Nonlinear regression modeling via regularized wavelets and smoothing parameter selection

Fujii, Toru
Graduate School of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

出版情報：Journal of multivariate analysis. 97 (9), pp.2023-2033, 2006-10. Elsevier
バージョン：
権利関係：

MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

Nonlinear regression modeling via regularized wavelets and smoothing parameter selection

T. Fujii & S. Konishi

MHF 2005-26

(Received July 4, 2005)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Nonlinear regression modeling via regularized wavelets and smoothing parameter selection

Toru FUJII

Graduate School of Mathematics, Kyushu University,
Hakozaki 6-10-1, Higashi-Ku, Fukuoka 812-8581, Japan.
fujii@math.kyushu-u.ac.jp

and

Sadanori KONISHI

Faculty of Mathematics, Kyushu University,
Hakozaki 6-10-1, Higashi-Ku, Fukuoka 812-8581, Japan.
konishi@math.kyushu-u.ac.jp

Abstract

We introduce regularized wavelet-based methods for nonlinear regression modeling when design points are not equispaced. 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. The use of the generalized cross-validation is discussed by showing that the estimated wavelet coefficients are a linear smoother. Monte Carlo simulations are conducted to examine the performance of the proposed wavelet-based modeling technique.

Keywords: Wavelets, irregular design points, regression modeling, linear shrinkage, automatic smoothing parameter selection.

1 Introduction

Smoothing methods in nonparametric regression have received considerable attention and many methods such as kernel, splines and basis expansions have been proposed for function estimation (see, for example, Green and Silverman [12], Eubank [10], Härdle [15], Hastie *et al.* [16] and references given therein). These procedures are known to be effective when underlying functions are sufficiently smooth.

In contrast wavelet methods provide a useful tool for analyzing data with intrinsically local properties and have drawn a large amount of attention in statistics. Wavelets have advantages over traditional Fourier expansions in the situations where the signal contains discontinuous and sharp spike, since they offer a simultaneous localization of a function in time and frequency domains.

Theoretical and practical developments in statistics have been made by Donoho *et al.* [8, 9], Hall and Patil [13], Johnston and Silverman [17] among others. These papers focused on density estimation and regression estimation for i.i.d. model, and demonstrated remarkable local adaptivity against large classes of irregular functions. It might be noticed that the vast majority of wavelet-based regression estimation have been conducted within the setting that the design points are decimal and equally spaced, and that smoothing methods with non-linear fashions such as "hard thresholding" and "soft thresholding" have been mainly used.

For the case that the design points are irregularly spaced, the corresponding design matrix is no longer orthogonal and wavelet decomposition procedure can not be directly applied. Several different approaches for irregular design points have been made by Hall and Patil [13], Hall and Turlach [14], Antoniadis and Fan [2] and Pensky and Vidakovic [22] among others.

The aim of the present paper is to propose linear shrinkage methods to wavelet smoothing within the setting of non-equally spaced and non-decimal design points. We first consider the wavelet density estimate to modify the irregularity of design points, which differs from the methods based on interpolation and averaging. The linearity of the proposed wavelet estimator makes it possible to select smoothing parameters by using the generalized cross-validation (Craven and Wahba [5]).

Second we propose nonlinear regression modeling via regularized wavelet-based methods when the design points are not equispaced. Choosing several tuning parameters is a crucial point in the model building process. We derive model selection criteria from an information-theoretic and Bayesian view points.

This paper is organized as follows. In section 2 we describe the wavelet-based regression model with the basic concept of wavelets. The modified linear shrinkage estimator is given. In section 3 we present a regularized wavelet-based method for nonlinear regression modeling when design points are not equispaced, and obtain 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

2.1 Nonlinear regression models

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

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

where the errors ε_l are the sequence of independent random variables with mean 0 and $\mathrm{Var}(\varepsilon_l) < \infty$, and $h(t) = E[x_0 \,|\, t_0 = t]$ 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 some class of functions spanned by a set of basis functions $\{\phi_k(t)\}$, for which we use wavelets bases in the situation that the design points are not equispaced.

2.2 Wavelets

We briefly describe the basic concepts of wavelets. Let $\phi(t)$ and $\psi(t)$ be respectively the father and mother wavelets.

Assume that $\phi(t)$ be the orthonormal function with compact support on $\mathbb{R}$, which satisfies

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

and

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

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 ψ by

$$\psi(t) = \sum_{k\in\mathbb{Z}} (-1)^k p_{1-k}\phi(2t-k), \tag{3}$$

where $\{p_k\}$ is the same sequence as in the father wavelet $\phi(t)$.

It follows that ψ has compact support on $\mathbb{R}$ and that

$$\int \psi(t)\,\mathrm{d}t = 0.$$

In addition, if $\sum_{k\in\mathbb{Z}}(-1)^k k^v p_k = 0$ $(1 \le v \le r)$ for some integer $r \ge 1$, then the moment condition $\int t^v \psi(t)\,\mathrm{d}t = 0$ $(1 \le v \le r)$ is satisfied.

There exist several families of wavelet basis. It remains an issue about which pair of wavelet bases should be chosen in nonlinear regression modeling (see e.g. Matsushima *et al.* [21] for this issue).

