Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 97 (2006) 2023–2033
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Nonlinear regression modeling via regularized wavelets and smoothing parameter selection

Toru Fujii[a], Sadanori Konishi[b,*]

[a] *Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan*
[b] *Faculty of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan*

Received 19 May 2005
Available online 10 July 2006

Abstract

We introduce regularized wavelet-based methods for nonlinear regression modeling when design points are not equally spaced. A crucial issue in the model building process is a choice of tuning parameters that control the smoothness of a fitted curve. We derive model selection criteria from an information-theoretic and also Bayesian approaches. Monte Carlo simulations are conducted to examine the performance of the proposed wavelet-based modeling technique.

AMS 2000 subject classification: 62J02; 62M10

Keywords: Automatic smoothing parameter selection; Irregular design points; Linear shrinkage; Regression modeling; Wavelets

1. Introduction

Smoothing methods in non-parametric regression have received considerable attention, and various methods have been proposed for function estimation including kernel-based smoothing, splines and other basis expansions (see, for example, [8,11] and references given therein). These methods are known to be effective when underlying functions are sufficiently smooth.

In contrast, wavelets provide useful methods for analyzing data with intrinsically local properties, such as discontinuities and sharp spikes, and have given rise to significant activity in the field of statistics. Wavelets form an orthonormal basis and enable *multiresolution analysis* by localizing a function in different phases of both time and frequency domains simultaneously,

* Corresponding author. Fax: +81 92 642 2779.
E-mail address: konishi@math.kyushu-u.ac.jp (S. Konishi).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2005.12.009

and thus offer some advantages over traditional Fourier expansions. Theoretical and practical developments of their use in statistics have been made by Donoho et al. [6,7], Hall and Patil [9], among others. These papers focused on density estimation and regression estimation with the use of nonlinear thresholding, and demonstrated remarkable local adaptivity for large classes of irregular functions.

It might be noticed that the vast majority of wavelet-based regression estimation has been conducted within the setting that given data are of decimal length and have equally spaced design points. For the case that the design points are not equally spaced, the corresponding design matrix is no longer orthogonal, and wavelet-based decomposition/reconstruction procedure cannot be directly applied. Several different approaches for this case of unequally spaced design points have been made by Hall and Patil [9], Hall and Turlach [10], Antoniadis and Fan [2] and Pensky and Vidakovic [15] among others.

The aim of the present paper is to propose linear shrinkage methods to wavelet smoothing within the setting of unequally spaced and non-decimal design points. We consider the modification of the irregularity of design points to be a wavelet-based density estimation, which differs from the methods based on interpolation and averaging.

A crucial point in the model building process is the selection of several tuning parameters. The linearity of the proposed wavelet estimator makes it possible to select smoothing parameters by using the generalized cross-validation. We also propose nonlinear regression modeling via regularized wavelet-based methods when the design points are not equally spaced, and then derive model selection criteria from an information-theoretic and Bayesian viewpoints.

This paper is organized as follows. In Section 2 we describe the wavelet-based regression model with the basic concepts of wavelets. A regularized wavelet-based method is given for nonlinear regression modeling when design points are not equally spaced. In Section 3 we present model selection criteria to choose smoothing parameters. Section 4 includes Monte Carlo simulations to investigate the performance of our modeling techniques and model selection criteria. Some concluding remarks are given in Section 5.

2. Wavelet methods

In this section, we present nonlinear regression models based on a regularized wavelet-based method.

2.1. Wavelets

Here, we briefly describe the basic concepts of wavelets. Let $\phi(t)$ and $\psi(t)$, respectively, be *father and mother wavelets*. Assume that $\phi(t)$ is an orthonormal function with compact support on $\mathbb{R}$, which satisfies

$$\int \phi(t)\,\mathrm{d}t = 1, \quad \int \phi(t)\phi(t-l)\,\mathrm{d}t = \delta_{0l},$$

for $l \in \mathbb{Z}$, and

$$\phi(t) = \sum_{k\in\mathbb{Z}} p_k \phi(2t-k), \tag{1}$$

where δ_{0l} is the Kronecker delta and $\{p_k\}$ is a finite sequence such that $\sum_{k\in\mathbb{Z}} p_k = 2$, $\sum_{k\in\mathbb{Z}} p_k p_{k+2l} = 2\delta_{0l}$ and $\sum_{k\in\mathbb{Z}} (-1)^k p_{1-k} = 0$.

Define the mother wavelet $\psi(t)$ by

$$\psi(t) = \sum_{k\in\mathbb{Z}} (-1)^k p_{1-k}\phi(2t-k), \tag{2}$$

where $\{p_k\}$ is the same sequence appearing in expression (1) for the father wavelet $\phi(t)$. It follows that ψ also has compact support on $\mathbb{R}$, and that $\int \psi(t)\,\mathrm{d}t = 0$. Further, if $\sum_{k\in\mathbb{Z}} (-1)^k k^v p_k = 0$ for $1 \leqslant v \leqslant r$ with some integer $r \geqslant 1$, then the moment condition $\int t^v \psi(t)\,\mathrm{d}t = 0$ is satisfied (for $1 \leqslant v \leqslant r$). There exist several families of wavelet bases. It remains an issue about which pair of wavelet bases should be chosen in nonlinear regression modeling (see e.g. [14] for this issue).

