2$$

be the energy functional for a fixed integer $\upsilon \geq 1$. For simplicity we use $\upsilon = 2$, though one can study similar problems for $\upsilon \geq 2$. Given $\lambda > 0$ and $\{\mathbf{X}_i, Y_i\}_{i=1}^n$, we consider the minimization problem:

$$\min_{s\in\mathcal{S}}\sum_{i=1}^n \{s(\mathbf{X}_i) - Y_i\}^2 + \lambda\mathcal{E}_\upsilon(s). \tag{2.2}$$

If $\widehat{m}_\lambda \in \mathcal{S}$ is the minimizer of (2.2), we call it the BPSOT estimator of m corresponding to λ. It is easy to see that $\widehat{m}_0 := \widehat{m}$ is the standard unpenalized least squares spline estimator. The tuning parameter λ controls the smoothness of the fitted spline function.

If $\mathcal{B} := B(\Omega)$ is the space of all bounded real-valued functions over $\Omega = \cup_{\tau\in\triangle}\tau$ equipped with the inner product $n\langle f, g\rangle_{n,\Omega} + \lambda\langle f, g\rangle_{\mathcal{E}_\upsilon}$ with

$$\langle f, g\rangle_{\mathcal{E}_\upsilon} = \sum_{i+j=\upsilon}\binom{\upsilon}{i}\sum_{\tau\in\triangle}\int_\tau (D^i_{x_1}D^j_{x_2}f)(D^i_{x_1}D^j_{x_2}g)dx_1dx_2,$$

then $P_\lambda : \mathcal{B} \mapsto \mathcal{S}$ defined by $P_\lambda Y = \widehat{m}_\lambda$ is a linear operator that is not in general a linear projection. Now we have $P_\lambda Y = P_\lambda m + P_\lambda\epsilon$, where $P_\lambda m$ and $P_\lambda\epsilon$ are the penalized spline estimators based on $\{m(\mathbf{X}_i)\}_{i=1}^n$ and $\{\epsilon_i\}_{i=1}^n$, respectively. Under some conditions (von Golitchek and Schumaker (2002a) and Huang (2003b)), P_0 is a bounded operator on $\mathcal{S}$ in the maximum norm. Indeed, these conditions can be described as follows: for every $s \in \mathcal{S}$ and every $\tau \in \triangle$, there exist a positive constant F_1, independent of s and τ, such that

$$F_1\|s\|_{\infty,\tau} \leq \Big\{\sum_{\mathbf{X}_i\in\tau,\ i=1,\cdots,n} s(\mathbf{X}_i)^2\Big\}^{1/2}, \quad \text{for all } \tau \in \triangle. \tag{2.3}$$

If F_2 is the largest among the numbers of observations in triangles $\tau \in \triangle$, we have

$$\Big\{\sum_{\mathbf{X}_i\in\tau,\ i=1,\cdots,n} s(\mathbf{X}_i)^2\Big\}^{1/2} \leq F_2\|s\|_{\infty,\tau}, \quad \text{for all } \tau \in \triangle, \tag{2.4}$$

where $\|s\|_{\infty,\tau}$ denotes the supremum norm of s over triangle τ. The constants F_1 and F_2 were used in von Golitchek and Schumaker (2002a) to describe one of the assumptions for the boundedness of $\|P_\lambda\|$ in the supremum norm. These are also associated with the Condition A.1 in Huang (2003b).

In the following, we write $s_{\lambda,m} = P_\lambda m$ and $s_{\lambda,\epsilon} = P_\lambda \epsilon$. Then, for the penalized spline estimator $\widehat{m}_\lambda$ in (2.2), we have the decomposition

$$\widehat{m}_\lambda(\mathbf{x}) - m(\mathbf{x}) = \{s_{\lambda,m}(\mathbf{x}) - m(\mathbf{x})\} + s_{\lambda,\epsilon}(\mathbf{x}), \tag{2.5}$$

where $s_{\lambda,m}(\mathbf{x}) - m(\mathbf{x})$ and $s_{\lambda,\epsilon}(\mathbf{x})$ are referred to as the bias and noise terms.

2.3. Penalty parameter selection

An important issue for application of the BPSOT smoothing is the selection of the penalty parameter λ. A standard possibility is to select λ using K-fold cross-validation. The original sample is randomly partitioned into K subsamples, and one subsample is retained as test set and the remaining $K-1$ subsamples are used as training set. The cross-validation process is then repeated K times (the folds), with each of the K subsamples used exactly once as the validation data. Let $k[i]$ be the part containing the ith observation. For any λ, let $\hat{m}_\lambda^{-k[i]}$ be the estimate of the mean with the measurements of the $k[i]$th part of the data points removed. Then the K-fold cross-validation score is

$$CV_\lambda = \sum_{i=1}^{n} \left\{ Y_i - \hat{m}_\lambda^{-k[i]}(\mathbf{X}_i) \right\}^2.$$

We select λ by minimizing CV_λ. In our simulation studies below, we have used $K = 10$ in the numerical examples.