As the translations about scale $j \in \mathbb{Z}$ and shift $k \in \mathbb{Z}$ of ϕ and ψ, define

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

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

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

and

$$\int \phi_{jk}(t)\psi_{lm}(t)\,\mathrm{d}t = 0.$$

for $j \leq l$.

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

$$h(t) = \sum_j \sum_k c_{jk}\psi_{jk}(t) = \sum_k c_k \phi_{j_0 k}(t) + \sum_{j \geq j_0} \sum_k c_{jk}\psi_{jk}(t), \tag{4}$$

with arbitral resolution level $j_0 \in \mathbb{Z}$. This is called the wavelet expansion of h in $\mathrm{L}_2(\mathbb{R})$. From the orthonormality, each coefficient in (4) is uniquely expressed by the L_2-products of h and ϕ_{jk}, and of h and ψ_{jk}, respectively as follows;

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

For details we refer Chui [4] and Daubechies [6].

2.3 Wavelet-based regression models

In the sequel, we discuss a wavelet-based regression modeling. Without loss of generality, we rescale the points $\{t_l\,;\, l = 1, \ldots, L\}$ to be contained in $[0, 1]$. In the regression model in equation (1), we first assume that 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 $j_1 \in \mathbb{Z}$. Then it follows from the orthonormality of $\{\phi_{j_1 k}(t),\, k \in \mathbb{Z}\}$ that each coefficient is uniquely determined as $\alpha_k = \int h(t)\phi_{j_1 k}(t)\,\mathrm{d}t$.

It is often the case in wavelet estimate that observational points $\{t_l\,;\, l = 1, \ldots, L\}$ are assumed to be decimal and equally spaced with respect to the computational aspects. For the case that the design points are irregularly spaced, Hall and Turlach [14] and Antoniadis and Fan [2] proposed to approximate the design points by the elements of some dyadic points $\{l/2^J\,;\, l = 1, \ldots, 2^J - 1\}$ with $2^J \geq L$, by using the wavelet interpolation. On the other hand, Hall and Patil [13] and Antoniadis and Pham [3] relaxed the restrictions by assuming that the observational points are independent random variables with identical density function $w(t)$. The estimator given by Hall and Patil [13] is $\hat{h}(t) = \hat{g}(t)/\hat{w}(t)$ in which the density $w(t)$ and $g(t) = h(t)w(t)$ are separately estimated using nonlinear wavelet estimate. This method might be unfavorable in practice 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 to the wavelet density estimate is to use the equation

$$\begin{aligned} \int h(t)\phi_{j_1 k}(t)\,\mathrm{d}t &= \int \frac{h(t)\phi_{j_1 k}(t)}{w(t)} w(t)\,\mathrm{d}t \\ &= E\left[\frac{1}{L}\sum_{l=1}^{L} \frac{x_l \phi_{j_1 k}(t_l)}{w(t_l)}\right], \end{aligned}$$

which yields $\tilde{\alpha}_k = L^{-1}\sum_{l=1}^{L} x_l\phi_{j_1k}(t_l)/\hat{w}(t_l)$. This type of empirical coefficient estimators was introduced by Pensky and Vidakovic [22], in which the kernel density estimate of $w(t)$ was used.

It follows from $\int w(t)\phi_{j_1k}(t)\,\mathrm{d}t = E[L^{-1}\sum_{l=1}^{L}\phi_{j_1k}(t_l)]$ that the wavelet density estimator is given by

$$\hat{w}(t) = \frac{1}{L}\sum_k\sum_{l=1}^{L}\phi_{j_1k}(t_l)\phi_{j_1k}(t). \tag{5}$$

Hence we use the estimators of wavelet coefficients given by

$$\tilde{\alpha}_k = \frac{1}{L}\sum_{l=1}^{L} x_l\phi_{j_1k}(t_l)/\hat{w}(t_l),$$

with $\hat{w}(t)$ in equation (5), for the regression model with non-equally spaced design points.

It is known as the discrete wavelet transform that the 2 scale relations of (2) and (3) yield the following decomposition

$$\sum_k \tilde{\alpha}_k\phi_{j_1k}(t) = \sum_k \tilde{c}_k\phi_{j_0k}(t) + \sum_{j=j_0}^{j_1-1}\sum_k \tilde{c}_{jk}\psi_{jk}(t),$$

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

Amato and Vuza [1] introduced the shrinkage rule in high resolution coefficients of $\psi_{jk}(t)$ with $j \geq j_0$ by smoothing parameter γ and level dependent constants $d_j = 2^{(j-j_0+1)}$ as follows;

$$\hat{h}^{\nu}(t) = \sum_k \tilde{c}_k\phi_{j_0k}(t) + \sum_{j=j_0}^{j_1-1}\frac{1}{1+\gamma d_j}\sum_k \tilde{c}_{jk}\psi_{jk}(t). \tag{6}$$

This shrinkage estimator differs from nonlinear thresholding rules named "hard thresholding" $\tilde{c}^*_{jk} = \tilde{c}_{jk}\,\delta(|\tilde{c}_{jk}| > \gamma)$ and "soft thresholding" $\tilde{c}^*_{jk} = \mathrm{sgn}(\tilde{c}_{jk})(|\tilde{c}_{jk}| - \gamma)\,\delta(|\tilde{c}_{jk}| > \gamma)$, which have been mainly used in wavelet based estimates (Donoho *et al.* [8, 9], Hall and Patil [13] among others).

