Biometrika Advance Access published June 24, 2015

(2015), 1–15 doi: 10.1093/biomet/asv009

Efficient computation of smoothing splines via adaptive basis sampling

BY PING MA

pingma@uga.edu

JIANHUA Z. HUANG AND NAN ZHANG

jianhua@stat.tamu.edu nanzhang@stat.tamu.edu

SUMMARY

Smoothing splines provide flexible nonparametric regression estimators. However, the high computational cost of smoothing splines for large datasets has hindered their wide application. In this article, we develop a new method, named adaptive basis sampling, for efficient computation of smoothing splines in super-large samples. Except for the univariate case where the Reinsch algorithm is applicable, a smoothing spline for a regression problem with sample size can be expressed as a linear combination of basis functions and its computational complexity is generally (3). We achieve a more scalable computation in the multivariate case by evaluating the smoothing spline using a smaller set of basis functions, obtained by an adaptive sampling scheme that uses values of the response variable. Our asymptotic analysis shows that smoothing splines computed via adaptive basis sampling converge to the true function at the same rate as full basis smoothing splines. Using simulation studies and a large-scale deep earth core-mantle boundary imaging study, we show that the proposed method outperforms a sampling method that does not use the values of response variables.

: Bayesian confidence interval; Core-mantle boundary; Nonparametric regression; Penalized least squares; Reproducing kernel Hilbert space; Sampling.

1. INTRODUCTION

Consider the nonparametric regression model

$$ = \eta( ) + \epsilon \qquad (= 1, \ldots,), \tag{1}$$

where is the th observation of the response variable, is the th observation of the predictor variable on the domain $\mathcal{X} \subset \mathbb{R}$ ($\geqslant 1$), η is the nonparametric function to be estimated, and the ϵ s are independent and identically distributed random errors with mean zero and unknown constant variance σ^2. A widely used method for estimating the unknown function η in (1) is via

minimization of the penalized least squares criterion

$$\text{PLS}(\eta) = \frac{1}{} \sum_{=1} \{ - \eta( )\}^2 + \lambda \quad (\eta), \tag{2}$$

where (η) is a quadratic functional quantifying the roughness of η. The first term on the right of (2) discourages lack of fit, and the second term penalizes the roughness of η. The penalty parameter λ controls the trade-off between the goodness-of-fit and smoothness of η. Multivariate penalty parameters can be introduced when estimating a multivariate function, but we focus on the single penalty case. See Wahba (1990), Gu (2013) and Wang (2011) for overviews of this method, including how to introduce multivariate penalty parameters.

The standard formulation of smoothing splines performs the minimization of (2) in a reproducing kernel Hilbert space $\mathcal{H} = \{\eta : (\eta) < \infty\}$, where $(\cdot)$ is a squared semi-norm. Let $\mathcal{N} = \{\eta : (\eta) = 0\}$ be the null space of (η) and assume that $\mathcal{N}$ is a finite-dimensional linear subspace of $\mathcal{H}$ with basis $\{\xi : = 1, \ldots, \}$, where $= \dim(\mathcal{N})$. Denote by $\mathcal{H}$ the orthogonal complement of $\mathcal{N}$ in $\mathcal{H}$ such that $\mathcal{H} = \mathcal{N} \oplus \mathcal{H}$. Let be the orthogonal projection operator from $\mathcal{H}$ onto $\mathcal{H}$. Then $(\cdot)$ is a well-defined squared norm of $\mathcal{H}$ and for any $\eta \in \mathcal{H}$, $(\eta) = (\eta) = \| \eta\|^2_{\mathcal{H}}$. With this norm, $\mathcal{H}$ is also a reproducing kernel Hilbert space, and we denote its reproducing kernel by $(\cdot, \cdot)$.

