



Automatic identification of curve shapes with applications to ultrasonic vocalization

Zhikun Gao^a, Yanlin Tang^b, Huixia Judy Wang^{a,*}, Guangying K. Wu^c, Jeff Lin^d

^a Department of Statistics, George Washington University, United States of America

^b Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, School of Statistics, East China Normal University, China

^c Division of Biomedical Physics, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland, United States of America

^d Department of Psychology, George Washington University, United States of America

ARTICLE INFO

Article history:

Received 19 August 2019

Received in revised form 10 March 2020

Accepted 11 March 2020

Available online 19 March 2020

Keywords:

Curve classification

Nonparametric regression

Penalization

Shape identification

Ultrasonic vocalization

ABSTRACT

Like human beings, many animals produce sounds for communication and social interactions. The vocalizations of mice have the characteristics of songs, consisting of syllables of different types determined by the frequency modulations and structure variations. To characterize the impact of social environments and genotypes on vocalizations, it is important to identify the patterns of syllables based on the shapes of frequency contours. Using existing hypothesis testing methods to determine the shape classes would require testing various null and alternative hypotheses for each curve, and is impractical for vocalization studies where the interest is on a large number of frequency contours. A new penalization-based method is proposed, which provides function estimation and automatic shape identification simultaneously. The method estimates the functional curve through quadratic B-spline approximation, and captures the shape feature by penalizing the positive and negative parts of the first two derivatives of the spline function in a group manner. It is shown that under some regularity conditions, the proposed method can identify the correct shape with probability approaching one, and the resulting nonparametric estimator can achieve the optimal convergence rate. Simulation shows that the proposed method gives more stable curve estimation and more accurate curve classification than the unconstrained B-spline estimator, and it is competitive to the shape-constrained estimator assuming prior knowledge of the curve shape. The proposed method is applied to the motivating vocalization study to examine the effect of Methyl-CpG binding protein 2 gene on the vocalizations of mice during courtship.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Like many animals, rodents produce sounds and more specifically ultrasonic vocalizations to communicate within species. The vocalization of mice is a complex behavior that is used primarily for parenting, territorial purposes and courtship. These vocalizations have the characteristics of songs, consisting of several different syllable types, determined by the shapes of the frequency contours. Under different societal circumstances such as mating and parenting, diverse

* Correspondence to: 761 Rome Hall, 801 22nd St NW, Washington, DC 20052, United States of America.

E-mail address: judywang@gwu.edu (H.J. Wang).

ultrasonic vocalizations with different temporal structure can be induced (Mahrt et al., 2013). For example, Fig. 2 plots the sonogram of a ultrasonic vocalization produced by a male mouse evoked by female urine. The plot suggests that the frequency contours of each syllable exhibit different patterns such as upward, downward, chevron and so on. Understanding the social information carried by vocalization, with respect to the motor commands that produced these vocalizations, is central for understanding how vocalization is used to guide animal behaviors for survival and reproduction. Because early deafness has profound effects on vocal development in humans as well as in mice, understanding how vocalization is produced and perceived will provide a framework toward which neural mechanisms underlying vocal development in humans can be studied. To characterize the social signals embedded in ultrasonic vocalizations, it is helpful to identify the patterns of syllables based on frequency modulations and structure variations. This motivated us to develop a new statistical method for simultaneous estimation and shape identification of smooth contours.

In the statistical literature, researchers have developed shape-constrained nonparametric estimation methods when prior knowledge on the shape of the function such as monotonicity and convexity is available (He and Shi, 1998; Hall and Huang, 2001; Dette et al., 2006; Yatchew and Härdle, 2006; Meyer, 2008; Carroll et al., 2011; Seijo and Sen, 2011; Wang and Ghosh, 2012; Feng et al., 2014; Papp and Alizadeh, 2014; Boente et al., 2020). All these methods require knowing the shape information of the functional curve in advance.

Some researchers proposed hypothesis testing methods to test the validity of specific shape constraints (Ghosal et al., 2000; Abrevaya and Jiang, 2005; Meyer, 2008; Wang and Meyer, 2011; Du et al., 2013; Ahkim et al., 2017). To determine the shape of a functional curve using hypothesis testing, we need to consider a large number of different null and alternative hypotheses and perform estimation based on different shape constraints. The hypothesis testing approach is feasible if we are interested in estimating few functional curves. However, the motivating mice ultrasonic vocalization study (see more details in Section 5) involves in total 6273 frequency contours (syllables), and it would be computationally impractical to determine the shapes of so many contours through hypothesis testing. The scientific interest of the motivating study is to classify the functional curves to pre-defined shape classes, such as upward, downward, chevron etc., and this can be regarded as semi-supervised clustering. Therefore, the existing unsupervised clustering methods for functional data, are not suitable (Jacques and Preda, 2013; Zeng et al., 2019).

To overcome the limitation of hypothesis testing, we propose a novel penalization-based method. Specifically, we approximate the nonparametric functional curve by quadratic B-splines, and penalize the positive and negative parts of the first and second derivatives at the knots, using an adaptive group penalty. Our new approach has the following advantages. First, the proposed method provides function estimation and shape identification simultaneously. Second, using the properties of quadratic B-splines, the method reduces the identification of the function shape over the domain of interest to the identification of signs of the derivatives at interior knots, which is computationally efficient. Third, the proposed method is consistent for shape identification in terms of preserving the signs of the first two derivatives of the function. Fourth, the resulting nonparametric estimator can achieve the optimal convergence rate obtained by Stone (1985) for piece-wise polynomial approximations. Theorem 4.1 of Wang and Shen (2013) showed that shape constraints, specifically convex constraints, do not affect the optimal rate of convergence of the shape-constrained estimator. Their proof is based on the prior information of the true shape. The proof in our paper is more challenging since the true shape is unknown and left to be identified through the proposed procedure. Through simulation studies we demonstrate that the proposed method gives more stable curve estimation and more accurate shape identification than the unconstrained B-spline estimator, and it is competitive to the shape-constrained estimator assuming prior knowledge of the functional shape.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed estimation method and computation. In Section 3, we present the asymptotic properties. In Section 4, we conduct simulation to compare the finite sample performance of the proposed estimator with the oracle estimator with known shape information and the unpenalized estimator without any shape constraints. In Section 5, we apply the proposed method to analyze the motivating mice data to classify ultrasonic vocalizations of male mice and study the effect of Mecp2 gene on the ultrasonic vocalizations. Section 6 concludes the paper with some discussion. All the technical proofs are deferred to Appendix A. The developed R package is available at <http://blogs.gwu.edu/judywang/software/>.

2. Penalized estimation for shape identification

2.1. The proposed method

Suppose that Y_i is the response associated with the univariate predictor X_i , which has a bounded support. We assume the following regression model

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where $f(\cdot)$ is an unknown smooth function, and ε_i are independent random errors with mean zero and variance $\sigma^2 < \infty$. Without loss of generality, we assume that $X_i \in [0, 1]$. In the motivating mice vocalization study (see Section 4), one main interest is to identify whether the continuous frequency contour $f(\cdot)$ has the upward, downward, chevron, reverse-chevron, wave or more complex shapes, which can be characterized by $f'(\cdot)$ and $f''(\cdot)$, the first and second derivatives of $f(\cdot)$. We consider the simple nonparametric regression model (1) with an univariate predictor because of the motivating

study. However, the proposed idea can also be extended to other models such as single index models and partially linear models to accommodate multiple predictors.

