Nonlinear regression modeling and spike detection via Gaussian basis expansions

Tateishi, Shohei
Graduate School of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Science and Engineering, Chuo University

出版情報：MI Preprint Series. 2010-31, 2010-09-24. 九州大学大学院数理学研究院
バージョン：
権利関係：

MI Preprint Series
Kyushu University
The Global COE Program
Math-for-Industry Education & Research Hub

Nonlinear regression modeling and spike detection via Gaussian basis expansions

Shohei Tateishi
and Sadanori Konishi

MI 2010-31

(Received September 24, 2010)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Nonlinear regression modeling and spike detection via Gaussian basis expansions

Shohei Tateishi* and Sadanori Konishi†

* *Graduate School of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.*

† *Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan.*

s-tateishi@math.kyushu-u.ac.jp konishi@math.chuo-u.ac.jp

Abstract

We consider the problem of constructing nonlinear regression models in the case that the structure of data has abrupt change points at unknown points. We propose two stage procedure where the spikes are detected by fused lasso signal approximator at the first stage, and the smooth curve are effectively estimated along with the technique of regularization method at the second. In order to select tuning parameters in the regularization method, we derive a model selection criterion from information-theoretic viewpoints. Simulation results and real data analysis demonstrate that our methodology performs well in various situations.

Key Words and Phrases: Basis expansion, Information criterion, Lasso, Nonlinear regression, Regularization, Spike detection.

1 Introduction

For analysis of data with complex structure, flexible model is absolutely imperative. As a useful tool to analyse such diverse phenomena, nonlinear regression model based on basis expansions is widely used. The essential idea behind basis expansions is to express a regression function as a linear combination of known functions, called basis functions (Bishop, 2006; Konishi and Kitagawa, 2008; Hastie *et al.*, 2009). In constructing the model, various functions are used to represent a regression function according to the structure of data or analysis objective. For example, natural cubic splines (Green and Silverman, 1994), B-splines (Eilers and Marx, 1996; de Boor, 2001; Imoto and Konishi, 2003) and radial basis functions (Kawano and Konishi, 2007; Ando *et al.*, 2008; Hastie *et al.*, 2009) involving Gaussian basis functions, thin plate splines and so on. In particular, Gaussian basis functions have been most commonly used to construct nonlinear regression

models. Gaussian basis functions have center and width parameters that have to be determined from observed data. The width parameters adjust the amount of overlapping among basis functions and notably play the essential role to capture the structure in the data over the region of the input space. Moody and Darken (1989) used the k-means clustering algorithm to determine the width parameters. In applying nonlinear regression models based on k-means-based Gaussian basis functions, it is assumed that the structure of phenomena are smooth.

However, the underlying true structure which is generating data cannot be smooth in practice at some points where jump discontinuity may occur. Thus, the application of a usual nonlinear regression model described above will lead difficulty of obtaining effective information from the data in which the mean structure is suddenly changed.

In order to overcome this problem, we propose the method of appropriately estimating a nonlinear structure with the spiky change points by applying fused lasso signal approximator (FLSA; Friedman *et al.*, 2007) which is a special version of fused lasso (Tibshirani *et al.*, 2005) to construction of Gaussian basis functions. We present a two-stage procedure to fit spiky regression curve.

In the first stage, we apply FLSA estimation procedure to get the information of locations of spikes. FLSA estimation encourages flatness between the resembled data and the abrupt change before and behind the jump. And then, we introduce the new Gaussian basis functions based on FLSA estimate. In the second stage, we fit the nonlinear regression model with the FLSA-based Gaussian basis functions by the method of regularization. The unknown parameters in regression model are estimated by maximizing the penalized log-likelihood function.

It is a crucial issue to determine the tuning parameters, including a smoothing parameter and a hyperparameter associated with Gaussian basis functions. To choose these parameters, we derive model selection criterion from information-theoretic viewpoint. The proposed nonlinear modeling procedure is investigated through the numerical examples.

This paper is organized as follows. Section 2 describes the framework of basis expansions and Gaussian basis function models. In Section 3 we present a new Gaussian basis functions based on FLSA. Section 4 provides nonlinear regression modeling strate-

gies based on proposed asymmetric Gaussian basis functions using maximum penalized likelihood estimation procedure. Section 5 gives a model selection criterion for evaluating statistical models estimated by the regularization method. In Section 6 we investigate the performance of our nonlinear regression modeling techniques through Monte Carlo simulations and real data example. Some concluding remarks are presented in Section 7.

2 Nonlinear regression model with basis expansions

Suppose that we have n independent observations $\{(y_\alpha, x_\alpha); \alpha = 1, 2, \cdots, n\}$, where y_α are random response variables and x_α are explanatory variables. We consider the regression model

$$y_\alpha = g(x_\alpha) + \epsilon_\alpha, \quad \alpha = 1, 2, \cdots, n, \tag{1}$$

where $g(\cdot)$ is an unknown smooth function and ϵ_α are independently, normally distributed with mean zero and variance σ^2. It is assumed that the function $g(\cdot)$ can be expressed as a linear combination of basis functions $b_j(x)$ $(j = 1, 2, \cdots, m)$ in the form

$$g(x; \boldsymbol{w}) = w_0 + \sum_{j=1}^{m} w_j b_j(x) = \boldsymbol{w}^T \boldsymbol{b}(x) \ , \tag{2}$$

where $\boldsymbol{b}(x) = (1, b_1(x), \cdots, b_m(x))^T$ is a vector of basis functions and $\boldsymbol{w} = (w_0,\ w_1, \cdots,\ w_m)^T$ is an unknown coefficient parameter vector. A variety of basis functions are used according to the structure of data.