The reproducing kernel Hilbert space provides a very general framework for nonparametric regression where the penalty term (η) can be chosen to serve different purposes. For univariate function estimation on a compact interval $\mathcal{X}$, one can use

$$(\eta) = \int_{\mathcal{X}} (\eta^{( )})^2 \, \mathrm{d} .$$

In particular, $= 2$ corresponds to the commonly-used second derivative penalty and the minimizer of (2) is a natural cubic spline. For estimating a multivariate function on a compact domain $\mathcal{X} \subset \mathbb{R}^{} (> 1)$, one can use the thin-plate spline penalty

$$(\eta) = \int \cdots \int_{\mathcal{X}} \sum_{_1 + \cdots + _{} = } \frac{!}{_1! \cdots _{}!} \left(\frac{\partial \eta}{\partial {}^{_1}_1 \cdots \partial {}^{_{}}_{}} \right)^2 \mathrm{d}_1 \cdots \mathrm{d}_{}, \tag{3}$$

where is the order of derivatives and is the number of predictor variables (Duchon, 1977). As a special case, when $= 2$ and $= 2$ we have

$$_{22}(\eta) = \iint_{\mathcal{X}} \left(\frac{\partial^2 \eta}{\partial {}^2_1} \right)^2 + \left(\frac{\partial^2 \eta}{\partial _1 \partial _2} \right)^2 + \left(\frac{\partial^2 \eta}{\partial {}^2_2} \right)^2 \mathrm{d}_1 \, \mathrm{d}_2.$$

See Gu (2013) for details about defining the penalty term and corresponding reproducing kernel Hilbert space for modelling a multivariate regression function using smoothing spline analysis of variance models.

Univariate smoothing splines can be computed in $( )$ operations by applying the Reinsch (1967) algorithm. In general, as we shall see in the next section, the computational cost of finding the minimizer of (2) is in the order of $(^3)$ and thus is very expensive for big datasets. To lower the computational cost, over the past decades, there have been efforts to find sparse sets of basis functions to approximate the minimizer of (2). Luo & Wahba (1997) and Zhang et al. (2004) applied variable selection techniques, but it is not clear whether the resulting estimators share the good asymptotic properties of standard smoothing splines. Gu & Kim (2002) and

Kim & Gu (2004) developed a simple random sampling approach for basis function selection and established a coherent theory for the convergence of their approximated smoothing splines. To overcome the computational burden of smoothing splines, pseudosplines (Hastie, 1996) and penalized splines (Ruppert et al., 2003) have also been proposed. Both use a small number of fixed basis functions to approximate the smoothing splines; they are similar in spirit to Gu & Kim (2002) and Kim & Gu (2004) but differ in the construction of the basis functions.

In this paper, extending the simple random sampling approach of Gu & Kim (2002) and Kim [illegible]

On the other hand, for any function η with the expansion (1), the penalty function $J(\eta)$ in (2) can also be written in a matrix form using the reproducing property of $K(\cdot,\cdot)$, i.e.,

$$\langle K(x_i,\cdot), K(x_j,\cdot)\rangle_{\mathcal{H}} = K(x_i,x_j).$$

Recall that $P:\mathcal{H}\to\mathcal{H}$ is a projection operator. For any η as in (4), $P\eta=\sum_{i=1}^{n}\alpha_i K(x_i,\cdot)$. Hence

$$\begin{aligned} J(\eta) = \|P\eta\|_{\mathcal{H}}^2 &= \left\langle \sum_{i=1}^{n}\alpha_i K(x_i,\cdot), \sum_{i=1}^{n}\alpha_i K(x_i,\cdot)\right\rangle_{\mathcal{H}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i K(x_i,x_j)\alpha_j = \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\boldsymbol{\alpha}. \end{aligned} \tag{6}$$

Combining (5) and (6), we see that the penalized least squares criterion (2) is reduced to

$$\mathrm{PLS}(\eta) = \frac{1}{n}(\mathbf{y}-\mathbf{K}\mathbf{d}-\mathbf{K}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\mathbf{K}\mathbf{d}-\mathbf{K}\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}\boldsymbol{\alpha}. \tag{7}$$