Let $B_{j,q}(\cdot)$ be the j th normalized B-spline basis function with order q , and we propose to approximate $f(\cdot)$ by quadratic ($q = 3$) B-spline basis functions. We consider using B-spline basis functions to approximate $f(\cdot)$ due to its computational efficiency and stability. Huang et al. (2004), Wang and Yang (2009) and Ma and Racine (2013) Let $0 = x_{-2}^* = x_{-1}^* = x_0^* < x_1^* < \dots < x_{k_n}^* < x_{k_n+1}^* = x_{k_n+2}^* = x_{k_n+3}^* = 1$ be an extended partition of $[0, 1]$, where k_n is the number of internal knots. In our implementation, we choose $x_j^*, j = 1, \dots, k_n$, as the $j/(k_n + 1)$ th sample quantile of $\{X_i, i = 1, \dots, n\}$. Let $B_3(\cdot) = (B_{1,3}(\cdot), \dots, B_{N,3}(\cdot))^T$ be the normalized quadratic B-splines based on the distinct knots $x_1^{**} = \{x_k^*\}_{k=0}^{k_n+1}$, where $N = k_n + 3$ is the number of basis functions. Let $S(3, x_1^{**})$ be the functional space spanned by $\{B_{j,3}\}_{j=1}^N$, then $f(\cdot)$ can be approximated by a spline function $s_n(x) \in S(3, x_1^{**})$, that is,

$$f(x) \approx s_n(x) = \sum_{j=1}^N B_{j,3}(x) \beta_j \doteq B_3(x)^T \beta, \quad (2)$$

where $\beta = (\beta_1, \dots, \beta_N)^T$ is the unknown coefficient vector. By the definition of the B-spline basis functions (de Boor, 1972; Zhou and Wolfe, 2000; Ma and Song, 2015), $f'(\cdot)$ and $f''(\cdot)$ can be approximated by

$$f'(x) \approx s'_n(x) = \sum_{j=1}^{N-1} 2 \frac{\beta_{j+1} - \beta_j}{x_j^* - x_{j-2}^*} B_{j,2}(x),$$

$$f''(x) \approx s''_n(x) = \sum_{j=1}^{N-2} 2(x_j^* - x_{j-1}^*)^{-1} \left(\frac{\beta_{j+2} - \beta_{j+1}}{x_{j+1}^* - x_{j-1}^*} - \frac{\beta_{j+1} - \beta_j}{x_j^* - x_{j-2}^*} \right) B_{j,1}(x),$$

where $B_1(\cdot) \doteq (B_{1,1}(\cdot), \dots, B_{N-2,1}(\cdot))^T$ and $B_2(\cdot) \doteq (B_{1,2}(\cdot), \dots, B_{N-1,2}(\cdot))^T$ are the basis functions corresponding to piece-wise constant and piece-wise linear B-splines.

To simplify the notations, we introduce two knot-dependent matrices M_1, M_2 to rewrite the derivatives of $f(\cdot)$ as

$$f'(x) \approx s'_n(x) = B_2(x)^T M_1^T \beta, \quad f''(x) \approx s''_n(x) = B_1(x)^T M_2^T M_1^T \beta,$$

where the transformation matrices M_1 and M_2 are defined as

$$M_1 = 2 \begin{pmatrix} \frac{-1}{x_1^* - x_{-1}^*} & 0 & 0 & \dots & 0 & 0 \\ \frac{1}{x_1^* - x_{-1}^*} & \frac{-1}{x_2^* - x_0^*} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{x_2^* - x_0^*} & \frac{-1}{x_3^* - x_1^*} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{x_{N-2}^* - x_{N-4}^*} & \frac{-1}{x_{N-1}^* - x_{N-3}^*} \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{x_{N-1}^* - x_{N-3}^*} \end{pmatrix}_{N \times (N-1)},$$

$$M_2 = \begin{pmatrix} \frac{-1}{x_1^* - x_0^*} & 0 & 0 & \dots & 0 & 0 \\ \frac{1}{x_1^* - x_0^*} & \frac{-1}{x_2^* - x_1^*} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{x_2^* - x_1^*} & \frac{-1}{x_3^* - x_2^*} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{x_{N-3}^* - x_{N-4}^*} & \frac{-1}{x_{N-2}^* - x_{N-3}^*} \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{x_{N-2}^* - x_{N-3}^*} \end{pmatrix}_{(N-1) \times (N-2)}.$$

Since $B_2(x)$ are piece-wise linear functions, thus for $f(\cdot)$ to be nondecreasing in $(0, 1)$, it suffices to require $B_2(\cdot)^T M_1^T \beta \geq 0$ at all distinct knots $\{x_k^*\}_{k=0}^{k_n+1}$. In addition, $B_1(x)$ are piece-wise constant functions and the domain of X is divided into meshes $[x_k^*, x_{k+1}^*]$, so $B_1(\cdot)^T M_2^T M_1^T \beta$ is piecewise constant and right continuous on knots $x_2^{**} = \{x_k^*\}_{k=0}^{k_n}$. Therefore, to achieve convexity, it suffices to require $B_1(\cdot)^T M_2^T M_1^T \beta \geq 0$ at x_2^{**} , that is, the right limits of the second derivative at all but the right boundary knots.

Define $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{k_n+1})^T$ with $\gamma_k = B_2(x_k^*)^T M_1^T \beta$ for $k = 0, \dots, k_n + 1$, and $\delta = (\delta_0, \delta_1, \dots, \delta_{k_n})^T$ with $\delta_k = B_1(x_k^*)^T M_2^T M_1^T \beta$ for $k = 0, \dots, k_n$, representing the right limit of the first two derivatives of $s_n(\cdot)$ at knots. Denote

the component-wise positive and negative parts of γ and δ , respectively, as $\gamma^+ \geq 0$ and $\gamma^- \geq 0$, and $\delta^+ \geq 0$ and $\delta^- \geq 0$, such that $\gamma = \gamma^+ - \gamma^-$ and $\delta = \delta^+ - \delta^-$. We propose to estimate β by

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \left[\{Y_i - B_3(X_i)^T \beta\}^2 + \lambda_{1n}(\omega_1 \|\gamma^+\|_{\infty} + \omega_2 \|\gamma^-\|_{\infty}) + \lambda_{2n}(\omega_3 \|\delta^+\|_{\infty} + \omega_4 \|\delta^-\|_{\infty}) \right], \quad (3)$$

where $\|\cdot\|_{\infty}$ is the L_{∞} norm, λ_{1n} and λ_{2n} are the regularization parameters for the first and second derivatives, and $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ are the adaptive group weights. In our implementation, the adaptive weights are chosen as

$$\omega_1 = \|\gamma^{u+}\|_{\infty}^{-1}, \quad \omega_2 = \|\gamma^{u-}\|_{\infty}^{-1}, \quad \omega_3 = \|\delta^{u+}\|_{\infty}^{-1}, \quad \omega_4 = \|\delta^{u-}\|_{\infty}^{-1}, \quad (4)$$

where $\gamma^{u+}, \gamma^{u-}, \delta^{u+}, \delta^{u-}$ are the initial estimators based on the unconstrained estimator β^u , that is, the minimizer of (3) with $\lambda_{1n} = \lambda_{2n} = 0$. The purpose of the adaptive weights is to assign a larger penalty to the vector that is more likely to be sparse, thus shrinking it toward zero at a faster rate.

We penalize γ^+, γ^- and δ^+, δ^- in a group manner with the L_{∞} penalty for two reasons. Firstly, it is intuitive, for instance, if the largest component in γ^+ is shrunk to zero, this implies that $\gamma^+ = 0$ and the function is non-increasing. Secondly, it is easy to cast the optimization problem into quadratic programming, so that any existing optimization algorithms can be applied. As discussed earlier, the properties of quadratic B-splines imply that $f(\cdot)$ is (i) non-increasing if all elements of γ^+ are shrunk to zero; (ii) concave if all elements in δ^+ are shrunk to zero; (iii) non-increasing and concave if both γ^+ and δ^+ contain only zeros; (v) of a complex shape form if none of $\gamma^+, \gamma^-, \delta^-$ and δ^+ are shrunk to zero.

For a given pair (λ_1, λ_2) , there exists a pair of regularization parameters (s_1, s_2) such that minimizing (3) is equivalent to solving the following constrained minimization problem

$$\begin{aligned} \min \quad & \sum_{i=1}^n \left\{ Y_i - \sum_{j=1}^N B_{j,3}(X_i) \beta_j \right\}^2, \\ \text{s.t.} \quad & \omega_1 \|\gamma^+\|_{\infty} + \omega_2 \|\gamma^-\|_{\infty} \leq s_1, \quad \omega_3 \|\delta^+\|_{\infty} + \omega_4 \|\delta^-\|_{\infty} \leq s_2. \end{aligned} \quad (5)$$

The minimization problem can be solved by using quadratic programming, see next subsection for more details. When s_1 and s_2 are large enough, minimizing (5) gives the unpenalized estimator β^u . On the other hand, letting $s_1 = s_2 = 0$ will lead to an estimator of β that gives a flat curve with derivatives equal to zero. By the definition of weights in (4), the tuning parameters s_1 and s_2 have the range of $[0, 2]$. We discuss the selection of tuning parameters in the next subsection.