3. Theoretical Results

To measure the smoothness of a function, we use the standard Sobolev space $W^{\ell,\infty}(\Omega) = \{f : |f|_{k,\infty,\Omega} < \infty, 0 \le k \le \ell\}$. Given random variables T_n for $n \ge 1$, we write $T_n = O_P(b_n)$ if $\lim_{c\to\infty} \limsup_n P(|T_n| \ge cb_n) = 0$. Similarly, we write $T_n = o_P(b_n)$ if $\lim_n P(|T_n| \ge cb_n) = 0$, for any constant $c > 0$. Also, we write $a_n \asymp b_n$ if there exist two positive constants c_1, c_2 such that $c_1|a_n| \le |b_n| \le c_2|a_n|$, for all $n \ge 1$.

3.1. Convergence rate

Our results rely on the following conditions.

(A1) The bivariate function $m \in W^{\ell+1,\infty}(\Omega)$ for an integer $\ell \ge 1$.

(A2) The noise ϵ satisfies $\lim_{\eta\to\infty} E\left[\epsilon^2 I(\epsilon > \eta)\right] = 0$. The standard deviation function $\sigma(\mathbf{x})$ is continuous on Ω and $0 < c_\sigma \le \inf_{\mathbf{x}\in\Omega} \sigma(\mathbf{x}) \le \sup_{\mathbf{x}\in\Omega} \sigma(\mathbf{x}) \le C_\sigma < \infty$.

(A3) The constant F_1 in (2.3) is positive.

(A4) The number N of the triangles and the sample size n satisfy $N = Cn^{\gamma}$ for some constant $C > 0$ and $\gamma < 1$.

Remark 1. We note that (A1) and (A2) are standard conditions for nonparametric models, while (A3) ensures the existence of a discrete least squares spline. When studying the convergence of bivariate penalized least squares splines, we need (A3) though one can get a decent penalized least squares spline fitting with F_1 zero for some triangles. Condition (A4) ensures the asymptotic equivalence of the theoretical and empirical inner products/norms defined in Section 2.2.

According to (2.5), the convergence rate of $\widehat{m}_\lambda$ depends on the size of the bias and noise terms. We proceed to bound them.

Proposition 1. *Under* (A1), (A3), *and* (A4), *if* $d \geq 3r+2$ *and* $\triangle$ *is a* β *quasi-uniform triangulation, then*

$$\|s_{\lambda,m} - m\|_{\infty,\Omega} = O_P\left\{\frac{\lambda}{n\,|\triangle|^3}|m|_{2,\infty,\Omega} + \left(1 + \frac{\lambda}{n\,|\triangle|^5}\right)\frac{F_2}{F_1}\,|\triangle|^{\ell+1}\,|m|_{\ell+1,\infty,\Omega}\right\}.$$

Proposition 2. *Under* (A2) *and* (A4), $\|s_{\lambda,\epsilon}\|_{L^2(\Omega)} = O_P\left(1/(\sqrt{n}|\triangle|)\right)$.

Proposition 3. *Under* (A2) *and* (A4), $\|s_{\lambda,\epsilon}\|_{\infty,\Omega} = O_P\{(\log n)^{1/2}/(\sqrt{n}|\triangle|) + \lambda/(n|\triangle|^3)\}$.

It is easy to see that $\|s_{\lambda,m} - m\|_{L^2(\Omega)}$ has the same order as $\|s_{\lambda,m} - m\|_{\infty,\Omega}$. Our result provides the convergence rate of the BPSOT estimator, $\widehat{m}_\lambda$, in terms of the L^2 and supremum norms.

Theorem 1. *Under* (A1)−(A3), *if* $d \geq 3r+2$ *and* $\triangle$ *is a* β *quasi-uniform triangulation, we have*

$$\begin{aligned}\|\widehat{m}_\lambda - m\|_{L^2(\Omega)} = O_P\Bigg\{&\frac{\lambda}{n\,|\triangle|^3}|m|_{2,\infty,\Omega} + \left(1 + \frac{\lambda}{n\,|\triangle|^5}\right)\frac{F_2}{F_1}|\triangle|^{\ell+1}|m|_{\ell+1,\infty,\Omega} \\ &+ \frac{1}{\sqrt{n}|\triangle|}\Bigg\}, \\ \|\widehat{m}_\lambda - m\|_{\infty,\Omega} = O_P\Bigg\{&\frac{\lambda}{n\,|\triangle|^3}|m|_{2,\infty,\Omega} + \left(1 + \frac{\lambda}{n\,|\triangle|^5}\right)\frac{F_2}{F_1}|\triangle|^{\ell+1}|m|_{\ell+1,\infty,\Omega} \\ &+ \frac{(\log n)^{1/2}}{\sqrt{n}|\triangle|} + \frac{\lambda}{n|\triangle|^3}\Bigg\}.\end{aligned}$$