Since $\mathrm{PLS}(\eta)$ is a quadratic form in both $\mathbf{d}$ and $\boldsymbol{\alpha}$, its minimizer has a closed-form expression. Differentiating (7) with respect to $\mathbf{d}$ and $\boldsymbol{\alpha}$ and setting the derivatives to zero, we obtain the linear system of equations

$$\begin{pmatrix} \mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n & \mathbf{1}_n^{\mathrm{T}}\mathbf{K} \\ \mathbf{K}^{\mathrm{T}}\mathbf{1}_n & \mathbf{K}^{\mathrm{T}}\mathbf{K} + n\lambda\mathbf{K} \end{pmatrix} \begin{pmatrix} \mathbf{d} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_n^{\mathrm{T}}\mathbf{y} \\ \mathbf{K}^{\mathrm{T}}\mathbf{y} \end{pmatrix}.$$

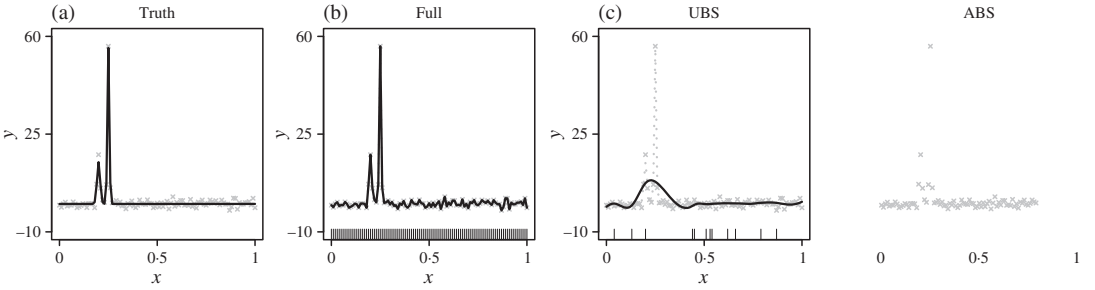
To solve this system, of size $n+1$, the computational cost is generally of the order $O(n^3)$, which can be prohibitive when the sample size n is large. From Theorem 1, the number of basis functions used to represent the solution is $q+1$, which grows with n. While the n basis functions for $\mathcal{N}$ are needed, it may not be necessary to use all n basis functions for $\mathcal{H}$. If a smaller number of basis functions can provide a good approximation of the smoothing spline solution, then a computationally efficient algorithm can be developed to handle cases with large sample size. We discuss two sampling approaches for selecting basis functions in the next section.

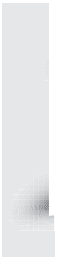
3. SAMPLING OF BASIS FUNCTIONS

3·1. Subsampling of the predictor variable

We first review an approach to selecting basis functions by randomly sampling the observations of the predictor variable and discuss its limitations, and then present our new sampling approach that involves the response variable.

From the representer theorem, each of the n basis functions for representing the function in $\mathcal{H}$ is uniquely associated with an observed value of the predictor variable. Thus a natural idea for selecting the basis functions is through randomly sampling the observed values of the predictor variable. Specifically, we draw a random sample of size n^* from the observed predictor values $\{x_i\}_{i=1}^{n}$, denoted as $\mathbf{x}^* = (x_1^*,\ldots,x_{n^*}^*)^{\mathrm{T}}$, and use the corresponding basis functions, $\{K(x_i^*,\cdot)\}_{i=1}^{n^*}$, to represent functions in $\mathcal{H}$. We then solve the penalized least squares problem in the effective model space $\mathcal{H}=\mathcal{N}\oplus\mathrm{span}\{K(x_i^*,\cdot), i=1,\ldots,n^*\}$. When n^* is much smaller than n, the computational cost can be significantly reduced.


(a)
Truth
(b)
Full
(c)
UBS
ABS
60
25
−10
y
0
0·5
1
x



3·3.