2.2. Computational details

In this subsection, we present the computational details, including the explicit quadratic programming formulation and the tuning selection criterion.

Recall that $\beta = (\beta_1, \dots, \beta_N)^T$, and we further denote $Y = (Y_1, \dots, Y_n)^T$, $X = (X_1, \dots, X_n)^T$,

$$B(X) = \begin{pmatrix} B_{1,3}(X_1) & B_{2,3}(X_1) & \cdots & B_{N,3}(X_1) \\ B_{1,3}(X_2) & B_{2,3}(X_2) & \cdots & B_{N,3}(X_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_{1,3}(X_n) & B_{2,3}(X_n) & \cdots & B_{N,3}(X_n) \end{pmatrix} \quad \text{and} \quad \alpha = \begin{pmatrix} \beta \\ \gamma^+ \\ \gamma^- \\ \delta^+ \\ \delta^- \end{pmatrix}.$$

Thus, the squared loss in expression (5) can be rewritten as

$$\{Y - B(X)\beta\}^T \{Y - B(X)\beta\} = (Y - D\alpha)^T (Y - D\alpha),$$

where $D = (B(X), O)$ with O denoting a $n \times (4N - 6)$ zero matrix and $N = k_n + 3$ is the number of basis functions. The constraints in (5) are equivalent to $\forall i, j \in \{0, 1, \dots, k_n + 1\}$, $\omega_1 \gamma_i^+ + \omega_2 \gamma_j^- \leq s_1$ and $\forall i', j' \in \{0, 1, \dots, k_n\}$, $\omega_3 \delta_{i'}^+ + \omega_4 \delta_{j'}^- \leq s_2$. By the definition of α , we can rewrite the constraints as

$$A\alpha \leq \begin{pmatrix} \bar{S}_1 \\ \bar{S}_2 \end{pmatrix},$$

where the first $(k_n + 2)^2$ rows of $A\alpha$ include all pair combinations of (γ_i^+, γ_j^-) , $i, j = 0, 1, \dots, k_n + 1$, and the rest $(k_n + 1)^2$ rows include all pair combinations of $(\delta_{i'}^+, \delta_{j'}^-)$, $i', j' = 0, 1, \dots, k_n$, and $\bar{S}_1 = s_1 \cdot \bar{1}_{(k_n+2)^2}$, $\bar{S}_2 = s_2 \cdot \bar{1}_{(k_n+1)^2}$.

In addition, by the definitions of $\{\gamma^+, \gamma^-, \delta^+, \delta^-\}$, we have the following constraints,

$$D_1 \alpha = \gamma^+ - \gamma^- - B_2(x_1^{**})^T M_1^T \beta = \vec{0},$$

$$D_2\alpha = \delta^+ - \delta^- - B_1(x_2^{**})^T M_2^T M_1^T \beta = \vec{0},$$

$$D_3\alpha \geq \vec{0},$$

where $D_1 = (-B_2(x_1^{**})^T M_1^T, I_1, I_1, O_1)$, $D_2 = (-B_1(x_2^{**})^T M_2^T M_1^T, O_2, I_2, I_2)$, $D_3 = (O_3, I_3)$ and $x_2^{**} = (x_0^*, \dots, x_{k_n}^*)^T$, i.e. x_2^{**} represents the whole distinct knot vector except the right boundary; $B_2(x_1^{**}) = (B_2(x_0^*), \dots, B_2(x_{k_n+1}^*))$ is a $(N-1) \times (k_n+2)$ matrix, with $B_2(x_j^*)$ being its column, and similar for $B_1(x_2^{**})$. The reason we omit the right end point is that $\delta_{k_n} \doteq B_1(x_{k_n}^*)^T M_2^T M_1^T \beta$ (right limit on the knot $x_{k_n}^*$) is enough to decide the second derivative in the middle of last mesh $[x_{k_n}^*, x_{k_n+1}^*]$. Here I_1, I_2 and I_3 are identity matrices of dimensions k_n+2 , k_n+1 and $4k_n+6$, and O_1, O_2 and O_3 are zero matrices of dimensions $(k_n+2) \times (2k_n+2)$, $(k_n+1) \times (2k_n+4)$ and $(k_n+3) \times (k_n+3)$, respectively. With these notations, we can rewrite the optimization problem (5) as

$$\min_{\alpha} (Y - D\alpha)^T (Y - D\alpha),$$

$$\text{s.t. } A\alpha \leq \begin{pmatrix} \vec{S}_1 \\ \vec{S}_2 \end{pmatrix}, \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} \alpha = \vec{0} \text{ and } D_3\alpha \geq \vec{0}. \quad (6)$$

This optimization problem can be solved by any existing quadratic programming algorithm. In our implementation, we use the function “solve.QP” in the R package *quadprog*.

The tuning parameters (s_1, s_2) are chosen by the covariance information criterion in Tibshirani and Knight (1999); see also Shen and Huang (2006) and Sklar et al. (2013). The covariance information criterion is defined as

$$\text{CIC}(s_1, s_2) = \frac{1}{n} \text{RSS}(s_1, s_2) + \frac{2}{n} C(s_1, s_2),$$

where $\text{RSS}(s_1, s_2) = \sum_{i=1}^n \{Y_i - \hat{f}_{(s_1, s_2)}(X_i)\}^2$ is the sum of squared residuals based on the penalized estimator $\hat{f}_{(s_1, s_2)}$ associated with tuning parameters (s_1, s_2) , and $C(s_1, s_2)$ is the covariance penalty defined as

$$C(s_1, s_2) = \sum_{i=1}^n \text{Cov}\{\hat{f}_{(s_1, s_2)}(X_i), Y_i\}.$$

The covariance information criterion is an unbiased estimate of $\text{MSE}\{\hat{f}_{(s_1, s_2)}\} + \sigma^2$. The covariance penalty $C(s_1, s_2)$ is a measurement of model complexity, and $C(s_1, s_2)/\sigma^2$ is equivalent to the generalized degree of freedom defined in Ye (1998) for general modeling procedures including nonparametric regression.

In our implementation, we estimate the covariance penalty by a bootstrap method (Tibshirani and Knight, 1999; Efron, 2004). Let $\{Y_i^*, i = 1, \dots, n\}$ be the b th bootstrap sample of $\{Y_i, i = 1, \dots, n\}$, where $b = 1, \dots, n_b$. Let $\{\hat{f}_b^*(X_i), i = 1, \dots, n\}$ be the penalized estimation based on the bootstrap data $\{(Y_{i,b}^*, X_i)\}, i = 1, \dots, n$ and the tuning parameters (s_1, s_2) . Then the bootstrap estimator of the covariance penalty for the i th observation is calculated as

$$\widehat{\text{Cov}}_i(s_1, s_2) = \frac{1}{n_b - 1} \sum_{b=1}^{n_b} \hat{f}_b^*(X_i)(Y_{i,b}^* - \bar{Y}_i^*),$$

where $\bar{Y}_i^* = \sum_{b=1}^{n_b} Y_{i,b}^*/n_b$. Then the covariance penalty function $C(s_1, s_2)$ can be estimated by $\widehat{C}(s_1, s_2) = \sum_{i=1}^n \widehat{\text{Cov}}_i(s_1, s_2)$. Finally, we choose the tuning parameters (s_1, s_2) by minimizing

$$\widehat{\text{CIC}}(s_1, s_2) = \frac{1}{n} \text{RSS}(s_1, s_2) + \frac{2}{n} \widehat{C}(s_1, s_2),$$

over a grid of (s_1, s_2) in $[0, 2] \times [0, 2]$.

3. Asymptotic properties

We establish the asymptotic properties of the proposed penalized estimator in terms of selection consistency and convergence rate. For notation, we denote $a_n \lesssim b_n$ if $a_n/b_n \leq c$ for some positive constant c , and $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. We assume the following conditions. Condition C1 requires that the variances of the errors are bounded, and C2 poses some condition on the distribution of X_i . Condition C3 is about the smoothness of the underlying function, which is standard in nonparametric regression when the second derivative is of interest.