According to the scale $j \in \mathbb{Z}$ and the shift $k \in \mathbb{Z}$, define the translations of ϕ and ψ, respectively, by

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{jk}(t) = 2^{j/2}\psi(2^j t - k).$$

It follows that $\phi_{jk}(t)$ and $\psi_{jk}(t)$ are orthonormal, i.e.

$$\int \phi_{jk}(t)\phi_{jm}(t)\,\mathrm{d}t = \delta_{km}, \quad \int \psi_{jk}(t)\psi_{lm}(t)\,\mathrm{d}t = \delta_{jl}\delta_{km},$$

for $j, k, l, m \in \mathbb{Z}$, and $\int \phi_{jk}(t)\psi_{lm}(t)\,\mathrm{d}t = 0$ for $j \leqslant l$.

By using $\phi_{jk}(t)$ and $\psi_{jk}(t)$ as basis functions, any function $h \in \mathrm{L}_2(\mathbb{R})$ can be expressed as a series expansion

$$h(t) = \sum_k c_k \phi_{j_0 k}(t) + \sum_{j \geqslant j_0}^{\infty} \sum_k c_{jk}\psi_{jk}(t), \tag{3}$$

with arbitrary resolution level $j_0 \in \mathbb{Z}$. This is called the *wavelet expansion* of h in $\mathrm{L}_2(\mathbb{R})$. Due to the orthonormality, each coefficient in Eq. (3) is uniquely expressed by the L_2-products of h and ϕ_{jk}, and of h and ψ_{jk}, given, respectively, by

$$c_k = \int h(t)\,\phi_{j_0 k}(t)\,\mathrm{d}t, \quad c_{jk} = \int h(t)\,\psi_{jk}(t)\,\mathrm{d}t.$$

For further details we refer Chui [4] and Daubechies [5].

2.2. *Wavelet-based regression models*

In the sequel, we discuss a wavelet-based regression modeling. Suppose we have L observations $\{(x_l, t_l);\ l = 1, \ldots, L\}$, where $x_1, \ldots, x_L$ are observed values at design points $t_1, \ldots, t_L$, respectively. It is assumed that the data are generated from a regression model

$$x_l = h(t_l) + \varepsilon_l, \quad l = 1, \ldots, L, \tag{4}$$

where the errors ε_l are a sequence of independent random variables with mean 0 and $\mathrm{Var}(\varepsilon_l) < \infty$, and $h(t) = E(X \mid T = t)$ is an unknown regression function. Then $h(t)$ is estimated from the data by using some smoothing techniques. The unknown function $h(t)$ is assumed to be included in a class of functions spanned by a set of basis functions $\{\phi_k(t)\}$, for which we use wavelet bases.

Without loss of generality, we rescale the points $\{t_l;\ l = 1, \ldots, L\}$ to be contained in the interval [0, 1]. In the regression model of Eq. (4), we first assume that the unknown function $h(t)$ may be expressed as

$$h(t) = \sum_{k=1}^{2^{j_1}} \alpha_k \phi_{j_1 k}(t),$$

where $\phi_{j_1 k}(t)$ are the father wavelet bases with some resolution level $j_1 \in \mathbb{Z}$. It then follows from the orthonormality of $\{\phi_{j_1 k}(t);\ k \in \mathbb{Z}\}$ that each coefficient is uniquely determined as $\alpha_k = \int h(t)\phi_{j_1 k}(t)\,\mathrm{d}t$.

2.3. Wavelet decomposition and shrinkage methods

It is known as the *discrete wavelet transformation* that the 2-scale relations of (1) and (2) yield the following decomposition:

$$\sum_{k=1}^{2^{j_1}} \alpha_k \phi_{j_1 k}(t) = \sum_{k=1}^{2^{j_0}} c_k \phi_{j_0 k}(t) + \sum_{j=j_0}^{j_1-1} \sum_{k=1}^{2^j} c_{jk} \psi_{jk}(t), \tag{5}$$

where $j_0 \in \mathbb{Z}$ indicates the lowest resolution level.

In considering computational methods, it is often the case in wavelet-based estimates that the design points $\{t_l;\ l = 1, \ldots, L\}$ are taken to be decimal and equally spaced. For these cases, Amato and Vuza [1] introduced the shrinkage method in high-resolution coefficients as $\tilde{c}^*_{jk} = \tilde{c}_{jk}/(1 + \gamma d_j)$, where $\tilde{c}_{jk}$ are decomposed from the initial coefficients $\tilde{\alpha}_k \equiv x_k$ for $k = 1, \ldots, 2^J$, $\gamma > 0$ is a smoothing parameter and $d_j = 2^{(j-j_0+1)}$ is a level-dependent constant. Note that the above shrinkage method of coefficients differs from nonlinear methods such as *hard thresholding* $\tilde{c}^*_{jk} = \tilde{c}_{jk}\,\delta(|\tilde{c}_{jk}| > \gamma)$ and *soft thresholding* $\tilde{c}^*_{jk} = \mathrm{sgn}(\tilde{c}_{jk})(|\tilde{c}_{jk}| - \gamma)\,\delta(|\tilde{c}_{jk}| > \gamma)$, both of which have been mainly used in wavelet-based estimates ([6,7,9] among others).