We now present the details of the computational algorithm when adaptive basis sampling is used to compute the smoothing spline estimator. Recall that the selected data points are denoted by $x^* = (x_1^*, \ldots, x_*^*)^T$. Under adaptive basis sampling, the minimizer of (2) is approximated by

$$\eta(\cdot) = \sum \cdot \xi(\cdot) + \sum^{*} \cdot \cdot (\cdot^*, \cdot).$$

We let x denote the $\times$ matrix with $(i, \cdot)$th entry $\xi_\cdot(x_i)$. Let $\cdot_*$ be a $\times$ * matrix with the $(i, \cdot)$th entry $\cdot\cdot(x_i, x_\cdot^*)$ and $\cdot_{**}$ be a $^* \times$ * matrix with the $(i, \cdot)$th entry $\cdot\cdot(x_i^*, x_\cdot^*)$. If we rearrange the original data by putting the selected data points x^* at the front, $\cdot_*$ is just the left part of $\cdot$ while $\cdot_{**}$ is the top-left corner of $\cdot$. The evaluations of η at locations x, η

and we minimize it as a function of the penalty parameter λ (Tenorio et al., 2011), using standard nonlinear optimization algorithms. We use the modified Newton algorithm developed by Dennis & Schnabel (1996).

Now we calculate the computational complexity, using the fact that $\ll * \ll$ to simplify the expressions. The construction of the linear system (8) is of the order $(*^2)$, the Cholesky decomposition takes $(*^3)$ flops, the subsequent forward and backward substitutions take $(*^2)$ flops respectively, and the evaluation of (9) requires the calculation of tr$\{ (\lambda)\}$, which takes $(*^2)$ flops. The overall computational cost is of the order $(*^2)$. The efficient computational scheme can also be used to compute Bayesian confidence intervals (Wahba, 1983); see the Supplementary Material for details.

4. CONVERGENCE RATES FOR FUNCTION ESTIMATION

4·1.

We first introduce an inner product associated with the marginal density $(\cdot)$ of the predictor variable . For any $_1$ and $_2$ in $\mathcal{L}_2(\mathcal{X})$, defi0 3 4 0 c 7 (g i . e 5 . 9 5 1 5 - 2 . 9 5 5 8 9 1 . 7 1 V 9 0 T d (

the standard $_2$ norm (Utreras, 1981). For thin-plate splines on a bounded domain of $\mathcal{X} \in \mathbb{R}$ with the penalty (3), Conditions 1 and 2 are satisfied with $= 2$ $/$. For tensor-product smoothing splines with penalty $(\eta) = \sum_{\beta=1} \theta_\beta^{-1} \| \ _\beta \eta \|^2_{\mathcal{H}^\beta}$, one can prove that Condition 1 holds using the argument in Example 9.2 of Gu (2013), and Condition 2 holds with $= 4 - \epsilon$, where $\epsilon > 0$ (Wahba, 1990).

3. For a constant $< \infty$, $\mathrm{var}\{\phi_\nu(\ )\phi_\mu(\ )\} \leqslant$ for all ν and μ.

Since ϕ_ν is an orthonormal system relative to $(\cdot, \cdot)$, i.e.,

$$(\phi_\nu, \phi_\mu) = \int_{\mathcal{X}} \phi_\nu(\ )\phi_\mu(\ ) \ (\ ) \, \mathrm{d} \ = \delta_{\nu\mu},$$

we have that

$$\mathrm{var}\{\phi_\nu(\ )\phi_\mu(\ )\} = \int_{\mathcal{X}} \phi_\nu^2(\ )\phi_\mu^2(\ ) \ (\ ) \, \mathrm{d} \ - \delta_{\nu\mu}.$$

Thus Condition 3 is equivalent to the requirement that $\int_{\mathcal{X}} \phi_\nu^2(\ )\phi_\mu^2(\ ) \ (\ ) \, \mathrm{d}$ is uniformly bounded for all ν and μ.

4·2.

This section presents our main results on convergence rates. All proofs are given in the Supplementary Material.