- C1. The random errors ε_i are independent with mean zero and the maximum variance satisfying $\max_{1 \leq i \leq n} \text{Var}(\varepsilon_i) \leq c_0 < \infty$ for some positive constant c_0 .
- C2. The distribution of X_i is absolutely continuous with density function g in $[0, 1]$. Furthermore, there exist two constants $c_1 > 0$ and $c_2 < \infty$ such that $c_1 \leq g(x) \leq c_2$ for any $x \in [0, 1]$.
- C3. The second order derivative of $f(\cdot)$ satisfies Hölder condition of order ν . That is, there exists a constant $c_3 > 0$ such that $|f''(s) - f''(t)| \leq c_3 |s - t|^\nu$, for any $0 \leq s, t \leq 1$.

Let $r = \min\{3, 2 + \nu\}$ and $k_n \sim n^{\frac{1}{2r+1}}$. Denote $\hat{\beta}$ as the proposed B-spline coefficient estimator, $\hat{f}(\cdot) = B_3(\cdot)^T \hat{\beta}$ as the estimated curve, and $\hat{\gamma} = B_2(x_1^{**})^T M_1^T \hat{\beta}$ and $\hat{\delta} = B_1(x_2^{**})^T M_2^T M_1^T \hat{\beta}$ as the estimated first two derivatives of the curve at the knots. We establish the consistency of the sign identification of the first two derivatives and the convergence rate of the proposed estimator in [Theorems 1–2](#). [Theorem 1](#) is for the case where the signs of $f'(\cdot)$ and $f''(\cdot)$ are common across the domain of X , and [Theorem 2](#) is for the case where the signs change across the domain.

Theorem 1. Assume that C1–C3 hold with $\nu > 0.5$, and $c_4 \lesssim \lambda_{1n}/\lambda_{2n} \lesssim k_n$ with some positive constants c_4 . Suppose that $f(x)$ is increasing and convex across $x \in [0, 1]$, then

- (I) if $(n^{\frac{1}{2r+1}} \log n)^{-1} \lambda_{2n} \rightarrow \infty$ and $n^{-1} \lambda_{2n} \rightarrow 0$, we have $P\{\|\hat{\gamma}^-\|_\infty = 0, \|\hat{\delta}^-\|_\infty = 0\} \rightarrow 1$;
 (II) if $n^{-\frac{r-3/2}{2r+1}} \lambda_{2n} \rightarrow 0$, we have $n^{-1} \sum_{i=1}^n \{\hat{f}(X_i) - f(X_i)\}^2 = O_p(n^{-\frac{2r}{2r+1}})$.

[Theorem 1\(I\)](#) shows that with probability approaching one, the proposed estimator can correctly identify the signs of the first two derivatives as long as the signs of $f'(\cdot)$ and $f''(\cdot)$ are common across the domain of X . [Theorem 1\(I\)](#) solely presents the correct identification for $f(\cdot)$ being increasing and convex for illustration, but the consistency also hold for other scenarios as long as the signs of the derivatives do not change across the domain. [Theorem 1\(II\)](#) shows that the proposed estimator can achieve the optimal convergence rate of the quadratic B-spline approximation.

Theorem 2. Assume that C1–C3 hold with $\nu > 0.5$, and $c_4 \lesssim \lambda_{1n}/\lambda_{2n} \lesssim k_n$ with some positive constants c_4 .

- (I) If $n^{-\frac{2r-3}{2r+1}} \lambda_{2n} \rightarrow 0$, we have
 (i) if $\exists x_1, x_2 \in (0, 1)$, s.t. $f'(x_1) \cdot f'(x_2) < 0$, then $P\{\|\hat{\gamma}^+\|_\infty = 0 \text{ or } \|\hat{\gamma}^-\|_\infty = 0\} \rightarrow 0$;
 (ii) if $\exists x_1, x_2 \in (0, 1)$, s.t. $f''(x_1) \cdot f''(x_2) < 0$, then $P\{\|\hat{\delta}^+\|_\infty = 0 \text{ or } \|\hat{\delta}^-\|_\infty = 0\} \rightarrow 0$.
 (II) If $n^{-\frac{r-3/2}{2r+1}} \lambda_{2n} \rightarrow 0$, we have $n^{-1} \sum_{i=1}^n \{\hat{f}(X_i) - f(X_i)\}^2 = O_p(n^{-\frac{2r}{2r+1}})$.

[Theorem 2\(I\)](#) suggests that if the derivatives of the function change signs across the domain, that is, the function is not monotonic or convex/concave across the domain, then with probability approaching one the proposed estimator will not identify the signs to be common. [Theorem 2\(II\)](#) shows that with a slower rate of λ_{2n} , the proposed estimator can achieve the optimal convergence rate of the quadratic B-spline approximation. This together with [Theorem 1\(II\)](#) suggests that the optimal convergence rate can be achieved regardless of the shape of the underlying curve. When λ_{2n} satisfies the conditions $(n^{\frac{1}{2r+1}} \log n)^{-1} \lambda_{2n} \rightarrow \infty$ and $n^{-\frac{r-3/2}{2r+1}} \lambda_{2n} \rightarrow 0$, results in both [Theorems 1–2](#) will hold.

4. Simulation study

We carry out a simulation study to assess the finite sample performance of the proposed penalized estimator. The simulation data is generated from model [\(1\)](#) with $X_i \sim U(0, 1)$ and $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$. We consider four different cases. In Case 1, we set $f(x) = 1 + 10(x - 0.3)^2$ being a convex function with $\sigma = 1$. In Case 2, we set $f(x) = 1 + 5 \exp(0.7x^2)$ being an increasing and convex function with $\sigma = 1$. In Case 3, we set $f(x) = 1 + 10 \sin(\pi x^2)$ being increasing and convex for $x \in (0, 1/\sqrt{2})$ and decreasing and concave for $x \in (1/\sqrt{2}, 1)$ with $\sigma = 4$. In Case 4, we set $f(x) = 1 + 5 \sin(2\pi x^{2.5}) + 10x$ being a complex form: it changes from increasing and convex to decreasing and concave, and then to increasing and convex, with $\sigma = 4$. We consider two sample sizes $n = 50$ and 200 , and the number of replicates is 300 for each scenario.

We consider four methods for comparison, (I) the unpenalized B-spline estimator, that is, the conventional B-spline estimator without any shape constraints or penalization, (II) the oracle estimator, that is, the shape-constrained B-spline estimator assuming correct shape knowledge of the underlying function, (III) the proposed penalized estimator, that is, the solution from optimization problem [\(5\)](#), and (IV) the cubic smoothing spline estimator ([Green and Silverman, 1994](#)). The first three methods are all based on quadratic B-splines. To assess the sensitivity of three B-spline based estimators against the choice of knots, we vary the number of internal knots as $k_n = \lfloor cn^{1/10} \rfloor$ with $c \in [3, 6]$, resulting in $k_n \in \{4, \dots, 8\}$ for $n = 50$ and $k_n \in \{5, \dots, 10\}$ for $n = 200$, and the internal knots are assigned to be equally spaced in $(0, 1)$. For the unpenalized estimator, we also include a variation with k_n selected by the Bayesian information criterion ([Huang and Yang, 2004](#)) over $k_n \geq 0$, and the chosen k_n is denoted as k_n^* . Estimation based on the smoothing spline is implemented by the function `smooth.spline` in the R package `splines` with the smoothing parameter selected by cross validation.

We first assess the estimation accuracy of different estimators by comparing the mean integrated squared error, $E\|\hat{f}(X) - f(X)\|_2^2$. [Fig. 1](#) summarizes the performance of different estimators against the number of internal knots k_n for $n = 50$; results for $n = 200$ can be found in the on line Supporting Information (Figure S.1). Results suggest that the unpenalized estimator is more sensitive to k_n than the penalized estimator and oracle estimator, and even with the data-driven k_n^* chosen by Bayesian information criterion, it still in general gives higher mean integrated squared errors. The penalized estimator is quite robust to k_n , chosen based on this rule of thumb across $c \in [3, 6]$ in terms of the estimation efficiency, and the performance is similar with the oracle estimator in most of the cases, and the former is even slightly better in Cases 1–2. One explanation of this seemingly surprising result is that the proposed penalization tends to shrink the derivatives at internal knots, and this leads to more smoothed curves, which have smaller finite-sample variance than the oracle estimator. The proposed penalized estimator is also consistently more accurate than the smoothing spline estimator.