One of the many basis functions is Gaussian basis function given by

$$b_j(x) = \exp\left\{-\frac{(x - c_j)^2}{2h_j^2}\right\}, \quad j = 1, 2, \cdots, m, \tag{3}$$

where c_j is the center of the basis function, h_j^2 is a parameter that determines the dispersion. However, basis functions (3) often yield inadequate results because of the lack of overlapping among basis functions. In order to overcome this problem, Ando *et al.* (2008) proposed the use of Gaussian basis functions with a hyperparameter, i.e. functions of the form

$$b_j(x) = \exp\left\{-\frac{(x - c_j)^2}{2\nu h_j^2}\right\}, \quad j = 1, 2, \cdots, m, \tag{4}$$

where ν is a hyperparameter that adjusts the dispersion of basis functions. Ando *et al.* (2008) showed that nonlinear models with these basis functions were effective in capturing the information from the data.

Unknown parameters in the Gaussian basis functions (4) include the centers c_j and dispersion parameters h_j^2. These parameters are generally determined by using unsupervised learning. For example, Moody and Darken (1989) determined the centers c_j and dispersion h_j^2 using the k-means clustering algorithm to avoid local minimum and identification problems. The data set of observations of the explanatory variables $\{x_1, \cdots, x_n\}$ is divided into m clusters $\{C_1, \cdots, C_m\}$; centers c_j and dispersions h_j^2 are determined by

$$\hat{c}_j = \frac{1}{n_j} \sum_{x_\alpha \in C_j} x_\alpha, \quad \hat{h}_j^2 = \frac{1}{n_j} \sum_{x_\alpha \in C_j} (x_\alpha - \hat{c}_j)^2, \tag{5}$$

where n_j is the number of observations included in the the jth cluster C_j. Replacing c_j and h_j^2 in equation (3) by $\hat{c}_j$ and $\hat{h}_j^2$ respectively, we obtain a set of m basis functions

$$b_j(x; \hat{c}_j, \hat{h}_j^2) = \exp\left\{ -\frac{(x - \hat{c}_j)^2}{2\nu \hat{h}_j^2} \right\}, \quad j = 1, 2, \cdots, m. \tag{6}$$

However, the models with these k-means-based Gaussian basis functions will lead to smooth curve estimates, even though spikes are present. Therefore, they will underestimate spikes and sudden change points will not be visible in resulting curve. In order to overcome this problem, we construct new Gaussian basis functions using fused lasso signal approximator (FLSA; Friedman *et al.*, 2007).

3 Gaussian basis functions based on FLSA

For n independent observations $\{(y_\alpha, x_\alpha); \alpha = 1, \cdots, n\}$, FLSA procedure minimizes

$$\frac{1}{2} \sum_{\alpha=1}^{n} (y_\alpha - \beta_\alpha)^2 + \lambda_1 \sum_{\alpha=1}^{n} |\beta_\alpha| + \lambda_2 \sum_{\alpha=1}^{n-1} |\beta_\alpha - \beta_{\alpha+1}|, \ \alpha = 1, \cdots, n, \tag{7}$$

where β_α is an estimate of response variable y_α taken at position α, and (λ_1, λ_2) are positive smoothing parameters to be chosen appropriately. The first penalty encourages sparsity in βs and second penalty encourages sparsity in their neighboring differences. As the value of λ_2 increases, the number of fused parameters increases. Applying FLSA, the data set of n observations of the explanatory variables $\{x_1, \cdots, x_n\}$ is divided into n_F

sets $\{U_1, \cdots, U_{n_F}\}$ where $U_j = \{x_\alpha | \alpha \in F_j\} = \{x_1^{F_j}, \cdots, x_{n_j}^{F_j}\}$, F_j is the jth subset of $\{1, \cdots, n\}$ that are considered to be fused at the λ_2, that is, $\hat{\beta}_{\alpha_1} = \hat{\beta}_{\alpha_2}$ $(\forall \alpha_1, \alpha_2 \in F_j)$ and n_F is the number of such sets.

We take the sets U_j $(j = 1, \cdots, n_F)$ the clusters and then we construct the asymmetric Gaussian basis functions based on FLSA as follows:

$$b_j(x) = \begin{cases} \exp\left\{-\dfrac{(x-c_j)^2}{2\nu h_{1j}^2}\right\}, & (x < c_j) \\ 1, & (x = c_j) \\ \exp\left\{-\dfrac{(x-c_j)^2}{2\nu h_{2j}^2}\right\}, & (x > c_j) \end{cases} \quad j = 1, 2, \cdots, m. \tag{8}$$

We take the number of clusters n_F as that of basis functions (i.e. $m = n_F$) and set basis functions on each center of clusters. And then we determine dispersion parameters using gradients of FLSA estimates; for fixed N_{ij}, centers c_j and dispersions h_j^2 are determined by

$$\hat{c}_j = \frac{1}{n_j} \sum_{x_\alpha \in U_j} x_\alpha, \quad \hat{h}_{ij}^2 = \frac{N_{ij}}{|G_{ij}|} \exp\left(-\frac{1}{2} N_{ij}^2\right), \quad i = 1, 2, \quad j = 1, \cdots, m, \tag{9}$$

where n_j is the number of observations included in the the jth clusters F_j, and we defined gradients G_{ij} as follows:

$$G_{1j} = \frac{\hat{\beta}_{F_j} - \hat{\beta}_{F_{j-1}}}{x_{n_j}^{F_j} - x_{n_{j-1}}^{F_{j-1}}}, \quad G_{2j} = \frac{\hat{\beta}_{F_{j+1}} - \hat{\beta}_{F_j}}{x_1^{F_{j+1}} - x_1^{F_j}}. \tag{10}$$

These dispersion parameters $\hat{h}_{ij}$ are the solutions that the gradients of Gaussian basis function (3) at $c_j - N_{ij} h_{ij}$ and $c_j + N_{2j} h_{ij}$ equal G_{1j} and G_{2j} respectively. N_{ij} are parameters which adjust the widths of basis functions (8). As value of N increases, the width of Gaussian basis function decreases. In this study, we fixed N as follows:

$$N_{ij} = \begin{cases} 3, & (\exp(-1/2)/|G_{ij}| < \delta) \\ 1, & (otherwise) \end{cases} \tag{11}$$

where δ is enough small.