In our adaptive sampling scheme, the search for the smoothing spline estimator is restricted to the effective model space $\mathcal{H}$. We first establish a lemma that justifies the use of the effective model space. Let $\mathcal{H} \ominus \mathcal{H}$ denote the orthogonal complement of $\mathcal{H}$ in the reproducing kernel Hilbert space $\mathcal{H}$.

LEMMA 1. $\lambda \to 0$ *λ7.8

using the full basis indicated by the representer theorem. The parameter γ in the condition yields a faster rate of convergence for certain functions: for the roughest η satisfying $J(\eta) < \infty$, we have $\gamma = 1$, whereas for the smoothest η, we have $\gamma = 2$; see Wahba (1985) for details.

Note that $J(\eta_0) = \sum_i \rho_i^{-1} (\eta_0, \phi_i)^2$. When $J(\eta_0) < \infty$, the condition in Theorem 3 holds with $\gamma = 1$, and the convergence rate is $O(n^{-r/(r+1)})$. When η_0 is in the Sobolev space $W^{m,2}$ on a bounded domain in $\mathbb{R}^d$, we have $r = 2m/d$ and Theorem 3 yields the convergence rate $n^{-2m/(2m+d)}$, which is the optimal rate of convergence (Stone, 1982). For the case $d = 1$, Claeskens et al. (2009) and Wang et al. (2011) showed that penalized splines can also achieve the optimal rate of convergence.

Theorem 3 helps determine the dimension of the effective model space $\mathcal{H}_\lambda$. With $\lambda \asymp n^{-r/(\gamma r+1)}$, Lemma 1 and Theorem 3 require that $M_*\lambda^{1/\gamma} \to \infty$, which suggests that a suitable choice of M_* should satisfy $M_* \asymp n^{1/(\gamma r+1)+\delta}$, where δ is an arbitrary small positive number. For univariate cubic smoothing splines with the penalty $J(\eta) = \int_0^1 (\eta'')^2$, $r = 4$ and $\lambda \asymp n^{-4/(4\gamma+1)}$, a suitable choice of the dimension of the effective model space is $M_* = n^{1/(4\gamma+1)+\delta}$, which lies in the interval $(n^{2/9+\delta}, n^{2/5+\delta})$ for γ taking values in $[1, 2]$. For tensor-product splines, $r = 4 - \epsilon$, where $\epsilon > 0$, a suitable choice of the dimension of effective model space is $M_* = n^{1/(4\gamma+1)+\delta}$, which is roughly in interval $(n^{2/9+\delta}, n^{2/5+\delta})$. In our simulation study and real data analysis, we take the dimension of the effective model space M_* to be between $5n^{2/9}$ and $20n^{2/9}$.

5. SIMULATION RESULTS

Using simulated multivariate regression functions, we compared the smoothing spline estimators based on adaptive basis sampling and uniform basis sampling in terms of estimation accuracy and computational time. We also compared adaptive basis sampling with fast bivariate P-splines, an efficient algorithm for bivariate spline smoothing (Xiao et al., 2013).

Some of our simulation set-ups involve the joint probability density of a d-dimensional nonparanormal distribution (Liu et al., 2009)

η_{copula}

3. a 4-d additive function, $\eta(\ ) = \eta_{\text{blocks}}(\ _{\langle 1\rangle}, \ _{\langle 2\rangle}) + \eta_{\text{copula}}(\ _{\langle 3\rangle}, \ _{\langle 4\rangle})$, where η_{blocks} and η_{copula} are as in set-ups 1 and 2;
4. a 6-d copula function, the function given in (10), with $\ = 6$ and $\alpha_{\ } = 0{\cdot}1$ for all $\ $. The domain of interest is $[-1, 1]^6$.