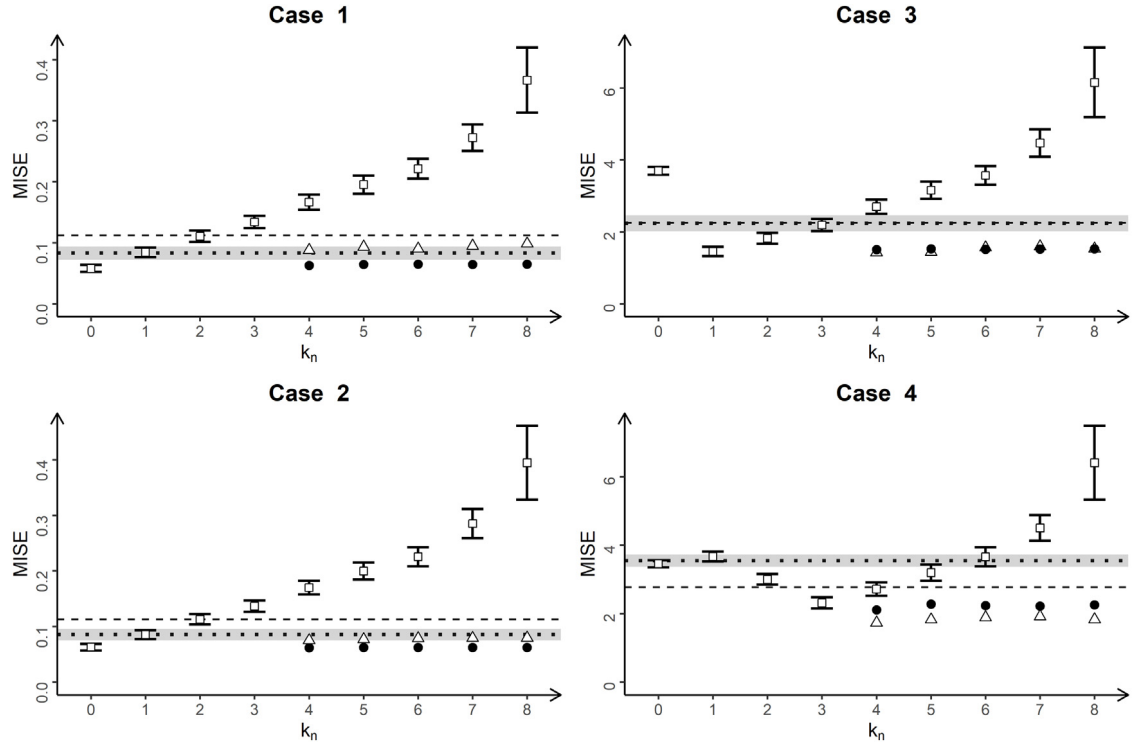


Fig. 1. Mean integrated squared errors for $n = 50$. Penalized estimator: solid circle; oracle estimator: open triangle; unpenalized estimator: open square; smoothing spline estimator: horizontal dashed line; unpenalized estimator with k_n^* : horizontal dotted line. The horizontal bars around the squares and the shaded area around the horizontal dotted line represent the 95% confidence intervals/belt of the unpenalized estimator with k_n varying and fixed at k_n^* , respectively.

We next compare the performance of different methods for shape identification, in terms of the correct identification of derivative signs and the shape-class classification. More specifically, for the identification of derivative signs, we calculate the average percentages that the signs of the first/second derivatives of $f(x)$ are correctly identified over a 50-point grid on $[0.02, 0.98]$. The shape-class classification is motivated by the ultrasonic vocalization study, where we aim to classify functional curves to some pre-defined shapes, such as upward, downward, chevron, reverse chevron, wave and complex. We report the percentages of replicates in which the shape class is correctly identified, where the true classes of the curves in Cases 1–4 are reverse chevron, upward, chevron and wave, respectively. To account for the identifiability issue of signs around zero, we exclude the regions of x for which $f^{(j)}(x)$ are close to zero, using

$$\mathbb{P} \left[\text{sign} \{ \hat{f}^{(j)}(x) \} = \text{sign} \{ f^{(j)}(x) \} \mid |f^{(j)}(x)| > c_j \right], \quad j = 1, 2,$$

where $c_j = \text{sd} \{ f^{(j)}(u_i) \} / \max_i \{ |f^{(j)}(u_i)| \}$ and $\{u_i\}_{i=1}^{50}$ is the grid in $[0.02, 0.98]$.

Table 1 summarizes the shape identification results for the unpenalized estimator with data-adaptive k_n^* and the proposed penalized estimators with $k_n = \lfloor 4n^{1/10} \rfloor$; results for $k_n = \lfloor cn^{1/10} \rfloor$ with $c \in [3, 6]$ can be found in the on line Supporting Information (Table S.1), which shows that the proposed method is quite robust to k_n , chosen based on this rule of thumb across $c \in [3, 6]$. Overall, the penalized and unpenalized estimators perform similarly in terms of derivative sign identification, and both are better than the smoothing spline estimator in sign identification of the second derivative. The penalized estimator performs much better in the shape-class classification, especially in Cases 2–4. Our investigation shows that the unpenalized estimator is sensitive to the choice of k_n^* , and the estimates are often unstable on the boundary of x if k_n^* is not chosen well; the distribution of k_n^* in the on line Supporting Information (Table S.2) shows that the selection of a suitable k_n^* is challenging especially when the convexity of the function changes across x . To help understand why the unpenalized estimator under-performs in terms of shape identification, we show some typical data sets from each case in Figure S.2 of the Supporting Information.

5. Analysis of the mice ultrasonic vocalization data

Rett Syndrome is a neurodevelopment disorder, for which the symptoms are characterized by pseudo-stationary stage of seizures, stereotypical hand wringing, autistic-like symptoms and debilitating respiratory problems leading to sudden

Table 1
Percentages of correct identification of derivative signs and shape classes.

Case	Method	$n = 50$			$n = 200$		
		PS ₁	PS ₂	PC	PS ₁	PS ₂	PC
1	UPE	92.5	93.4	79.0	97.5	99.0	98.5
	PE	94.2	94.9	87.0	98.7	97.2	99.5
	SME	91.0	78.9	61.5	93.7	81.7	64.4
2	UPE	88.8	92.7	32.5	89.4	93.7	18.5
	PE	98.1	91.2	74.0	98.5	93.5	91.5
	SME	87.9	82.7	64.9	94.5	85.6	80.8
3	UPE	87.7	81.0	54.5	91.3	93.8	48.0
	PE	93.7	79.4	86.5	96.4	85.8	88.5
	SME	87.9	73.6	56.7	92.2	79.2	58.2
4	UPE	82.4	54.3	23.5	89.3	75.8	91.5
	PE	85.1	68.8	65.0	91.9	71.3	97.0
	SME	83.0	69.9	21.2	91.4	73.7	57.7

UPE: unpenalized estimator; PE: penalized estimator; SME: smoothing spline estimator; PS_j: average percentages that the signs of the j th derivatives of $f(x)$ are correctly identified over x , $j = 1, 2$; PC: percentages of replicates in which the shape class is correctly identified.

unexplained death (Hagberg et al., 1983). Methyl-CpG binding protein 2 is a chromosome binding protein and regulator for brain development. Mutations of methyl-CpG-binding protein 2 have been shown to cause Rett Syndrome in terms of language and motor deficits (Amir et al., 1999). To understand the disruptions of mutated methyl-CpG-binding protein 2 in brain development and function, mice models of Rett Syndrome have been created with mutations (Shahbazian et al., 2002; Collins et al., 2004; Santos et al., 2007; Goffin et al., 2012). Like human beings, mice could generate different ultrasonic vocalizations to communicate under different social scenarios such as isolation, aggression and courtship (Scattoni et al., 2009). In this analysis, we study the effect of methyl-CpG-binding protein 2 on the ultrasonic vocalizations of mice under the courtship paradigm.

The experiment was conducted in the Department of Psychology at George Washington University. The data includes seven methyl-CpG-binding protein 2-null male mice with the gene knocked out, and their eight wild type siblings. On the testing day, the ultrasonic vocalizations of each subject was recorded for 300 s with the presence of female mice urine in the cage. The ultrasonic vocalizations have the characteristics of songs, consisting of syllables of different lengths and types. Signals with more than 30 ms gaps between spectrally pure time points were considered as separate syllables. This resulted in total 723 syllables vocalized by knocked out mice and 5550 syllables by wild type mice.