Replacing c_j and h_j^2 in equation (8) by $\hat{c}_j$ and $\hat{h}_j^2$ respectively, we obtain a set of m basis functions

$$b_j(x) = \begin{cases} \exp\left\{-\dfrac{(x-\hat{c}_j)^2}{2\nu \hat{h}_{1j}^2}\right\}, & (x < \hat{c}_j) \\ 1, & (x = \hat{c}_j) \\ \exp\left\{-\dfrac{(x-\hat{c}_j)^2}{2\nu \hat{h}_{2j}^2}\right\}, & (x > \hat{c}_j) \end{cases} \quad j = 1, 2, \cdots, m. \tag{12}$$

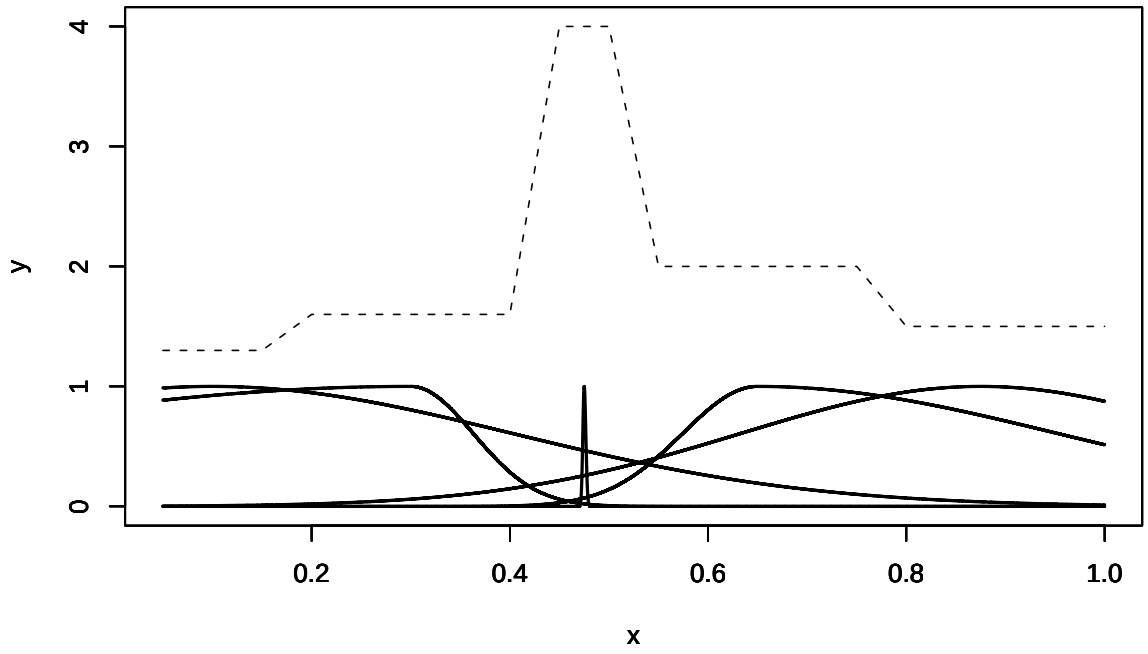

Fig. 1: Simple example of constructing our proposed Gaussian basis functions for $\nu = 1$ (solid line) based on FLSA estimates (dashed line).

The width of Gaussian basis function $b(x)$ will be very small if the absolute gradients of FLSA $|G_{1j}|$ and $|G_{2j}|$ are very large. Figure 1 shows FLSA estimates and our proposed Gaussian basis functions. Actually, in Figure 1, the central basis function become very narrow compared with the others.

4 Maximum penalized likelihood estimation

For n independent observations $\{(y_\alpha, x_\alpha); \alpha = 1, \cdots, n\}$, the nonlinear regression model based on Gaussian basis functions $b_j(x)$ $(j = 1, \cdots, n)$ given in Section 3 is expressed as

$$y_\alpha = \boldsymbol{w}^T \boldsymbol{b}(x_\alpha) + \epsilon_\alpha, \ \alpha = 1, \cdots, n, \tag{13}$$

where $\boldsymbol{b}(x_\alpha) = (1, b_1(x_\alpha), \cdots, b_m(x_\alpha))^T$, $\boldsymbol{w} = (w_0, w_1, \cdots, w_m)^T$ and ϵ_α are error terms. If the error terms ϵ_α are independently and normally distributed with mean 0 and variance

σ^2, the nonlinear regression model (13) has a probability density function

$$f(y_\alpha|\boldsymbol{w},\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\{y_\alpha-\boldsymbol{w}^T\boldsymbol{b}(x_\alpha)\}^2}{2\sigma^2}\right],\ \alpha=1,\cdots,n. \tag{14}$$

And then, the coefficient parameters w_j $(j=0,1,\cdots,m)$ are estimated by the maximum penalized likelihood method.