Remark 2. Assume $F_2/F_1 = O(1)$ as in (A3') below. When $\lambda = 0$, if $N \asymp n^{1/(\ell+2)}$, then the unpenalized spline estimator $\widehat{m}_0$ satisfies $\|\widehat{m}_0 - m\|^2_{L^2(\Omega)} = O_P(n^{-(\ell+1)/(\ell+2)})$, the optimal convergence rate given in Stone (1982). If $N \asymp$

$(n/\log n)^{1/(\ell+2)}$, then $\|\widehat{m}_0 - m\|^2_{\infty,\Omega} = O_P\{(n^{-1}\log n)^{(\ell+1)/(\ell+2)}\}$, the optimal rate of convergence for the supremum norm. For the penalized estimator $\widehat{m}_\lambda$ ($\lambda > 0$), if $\lambda = o(n^{\ell/2(\ell+2)})$ and N is of the same order as in the unpenalized case, we still have the optimal convergence rate.

3.2. Asymptotic normality

To derive the asymptotic normality of the BPSOT estimator, we bring in further conditions.

(A3') The constants F_1 and F_2 at (2.3) and (2.4) satisfy $F_2/F_1 = O(1)$.

(A4') $N = Cn^\gamma$ for some constant $C > 0$ and $1/(\ell+2) < \gamma < 1$.

(A5) $\lambda = o(n^{1/2}N^{-1})$.

Remark 3. Recall that F_2 is the maximum of the number of locations in triangles in $\triangle$, and (A3') suggests that we should not put too many observations in one triangle as the larger F_2 the larger the approximation constant. Compared with (A4), (A4') further requires that the number of triangles be above some minimum depending upon the degree of the spline, which is similar to the requirement of Li and Ruppert (2008) in the univariate case. In smoothing, there is usually a fundamental trade-off between the bias and noise terms of the estimate. However, deriving an explicit expression for the asymptotic bias of the BPSOT estimator is a challenging and unsolved problem even in pure approximation theory for noise-free data. Two factors that affect the smoothness of the fitted surface are the number of triangles and the penalty term. When the triangulation is finer, the bias is lower but the variance is higher. Larger values of λ give more weight to the penalty term, leading to fitted surfaces with smaller variance but higher bias. Our strategy here is to reduce the bias through "undersmoothing" and "choosing smaller λ". Conditions (A3'), (A4'), and (A5) give a sufficient condition for the bias term to be negligible compared with the noise term.

Let $\mathbb{X}$ be the collections of $\mathbf{X}_i$, $i = 1, \ldots, n$. The next theorem gives upper and lower bounds on the conditional variance of the noise term $s_{\lambda,\epsilon}(\mathbf{x})$ for each $\mathbf{x} \in \Omega$.

Theorem 2. *Fix $\mathbf{x} \in \Omega$. Under (A1) – (A4), if $\lambda = o(n/N^2)$ we have, with probability approaching 1 as $n \to \infty$,*

$$\frac{C_1 c_\sigma^2}{n(1 + (n|\triangle|^4)^{-1}\lambda)^2|\triangle|^2} \le \mathrm{Var}\,(s_{\lambda,\epsilon}(\mathbf{x})|\mathbb{X}) \le \frac{C_2 C_\sigma^2}{n|\triangle|^2}$$

for positive constants C_1 and C_2 that depend only on β.

Remark 4. According to Theorem 1.1 of Lai and Schumaker (2007),

$$\begin{aligned}\|g\|_\infty = \|g\|_{\infty,\tau} &\leq K A_\tau^{-1/2} \|g\|_{L^2(\tau)} \leq \frac{K}{2\rho_\triangle} \|g\|_{L^2(\tau)} \leq \frac{K}{2\rho_\triangle} \|g\|_{L^2(\Omega)} \\ &\leq \frac{K_\beta}{|\triangle|} \|g\|_{L^2(\Omega)},\end{aligned}$$

where A_τ stands for the area of triangle τ and K_β is a positive constant dependent on β. This implies that $H_n := \sup_{g\in\mathcal{S}} \|g\|_{\infty,\Omega} / \|g\|_{L^2(\Omega)} \leq K_\beta / |\triangle|$. Thus, for $\lambda = 0$, our result is comparable with Corollary 3.1 in Huang (2003a), where the supreme conditional variance is of the order $n^{-1} H_n^2 \sigma^2 \{1 + o_P(1)\}$.

Theorem 3. *Under* (A1), (A2), (A3'), (A4') *and* (A5), *as* $n \to \infty$, *for each* $\mathbf{x} \in \Omega$,

$$\frac{\widehat{m}_\lambda(\mathbf{x}) - m(\mathbf{x})}{\sqrt{\mathrm{Var}(\widehat{m}_\lambda(\mathbf{x}) | \mathbb{X})}} \Rightarrow N(0,1).$$

This asymptotic distribution can be used to construct asymptotic confidence intervals. For example, if we estimate $m(\mathbf{x})$ using piecewise constant splines, Lemma 5 in Section S3 (online supplement) gives the size of the pointwise variance

$$\mathrm{Var}\{\widehat{m}_\lambda(\mathbf{x})\} = \frac{\sigma^2(\mathbf{x})}{n f(\mathbf{x}) A_\tau}(1 + o(1)), \qquad \mathbf{x} \in \Omega,$$

so that an asymptotic $100(1-\alpha)\%$ pointwise confidence envelope for $m(\mathbf{x})$ over Ω is

$$\widehat{m}_\lambda(\mathbf{x}) \pm z_{\alpha/2} \frac{\sigma(\mathbf{x})}{\{n f(\mathbf{x}) A_\tau\}^{1/2}},$$

where f stands for the density function of $\mathbf{X}$.