For illustration, Fig. 2 shows an example sonogram of ultrasonic vocalizations generated by one knocked out mouse in the first 0.8 s. The y-axis is the transformed frequency contour in Cent scale, calculated as $12 \log_2(F/25 \text{ kHz})$, where F is the frequency in unit Hz. During the first 0.8 s, the mouse produced 12 syllables of different shapes.

As suggested in Fig. 2 and previous studies, Scattoni et al. (2008) and Grimsley et al. (2011) the frequency contours of typical ultrasonic vocalizations can be classified into 10 classes: “harmonic”, “short”, “jump”, “flat”, “upward”, “downward”, “chevron”, “reverse chevron”, “wave” and “complex”. We first identify the four classes “harmonic”, “short”, “jump”, “flat” based on some commonly used hard rules. Specifically, we identify a frequency contour with spectral energy with 0.5 octaves from the fundamental frequency as “harmonic”; with duration less than 10 ms as “short”; with more than 2 cents difference in consecutive frequencies as “jump” and with the ratio of range over duration less than 0.03 as “flat”. Next we apply our proposed penalization method to classify the remaining smooth contours into the six classes “upward”, “downward”, “chevron”, “reverse chevron”, “wave” and “complex”. For each syllable, we choose $k_n = \lfloor 4n^{1/10} \rfloor$ number of equally spaced internal knots, where n is the number of measurements (duration) of the frequency contour. The shape of the syllable is identified based on the signs of the first derivative of the penalized curve estimator at the internal knots. A syllable is identified as “upward”/“downward” if the first derivative of the estimated curve is positive/negative at all internal knots, as “chevron”/“reverse chevron” if the first derivative changes once from positive/negative to negative/positive, as “wave” if the sign of the first derivative changes twice, as “complex” if the sign changes three times or more.

For comparison, we also consider using the bootstrap test procedure in Du et al. (2013), which is based on kernel smoothing, to identify shapes through testing the validity of various shape constraints. We use four example syllables to demonstrate the differences of the penalization-based and testing-based methods. Fig. 3 shows the frequency contours, the penalized curve estimation and the unconstrained Nadaraya–Watson estimation (Du et al., 2013). Table 2 summarizes the hypothesis testing results (based on 200 bootstrap), and the computing times of the testing and penalization methods. The penalization method classifies the four syllables to the classes “downward”, “chevron”, “reverse chevron” and “wave”, respectively. For Syllable 1, the testing method fails to reject the two shape constraints: “downward” (decreasing across t) and “reverse chevron” (decreasing for $t \in [1, 64]$ and increasing for $t \in [65, 80]$). If we choose the shape as the one giving the largest p -value, the testing method will classify the four syllables to “reverse chevron”, “chevron”, “reverse chevron”, and “reverse chevron”, which appear to be misleading for Syllables 1 and 4. On the other hand, choosing the

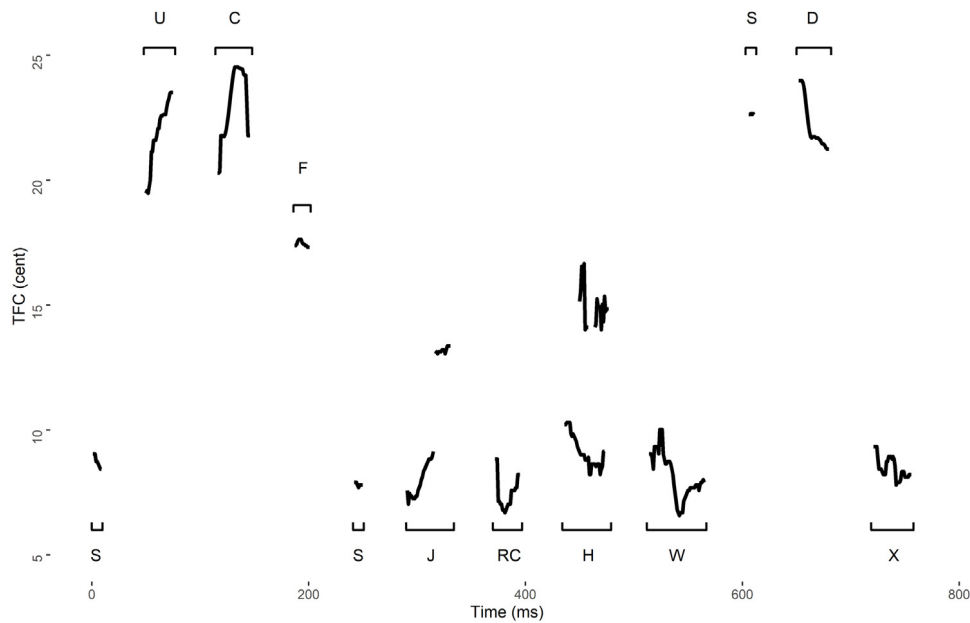


Fig. 2. Sonogram of an example ultrasonic vocalization of a male mice evoked by female urine during the first 0.8 s. The curves are the transformed frequency contours of different syllables, and the letters above or below the syllables are the identified shape classes.

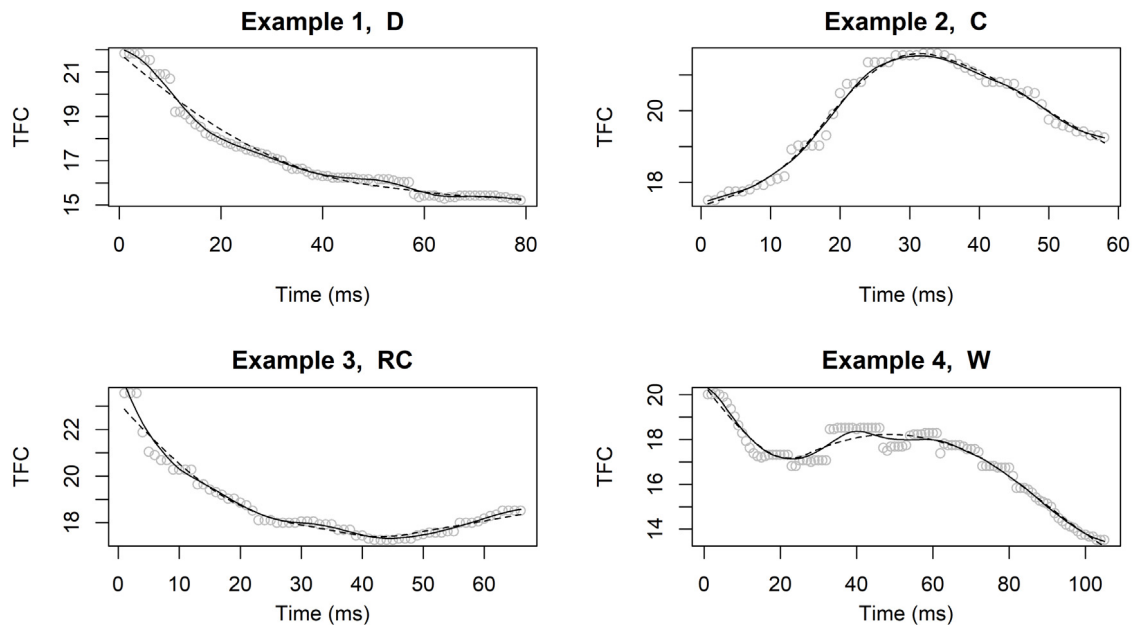


Fig. 3. The proposed penalized estimation (dashed) and the unconstrained Nadaraya-Watson estimation (solid) for four example syllables from the ultrasonic vocalization study. D: down; C: chevron; RC: reverse chevron; W: wave.

one giving the smallest test statistic will often classify syllables to the “complex” class, since the test statistic measures the distance of probability weights associated with the shape-constrained and unconstrained estimations, and tends to be smaller for more complex shape constraints. In addition, the testing method requires more than 10 times of computing time than the penalization method. All these suggest that the testing procedure is not suitable for identifying shapes for the ultrasonic vocalization study that involves thousands of contours, since the testing procedure may be unreliable and it also requires user’s manual intervention by specifying the exact shape constraints and regions for each contour.