For the case that the design points are not decimal and unequally spaced, Hall and Turlach [10] and Antoniadis and Fan [2] relaxed the restrictions by approximating the design points by the elements of some dyadic points $\{l/2^J;\ l = 1, \ldots, 2^J\}$ with $2^J \geqslant L$, by using the wavelet interpolation. Hall and Patil [9] and Antoniadis and Pham [3] approached this problem instead by assuming that the design points are independent random variables each with identical density function $w(t)$. The estimator given by Hall and Patil [9] is $\hat{h}(t) = \hat{g}(t)/\hat{w}(t)$ in which the density $w(t)$ and $g(t) = h(t)w(t)$ are estimated separately by a nonlinear wavelet estimate. This method may, however, have practical limitations because one needs to determine the degree of smoothness separately for $w(t)$ and $g(t)$, and in consequence the behavior of the estimator $\hat{h}(t)$ could be unstable.

Another approach for unequally spaced design points is to estimate the coefficients α_k in (5) based on the equation

$$\begin{aligned}\int h(t)\phi_{j_1 k}(t)\,\mathrm{d}t &= \int \frac{h(t)\phi_{j_1 k}(t)}{w(t)} w(t)\,\mathrm{d}t \\ &= E\left\{\frac{E(X \mid T)\phi_{j_1 k}(T)}{w(T)}\right\},\end{aligned}$$

which yields $\tilde{\alpha}_k = L^{-1}\sum_{l=1}^{L} x_l \phi_{j_1 k}(t_l)/\hat{w}(t_l)$. This type of empirical coefficient estimators was introduced by Pensky and Vidakovic [15], who used the kernel density estimate of $w(t)$.

We propose to use the wavelet density estimate of $w(t)$ for the above empirical coefficients. It follows from $\int w(t)\phi_{j_1 k}(t)\,\mathrm{d}t = E\{\phi_{j_1 k}(T)\}$ that a wavelet density estimator is given by

$$\hat{w}(t) = \sum_{k=1}^{2^{j_1}} \frac{1}{L} \sum_{l=1}^{L} \phi_{j_1 k}(t_l)\, \phi_{j_1 k}(t). \tag{6}$$

Hence we use the estimators of wavelet coefficients given by

$$\tilde{\alpha}_k = \frac{1}{L} \sum_{l=1}^{L} \frac{x_l \phi_{j_1 k}(t_l)}{\hat{w}(t_l)}, \tag{7}$$

as the basis for a regression model with unequally spaced design points, where $\hat{w}(t)$ is given by Eq. (6). Then by extending the shrinkage method introduced by Amato and Vuza [1], we have an estimator of $h(t)$ in the form

$$\begin{aligned} \hat{h}_v(t) &= \sum_k \tilde{c}_k \phi_{j_0 k}(t) + \sum_{j=j_0}^{j_1-1} \frac{1}{1+\gamma d_j} \sum_k \tilde{c}_{jk} \psi_{jk}(t) \\ &= \sum_k \hat{\alpha}_k \phi_{j_1 k}(t), \end{aligned} \tag{8}$$

where $v = (j_0, j_1, \gamma)$ is a set of smoothing parameters, the coefficients $\tilde{c}_k$ and $\tilde{c}_{jk}$ are decomposed from the empirical coefficients proposed in Eq. (7), and $\hat{\alpha}_k$ are reconstructed by using the inverse of the discrete wavelet transformation.

Let $\hat{h}_v(\boldsymbol{t}) = (\hat{h}_v(t_1), \ldots, \hat{h}_v(t_L))^T$ be an L-dimensional vector, and let B be the $L \times 2^{j_1}$ basis matrix whose elements are given by $B_{(lk)} = \phi_{j_1 k}(t_l)$. It follows that the estimator in Eq. (8) can be written as $\hat{h}_v(\boldsymbol{t}) = B\widehat{\boldsymbol{\alpha}}$, and that

$$\begin{aligned} \widehat{\boldsymbol{\alpha}} &= \mathcal{W}(I + \gamma\,\mathcal{S})^{-1}\mathcal{W}^T B^T \Omega \boldsymbol{x} \\ &= (I + \gamma\,\mathcal{W}\mathcal{S}\mathcal{W}^T)^{-1} B^T \Omega \boldsymbol{x}, \end{aligned} \tag{9}$$

where $\mathcal{S} = \mathrm{diag}(\mathbf{0}_{2^{j_0}}, d_{j_0}\mathbf{1}_{2^{j_0}}, \ldots, d_{j_1-1}\mathbf{1}_{2^{j_1-1}})$ denotes the shrinkage matrix, and $\Omega = \mathrm{diag}(BB^T\mathbf{1}_L)^{-1} = L^{-1}\mathrm{diag}(\hat{w}(t_1)^{-1}, \ldots, \hat{w}(t_L)^{-1})$. The matrix $\mathcal{W}^T$ denotes the discrete wavelet transformation, which translates the coefficients $(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{2^{j_1}})^T$ of $\{\phi_{j_1 k}(t)\,;\, k \in \mathbb{Z}\}$ into the coefficients of $\{\phi_{j_0 k}(t)\,;\, k \in \mathbb{Z}\}$ and $\{\psi_{jk}(t)\,;\, k \in \mathbb{Z}\}_{j=j_0}^{j_1-1}$ in the wavelet expansion (5), and consists of the 2-scale sequences $\{p_k\}$ and $\{(-1)^k p_{1-k}\}$ of wavelet bases (see [17] for further detail). It may be noted that $B^T\Omega\boldsymbol{x} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{2^{j_1}})^T$ coincides with Eq. (7).