For all four settings, we computed the smoothing spline estimator using the full basis, and using the bases chosen by adaptive basis sampling and uniform basis sampling. For adaptive basis sampling, the number of slices was chosen based on the Scott (1992

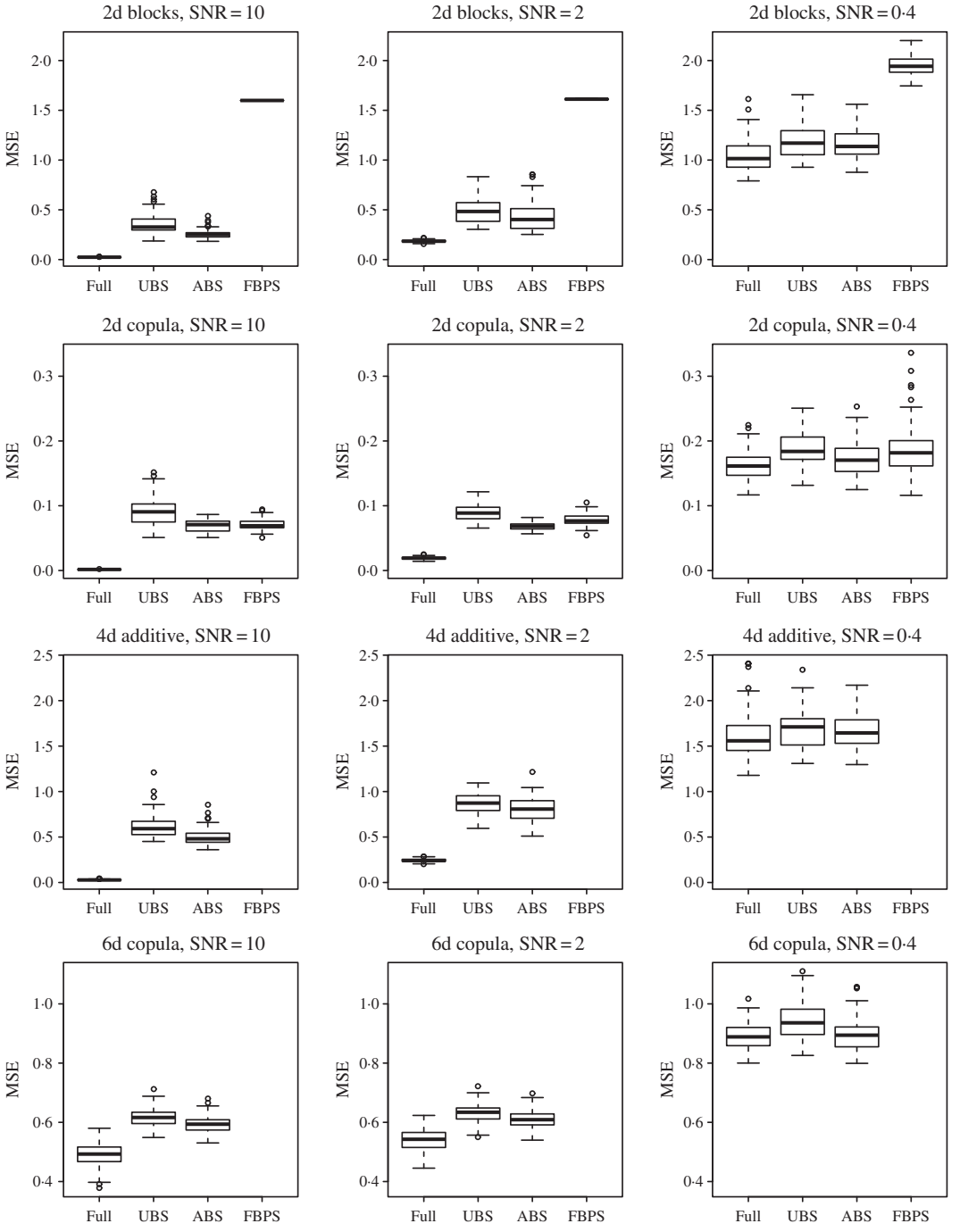


Fig. 3. Boxplots of the mean squared errors for four multivariate test functions under three signal-to-noise ratios, SNR, (10, 2, 0·4), based on 100 simulation runs. Full, UBS and ABS stand for smoothing spline estimators with full basis, uniform basis sampling and adaptive basis sampling. FBPS is fast bivariate P-splines.