Theorem 3 can be strengthened to hold uniformly.

Theorem 4. *Set*

$$D_n = \sup_z \left| P\left\{ \frac{\widehat{m}_\lambda(\mathbf{x}) - m(\mathbf{x})}{\sqrt{\mathrm{Var}(\widehat{m}_\lambda(\mathbf{x}))}} \leq z \,\middle|\, \mathbb{X} \right\} - \Phi(z) \right|,$$

where $\Phi(z)$ *is the standard normal. Under the conditions of Theorem 3, if* $\lim_{\eta\to\infty} E\left[\epsilon^2 I\{\epsilon^2 \geq \eta E(\epsilon^2)\}\right] / E(\epsilon^2) = 0$, *one has* $\sup_{\mathbf{x}\in\Omega} P(D_n > \eta) = o(1)$ *for any* $\eta > 0$, *and then*

$$\sup_{\mathbf{x}\in\Omega, z\in\mathbb{R}} \left| P\left\{ m(\mathbf{x}) \geq \widehat{m}_\lambda(\mathbf{x}) - z\sqrt{\mathrm{Var}(\widehat{m}_\lambda(\mathbf{x}) | \mathbb{X})} \right\} - \Phi(z) \right| = o(1).$$

When $\lambda = 0$, the result recovers Theorem 4.1 in Huang (2003b). This uniform asymptotic normality result can be used to construct asymptotic confidence envelops whose coverage probability converges uniformly to its nominal level.

4. Simulation

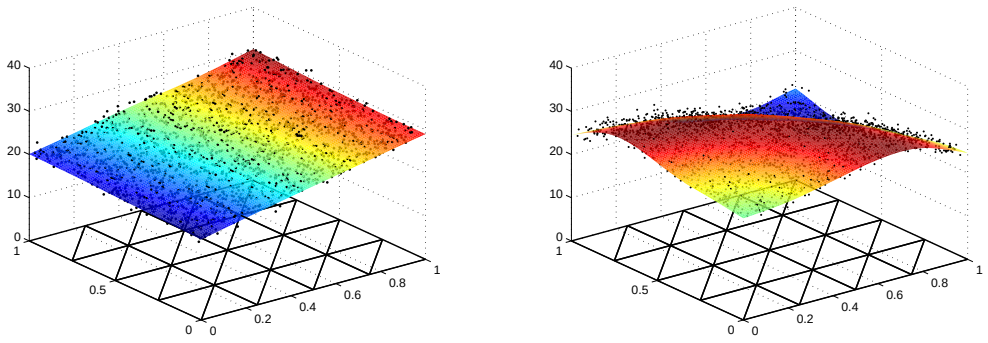
In this section, we provide numerical results from our preliminary computational experiments to compare the performance of BPSOT with thin plate spline (TPS). We wrote our BPSOT codes in MATLAB based on the computational algorithms given in Awanou, Lai, and Wenston (2006), and these programs are available on `www.math.uga.edu/~mjlai/2Dsplines.html`.

The TPS smoothing method has been a standard tool for high-dimensional data smoothing. In our examples, we also implemented the TPS for comparison. We used the function `Tps()` in R package `fields` (Furrer, Nychka, and Sainand (2011)) to produce the TPS fit (TPS1). We also included a fast (low-rank approximation) version TPS2 (Wood (2003)) using the function `gam()` in R package `mgcv`. In the examples, we report the simulation results for BPS, TPS1, and TPS2 with various ranks.

To obtain the BPSOT estimators, one first needs to determine the spline space S_d^r. A general principle here is to use a degree $d \geq 3r + 2$ for $r \geq 1$. One could also use a degree $d < 5$ for $r = 1$ if the triangulation $\triangle$ is special; see a summary of special triangulations in Lai and Schumaker (2007). Throughout the simulation study, we set $r = 1$. In the examples, we first found a polygon containing all the design points; next we triangulated it to obtain an initial triangulation and repeatedly applied a suitable number of uniform refinement (see details below). To select our penalty parameter λ, we used K-fold cross validation. For TPS1 and TPS2, λ was selected by GCV. Each smoothing method was evaluated on 100 randomly generated independent datasets.