2.4. Regularized wavelet-based methods

We now show that the proposed estimator (8) and its vector form (9) can also be formulated as the solution of the regularized log-likelihood function.

Suppose that the errors ε_l in (4) are independently, normally distributed with mean 0 and variance σ^2. Then the regression model can be expressed as

$$f(x_l \mid t_l \,;\, \boldsymbol{\alpha},\, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_l - \boldsymbol{b}_l^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\}, \quad l = 1, \ldots, L,$$

where $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L)^T$ is the vector of wavelet bases and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{2^{j_1}})^T$ is the vector of the corresponding wavelet coefficients. We estimate the coefficients of wavelet bases by maximizing the regularized log-likelihood function

$$\begin{aligned} \ell_{\gamma^*}(\boldsymbol{\alpha}, \sigma^2) &= \sum_{l=1}^{L} \omega_l \log f(x_l \mid t_l \,;\, \boldsymbol{\alpha},\, \sigma^2) - \frac{L\gamma^*}{2} \boldsymbol{\alpha}^T K \boldsymbol{\alpha} \\ &= -\frac{1}{2} \sum_{l=1}^{L} \omega_l \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{l=1}^{L} \omega_l (x_l - \boldsymbol{b}_l^T \boldsymbol{\alpha})^2 - \frac{L\gamma^*}{2} \boldsymbol{\alpha}^T K \boldsymbol{\alpha}, \end{aligned} \tag{10}$$

where the weights ω_l are the lth diagonal elements of Ω, which is based on the densities estimation of the design points. Maximization of the function expressed by Eq. (10) yields

$$\widehat{\boldsymbol{\alpha}} = (B^T \Omega B + L\hat{\sigma}^2 \gamma^* K)^{-1} B^T \Omega \boldsymbol{x}, \quad \hat{\sigma}^2 = \frac{1}{\mathrm{tr}(\Omega)} (\boldsymbol{x} - B\widehat{\boldsymbol{\alpha}})^T \Omega (\boldsymbol{x} - B\widehat{\boldsymbol{\alpha}}).$$

The estimator $\widehat{\boldsymbol{\alpha}}$ is equivalent to the expression of Eq. (9) upon substituting $\gamma = L\hat{\sigma}^2\gamma^*$ and $K = \mathcal{W}\mathcal{S}\mathcal{W}^T$, and the replacement of $B^T \Omega B$ by the identity matrix I.

Noting that $\mathcal{W}^T \boldsymbol{\alpha}$ is the discrete wavelet transformation that gives the vector of coefficients in the wavelet expansion of $\{\phi_{j_0 k}(t)\}$ and $\{\psi_{jk}(t)\}_{j=j_0}^{j_1-1}$ from $\boldsymbol{\alpha}$, the penalty term in Eq. (10) can be expressed as $\boldsymbol{\alpha}^T \mathcal{W}\mathcal{S}\mathcal{W}^T \boldsymbol{\alpha} = \sum_{j=j_0}^{j_1-1} d_j \|\boldsymbol{\alpha}_j^*\|_2^2$, where $\boldsymbol{\alpha}_j^*$ denotes the vector of coefficients corresponding to $\{\psi_{jk}(t)\}$ and the constant $d_j = 2^{(j-j_0+1)}$ is proportional to $\int |\psi_{jk}''(t)|^2 \,\mathrm{d}t$, which may be considered as the degree of oscillation in $\psi_{jk}(t)$.

3. Selection of smoothing parameters

In this section, we give model selection criteria for the choice of smoothing parameters.

3.1. Generalized information criterion

It may be noted that the estimator $\widehat{\boldsymbol{\vartheta}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{2^{j_1}},\, \hat{\sigma}^2)^T$ obtained by maximizing Eq. (10) can be regarded as an M-estimator defined to be the solution of the implicit equations $\sum_{l=1}^{L} \boldsymbol{\varphi}(x_{il} \mid t_l \,;\, \boldsymbol{\vartheta}) = \boldsymbol{0}$ with

$$\boldsymbol{\varphi}(x_l \mid t_l \,;\, \boldsymbol{\vartheta}) = \frac{\partial}{\partial \boldsymbol{\vartheta}} \left\{ \omega_l \log f(x_l \mid t_l \,;\, \boldsymbol{\vartheta}) - \frac{\gamma^*}{2} \boldsymbol{\alpha}^T K \boldsymbol{\alpha} \right\}, \quad l = 1, \ldots, L.$$