Table 3 summarizes the percentages of syllables classified by our proposed penalization method for the knocked out and wild type mice separately. The variation of vocalization between methyl-CpG-binding protein 2-null and wild type

Table 2

The hypothesis testing results, and the computing times of the method in Du et al. (2013) and the proposed penalization method from analyzing four example syllables. The H_0 denotes the type of shape constraint to be tested.

Syllable	Du et al. (2013)				Proposed
	H_0	Statistic	P -value	Time (s)	Time (s)
1	D*	0.0003	0.180	414	38
	U	0.6206	0.010		
	RC	0.0001	0.955		
2	D	0.3257	0.155	333	28
	U	0.0446	0.010		
	C*	0.0000	0.849		
3	D	0.0240	0.040	438	30
	U	0.2826	0.050		
	RC*	0.0001	0.595		
	X	0.0007	0.180		
4	D	0.0737	0.000	671	38
	U	0.8165	0.000		
	C	0.0991	0.200		
	RC	0.6734	0.740		
	W*	0.0030	0.190		
	X	0.0030	0.430		

D: down; U: upward; C: chevron; RC: reverse chevron; W: wave; X: complex. Bold: classified by Du et al. (2013), chosen as the shape associated with the largest p -value; with superscript *: classified by the proposed penalization method.

Table 3

Percentages of syllables classified into each shape class for the knocked out and wild type groups separately.

Type	H	S	J	F	U	D	C	RC	W	X	#
KO	10	29	31	4	4	2	8	3	6	3	723
WT	4	19	30	6	8	5	14	3	7	3	5550

KO: knocked out; WT: wild type; H: harmonic; S: short; J: jump; F: flat; U: up; D: down; C: chevron; RC: reverse chevron; W: wave; X: complex. #: Total amount of syllables. The last column is the total number of syllables.

mice are visible. First, methyl-CpG-binding protein 2-null mice produced significantly much less vocal signs than the wild type mice. Second, In terms of the distribution of syllable types, methyl-CpG-binding protein 2-null mice produced more short syllables, but less “upward”, “downward”, and “chevron” types of syllables than wild type mice.

To further assess the impact of methyl-CpG-binding protein 2 on the vocalization of mice, we summarize the average of mean and range of frequencies, and the average duration of syllables within each class (excluding the harmonic class) for the knocked out and wild type mice, separately, in Table 4. Results suggest that methyl-CpG-binding protein 2-null mice have much smaller average frequency, narrower frequency range and shorter duration of syllables than wild type mice, especially for the smooth contours in the “upward”, “downward”, “chevron”, “reverse chevron”, “wave” and “complex” classes, which are informative due to their modulation in frequency. The reduced vocalization of methyl-CpG-binding protein 2-null mice can be seen as an implication of isolation and defect during communication. As observed in the earlier study (Chahrouh and Zoghbi, 2007; Guy et al., 2011; Marschik et al., 2012), the brain regions responsible for movement and cognitive functions could be compromised in animals and patients with methyl-CpG-binding protein 2 mutations, which in turn may undermine the production of vocal signals of the methyl-CpG-binding protein 2-null mice. The findings from our study will shed light on understanding the influence of methyl-CpG-binding protein 2 on language development, and pinning down the target of methyl-CpG-binding protein 2-affected areas for communication.

6. Conclusion and discussion

In some practical settings, information about the relationship between the responses and predictors may be available. Some popular examples include the study of utility functions, cost functions, and profit functions in economics (Gallant and Golub, 1984; Terrell, 1996), the study of dose response curve in the phase I clinical trials, growth curves of animals and plants in ecology, and the estimation of the hazard rate and the failure rate in reliability and survival analysis (Molitor and Sun, 2002; Chang et al., 2007). Even though this paper is motivated by mice vocalization, the proposed method can also be applied to the aforementioned areas such as economics, social sciences, biology, clinical trials to confirm the shape assumption, or to help identify the shape and improve the estimation efficiency when the exact shape information is not known in advance.

In this paper, we adopt the quadratic B-spline basis functions, so that $f'(\cdot)$ is approximated by a piece-wise linear function, and $f''(\cdot)$ is approximated by a piece-wise constant function, which makes the shape identification easier. One may consider adopting higher order B-splines, such as cubic B-splines, to obtain smother curve estimation, but the

Table 4

The average mean and range of transformed frequency contour and the average duration of syllables within each identified class. Values in the parentheses are standard errors.

Type	S	J	F	U	D	C	RC	W	X
Average mean of transformed frequency contour									
KO	16.1 (0.6)	16.3 (0.3)	14.3 (0.4)	19.9 (0.7)	18.5 (0.8)	16.7 (0.7)	12.8 (0.7)	15.3 (1.2)	12.6 (1.3)
WT	18.1 (0.1)	18.2 (0.1)	18.2 (0.2)	19.8 (0.1)	18.5 (0.1)	19.1 (0.1)	17.9 (0.4)	17.9 (0.3)	17.7 (0.3)
Average range of transformed frequency contour									
KO	1.9 (0.3)	6.8 (0.2)	0.4 (0.1)	1.8 (0.1)	2.2 (0.2)	2.3 (0.2)	1.6 (0.2)	2.5 (0.3)	2.3 (0.1)
WT	1.0 (0.1)	6.2 (0.1)	0.5 (0.1)	2.1 (0.1)	2.9 (0.1)	3.0 (0.1)	2.1 (0.1)	2.9 (0.1)	4.8 (0.2)
Average duration of syllabus									
KO	7.4 (0.1)	27 (1.6)	20 (3.4)	17 (1.1)	22 (5.1)	22 (1.8)	19 (2.0)	31 (4.7)	32 (6.3)
WT	7.3 (0.1)	49 (0.9)	21 (0.7)	19 (0.3)	32 (1.2)	35 (1.0)	29 (2.0)	45 (2.1)	90 (4.2)

KO: knocked out; WT: wild type; S: short; J: jump; F: flat; U: up; D: down; C: chevron; RC: reverse chevron; W: wave; X: complex.

computation for shape penalization will be much more complicated since it would require penalization at all points instead of only at the knots required by quadratic B-splines.

Acknowledgments

The authors would like to thank two reviewers, an associate editor, and the editor for constructive comments and helpful suggestions. The research was partly supported by National Science Foundation, United States of America grant DMS-1712760, the OSR-2015-CRG4-2582 grant from KAUST, Saudi Arabia, the National Natural Science Foundation of China grant 11871376, Shanghai Pujiang Program, China 18PJ1409800, the Whitehall Foundation, United States of America, Brain and Behavior Research Foundation, United States of America, and the IR/D program from the National Science Foundation, United States of America. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix A. Proof of Theorems 1–2

A.1. Notations

Recall that knots $x_1^{**} = (x_2^{**T}, x_{k_n+1}^*)^T$, and B-spline derivatives (right limits) at knots

$$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{k_n+1})^T \doteq B_2(x_1^{**})^T M_1^T \beta, \quad \delta = (\delta_0, \delta_1, \dots, \delta_{k_n})^T \doteq B_1(x_2^{**})^T M_2^T M_1^T \beta. \quad (\text{A.1})$$

For notational ease, we let $B \doteq B_3(X)$, $B' \doteq B'(x_1^{**}) \doteq B_2(x_1^{**})^T M_1^T$, $B'' \doteq B''(x_2^{**}) \doteq B_1(x_2^{**})^T M_2^T M_1^T$. Decompose β , $B'(x_1^{**})$, $B''(x_2^{**})$ as: $\beta \doteq (\beta_1, \tilde{\beta}^T)^T$, $B'(x_1^{**}) \doteq (B'_{(1)}, \tilde{B}')$, $B''(x_2^{**}) \doteq (B''_{(1)}, \tilde{B}'')$, with β_1 , $B'_{(1)}$, $B''_{(1)}$ being the first element or column. Then,

$$\begin{pmatrix} \gamma \\ \delta \end{pmatrix} \doteq \begin{pmatrix} B'_{(1)}\beta_1 + \tilde{B}'\tilde{\beta} \\ B''_{(1)}\beta_1 + \tilde{B}''\tilde{\beta} \end{pmatrix}. \quad (\text{A.2})$$

For notational ease, we define

$$Q(\beta) = (Y - B\beta)^T(Y - B\beta) + \lambda_{1n}(\omega_1\|\gamma^+\|_\infty + \omega_2\|\gamma^-\|_\infty) + \lambda_{2n}(\omega_3\|\delta^+\|_\infty + \omega_4\|\delta^-\|_\infty). \quad (\text{A.3})$$

We use $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ to denote the minimum and maximum eigenvalues of a matrix.