Example 1. We generated the data from a rectangular domain $[0, 1]^2$ using the regression model $Y = m(X_1, X_2) + \epsilon$. We considered a linear regression $m(x_1, x_2) = 10x_1 + x_2 + 19$, and the sinusoid $m(x_1, x_2) = 24 + 5\sin\{\pi(x_1^2 + x_2^2)\}$, with standard Gaussian random noises. We created a 101×101-point grid with values evenly spaced between 0 and 1. We obtained the true signal and noisy observation for each coordinate pair (x_1, x_2) lying on the grid in the unit square. Next we took a random sample of size n =1,000, 2,000, 3,000, 4,000 from the 101×101 points.

In this example, we used the triangulation given at the bottom of Figure 3. There are 32 triangles and 25 vertices in each triangulation. For each case we solved for the BPSOT estimator, $\hat{m}_\lambda$, with respect to different sample sizes. Figure 3 illustrates an example of the BPSOT estimated surface based on n =2,000 observations.



(a) (b)

Figure 3. The triangulations and BPSOT estimators (a) the linear surface and (b) of the sinusoid surface.

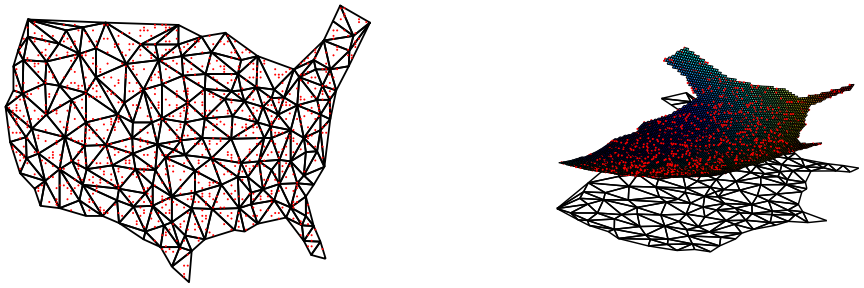
Table 1. Averages of the RMSEs in Example 1

Model	Method	Sample Size			
		1000	2000	3000	4000
Linear	BPSOT ($d = 5$)	0.0502	0.0403	0.0307	0.0287
	TPS1	0.0540	0.0438	0.0307	0.0298
	TPS2 ($k = 10$)	0.0524	0.0412	0.0314	0.0301
	TPS2 ($k = 30$)	0.0530	0.0439	0.0322	0.0307
	TPS2 ($k = 50$)	0.0539	0.0431	0.0321	0.0310
Sinusoid	BPSOT ($d = 5$)	0.2142	0.1705	0.1369	0.1205
	TPS1	0.2198	0.1715	0.1443	0.1303
	TPS2 ($k = 10$)	0.7498	0.7443	0.7420	0.7409
	TPS2 ($k = 30$)	0.2342	0.2016	0.1846	0.1773
	TPS2 ($k = 50$)	0.2143	0.1689	0.1429	0.1315

To see the accuracy of the data fitting, we evaluated the BPSOT fit over the 101×101 grid points on the domain compared with the true function. We ran 100 Monte Carlo samples under each model, and for each replication we calculated the root mean square errors (RMSE)

$$\text{RMSE} = \Big[\frac{1}{101^2}\sum_{i=1}^{101^2}\{\hat{m}_\lambda(\mathbf{x}_i) - m(\mathbf{x}_i)\}^2\Big]^{1/2}.$$

Table 1 shows the average RMSE based on the 100 replications for BPSOT, TPS1, and TPS2. For TPS2, we smoothed the surface using rank $k = 10, 30, 50$ thin plate regression spline bases. From Table 1, one sees that the RMSE for all the methods decreases as sample size increases. As expected, the BPSOT is



(a) (b)

Figure 4. (a) a triangulation for the domain delimited by the US frontiers with $n = 2{,}000$ random locations; (b) the BPSOT surface fitting.

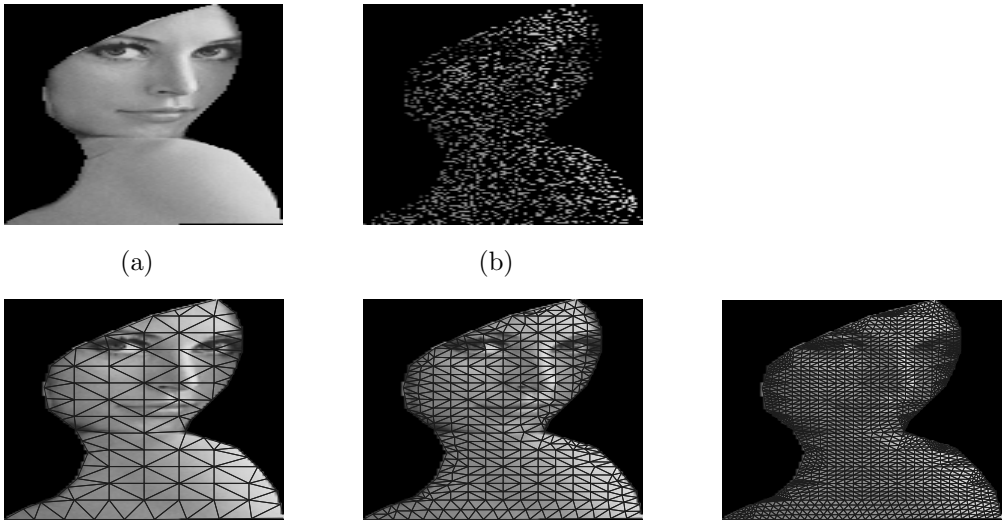
Table 2. Averages of the RMSEs in Example 2