Hence, by using the result given in Konishi and Kitagawa [13, p. 889], we obtain the model selection criterion for evaluating the statistical model $f(x_l \mid t_l \,;\, \widehat{\boldsymbol{\alpha}},\, \hat{\sigma}^2)$ estimated by the regularized wavelet-based method,

$$\mathrm{GIC} = -2\sum_{l=1}^{L} \log f(x_l \mid t_l\,;\,\widehat{\boldsymbol{\vartheta}}) + 2\,\mathrm{tr}\{R(\boldsymbol{\varphi})^{-1} Q(\boldsymbol{\varphi})\},$$

where the $(2^{j_1}+1)\times(2^{j_1}+1)$ matrices R and Q are given by

$$R(\boldsymbol{\varphi}) = \frac{1}{L\hat{\sigma}^2}\begin{bmatrix} B^T \Omega B + \gamma K & \dfrac{1}{\hat{\sigma}^2} B^T \Lambda \boldsymbol{\omega} \\ \dfrac{1}{\hat{\sigma}^2}\boldsymbol{\omega}^T \Lambda B & \dfrac{1}{2\hat{\sigma}^2}\mathrm{tr}(\Omega) \end{bmatrix},$$

$$Q(\boldsymbol{\varphi}) = \frac{1}{L\hat{\sigma}^4}\begin{bmatrix} \left(B^T \Omega \Lambda^2 - \dfrac{\gamma}{L} K \widehat{\boldsymbol{\alpha}} \mathbf{1}_L^T \Lambda\right) B & \dfrac{1}{2} B^T \left(\dfrac{\Lambda^3}{\hat{\sigma}^2} - \Lambda\right)\boldsymbol{\omega} + \dfrac{\gamma(L - \mathrm{tr}(\Omega))}{2L} K\widehat{\boldsymbol{\alpha}} \\ \dfrac{1}{2}\boldsymbol{\omega}^T \left(\dfrac{\Lambda^3}{\hat{\sigma}^2} - \Lambda\right) B & \dfrac{1}{4\hat{\sigma}^4}\boldsymbol{\omega}^T \Lambda^4 \mathbf{1}_L - \dfrac{1}{4}\mathrm{tr}(\Omega) \end{bmatrix}, \tag{11}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_L)^T$, $\Lambda = \mathrm{diag}(x_1 - \boldsymbol{b}_1^T\widehat{\boldsymbol{\alpha}}, \ldots, x_L - \boldsymbol{b}_L^T\widehat{\boldsymbol{\alpha}})$ and $K = \mathcal{WSW}^T$.

3.2. Bayesian information criterion

Konishi et al. [12] extended Schwarz's BIC [16] to the evaluation of models fitted by the maximum penalized likelihood method. Using the result given in Konishi et al. [12, p. 30] and taking the prior density for the unknown parameter vector $\boldsymbol{\vartheta}$ to be a multivariate normal distribution given by

$$\pi(\boldsymbol{\vartheta} \mid \gamma) = (2\pi)^{-(p-k)/2}(L\gamma)^{(p-k)/2}|K_p|_+^{1/2}\exp\left(-\frac{L\gamma}{2}\boldsymbol{\vartheta}^T K_p \boldsymbol{\vartheta}\right),$$

where K_p is a $p \times p$ matrix of rank $p-k$ and $|K_p|_+$ denotes the product of $p-k$ non-zero eigenvalues of K_p, we have

$$\begin{aligned}\mathrm{GBIC} = {} & -2\sum_{l=1}^{L}\log f(x_l \mid t_l\,;\,\widehat{\boldsymbol{\vartheta}}) + \frac{\gamma}{\hat{\sigma}^2}\widehat{\boldsymbol{\alpha}}^T \mathcal{WSW}^T \widehat{\boldsymbol{\alpha}} \\ & + (2^{j_0}+1)\log\frac{L}{2\pi} - (2^{j_1} - 2^{j_0})\log\frac{\gamma}{L\hat{\sigma}^2} + \log|R| - \log|\mathcal{WSW}^T|_+,\end{aligned}$$

where R is given by (11).

We choose the optimal values of the smoothing parameters included in wavelet estimator (9) by minimizing either GIC or GBIC.

3.3. Cross-validation and Mallow's C_p statistic

There exist other ways of selecting the smoothing parameters such as the cross-validation in the form

$$\mathrm{CV}(v) = \frac{1}{L}\sum_{l=1}^{L}\left\{\frac{x_l - \hat{h}_v(t_l)}{1 - H_{v\,(l,l)}}\right\}^2,$$

where $H_{v(l,l)}$ are diagonal elements of the so-called smoother matrix given by $H_v = B\mathcal{W}(I+\gamma\mathcal{S})^{-1}\mathcal{W}^T B^T \Omega$. By using the smoother matrix H_v the generalized cross-validation is given in the form

$$\mathrm{GCV}(v) = \sum_{l=1}^{L} \frac{L\{x_l - \hat{h}_v(t_l)\}^2}{\{\mathrm{tr}(I - H_v)\}^2}.$$

Mallow's C_p statistic is given as follows:

$$C_p(v) = \sum_{l=1}^{L} \{x_l - \hat{h}_v(t_l)\}^2 + 2\hat{\sigma}^2 \mathrm{tr}(H_v),$$

where $\hat{\sigma}^2 = \sum_{l=2}^{L-1} \tilde{\varepsilon}_l^2/(L-2)$ and $\tilde{\varepsilon}_l = (x_l - a_l x_{l-1} + b_l x_{l+1})/(1 + a_l^2 + b_l^2)^{1/2}$, $a_l = (t_{l+1} - t_l)/(t_{l+1} - t_{l-1})$ and $b_l = (t_l - t_{l-1})/(t_{l+1} - t_{l-1})$. See, for example, Eubank [8].