2.4 Tuning parameter selection

We denote $\hat{h}^{\nu}(\boldsymbol{t})$ by the L-dimensional vector of the fitted values $\hat{h}^{\nu}(t_l)$ at each design point t_l. It follows from (6) that $\hat{h}^{\nu}(\boldsymbol{t})$ can be written as

$$\hat{h}^{\nu}(\boldsymbol{t}) = B\widehat{\boldsymbol{\alpha}}, \quad \widehat{\boldsymbol{\alpha}} = \mathcal{W}\mathcal{S}_{\gamma}\mathcal{W}^T B^T\mathrm{diag}(BB^T\mathbf{1}_L)^{-1}\boldsymbol{x}, \tag{7}$$

where B is the basis matrix with each element $\{B\}_{lk} = \phi_{j_1k}(t_l)$, $(l = 1,\ldots,L,\ k = 1,\ldots,2^{j_1})$, $\mathcal{W}$ is the matrix of inverse discrete wavelet transform, translating the coefficients of wavelet expansion with $\{\phi_{j_0k}(t)\,;\,k \in \mathbb{Z}\}$ and $\{\psi_{jk}(t)\,;\,k \in \mathbb{Z}\}_{j=j_0}^{j_1-1}$ into the coefficients of $\{\phi_{j_1k}(t)\,;\,k \in \mathbb{Z}\}$'s, and $\mathcal{S}_{\gamma} = \mathrm{diag}\{\mathbf{1}_{2^{j_0}}\,,\,(1+\gamma d_{j_0})^{-1}\mathbf{1}_{2^{j_0}}\,,\,\ldots\,,\,(1+\gamma d_{j_1-1})^{-1}\mathbf{1}_{2^{j_1-1}}\}$ denotes the shrinkage matrix. The matrix $\mathcal{W}$ is orthogonal and consists of the 2 scale sequences $\{p_k\}$ and $\{(-1)^k p_{1-k}\}$ of wavelet basis. See Vidakovic [24] for detail.

For the selection of smoothing parameters $\nu = (j_0, j_1, \gamma)$, we use the cross-validation in the form

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

where $H_{\nu\,(l,l)}$ are diagonal elements of the so-called smoother matrix given by

$$H_{\nu} = B\mathcal{W}\mathcal{S}_{\gamma}\mathcal{W}^T B^T\mathrm{diag}(BB^T\mathbf{1}_L)^{-1}.$$

By using the smoother matrix H_{ν} the generalized cross-validation (Craven and Wahba [5]) is given in the form

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

Another type of criterion is Mallow's C_p statistic (Mallows [20])

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

where $\hat{\sigma}^2 = \sum_{l=2}^{L-1} \tilde{\varepsilon}_l^2/(L-2)$ and $\tilde{\varepsilon}_l = (x_l - a_l x_{l-1} + b_l x_{l+1})/(1+a_l^2+b_l^2)^{1/2}$, $a_l = (t_{l+1}-t_l)/(t_{l+1}-t_{l-1})$, $b_l = (t_l - t_{l-1})/(t_{l+1}-t_{l-1})$. Fujikoshi and Satoh [11] investigated the asymptotic properties of C_p and AIC in Gaussian linear regression models.

3 Regularized wavelet-based methods

In this section, we present nonlinear regression models based on regularized wavelet-based method, and give model selection criteria for the choice of smoothing parameters.

3.1 Estimation

It is assumed that errors ε_l in (1) are independently, normally distributed with mean 0 and variance σ^2. The regression model is then expressed as

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

where $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L)^T$ is the vector of wavelet basis 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

$$\begin{aligned} \ell_{\gamma^*}(\boldsymbol{\alpha}, \sigma^2) &= \sum_{l=1}^{L} \omega_l \log f(x_l \,|\, t_l\,;\, \boldsymbol{\alpha},\, \sigma^2) - \frac{L\gamma^*}{2} \boldsymbol{\alpha}^T K \boldsymbol{\alpha}, \qquad (8) \\ &= -\frac{1}{2} \sum_{l=1}^{L} \omega_l \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{l=1}^{L} \omega_l (x_l - \boldsymbol{b}_l^T \boldsymbol{\alpha})^2 - \frac{L\gamma^*}{2} \boldsymbol{\alpha}^T K \boldsymbol{\alpha}. \end{aligned}$$

where the weights $\omega_l = \hat{w}^{-1}(t_l)$ are the densities estimated at irregular design points. The maximization of equation (8) yields

$$\widehat{\boldsymbol{\alpha}} = (B^T \varOmega B + L\hat{\sigma}^2 \gamma^* K)^{-1} B^T \varOmega\, \boldsymbol{x}, \quad \hat{\sigma}^2 = \frac{1}{\mathrm{tr}\,(\varOmega)} (\boldsymbol{x} - B\widehat{\boldsymbol{\alpha}})^T \varOmega (\boldsymbol{x} - B\widehat{\boldsymbol{\alpha}}), \qquad (9)$$

where $\varOmega = \mathrm{diag}(\omega_1, \ldots, \omega_L) = \mathrm{diag}(BB^T \mathbf{1}_L)^{-1}$.