The maximum likelihood estimates of the coefficient vectors $\boldsymbol{w}$ and σ^2 are respectively given by

$$\hat{\boldsymbol{w}}=(B^TB)^{-1}B^T\boldsymbol{y},\qquad \hat{\sigma}^2=\frac{1}{n}(\boldsymbol{y}-B\hat{\boldsymbol{w}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}),$$

where B= $(\boldsymbol{b}(x_1)^T,\cdots,\boldsymbol{b}(x_n)^T)^T$ and $\boldsymbol{y}=(y_1,\cdots,y_n)^T$. However, when fitting a nonlinear model to data with a complex structure the maximum likelihood method often yields unstable estimates and leads to overfitting. We therefore estimate $\boldsymbol{w}$ and σ^2 by the method of regularization. Instead of using the log-likelihood function, we consider maximizing the penalized log-likelihood function imposing ridge penalty (Hoerl and Kennard,1970)

$$l_\gamma(\boldsymbol{\theta})=\sum_{\alpha=1}^{n}\log f(y_\alpha|\boldsymbol{w},\sigma^2)-\frac{n\gamma}{2}\boldsymbol{w}^TK\boldsymbol{w}, \tag{15}$$

where $\boldsymbol{\theta}=(\boldsymbol{w}^T,\sigma^2)^T$, $\gamma\,(>0)$ is a smoothing parameter that controls the smoothness of the fitted model and K is a known $(m+1)$th square matrix (Konishi and Kitagawa, 2008). The typical form of K is given by $K=I_{m+1}$ for the identity matrix or $K=D_2^TD_2$ for a second-order difference matrix. Then, the maximum penalized likelihood estimates of $\boldsymbol{w}$ and σ^2 are respectively given by

$$\hat{\boldsymbol{w}}=(B^TB+n\gamma\hat{\sigma}^2K)^{-1}B^T\boldsymbol{y},\quad \hat{\sigma}^2=\frac{1}{n}(\boldsymbol{y}-B\hat{\boldsymbol{w}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}). \tag{16}$$

Note that these estimators depend on each other. Therefore, we provide an initial value for the variance $\sigma^2_{x(0)}$ first, then $\hat{\boldsymbol{w}}$ and $\hat{\sigma}^2_x$ are updated until convergence. The ridge estimators continuously shrink the coefficients as γ increases.

5 Model selection criterion

The statistical model estimated by the regularization method depends upon the value of the smoothing parameter γ and the value of the hyperparameter ν in the Gaussian basis functions. It is a crucial issue to determine these values appropriately.

Konishi and Kitagawa (1996, 2008) introduced evaluation criteria of statistical models that can be applied to the evaluation of statistical models estimated by various types of estimation procedures such as the robust and penalized likelihood procedures. By using the result, the model selection criterion for evaluating the statistical model constructed by Gaussian basis functions is given by

$$\mathrm{GIC} = n\{\log(2\pi) + 1\} + n\log\hat{\sigma}^2 + 2\mathrm{tr}\{R^{-1}Q\}, \tag{17}$$

where R and Q are $(m+2)$th square matrices and are, respectively, given by

$$R = \frac{1}{n\hat{\sigma}^2}\begin{bmatrix} B^TB + n\gamma\hat{\sigma}^2K & \frac{1}{\hat{\sigma}^2}B^T\Lambda\mathbf{1}_n \\ \frac{1}{\hat{\sigma}^2}\mathbf{1}_n^T\Lambda B & \frac{n}{2\hat{\sigma}^2} \end{bmatrix}, \tag{18}$$

$$Q = \frac{1}{n\hat{\sigma}^2}\begin{bmatrix} \frac{1}{\hat{\sigma}^2}B^T\Lambda^2B - \gamma K\hat{\boldsymbol{w}}\mathbf{1}_n^T\Lambda B & \frac{1}{2\hat{\sigma}^4}B^T\Lambda^3\mathbf{1}_n - \frac{1}{2\hat{\sigma}^2}B^T\Lambda\mathbf{1}_n \\ \frac{1}{2\hat{\sigma}^4}\mathbf{1}_n^T\Lambda^3B - \frac{1}{2\hat{\sigma}^2}\mathbf{1}_n^T\Lambda B & \frac{1}{4\hat{\sigma}^6}\mathbf{1}_n^T\Lambda^4\mathbf{1}_n - \frac{n}{4\hat{\sigma}^2} \end{bmatrix} \tag{19}$$

with $\mathbf{1}_n = (1, \cdots, 1)^T$ and $\Lambda = \mathrm{diag}(y_1 - \hat{\boldsymbol{w}}^T\boldsymbol{b}(x_1), \cdots, y_n - \hat{\boldsymbol{w}}^T\boldsymbol{b}(x_n))$. We obtain the optimal tuning parameters that minimize GIC.

6 Numerical examples

In this section, we investigate the performance of our nonlinear regression modeling techniques through Monte Carlo simulations and real data example. In FLSA procedure, we fixed $\lambda_1 = 0$ and select the value of λ_2 by using cross validation (CV) and we use an identity matrix as K in (15). In nonlinear regression modeling procedure with our proposed Gaussian basis functions based on FLSA (FLSA-G) and k-means-clustering (KM-G), the model selection criterion GIC was used for choosing optimal values of (γ, ν) and (γ, ν, m) respectively.