4. Numerical examples

We consider two example problems to investigate the property of the proposed nonlinear regression modeling. The first example problem is given by real data, while the second arise from a Monte Carlo simulation.

We first consider the problem of choosing the smoothing parameters by the analysis of the motorcycle impact data. By using our regularized wavelet procedure, we estimate the regression function $h(t)$ from given data, in which smoothing parameters $v = (j_0, j_1, \gamma)$ are selected by using five different criteria CV, GCV, C_p, GIC and GBIC given in Section 3.

We used the wavelet bases of a symmlet-5 which satisfies the fifth-order moment condition. The same resolution parameters $\hat{j}_0 = 1$ and $\hat{j}_1 = 4$ were chosen for all criteria, but the values of the smoothing parameters were slightly different; CV, GCV, C_p, GIC and GBIC selected the values $\hat{\gamma} \times 10^2 = 1.215$, 1.420, 1.523, .908 and 3.093, respectively. Fig. 1 shows the curve estimated by

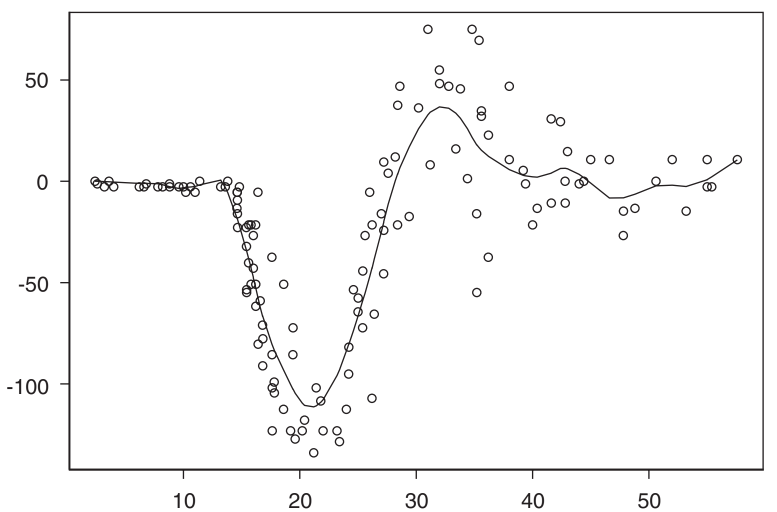

Fig. 1. The motorcycle impact data and the curve estimated by using GBIC.

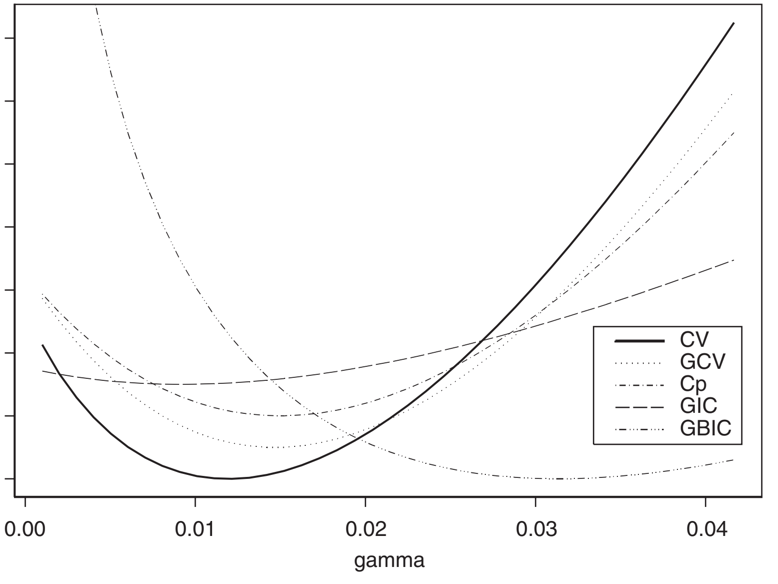

Fig. 2. The curves with respect to the smoothing parameter γ for the five criteria.