On the other hand, the wavelet estimator in equation (7) can be rewritten, using $\mathcal{W}^T = \mathcal{W}^{-1}$, as

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

where $\mathcal{S} = \mathrm{diag}(\mathbf{0}_{2^{j_0}}\,,\, d_{j_0} \mathbf{1}_{2^{j_0}}\,,\, \ldots\,,\, d_{j_1-1} \mathbf{1}_{2^{j_1-1}})$. This estimator coincides with the result in equation (9) with $\gamma = L\hat{\sigma}^2\gamma^*$, $K = \mathcal{W}\mathcal{S}\mathcal{W}^T$ and the replacement of $B^T \varOmega B$ by identity matrix I.

Noting that $\mathcal{W}^T \boldsymbol{\alpha}$ is the discrete wavelet transform 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 equation (8) 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 is considered as the degree of oscillation in $\psi_{jk}(t)$.

3.2 Selection of smoothing parameters

It might be noted that the estimator $\widehat{\boldsymbol{\vartheta}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{2^{j_1}}, \hat{\sigma}^2)^T$ obtained by maximization of equation (8) can be considered as an M-estimator defined as the solution of the implicit equations $\sum_{l=1}^{L} \boldsymbol{\varphi}(x_{il} \,|\, t_l \,;\, \boldsymbol{\vartheta}) = \mathbf{0}$ with

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

Hence by using the result given in Konishi and Kitagawa ([19], p. 889), we have the model selection criterion for evaluating the statistical model $f(x_l \,|\, t_l \,;\, \widehat{\boldsymbol{\alpha}}, \hat{\sigma}^2)$ estimated by the regularized wavelet-based method in the following;

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

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

$$R(\boldsymbol{\varphi}) = \frac{1}{L\hat{\sigma}^2} \begin{bmatrix} B^T \mathit{\Omega} B + \gamma K & \dfrac{1}{\hat{\sigma}^2} B^T \mathit{\Lambda} \boldsymbol{\omega} \\ \dfrac{1}{\hat{\sigma}^2} \boldsymbol{\omega}^T \mathit{\Lambda} B & \dfrac{1}{2\hat{\sigma}^2} \mathrm{tr}(\mathit{\Omega}) \end{bmatrix}, \tag{10}$$

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

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

Konishi *et al.* [18] extended Schwarz's BIC (Schwarz [23]) to the evaluation of models fitted by the maximum penalized likelihood method. Using the result given in Konishi *et al.* ([18], p. 30) and taking the prior density for the unknown parameter vector $\boldsymbol{\vartheta}$ to be a multivariate normal distribution given by

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

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

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

where R is given by (10).

We choose the optimal values of the smoothing parameters included in wavelet estimator (7) by minimizing GIC or GBIC criterion.

4 Numerical examples

We use a real data example and Monte Carlo simulation to investigate the properties of the proposed nonlinear regression modeling. At first, we consider the problem of choosing the smoothing parameters through the analysis of the motorcycle impact data. By using our regularized wavelet procedure, we estimate regression function $h(t)$ from given data, in which the the smoothing parameters $\nu = (j_0, j_1, \gamma)$ are selected by using five different criteria CV, GCV, C_p, GIC and GBIC. We used wavelet basis of the symmlet-5 which satisfies the 5th order moment condition. The same resolution parameters $\hat{j_0} = 1$ and $\hat{j_1} = 4$ were chosen by all criteria, but the smoothing parameters $\hat{\gamma}$ 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. Figure 1 shows the curve estimated by GBIC, and Figure 2 shows the comparison of the curves with respect to the smoothing parameter γ for five criteria, in which the vertical ranges are different for each criterion.

The second example is the analysis of the simulated data in which the true regression curve is given. The "heavisine" is a sinusoid function with jumps at .3 and .72 which has been studied in Donoho and Johnston [7] and it is explicitly given by

$$h(t) = 4\sin(4\pi t) - \mathrm{sgn}(t - .3) - \mathrm{sgn}(.72 - t).$$

For 1000 trials of Monte Carlo simulation, we repeatedly produced the data $\{(x_l, t_l)\,;\, l = 1\ldots, 100\}$ with true regression model $x_l = h(t_l) + \varepsilon_i$, in which the errors $\{\varepsilon_i\}$ were generated independently from normal distribution with the variances $\sigma^2 = .2$ and $\sigma^2 = .4$. In all trials, we used a fixed set of the design points $\{t_l\,;\, l = 1, \ldots, 100\}$ of which we produced from uniform distribution on $[0, 1]$. Figure 3 illustrates the plots of generated data with corresponding true regression function.

We compared the average squared errors (ASE) ASE $= \sum_{l=1}^{100}\{h(t_l) - \hat{h}(t_l)\}^2$ for $\hat{h}$ estimated by using the criteria CV, GCV, C_p, GIC and GBIC. The most frequently selected resolution parameters were $\hat{j_1} = 5$ and $\hat{j_0} = 3$ in all situations, so we fixed these parameters in simulation. Table 1 summarizes the Monte Carlo results for heavisine regression function, in which the notation MEAN and SD denote the average value and standard deviation of smoothing parameter $\hat{\gamma}$ over the trials. Figure 4 shows the boxplot of $\hat{\gamma}$ for each criterion. The goodness of fit can be compared by the averaged ASE values in the Table. For both cases of $\sigma^2 = .2$ and $\sigma^2 = .4$, the GBIC gave the smallest ASE values in the criteria, and the comparison of SD indicates that both GIC and GBIC gave stable estimates of parameter γ.