Method	Sample Size			
BPSOT ($d = 5$)	0.2232	0.1688	0.1276	0.1152
TPS1	0.2793	0.1729	0.1464	0.1319
TPS2 ($k = 10$)	0.5785	0.5575	0.5516	0.5495
TPS2 ($k = 30$)	0.2381	0.2004	0.1868	0.1772
TPS2 ($k = 50$)	0.2222	0.1680	0.1440	0.1313
TPS2 ($k = 100$)	0.2274	0.1708	0.1443	0.1300

statistically indistinguishable from TPS1 and TPS2 in terms of RMSE because the data were generated from a rectangular domain. From our simulation, we also found TPS2 far superior in terms of computing speed. For the same accuracy, TPS2 could be 100 times faster than BPSOT and/or TPS1.

Example 2. In this example, we compared the BPSOT, TPS1, and TPS2 estimators on a 101×101-point grid with values equally spaced on the rectangular domain containing the U.S. frontiers. We removed those points lying outside the U.S. leaving 6,132 points. We took the test function $m(x_1, x_2) = \{(x_1 - 30)^2 + (x_2 - 40)^2\}/25 + 20$, and sampled it at n randomly chosen points, with function values at these points were perturbed with standard Gaussian random noise. We generated 100 replicate data sets of size n =1,000, 2,000, 3,000, 4,000. We used the triangulation $\triangle$ in Figure 4(a) to obtain the BPSOT smoother. There are 302 triangles and 174 vertices in this triangulation.

We examined the three smoothing methods, and we used rank $k = 10, 30, 50,$ 100 for TPS2. An example of the estimated BPSOT surface is shown in Figure 4(b), created based on sample size 2,000.



(a) (b)

(c) (d) (e)

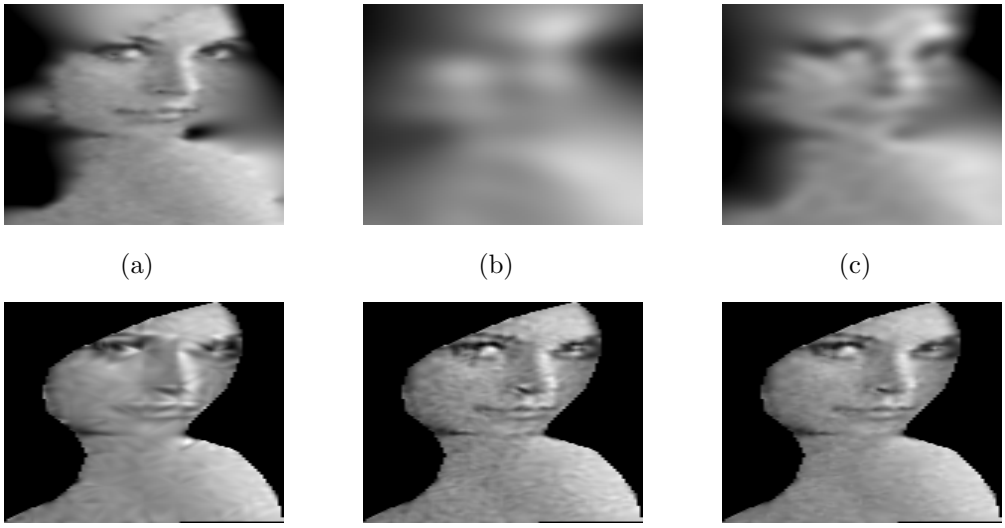
Figure 5. The face of Lena and its triangulation: (a) noise-free image; (b) noisy image based on $\sigma = 5$ and n =2,000; (c) triangulation $\triangle_0$, (d) triangulation $\triangle_1$, and (e) triangulation $\triangle_2$.

Table 2 reports the average RMSE of the methods over the equally-spaced grid points located inside the U.S.. BPSOT is clearly superior, in terms of RMSE, to TPS1 and TPS2.

Example 3. We consider smoothing a dataset of an image, the standard test image Lena Sjooblom, as an example. The noise-free image is 155×94 pixel and we cropped the picture as illustrated in Figure 5 (a). Among the 8,401 pixels on the cropped picture, we randomly selected n pixels and added Gaussian noises with $\sigma = 5$ to the gray scale values over the selected pixels. We replicated this 100 times to obtain 100 noisy images. An example of the noisy image is shown in Figure 6(b) with n =2,000.