6.1 Monte Carlo simulations

We conducted Monte Carlo simulations to investigate the effectiveness of our proposed nonlinear regression modeling. For the first simulation study, repeated random samples $\{(x_\alpha, y_\alpha); \alpha = 1, \cdots, n\}$ with $n = 100$ were generated from a true regression model $y_\alpha = u(x_\alpha) + v_S(x_\alpha) + \epsilon_\alpha$, where $u(x)$ is smooth function and $v_S(x)$ is spike function that has

peak region $S \subset \{x_1, \cdots, x_n\}$. The $x_\alpha = \alpha/n$ are n equally spaced design points in the interval $[0, 1]$ and the errors ϵ_α are independently, normally distributed with mean 0 and standard deviation $\tau = 0.2R_u$ with R_u being the range of $u(x)$ over $x \in [0, 1]$. We considered the following four cases for the true regression model:

$$\text{(a)} \quad \begin{cases} u(x) &= 0.6\exp\{-(x-0.3)^2/0.01\} + 1.4\exp\{-(x-0.7)^2/0.02\}, \\ v_S(x) &= -1.0 \cdot \mathrm{I}(x \in [0.44, 0.47]) - 0.2 \cdot \mathrm{I}(x \in [0.45, 0.46]), \\ S &= \{0.44, 0.45, 0.46, 0.47\}, \end{cases}$$

$$\text{(b)} \quad \begin{cases} u(x) &= \sin(4\pi x), \\ v_S(x) &= -1.5 \cdot \mathrm{I}(x \in [0.36, 0.39] \cup [0.86, 0.89]) - 0.5 \cdot \mathrm{I}(x \in [0.37, 0.38] \cup [0.87, 0.88]), \\ S &= \{0.36, 0.37, 0.38, 0.39, 0.86, 0.87, 0.88, 0.89\}, \end{cases}$$

$$\text{(c)} \quad \begin{cases} u(x) &= \sin(4\pi x) + 2\cos(5\pi x), \\ v_S(x) &= -1.5 \cdot \mathrm{I}(x \in [0.12, 0.15]) - 0.3 \cdot \mathrm{I}(x \in [0.13, 0.14]) \\ & \quad +1.5 \cdot \mathrm{I}(x \in [0.55, 0.58] \cup [0.85, 0.88]) + 0.3 \cdot \mathrm{I}(x \in [0.56, 0.57] \cup [0.86, 0.87]), \\ S &= \{0.12, 0.13, 0.14, 0.15, 0.55, 0.56, 0.57, 0.58, 0.85, 0.86, 0.87, 0.88\}, \end{cases}$$

$$\text{(d)} \quad \begin{cases} u(x) &= \exp(-x)\sin(5\pi\exp(-x)), \\ v_S(x) &= 0, \\ S &= \emptyset. \end{cases}$$

where $\mathrm{I}(\cdot)$ is an indicator function of the event A, that is, $\mathrm{I}(A) = 1$ if A is true and 0 otherwise. We compared the performance of FLSA-G with that of KM-G and FLSA estimation procedure (FLSA).

Figure 2 compares true curves with fitted ones. The estimated curves from the 1st row to 4th row correspond to the true curves from (a) to (d). We observe that our modeling strategy is effective in capturing the true data structures well.

We also performed 1000 repetitions, then calculated averages of mean squared errors on peak region (AMSE.p), that on smooth region (AMSE.s) and that on global region (AMSE.g) defined by MSE.p $= \sum_{x_\alpha \in S}\{(u(x_\alpha) + v(x_\alpha)) - \hat{y}_\alpha\}^2/n_S$, MSE.s $= \sum_{x_\alpha \in U \setminus S}\{(u(x_\alpha) + v(x_\alpha)) - \hat{y}_\alpha\}^2/(n - n_S)$ and MSE.g $= \sum_{\alpha=1}^{n}\{(u(x_\alpha) + v(x_\alpha)) - \hat{y}_\alpha\}^2/n$ respectively to assess the goodness of fit.

Table 1 shows summaries of the simulation results from (a) to (d). In all situations, our proposed modeling procedure minimized all of the AMSE.p, AMSE.s and AMSE.g, thus improving the accuracy of prediction. In the case of (d), the performance of our proposed method (FLSA-G) was superior to that of Km-G even though true model has no spiky change point.