Table 1: Monte Carlo results for heavisine function (1000 repetitions)

	CV	GCV	C_p	GIC	GBIC
$\sigma^2 = 0.2\,;$	$j_1 = 5,\ j_0 = 3$				
MEAN of $\hat{\gamma} \times 10^3$	9.580	8.771	9.735	4.099	7.960
SD of $\hat{\gamma} \times 10^3$	3.941	1.871	2.078	1.245	1.377
ASE of $\hat{h}(t) \times 10^2$	2.010	1.992	1.999	1.992	1.983
$\sigma^2 = 0.4\,;$	$j_1 = 5,\ j_0 = 3$				
MEAN of $\hat{\gamma} \times 10^2$	2.307	2.643	2.787	1.410	1.790
SD of $\hat{\gamma} \times 10^3$	12.090	8.689	9.513	5.342	3.924
ASE of $\hat{h}(t) \times 10^2$	4.957	4.908	4.922	4.992	4.876

5 Concluding remarks

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

References

[1] U. Amato, D.T. Vuza, Wavelet approximation of a function from samples affected by noise, Rev. Roum. Math. Pures Appl. 42 (1997) 481–493.

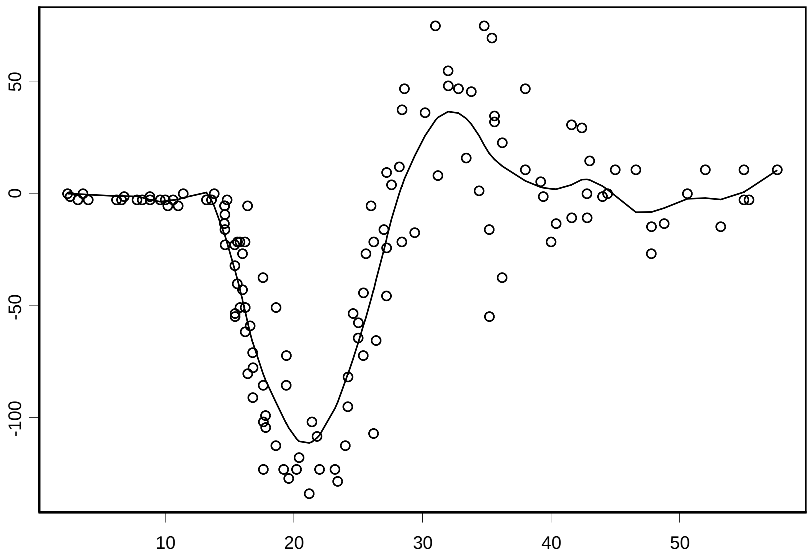


Figure 1: The motorcycle impact data and the curve estimated by using GBIC.

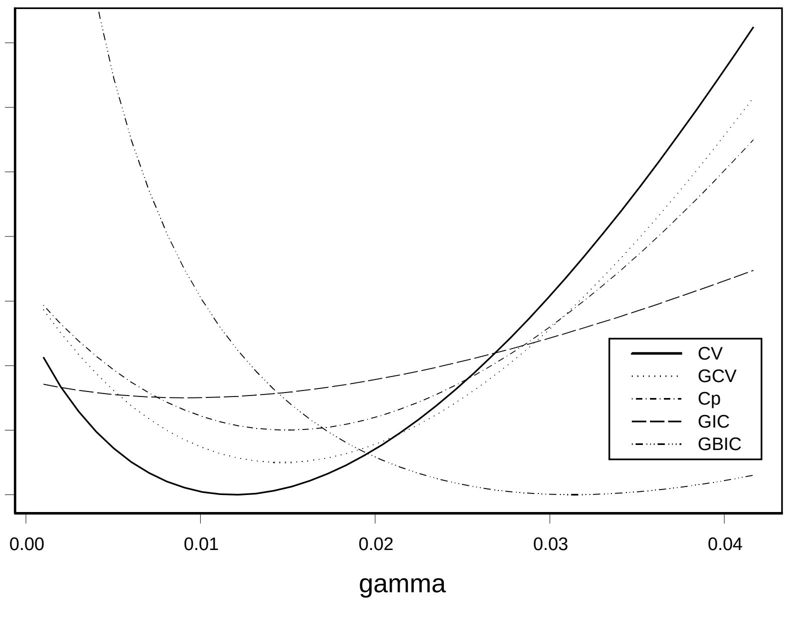


Figure 2: The curves with respect to the smoothing parameter γ for the five criteria.