We triangulated the domain to get the initial triangulation $\triangle_0$ shown in Figure 5(c), then uniformly refined it to get $\triangle_1$ and $\triangle_2$ as shown in Figure 5(d) and (e). We used the BPSOT method with triangulations $\triangle_0$, $\triangle_1$, and $\triangle_2$ to smooth the noisy gray scale image values. After finding the penalized splines, we evaluated the fittings over all 8401 data points and computed the average RMSE and PSNR (Peak signal-to-noise ratio) PSNR $= 20 \log_{10}(255/\text{RMSE})$ to measure the goodness of fit. The resulting BPSOT images for one random sample are shown in Figure 6(d), (e), and (f) based on three triangulations.

For the same noisy image we used TPS1 and TPS2, the recovered images are shown in Figure 6(a)−(c). According to the definition of TPS, it uses a



(a) (b) (c)

(d) (e) (f)

Figure 6. (a) TPS1 fit; (b) TPS2 fit ($k = 50$); (c) TPS2 fit ($k = 200$); (d) BPSOT ($d = 5$) fit based on $\triangle_0$; (e) BPSOT ($d = 5$) fit based on $\triangle_1$; and (f) BPSOT ($d = 5$) fit based on $\triangle_2$.

minimization defined over the entire $\mathbb{R}^2$ space. When the observations are located within a bounded domain Ω with irregular boundary, there is a leakage of the total thin-plate energy to the outside of Ω. This explains the blur of the fitted image near the boundary of the data domain as shown in Figure 6(a)$-$(c). Although one can define a TPS by minimizing thin-plate energy over Ω instead of $\mathbb{R}^2$, one has to find the kernel function that depends on Ω. When Ω has an irregular boundary, such a kernel is difficult to find. This explains why BPSOT are more flexible and convenient for data fitting than the TPS.

Tables 3 and 4 present the RMSE and PSNR of the three smoothing methods for sample sizes 1,000, 2,000 and 3,000. For BPSOT, the triangulation $\triangle_2$ was applied with various degree $d \geq 5$. For TPS2, we considered the ranks $k = 10, 30, 50, 100, 200$. From the two tables, one sees that TPS1 and BPSOT have a similar performance when a small degree d of BPSOT was applied. The performance of BPSOT improved when d increased, and BPSOT outperformed TPS1 for sufficiently large d. In contrast, TPS2 performed much worse than BPSOT and TPS1 even when we increased the rank k to be 200.

From our simulation, we also observed that the computing time of the TPS1 depended critically on sample size while the computing time of BPSOT was primarily based on the number of triangles and the spline degree. In this image example, TPS1 could smooth a data set of size less than 5,000 reasonably quickly,

Table 3. Averages of the RMSEs in Example 3

Method	Sample Size		
	1000	2000	3000
BPSOT (d=5)	12.8019	9.8526	8.2467
BPSOT ($d = 6$)	12.7094	9.949*	8.100*
BPSOT ($d = 7$)	12.6633	9.866*	8.098*
BPSOT ($d = 8$)	12.6202	9.785*	8.094*
BPSOT ($d = 9$)	12.6168	9.659*	8.083*
BPSOT ($d = 10$)			8.099*
BPSOT ($d = 11$)			8.083*
TPS1	12.6844	9.7028	8.0880
TPS2 ($k = 10$)	27.2443	27.1591	27.1381
TPS2 ($k = 30$)	20.9694	20.7907	20.7922
TPS2 ($k = 50$)	19.2849	19.0611	18.9436
TPS2 ($k = 100$)	17.1238	16.6686	16.4991
TPS2 ($k = 200$)	15.2322	14.3763	14.1055

* $\lambda = 0.05$ was fixed so that the simulation could be done much faster than using the K-fold cross validation.

Table 4. Averages of the PSNRs in Example 3

Method	Sample Size		
	1000	2000	3000
BPSOT ($d = 5$)	25.993	28.165	29.809
BPSOT ($d = 6$)	26.056	28.180*	29.964*
BPSOT ($d = 7$)	26.088	28.252*	29.966*
BPSOT ($d = 8$)	26.111	28.329*	29.975*
BPSOT ($d = 9$)	26.120	28.436*	29.981*
TPS1	26.075	28.398	29.978
TPS2 ($k = 10$)	19.425	19.452	19.459
TPS2 ($k = 30$)	21.699	21.774	21.773
TPS2 ($k = 50$)	22.427	22.528	22.582
TPS2 ($k = 100$)	23.459	23.693	23.782
TPS2 ($k = 200$)	24.478	24.979	25.143

* $\lambda = 0.05$ was fixed so that the simulation could be done much faster than using the K-fold cross validation.

but crashed when the size of the dataset increased to 8,000. Our BPSOT program can smooth the 8,000 observations easily.

5. Discussion

The BPSOT-smoother is a good choice when data are located in domains with complex boundaries and/or possible holes. In this paper, we have established the consistency and asymptotic normality of the BPSOT estimators, and

studied their convergence rates. The proposed BPSOT estimator worked well in smoothing various simulated datasets.

Triangulation is a convenient tool to partition domain of arbitrary shape. In practice, one would like to use a minimal number of triangles to fit the data well. A basic strategy is to find a polygonal domain Ω that includes all the data locations, and to create a triangulation $\triangle$ of Ω such that each triangle is as regular as possible in the sense that a small value of the shape parameter β at (2.1) holds for all triangles.