1300 earthquakes with magnitude mb $>5{\cdot}2$ that occurred between 1988 and 2002, and were recorded at one or more of a total of nearly 1200 stations in central America. Along a 2500 km strip, they then constructed point images of core-mantle boundary regions using a generalized Radon transform. They constructed 163 713 point images at various depths and locations of the

Table 1. [illegible] 100 [illegible]

True function	SNR	Full basis	UBS	ABS	FBPS
2d blocks	10	399 (12)	5·20 (0·12)	5·14 (0·10)	1·38 (0·03)
	2	408 (9)	7·16 (0·35)	6·40 (0·23)	1·41 (0·02)
	0·4	361 (7)	5·00 (0·17)	4·99 (0·17)	1·51 (0·02)
2d copula	10	260 (3)	6·56 (0·20)	6·41 (0·21)	1·63 (0·03)
	2	301 (6)	6·86 (0·18)	6·71 (0·33)	1·59 (0·03)
	0·4	317 (8)	4·69 (0·16)	4·79 (0·14)	1·58 (0·03)
4d blocks + copula	10	1247 (26)	15·16 (0·60)	13·84 (0·59)	–
	2	1222 (25)	16·62 (0·96)	15·54 (0·76)	–
	0·4	1135 (19)	13·16 (0·66)	13·27 (0·60)	–
6d copula	10	9336 (223)	162·88 (7·27)	145·14 (7·32)	–
	2	9572 (283)	179·12 (7·52)	181·27 (6·60)	–
	0·4	7639 (161)	143·10 (6·80)	135·01 (6·81)	–

SNR, signal-to-noise ratio; UBS, uniform basis sampling; ABS, adaptive basis sampling; FBPS, fast bivariate P-splines.

strip. At each depth and location, the point images constructed contain many noisy replicates resulting from different reflection angles of the seismic waves, so further statistical analysis is necessary to estimate the true image. In order to be computationally feasible, they estimated the true image using smoothing splines at each location and interpolated the estimated images from all locations to get the three-dimensional image. The image shows peaks of very different magnitudes at several unexpected locations (van der Hilst et al., 2007).

In this section, we apply a smoothing spline with adaptive basis sampling directly to all point images to estimate the three-dimensional image. We let $_{i.}$ denote the point image at the i th distance, $_{\langle 1\rangle}$, and the $.$ th depth, $_{\langle 2\rangle}$. We consider the following model for the point images

$$_{i.} = \eta(_{\langle 1\rangle i}, _{\langle 2\rangle .}) + \epsilon_{i.}.$$

Since the sample size is $= 163\,713$, the regular tensor-product smoothing spline is computationally prohibitive. Instead, we apply our cubic tensor-product smoothing spline with adaptive basis sampling to the dataset with $-$ 10 slices and let the dimension of the effective model space be $^* = 155$. Define $._1(\cdot) = \cdot - 0{\cdot}5$,

$$._2(\cdot) = \frac{1}{2}\left\{._1^2(\cdot) - \frac{1}{12}\right\}, \qquad ._4(\cdot) = \frac{1}{24}\left\{._1^4(\cdot) - \frac{._1^2(\cdot)}{2} + \frac{7}{240}\right\},$$

and $(\cdot_1, \cdot_2) =._2(\cdot_1)._2(\cdot_2) -._4(|\cdot_1 - \cdot_2|)$. The cubic tensor-product smoothing spline estimator with adaptive basis sampling has the form

$$\eta(\cdot) = \sum_{\nu=1}^{4} _{\nu}\phi_{\nu}(\cdot) + \sum_{.=1}^{^*} ._. (\cdot, _{.}^*, _{.}),$$


40
30
20
10
0
−10
−20
−30

COOK, R. D. (1998). . New York: Wiley.