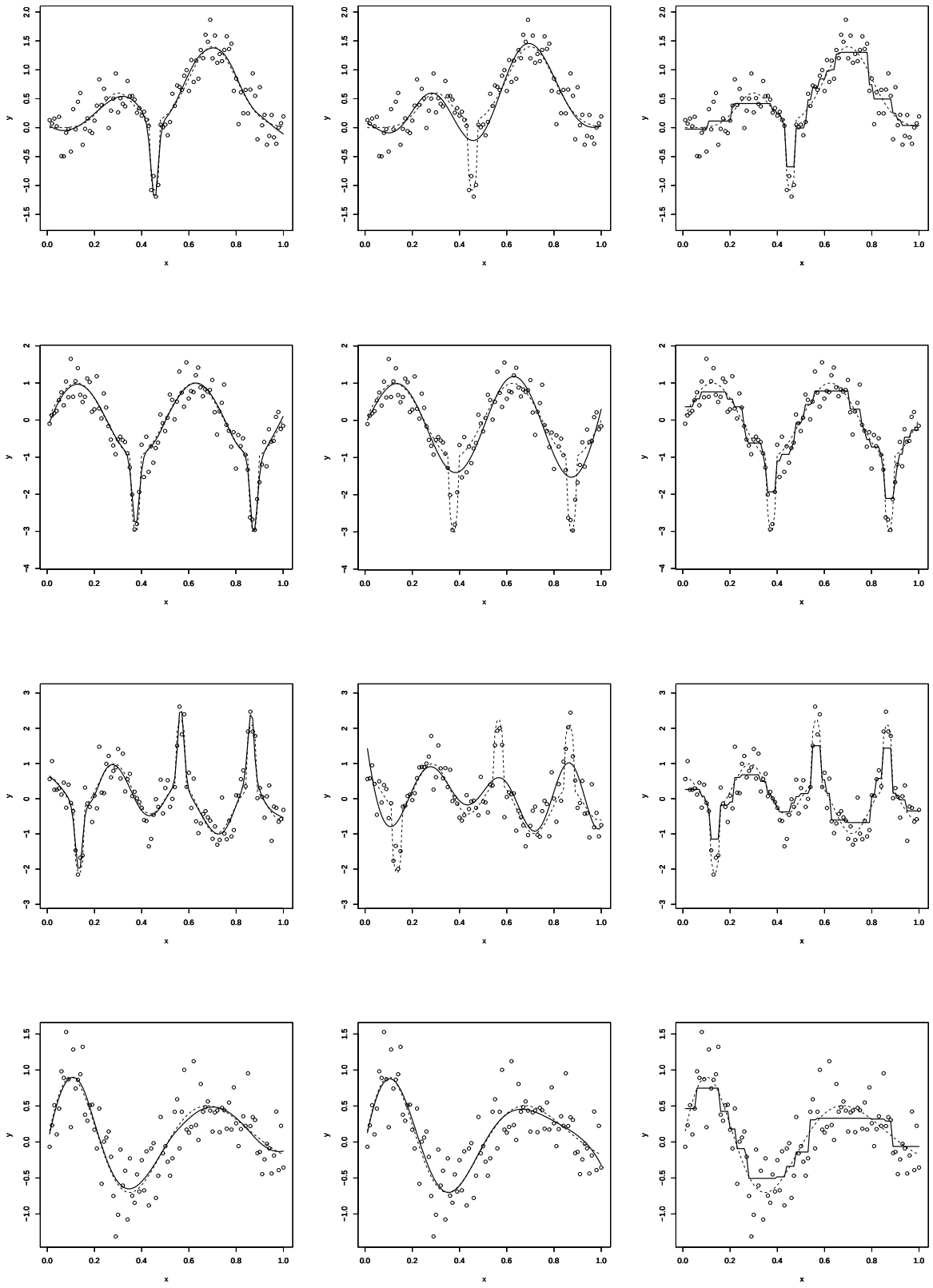

Fig. 2: These panels show true curves (dashed line) and estimated curves (solid line) obtained by FLSA-G (left), KM-G (center) and FLSA (right) for each true function.

Table 1: Comparison of results for curve fitting.

function		mean m	mean γ	mean ν	AMSE.p	AMSE.s	AMSE.g
(a)	FLSA-G	17.50	3.96×10^{-5}	58.26	1.57×10^{-1}	1.45×10^{-2}	2.02×10^{-2}
	KM-G	9.22	2.98×10^{-2}	4.26	3.96×10^{-1}	2.12×10^{-2}	3.62×10^{-2}
	FLSA	$--$	$--$	$--$	2.42×10^{-1}	1.77×10^{-2}	2.67×10^{-2}
(b)	FLSA-G	22.88	4.16×10^{-5}	190.03	1.95×10^{-1}	2.98×10^{-2}	4.30×10^{-2}
	KM-G	7.00	4.07×10^{-3}	3.60	1.39	1.08×10^{-1}	2.10×10^{-1}
	FLSA	$--$	$--$	$--$	3.12×10^{-1}	4.65×10^{-2}	6.78×10^{-2}
(c)	FLSA-G	21.67	1.59×10^{-5}	189.50	1.80×10^{-1}	4.91×10^{-2}	6.48×10^{-2}
	KM-G	10.52	3.05×10^{-3}	5.62	9.35×10^{-1}	1.07×10^{-1}	2.07×10^{-1}
	FLSA	$--$	$--$	$--$	3.49×10^{-1}	5.47×10^{-2}	9.00×10^{-2}
(d)	FLSA-G	16.86	1.04×10^{-4}	65.69	$--$	$--$	9.92×10^{-3}
	KM-G	7.23	9.21×10^{-3}	15.74	$--$	$--$	1.20×10^{-2}
	FLSA	$--$	$--$	$--$	$--$	$--$	1.99×10^{-2}

6.2 Real data example

We investigate the performance using the nursing time of beluga whale data set (Simonoff ,1996; Russell *et al.*, 1997). The data consists of 228 measurements of nursing time in seconds of a newborn male beluga whale calf named Hudson born in captivity at the New York Aquarium. This data was observed every six hours for 57 days after his birth. Russell *et al.* (1997) noted that the nursing time typically peaked at around 7-10 days postpartum from the biological view. We apply our proposed method, nonlinear regression modeling procedure with k-means-based Gaussian basis functions and FLSA estimation procedure to the Beluga data.

Figure 2 shows the result curves obtained by the three methods. Using our method, spiky change appears in around the 9th day of his life as Cheng and Raimondo (2008). We observe that our modeling procedure captures the abrupt change more remarkably than the others and lead smooth curve .