Acknowledgement

The research work of the first author was partially supported by NSF grant DMS-07-13807. The research work of the second author was supported in part by NSF grants DMS-09-05730 and DMS-11-06816. The authors thank the Editor, an associate editor, and two reviewers for their constructive comments.

References

Awanou, G., Lai, M. J. and Wenston, P. (2006). The multivariate spline method for numerical solution of partial differential equations and scattered data interpolation. In *Wavelets and Splines* (Edited by G. Chen and M. J. Lai), 24-74, Nashboro Press.

Baramidze, V., Lai, M. J. and Shum, C. K. (2006). Spherical splines for data interpolation and fitting. *SIAM J. Scientific Computing* **28**, 241-259.

Claeskens, G., Krivobokova, T. and Opsomer, J. (2008). Asymptotic properties of penalized spline regression estimators. *Biometrika* **96**, 529-544.

Ettinger, B., Guillas, S. and Lai, M. J. (2012). Bivariate splines for functional regression models with application to ozone forecasting. *Environmetrics* **23**, 317-328.

Furrer, R., Nychka, D. and Sainand, S. (2011). Package 'fields'. R package version 6.6.1. `http://cran.r-project.org/web/packages/fields/fields.pdf`.

Guillas, S. and Lai, M. J. (2010). Bivariate splines for spatial functional regression models. *J. Nonparametr. Stat.* **22**, 477-497.

Hall, P. and Opsomer, J. D. (2005). Theory for penalized spline regression. *Biometrika* **92**, 105-118.

Hansen, M., Kooperberg, C. and Sardy, S. (1998). Triogram models. *J. Amer. Statist. Assoc.* **93**, 101-119.

Huang, J. Z. (1998). Projection estimation for multiple regression with application to functional ANOVA models. *Ann. Statist.* **26**, 242-272.

Huang, J. Z. (2003a). Asymptotics for polynomial spline regression under weak conditions. *Statist. Prob. Lett.* **65**, 207-216.

Huang, J. Z. (2003b). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31**, 1600-1635.

Kooperberg, C., Stone, C. J. and Truong, Y. K. (1995a). The L_2 rate of convergence for hazard regression. *Scand. J. Statist.* **22**, 143-157.

Kooperberg, C., Stone, C. J. and Truong, Y. K. (1995b). Rate of convergence for logspline spectral density estimation. *J. Time Ser. Anal.* **16**, 389-401.

Lai, M. J. (2008). Multivariate splines for data fitting and approximation. *Approximation Theory XII* (Edited by M. Neamtu and L. L. Schumaker), 210-228. Nashboro Press.

Lai, M. J. and Schumaker, L. L. (1998). Approximation power of bivariate splines. *Adv. Comput. Math.* **9**, 251-279.

Lai, M. J. and Schumaker, L. L. (2007). *Spline Functions on Triangulations*. Cambridge Univ. Press.

Lai, M. J., Shum, C. K., Baramidze, V. and Wenston, P. (2009). Triangulated spherical splines for geopotential reconstruction. *J. Geodesy* **83**, 695-708.

Lai, M. J. and Wenston, P. (2004). Bivariate Splines for Fluid Flows, *Computers and Fluids*, **33**, 1047-1073

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95**, 415-436.

Marx, B. D. and Eilers, P. H. C. (2005). Multidimensional penalized signal regression. *Technometrics* **47**, 13-22.

Persson, P. O. and Strang, G. (2004). A simple mesh generator in MATLAB. *Siam Rev.* **46**, 329-345.

Ramsay, T. (2002). Spline smoothing over difficult regions. *J. Roy. Statist. Soc. Ser. B* **64**, 307-319.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *J. Comp. Graph. Statist.* **11**, 735-757.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 689-705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 590-606.

Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* **22**, 118-184.

von Golitchek, M. and Schumaker, L. L. (2002a). Bounds on projections onto bivariate polynomial spline spaces with stable local bases. *Constr. Approx.* **18**, 241-254.

von Golitchek, M. and Schumaker, L. L. (2002b). Penalized least squares fitting. *Serdica Math. J.* **18**, 1001-1020.

Wand, M. P. (1999). On the optimal amount of smoothing in penalised spline regression. *Biometrika* **86**, 936-940.

Wood, S. N. (2003). Thin plate regression splines. *J. Roy. Statist. Soc. Ser. B*, **65**, 95-114.

Xiao, L., Li, Y. and Ruppert, D. (2010). Bivariate penalized splines. Manuscript. `http://arxiv.org/pdf/1011.4916.pdf`.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26**, 1760-1782.

Department of Mathematics, University of Georgia, Athens, GA 30602, U.S.A.

E-mail: mjlai@math.uga.edu

Department of Statistics, University of Georgia, Athens, GA 30602, U.S.A.

E-mail: lilywang@uga.edu

(Received November 2010; accepted August 2012)