CRAVEN, P. & WAHBA, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. **31**, 377–403.

DENNIS, J. E. & SCHNABEL, R. B. (1996). . Philadelphia: SIAM.

DONOHO, D. L. & JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. **81**, 425–55.

DUCHON, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In , Eds. W. Schempp & K. Zeller. Berlin: Springer-Verlag, pp. 85–100.

GOLUB, G. & VAN LOAN, C. (1989). . Baltimore, MD: The Johns Hopkins University Press, 2nd ed.

GU, C. (2013). . New York: Springer, 2nd ed.

GU, C. & KIM, Y.-J. (2002). Penalized likelihood regression: General formulation and efficient approximation. **30**, 619–28.

GU, C. & QIU, C. (1994). Penalized likelihood regression: A simple asymptotic analysis. **4**, 297–304.

HASTIE, T. J. (1996). Pseudosplines. B **58**, 379–96.

KIM, Y.-J. & GU, C. (2004). Smoothing spline Gaussian regression: more scalable computation via efficient approximation. B **66**, 337–56.

LI, K. C. (1991). Sliced inverse regression for dimension reduction. **86**, 316–27.

LIU, H., LAFFERTY, J. & WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. **10**, 2295–28.

LIU, Z. & GUO, W. (2010). Data driven adaptive spline smoothing. **20**, 1143–63.

LUO, Z. & WAHBA, G. (1997). Hybrid adaptive splines. **92**, 107–16.

MA, P., WANG, P., TENORIO, L., DE HOOP, M. V. & VAN DER HILST, R. D. (2007). Imaging of structure at and near the core-mantle boundary using a generalized Radon transform: 2. Statistical inference of singularities. **112**, B08303.

PINTORE, A., SPECKMAN, P. & HOLMES, C. C. (2006). Spatially adaptive smoothing splines. **93**, 113–25.

REINSCH, C. H. (1967). Smoothing by spline functions. **10**, 177–83.

RUPPERT, D., WAND, M. & CARROLL, R. J. (2003). . Cambridge: Cambridge University Press.

SCOTT, D. W. (1992). . New York: Wiley.

SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. **10**, 795–810.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. **10**, 1040–53.

TENORIO, L., ANDERSSON, F., DE HOOP, M. & MA, P. (2011). Data analysis tools for uncertainty quantification of inverse problems. **27**, 045001.

UTRERAS, F. (1981). Optimal smoothing of noisy data using spline functions. **2**, 349–62.

VAN DER HILST, R. D., DE HOOP, M. V., WANG, P., SHIM, S. H., MA, P. & TENORIO, L. (2007). Seismo-stratigraphy and thermal structure of Earth's core-mantle boundary region. **315**, 1813–17.

WAHBA, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. B **45**, 133–50.

WAHBA, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. **13**, 1378–1402.

WAHBA, G. (1990). . Philadelphia: SIAM.

WANG, P., DE HOOP, M. V., VAN DER HILST, R. D., MA, P. & TENORIO, L. (2006). Imaging of structure at and near the core mantle boundary using a generalized radon transform: 1. Construction of image gathers. **111**, B12304.

WANG, X., SHEN, J. & RUPPERT, D. (2011). On the asymptotics of penalized spline smoothing. **5**, 1–17.

WANG, X., DU, P. & SHEN, J. (2013). Smoothing splines with varying smoothing parameter. **100**, 955–70.

WANG, Y. (2011). . Boca Raton: Chapman and Hall.

WEINBERGER, H. F. (1974). . Philadelphia: SIAM.

XIAO, L., LI, Y. & RUPPERT, D. (2013). Fast bivariate P-splines: The sandwich smoother. B **75**, 577–99.

ZHANG, H. H., WAHBA, G., LIN, Y., VOELKER, M., FERRIS, M., KLEIN, R. & KLEIN, B. (2004). Variable selection and model building via likelihood basis pursuit. **99**, 659–72.

[2013 , 2015]