7 Concluding remarks

We have proposed a nonlinear regression modeling procedure along with the technique of the method of regularization. We have introduced new asymmetric Gaussian basis

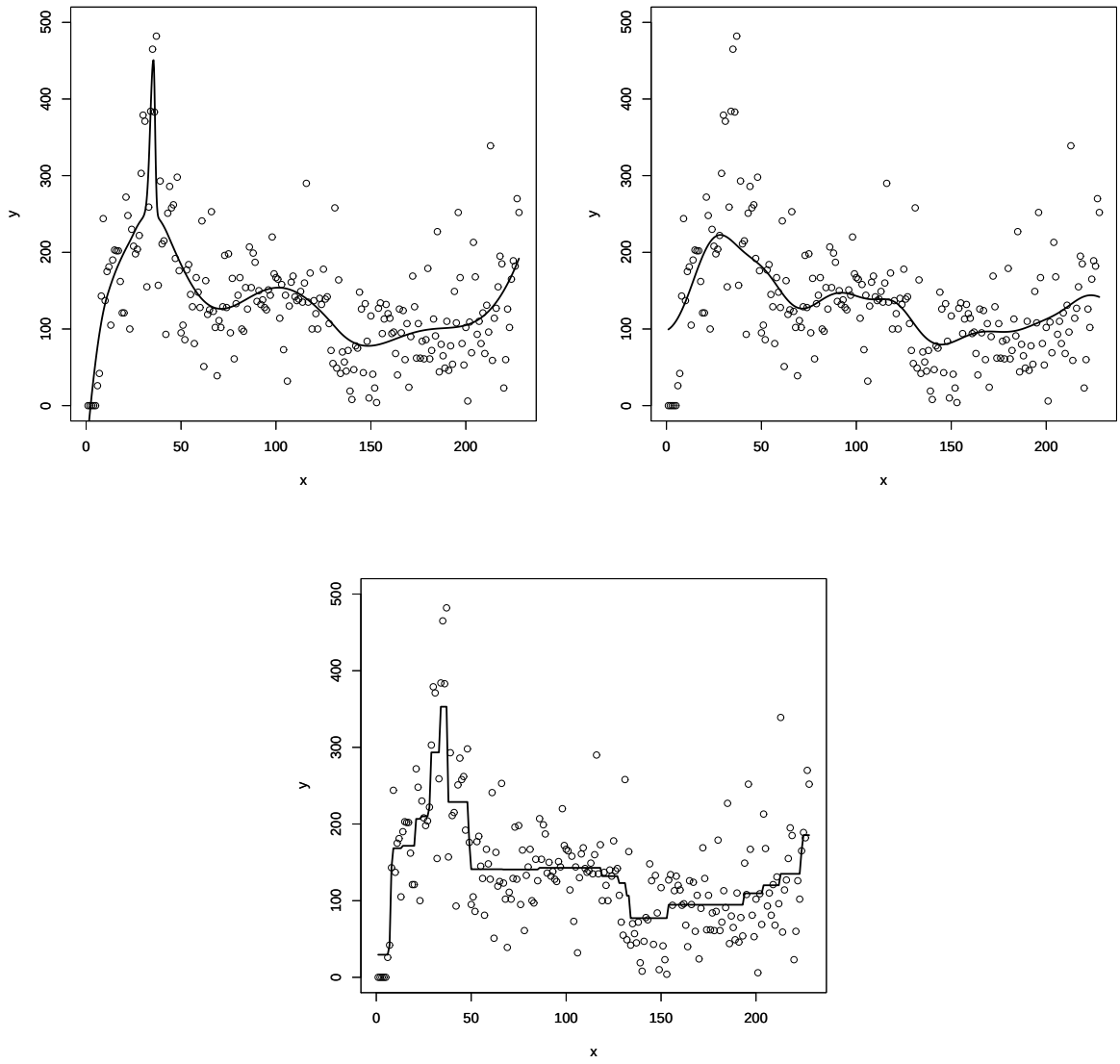

Fig. 3: Beluga data and estimated curves obtained by FLSA-G (Upper left), KM-G (Upper right) and FLSA (Lower).

functions based on FLSA estimates taking the information of spikes into account. The nonlinear regression model with these basis functions is able to capture the sudden changes in nonlinear structure of data. In order to choose tuning parameters, we presented the model selection criterion from information-theoretic approach. The simulation results reported here demonstrate the effectiveness of the proposed modeling strategy in terms of prediction accuracy and prominent visualization of spikes.

References

[1] Ando, T., Konishi, S. and Imoto, S. (2008). Nonlinear regression modeling via regularized radial basis function networks. *Journal of Statistical Planning and Inference* **138**, 3616–3633.

[2] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.

[3] Cheng, M. Y. and Raimondo, M. (2008). Kernel methods for optimal change-points estimation in derivatives. *Journal of Computational and Graphical Statistics*, **17**, 1–20.

[4] de Boor, C. (2001). *A Practical Guide to Splines.* Springer.

[5] Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.*, **11**, 89–121.

[6] Friedman, J., Hastie, T., Höfling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Statist.*, **1**, 302–332.

[7] Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach.* Chapman and Hall, London.

[8] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning.* 2nd ed. New York: Springer.

[9] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems, *Technometrics* **12**, 55–67.

[10] Imoto, S. and Konishi, S. (2003). Selection of smoothing parameter in B-spline nonparametric regression models using information criteria. *Annals of the Institute of Statistical Mathematics* **55**, 671–687.

[11] Kawano, S. and Konishi, S. (2007). Nonlinear regression modeling via regularized Gaussian basis functions. *Bull. Inform. Cybern.*, **39**, 83–96.

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

[13] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer.

[14] Moody, J. and Darken, C. J. (1989). Fast learning in networks of locally-turned processing units. *Neural Computation*. **1**, 281–294.

[15] Russell, J. M., Simonoff, J. S. and Nightingale, J. (1997). Nursing behaviors of beluga calves (Delphinapterus leucas) born in capativity. *Zoo Biology*, **16**, 247–262.