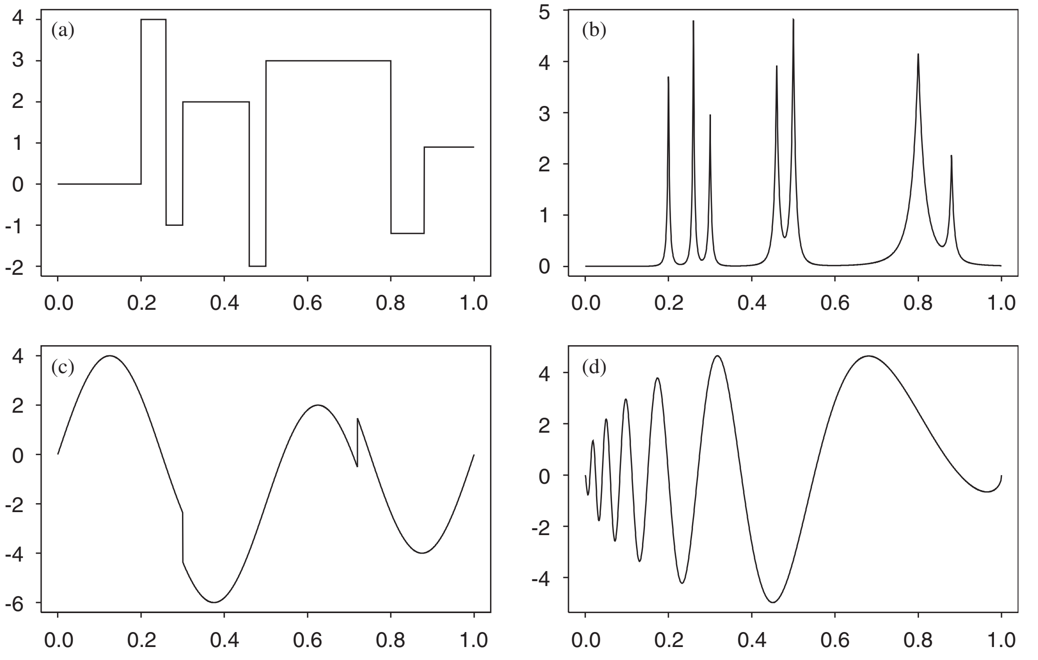

Fig. 3. (a) Blocks is a piecewise constant function $h(t) = \sum_{k=1}^{7} h_k K(t - t_k)$, where $K(t) = (1 + \mathrm{sgn}(t))/2$, $\{t_k\} = \{.2, .26, .30, .46, .50, .80, .88\}$ and $\{h_k\} = \{4, -5, 3, -4, 5, -4.2, 2.1\}$. (b) Bumps is a sum of bumps $h(t) = \sum_{k=1}^{7} h_k K((t - t_k)/s_k)$, where $K(t) = (1 + |t|)^{-4}$, $\{t_k\}$ is the same as be used in Blocks, $h_k = \{4, 5, 3, 4, 5, 4.2, 2.1\}$, and $\{s_k\} = \{01, .01, .012, .02, .02, .06, .02\}$. (c) Heavisine is a sinusoid function $h(t) = 4\sin(4\pi t) - \mathrm{sgn}(t - .3) - \mathrm{sgn}(.72 - t)$, which have the jumps at .3 and .72. (d) Doppler is a sinusoid function $h(t) = 10\sqrt{t(1-t)}\sin\{2\pi \cdot 1.05/(t + .15)\}$, with spatially varying degree of frequency.

Table 1
Monte Carlo results for irregular functions (1000 repetitions for each situation)

	σ/R_x		CV	GCV	C_p	GIC	GBIC
Blocks		MEAN of $\hat{\gamma}$ ($\times 10^{-4}$)	0.528	3.057	4.080	0.348	0.536
	0.05	SD of $\hat{\gamma}$ ($\times 10^{-4}$)	0.097	2.346	2.330	0.464	0.795
		ASE of $\hat{h}$ ($\times 10^{-1}$)	1.464	1.463	1.469	1.464	1.463
		MEAN of $\hat{\gamma}$ ($\times 10^{-4}$)	0.812	4.542	5.058	1.034	1.012
	0.1	SD of $\hat{\gamma}$ ($\times 10^{-4}$)	1.123	3.288	3.278	1.580	1.643
		ASE of $\hat{h}$ ($\times 10^{-1}$)	4.563	4.510	4.512	4.530	4.530
Bumps		MEAN of $\hat{\gamma}$ ($\times 10^{-2}$)	2.182	2.037	2.158	1.234	1.317
	0.05	SD of $\hat{\gamma}$ ($\times 10^{-2}$)	0.634	0.498	0.529	0.259	0.210
		ASE of $\hat{h}$ ($\times 10^{-1}$)	1.339	1.330	1.336	1.295	1.298
		MEAN of $\hat{\gamma}$ ($\times 10^{-2}$)	5.982	5.488	5.594	4.448	2.360
	0.1	SD of $\hat{\gamma}$ ($\times 10^{-2}$)	2.432	2.235	2.324	2.527	0.571
		ASE of $\hat{h}$ ($\times 10^{-1}$)	2.487	2.469	2.472	2.463	2.445
Heavisine		MEAN of $\hat{\gamma}$ ($\times 10^{-3}$)	9.580	8.771	9.735	4.099	7.960
	0.025	SD of $\hat{\gamma}$ ($\times 10^{-3}$)	3.941	1.871	2.078	1.245	1.377
		ASE of $\hat{h}$ ($\times 10^{-2}$)	2.010	1.992	1.999	1.992	1.983
		MEAN of $\hat{\gamma}$ ($\times 10^{-3}$)	23.07	26.43	27.87	14.10	17.90
	0.05	SD of $\hat{\gamma}$ ($\times 10^{-3}$)	12.090	8.689	9.513	5.342	3.924
		ASE of $\hat{h}$ ($\times 10^{-2}$)	4.957	4.908	4.922	4.992	4.876
Doppler		MEAN of $\hat{\gamma}$ ($\times 10^{-3}$)	5.110	2.287	2.321	1.701	2.436
	0.05	SD of $\hat{\gamma}$ ($\times 10^{-3}$)	1.405	0.392	0.377	0.355	0.380
		ASE of $\hat{h}$ ($\times 10^{-1}$)	1.396	1.345	1.344	1.341	1.346
		MEAN of $\hat{\gamma}$ ($\times 10^{-3}$)	9.135	6.135	6.209	4.095	5.741
	0.1	SD of $\hat{\gamma}$ ($\times 10^{-3}$)	3.020	1.356	1.446	1.058	1.040
		ASE of $\hat{h}$ ($\times 10^{-1}$)	2.863	2.801	2.801	2.795	2.793

We fixed the resolution parameters $(j_0, j_1) = (4, 6)$ for the function Blocks and $(3, 5)$ for the other functions. Sample size $L = 100$ is fixed for all situations.