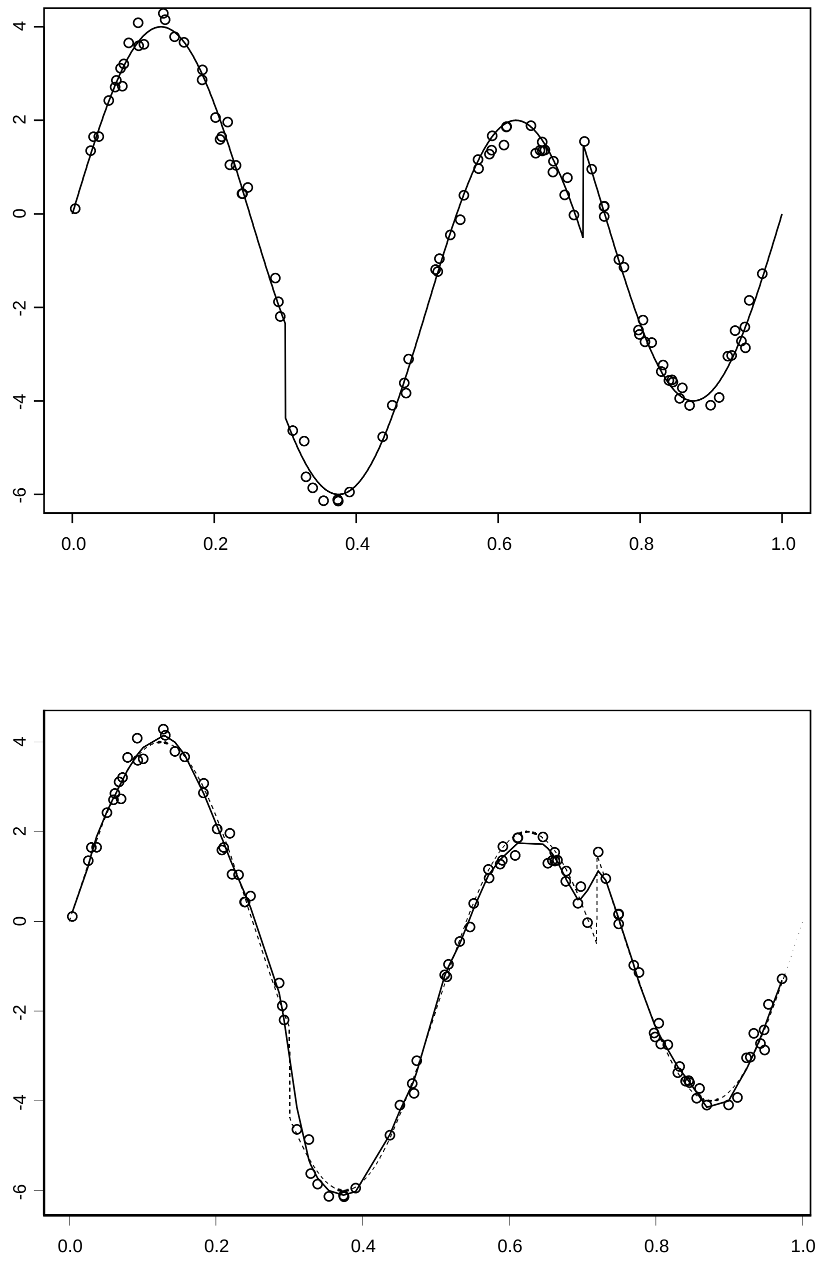
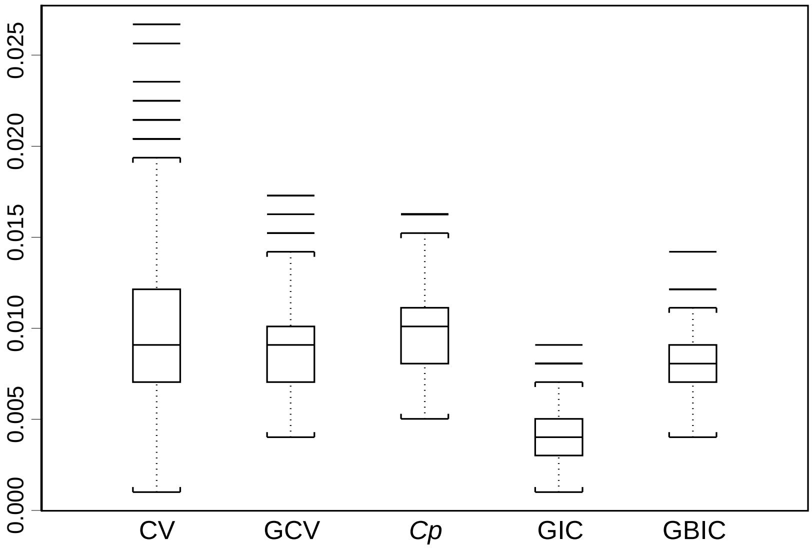


Figure 3: (upper) true heavisine function and the irregularly spaced data ($L = 100$, $\sigma^2 = .2$), (lower) The wavelet estimate based on GBIC (solid line).