[16] Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer-Verlag, New York.

[17] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. Roy. Statist. Soc. Ser. B*, **67**, 91–108.

List of MI Preprint Series, Kyushu University

The Global COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata

MI2008-2 Eiji ONODERA
The intial value problem for a third-order dispersive flow into compact almost Hermitian manifolds

MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space

MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme

MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in nfinite extensions over a p-adic field

MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields

MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited

MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition

MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials

MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors

MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the L_1 regularization

MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions

MI2009-5 Toshiro HIRANOUCHI & Yuichiro TAGUCHII
Flat modules and Groebner bases over truncated discrete valuation rings

MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations

MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow

MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization

MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity

MI2009-10 Shingo SAITO
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption

MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve

MI2009-12 Tetsu MASUDA
Hypergeometric τ-functions of the q-Painlevé system of type $E_8^{(1)}$

MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications

MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on L^p spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE
Spectrum in multi-species asymmetric simple exclusion process on a ring

MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO
Non-linear algebraic differential equations satisfied by certain family of elliptic functions

MI2009-18 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of an Elliptical Flow Subjected to a Coriolis Force

MI2009-19 Mitsunori KAYANO & Sadanori KONISHI
Sparse functional principal component analysis via regularized basis expansions and its application

MI2009-20 Shuichi KAWANO & Sadanori KONISHI
Semi-supervised logistic discrimination via regularized Gaussian basis expansions

MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations

MI2009-22 Eiji ONODERA
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces

MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions

MI2009-24 Yu KAWAKAMI
Recent progress in value distribution of the hyperbolic Gauss map

MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for H_0^2-projection

MI2009-26 Manabu YOSHIDA
Ramification of local fields and Fontaine's property (Pm)

MI2009-27 Yu KAWAKAMI
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space

MI2009-28 Masahisa TABATA
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme

MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance

MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke's zeros and higher depth determinants

MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type

MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect

MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type$(A_2 + A_1)^{(1)}$

MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations

MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models

MI2010-4 Nobuyuki IKEDA & Setsuo TANIGUCHI
The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons

MI2010-5 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and detecting change point via the relevance vector machine

MI2010-6 Shuichi KAWANO, Toshihiro MISUMI & Sadanori KONISHI
Semi-supervised logistic discrimination via graph-based regularization

MI2010-7 Teruhisa TSUDA
UC hierarchy and monodromy preserving deformation

MI2010-8 Takahiro ITO
Abstract collision systems on groups

MI2010-9 Hiroshi YOSHIDA, Kinji KIMURA, Naoki YOSHIDA, Junko TANAKA & Yoshihiro MIWA
An algebraic approach to underdetermined experiments

MI2010-10 Kei HIROSE & Sadanori KONISHI
Variable selection via the grouped weighted lasso for factor analysis models

MI2010-11 Katsusuke NABESHIMA & Hiroshi YOSHIDA
Derivation of specific conditions with Comprehensive Groebner Systems

MI2010-12 Yoshiyuki KAGEI, Yu NAGAFUCHI & Takeshi SUDOU
Decay estimates on solutions of the linearized compressible Navier-Stokes equation around a Poiseuille type flow

MI2010-13 Reiichiro KAWAI & Hiroki MASUDA
On simulation of tempered stable random variates

MI2010-14 Yoshiyasu OZEKI
Non-existence of certain Galois representations with a uniform tame inertia weight

MI2010-15 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of a Rotating Flow Driven by Precession of Arbitrary Frequency

MI2010-16 Yu KAWAKAMI & Daisuke NAKAJO
The value distribution of the Gauss map of improper affine spheres

MI2010-17 Kazunori YASUTAKE
On the classification of rank 2 almost Fano bundles on projective space

MI2010-18 Toshimitsu TAKAESU
Scaling limits for the system of semi-relativistic particles coupled to a scalar bose field

MI2010-19 Reiichiro KAWAI & Hiroki MASUDA
Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling

MI2010-20 Yasuhide FUKUMOTO, Makoto HIROTA & Youichi MIE
Lagrangian approach to weakly nonlinear stability of an elliptical flow

MI2010-21 Hiroki MASUDA
Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test

MI2010-22 Toshimitsu TAKAESU
A Generalized Scaling Limit and its Application to the Semi-Relativistic Particles System Coupled to a Bose Field with Removing Ultraviolet Cutoffs

MI2010-23 Takahiro ITO, Mitsuhiko FUJIO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Composition, union and division of cellular automata on groups

MI2010-24 Toshimitsu TAKAESU
A Hardy's Uncertainty Principle Lemma in Weak Commutation Relations of Heisenberg-Lie Algebra

MI2010-25 Toshimitsu TAKAESU
On the Essential Self-Adjointness of Anti-Commutative Operators

MI2010-26 Reiichiro KAWAI & Hiroki MASUDA
On the local asymptotic behavior of the likelihood function for Meixner Lévy processes under high-frequency sampling

MI2010-27 Chikashi ARITA & Daichi YANAGISAWA
Exclusive Queueing Process with Discrete Time

MI2010-28 Jun-ichi INOGUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Motion and Bäcklund transformations of discrete plane curves

MI2010-29 Takanori YASUDA, Masaya YASUDA, Takeshi SHIMOYAMA & Jun KOGURE
On the Number of the Pairing-friendly Curves

MI2010-30 Chikashi ARITA & Kohei MOTEGI
Spin-spin correlation functions of the q-VBS state of an integer spin model

MI2010-31 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and spike detection via Gaussian basis expansions