GBIC, while Fig. 2 shows the comparison of the curves with respect to the smoothing parameter γ for the five criteria, in which it may be noted that the vertical ranges differs for each criterion.

The second example is the analysis of the simulated data for which the true regression function $h(t)$ is given. We used the irregular functions "Blocks", "Bumps", "Heavisine" and "Doppler" in Donoho and Johnston [6], which we had added slight modifications. Fig. 3 shows plots of these irregular functions.

In the Monte Carlo experiment, we repeatedly simulated the data $\{(x_l, t_l)\,;\, l = 1, \ldots, 100\}$ using the true regression model $x_l = h(t_l) + \varepsilon_i$. The design points $\{t_l\,;\, l = 1, \ldots, 100\}$ were generated independently from a uniform distribution on [0, 1], and the errors $\{\varepsilon_i\}$ were generated independently from normal distributions with the standard deviations $\sigma = 0.025R_x$ and $0.05R_x$ for the function Heavisine, and $\sigma = 0.05R_x$ and $0.1R_x$ for the other functions, where R_x denotes the range of each $h(t)$ over $t \in [0, 1]$.

In all trials, we estimated $h(t)$ by the proposed method using symmlet-5 as a wavelet basis. The most frequently selected resolution parameters were $\hat{j}_1 = 6$ and $\hat{j}_0 = 4$ for the function Blocks, and $\hat{j}_1 = 5$ and $\hat{j}_0 = 3$ for the other functions, so we fixed these parameters and compared the estimator $\hat{h}(t)$ with respect to the smoothing parameter $\hat{\gamma}$.

Table 1 summarizes the results for the 1000 trials of the Monte Carlo simulations for each true function, in which the notations MEAN and SD denote the average values and standard deviations of the smoothing parameter $\hat{\gamma}$. We calculated the average squared errors (ASE), ASE $= \sum_{l=1}^{100}\{h(t_l) - \hat{h}(t_l)\}^2/100$ for $\hat{h}$ estimated by using the criteria CV, GCV, C_p, GIC and GBIC.

The goodness of fit can be assessed by the ASE values in the table. In all situations, GCV and C_p give the similar results. The GBIC, except for a small number of exceptions, achieves the smallest ASE values, while the GIC tends to yield a relatively small ASE for most cases. The comparison of SD indicates that both GIC and GBIC give stable estimates of a smoothing parameter, especially in contrast with those of CV.

5. Concluding remarks

The main aim of the present paper is to introduce nonlinear regression modeling strategies based on a regularized wavelet method when the design points are not equally spaced. In order to select the optimum values of smoothing parameters, we obtain the model selection criteria GIC and GBIC. We observe that our regularized wavelet-based nonlinear modeling strategies with GIC and GBIC perform well for analyzing noisy data with unequally spaced design points.

Acknowledgments

The authors would like to thank the editor and the anonymous referees for their helpful comments and suggestions. Especially, one of the referees provided informative and insightful comments that improved the quality of the paper considerably.

References

[1] U. Amato, D.T. Vuza, Wavelet approximation of a function from samples affected by noise, Rev. Roumaine Math. Pures Appl. 42 (1997) 481–493.
[2] A. Antoniadis, J. Fan, Regularization of wavelet approximations, J. Amer. Statist. Assoc. 96 (2001) 939–967.
[3] A. Antoniadis, D.T. Pham, Wavelet regression for random or irregular design, Computat. Statist. Data Anal. 28 (1998) 353–369.
[4] C.K. Chui, An Introduction to Wavelets, Academic Press, Boston, 1992.
[5] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.
[6] D.L. Donoho, I.M. Johnston, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. 90 (1995) 1200–1224.
[7] D.L. Donoho, I.M. Johnston, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, Ann. Statist. 24 (1996) 508–539.
[8] R.L. Eubank, Nonparametric regression and Spline Smoothing, second ed., Marcel Dekker, New York, 1999.
[9] P. Hall, P. Patil, On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods, J. Roy. Statist. Soc. Ser. B 58 (1996) 361–377.
[10] P. Hall, B.A. Turlach, Interpolation methods for nonlinear wavelet regression with irregularly spaced design, Ann Statist. 25 (1997) 1912–1925.
[11] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, 2003 (fourth corrected printing).
[12] S. Konishi, T. Ando, S. Imoto, Bayesian information criteria and smoothing parameter selection in radial basis function networks, Biometrika 91 (2004) 27–43.
[13] S. Konishi, G. Kitagawa, Generalised information criteria in model selection, Biometrika 83 (1996) 875–890.
[14] Y. Matsushima, S. Shirahata, W. Sakamoto, Selecting wavelet basis functions by AIC, Japan. J. Appl. Statist. 33 (2004) 201–219 (in Japanese).
[15] M. Pensky, B. Vidakovic, On non-equally spaced wavelet regression, Ann. Inst. Statist. Math. 53 (2001) 681–690.
[16] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.
[17] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, New York, 1999.