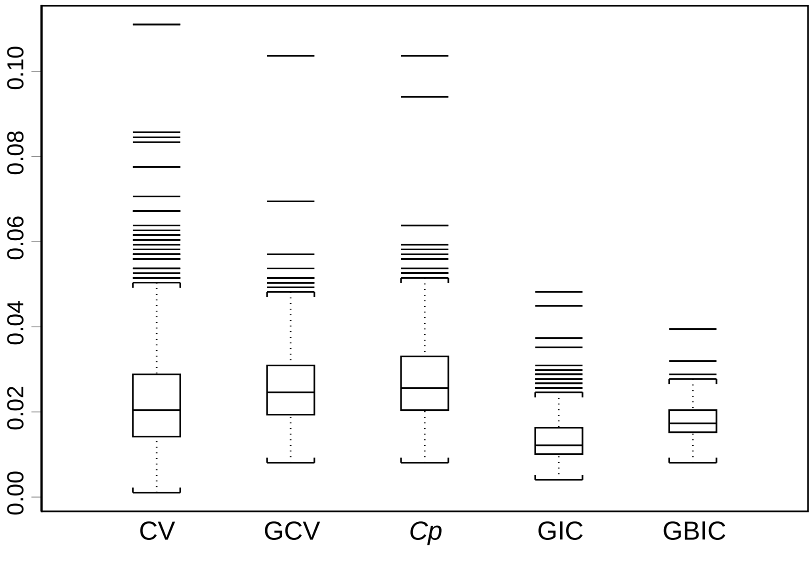


Figure 4: The boxplot of $\hat{\gamma}$ for the data generated from heavisine function with noise variance $\sigma^2 = .2$ (upper) and $\sigma^2 = .4$ (lower). The resolution parameters were fixed at $j_1 = 5$ and $j_0 = 3$.

[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] P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numer. Math. 31 (1979) 377–403.

[6] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.

[7] D.L. Donoho, I.M. Johnston, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. 90 (1995) 1200–1224.

[8] D.L. Donoho, I.M. Johnston, G. Kerkyacharian, D. Picard, Wavelet shrinkage: Asymptopia?, J. Roy. Statist. Soc. B 57 (1995) 301–369.

[9] D.L. Donoho, I.M. Johnston, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, Ann. Statist. 24 (1996) 508–539.

[10] R.L. Eubank, Nonparametric Regression and Spline Smoothing, 2nd Edition, Marcel Dekker, New York, 1999.

[11] Y. Fujikoshi, K. Satoh, Modified AIC and C_p in multivariate linear regression, Biometrika 84 (1997) 707–716.

[12] P.J. Green, B.W. Silverman, Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman & Hall, London, 1994.

[13] 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.

[14] P. Hall, B.A. Turlach, Interpolation methods for nonlinear wavelet regression with irregularly spaced design, Ann Statist. 25 (1997) 1912–1925

[15] W. Härdle, Applied Nonparametric Regression, Cambridge University Press, Cambridge, 1990.

[16] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, 2003.

[17] I.M. Johnston, B.W. Silverman, Wavelet methods for data with correlated noise, J. Roy. Statist. Soc. Ser. B 59 (1997) 319–351.

[18] S. Konishi, T. Ando, S. Imoto, Baysian information criteria and smoothing parameter selection in radial basis function networks, Biometrika 91 (2004) 27–43.

[19] S. Konishi, G. Kitagawa, Generalised information criteria in model selection, Biometrika 83 (1996) 875–890.

[20] C.L. Mallows, Some comments on C_p, Technometrics 15 (1973) 661–675.

[21] Y. Matsushima, S. Shirahata, W. Sakamoto, Wavelet basis selection via AIC (in Japanese), Japanese J. Appl. Statist. 33 (2004) 201–219.

[22] M. Pensky, B. Vidakovic, On non-equally spaced wavelet regression, Ann. Inst. Statist. Math. 53 (2001) 681–690.

[23] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.

[24] B. Vidakovic, Statistical Modeling by Wavelets, John Wiley & Sons, New York, 1999.

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF2003-1 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Yoshitaka WATANABE
A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems

MHF2003-2 Masahisa TABATA & Daisuke TAGAMI
Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients

MHF2003-3 Tomohiro ANDO, Sadanori KONISHI & Seiya IMOTO
Adaptive learning machines for nonlinear classification and Bayesian information criteria

MHF2003-4 Kazuhiro YOKOYAMA
On systems of algebraic equations with parametric exponents

MHF2003-5 Masao ISHIKAWA & Masato WAKAYAMA
Applications of Minor Summation Formulas III, Plücker relations, Lattice paths and Pfaffian identities

MHF2003-6 Atsushi SUZUKI & Masahisa TABATA
Finite element matrices in congruent subdomains and their effective use for large-scale computations

MHF2003-7 Setsuo TANIGUCHI
Stochastic oscillatory integrals - asymptotic and exact expressions for quadratic phase functions -

MHF2003-8 Shoki MIYAMOTO & Atsushi YOSHIKAWA
Computable sequences in the Sobolev spaces

MHF2003-9 Toru FUJII & Takashi YANAGAWA
Wavelet based estimate for non-linear and non-stationary auto-regressive model

MHF2003-10 Atsushi YOSHIKAWA
Maple and wave-front tracking — an experiment

MHF2003-11 Masanobu KANEKO
On the local factor of the zeta function of quadratic orders

MHF2003-12 Hidefumi KAWASAKI
Conjugate-set game for a nonlinear programming problem

MHF2004-15 Kouji HASHIMOTO, Ryohei ABE, Mitsuhiro T. NAKAO & Yoshitaka WATANABE
Numerical verification methods of solutions for nonlinear singularly perturbed problem

MHF2004-16 Ken-ichi MARUNO & Gino BIONDINI
Resonance and web structure in discrete soliton systems: the two-dimensional Toda lattice and its fully discrete and ultra-discrete versions

MHF2004-17 Ryuei NISHII & Shinto EGUCHI
Supervised image classification in Markov random field models with Jeffreys divergence

MHF2004-18 Kouji HASHIMOTO, Kenta KOBAYASHI & Mitsuhiro T. NAKAO
Numerical verification methods of solutions for the free boundary problem

MHF2004-19 Hiroki MASUDA
Ergodicity and exponential β-mixing bounds for a strong solution of Lévy-driven stochastic differential equations

MHF2004-20 Setsuo TANIGUCHI
The Brownian sheet and the reflectionless potentials

MHF2004-21 Ryuei NISHII & Shinto EGUCHI
Supervised image classification based on AdaBoost with contextual weak classifiers

MHF2004-22 Hideki KOSAKI
On intersections of domains of unbounded positive operators

MHF2004-23 Masahisa TABATA & Shoichi FUJIMA
Robustness of a characteristic finite element scheme of second order in time increment

MHF2004-24 Ken-ichi MARUNO, Adrian ANKIEWICZ & Nail AKHMEDIEV
Dissipative solitons of the discrete complex cubic-quintic Ginzburg-Landau equation

MHF2004-25 Raimundas VIDŪNAS
Degenerate Gauss hypergeometric functions

MHF2004-26 Ryo IKOTA
The boundedness of propagation speeds of disturbances for reaction-diffusion systems

MHF2004-27 Ryusuke KON
Convex dominates concave: an exclusion principle in discrete-time Kolmogorov systems

MHF2004-28 Ryusuke KON
Multiple attractors in host-parasitoid interactions: coexistence and extinction

MHF2004-29 Kentaro IHARA, Masanobu KANEKO & Don ZAGIER
Derivation and double shuffle relations for multiple zeta values

MHF2004-30 Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Generalized partitioned quantum cellular automata and quantization of classical CA

MHF2005-1 Hideki KOSAKI
Matrix trace inequalities related to uncertainty principle

MHF2005-2 Masahisa TABATA
Discrepancy between theory and real computation on the stability of some finite element schemes

MHF2005-3 Yuko ARAKI & Sadanori KONISHI
Functional regression modeling via regularized basis expansions and model selection

MHF2005-4 Yuko ARAKI & Sadanori KONISHI
Functional discriminant analysis via regularized basis expansions

MHF2005-5 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Point configurations, Cremona transformations and the elliptic difference Painlevé equations

MHF2005-6 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Construction of hypergeometric solutions to the q - Painlevé equations

MHF2005-7 Hiroki MASUDA
Simple estimators for non-linear Markovian trend from sampled data:
I. ergodic cases

MHF2005-8 Hiroki MASUDA & Nakahiro YOSHIDA
Edgeworth expansion for a class of Ornstein-Uhlenbeck-based models

MHF2005-9 Masayuki UCHIDA
Approximate martingale estimating functions under small perturbations of dynamical systems

MHF2005-10 Ryo MATSUZAKI & Masayuki UCHIDA
One-step estimators for diffusion processes with small dispersion parameters from discrete observations

MHF2005-11 Junichi MATSUKUBO, Ryo MATSUZAKI & Masayuki UCHIDA
Estimation for a discretely observed small diffusion process with a linear drift

MHF2005-12 Masayuki UCHIDA & Nakahiro YOSHIDA
AIC for ergodic diffusion processes from discrete observations

MHF2005-13 Hiromichi GOTO & Kenji KAJIWARA
Generating function related to the Okamoto polynomials for the Painlevé IV equation

MHF2005-14 Masato KIMURA & Shin-ichi NAGATA
Precise asymptotic behaviour of the first eigenvalue of Sturm-Liouville problems with large drift

MHF2005-15 Daisuke TAGAMI & Masahisa TABATA
Numerical computations of a melting glass convection in the furnace

MHF2005-16 Raimundas VIDŪNAS
Normalized Leonard pairs and Askey-Wilson relations

MHF2005-17 Raimundas VIDŪNAS
Askey-Wilson relations and Leonard pairs

MHF2005-18 Kenji KAJIWARA & Atsushi MUKAIHIRA
Soliton solutions for the non-autonomous discrete-time Toda lattice equation

MHF2005-19 Yuu HARIYA
Construction of Gibbs measures for 1-dimensional continuum fields

MHF2005-20 Yuu HARIYA
Integration by parts formulae for the Wiener measure restricted to subsets in $\mathbb{R}^d$

MHF2005-21 Yuu HARIYA
A time-change approach to Kotani's extension of Yor's formula

MHF2005-22 Tadahisa FUNAKI, Yuu HARIYA & Mark YOR
Wiener integrals for centered powers of Bessel processes, I

MHF2005-23 Masahisa TABATA & Satoshi KAIZU
Finite element schemes for two-fluids flow problems

MHF2005-24 Ken-ichi MARUNO & Yasuhiro OHTA
Determinant form of dark soliton solutions of the discrete nonlinear Schrödinger equation

MHF2005-25 Alexander V. KITAEV & Raimundas VIDŪNAS
Quadratic transformations of the sixth Painlevé equation

MHF2005-26 Toru FUJII & Sadanori KONISHI
Nonlinear regression modeling via regularized wavelets and smoothing parameter selection