A.2. Some useful lemmas

We first introduce some useful lemmas. [Lemma 1](#) states some inequalities, and the proof is routine. [Lemma 2](#) gives a property of B-spline expansion, which is the same as Lemma A.1 in [Tang et al. \(2012\)](#). [Lemma 3](#) presents the convergence rate of an infeasible intermediate estimator, which is derived by combining [Lemma 2](#) and Lemma A.4 in [Tang et al. \(2012\)](#). [Lemma 4](#) establishes the asymptotic order of the adaptive weights, which is a special case of the Proposition A.1 in [Wang and Yang \(2009\)](#). [Lemma 5](#) studies the difference between the proposed estimator and the infeasible intermediate estimator. [Lemma 6](#) gives the detailed structure of the transformation matrices. The proof of [Lemmas 5–6](#) can be found at the end of [Appendix A](#).

Lemma 1. Let a and b be two arbitrary vectors of the same length, then

$$| \|(a+b)^+\|_\infty - \|a^+\|_\infty | \leq \|b\|_\infty \quad \text{and} \quad | \|(a+b)^-\|_\infty - \|a^-\|_\infty | \leq \|b\|_\infty.$$

Lemma 2. Suppose that C2 and C3 hold. The eigenvalues of $n^{-1}k_n B(X)^T B(X)$ are uniformly bounded away from zero and infinity in probability.

Lemma 3. Suppose that C1–C3 hold with $\nu > 0.5$. Let $\beta^c = \{B(X)^T B(X)\}^{-1} B(X)^T E(Y|X)$. For $k_n \sim n^{1/(2r+1)}$, we have $\|B(X)\beta^c - f(X)\|_2 = O_p[n^{1/(2(2r+1))}]$.

Lemma 4. Suppose that C1–C3 hold with $\nu > 0.5$. Let $r = \min\{3, 2 + \nu\}$. For $k_n \sim n^{\frac{1}{2r+1}}$,

$$\sup_{x \in [0,1]} |B'(x)\beta^u - f'(x)| = O_p(n^{-\frac{r-1}{2r+1}} \log n), \quad \sup_{x \in [0,1]} |B''(x)\beta^u - f''(x)| = O_p(n^{-\frac{r-2}{2r+1}} \log n),$$

where β^u is the unpenalized estimator.

Lemma 5. For $Y = (Y_1, \dots, Y_n)^T$ and $X = (X_1, \dots, X_n)^T$, define $\beta^c = \{B(X)^T B(X)\}^{-1} B(X)^T E(Y|X)$. Suppose that C1–C3 hold with $\nu > 0.5$. For $k_n \sim n^{1/(2r+1)}$ and $c_4 \lesssim \lambda_{1n}/\lambda_{2n} \lesssim k_n$, we have $\|\hat{\beta} - \beta^c\|_2 = O_p[\max\{n^{-r/(2r+1)}, \lambda_{2n} n^{-(2r-2)/(2r+1)}\}]$.

Lemma 6. The structure and components of matrices $\tilde{B}'$ and $\tilde{B}''$ defined in (A.2) are fully determined once x_1^{**} is given. The rank of $\tilde{B}'$ is full and $\delta = \tilde{B}''(\tilde{B}')^{-1}\gamma$. Specifically, for equally spaced knots, let $\kappa = k_n + 1$, we have

$$\tilde{B}''(\tilde{B}')^{-1} = \begin{pmatrix} -\kappa & \kappa & 0 & \cdots & 0 & 0 \\ 0 & -\kappa & \kappa & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \kappa & 0 \\ 0 & 0 & 0 & \cdots & -\kappa & \kappa \end{pmatrix}_{(k_n+1) \times (k_n+2)}.$$

A.3. Proof of Theorem 1

We only prove Theorem 1 for $0.5 < \nu \leq 1$, so that $2.5 + \eta \leq r \leq 3$ for some positive constant η , and the proof for $\nu > 1$ is similar.

(I) Proof of part (I) of Theorem 1. We use the method of contradiction to prove the consistency in shape identification. For $f(\cdot)$ being strictly increasing and strictly convex, the false shape identification implies that $\lim_{n \rightarrow \infty} P\{\|(B'\hat{\beta})^-\|_\infty > 0\} = \eta_1$ for some constant $0 < \eta_1 < 1$, or $\lim_{n \rightarrow \infty} P\{\|(B''\hat{\beta})^-\|_\infty > 0\} = \eta_2$ for some constant $0 < \eta_2 < 1$, or both cases. Each situation is discussed as follows.

(I.1) Assume $\lim_{n \rightarrow \infty} P\{\|(B'(x_1^{**})\hat{\beta})^-\|_\infty > 0\} = \eta_1$ for some constant $0 < \eta_1 < 1$. We start with the case that the estimated first derivative is negative only at the j th knot, i.e.,

$$\hat{\gamma}_j = (\tilde{B}'\hat{\beta})_j < 0 \quad \text{and} \quad \hat{\gamma}_i \geq 0, \quad \forall i \neq j, \quad (\text{A.4})$$

where $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_j)^T$. Define $\gamma^* = (\gamma_0^*, \gamma_1^*, \dots, \gamma_{k_n+1}^*)^T$, where $\gamma_j^* = 0$ and $\gamma_i^* = \hat{\gamma}_i, \forall i \neq j$. From Lemma 6, we have that $\tilde{B}'$ is of full rank. Through equations (A.2), we can construct a vector β^* corresponding to γ^* as

$$\beta^* = \begin{pmatrix} \beta_1^* \\ \beta_j^* \end{pmatrix} \doteq \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} \tilde{0}^T \\ (\tilde{B}')^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \gamma^* \end{pmatrix},$$

where $\beta_1^* = \hat{\beta}_1$. Again, by Lemma 6, we have the corresponding second derivatives as

$$\hat{\delta} = \tilde{B}''(\tilde{B}')^{-1}\hat{\gamma}, \quad \delta^* = \tilde{B}''(\tilde{B}')^{-1}\gamma^*.$$

From the definition of $\tilde{B}''(\tilde{B}')^{-1}$ in Lemma 6, we know that δ^* and $\hat{\delta}$ differ only at the $(j-1)$ th and j th knots as

$$\begin{aligned} 0 &\geq \delta_{j-1}^* = -\kappa_{j-1,j-1}\gamma_{j-1}^* + \kappa_{j-1,j}\gamma_j^* = -\kappa_{j-1,j-1}\hat{\gamma}_{j-1} + 0 = \hat{\delta}_{j-1}, \\ 0 &< \delta_j^* = -\kappa_{j,j}\gamma_j^* + \kappa_{j,j+1}\gamma_{j+1}^* = 0 + \kappa_{j,j+1}\hat{\gamma}_{j+1} \leq -\kappa_{j,j}\hat{\gamma}_j + \kappa_{j,j+1}\hat{\gamma}_{j+1} = \hat{\delta}_j, \end{aligned}$$

which implies $\|\delta^{*-}\|_\infty \leq \|\hat{\delta}^-\|_\infty$ and $\|\delta^{*+}\|_\infty \leq \|\hat{\delta}^+\|_\infty$. Hence for β^* , we know that

$$\begin{aligned} \beta_1^* &= \hat{\beta}_1, \quad (B'\beta^*)^+ = (B'\hat{\beta})^+, \quad (B'\beta^*)^- = \tilde{0}, \\ \|(B''\beta^*)^+\|_\infty &\leq \|(B''\hat{\beta})^+\|_\infty, \quad \|(B''\beta^*)^-\|_\infty \leq \|(B''\hat{\beta})^-\|_\infty. \end{aligned} \quad (\text{A.5})$$

As $\hat{\beta}$ is the minimizer of $Q(\beta)$ in (A.3), we prove $Q(\hat{\beta}) - Q(\beta^*) > 0$ for contradiction. We decompose $Q(\hat{\beta}) - Q(\beta^*)$ into M_I and M_p as

$$M_I = (Y - B\hat{\beta})^T(Y - B\hat{\beta}) - (Y - B\beta^*)^T(Y - B\beta^*),$$

$$M_p = \lambda_{2n}[\omega_3\{\|(B''\hat{\beta})^+\|_\infty - \|(B''\beta^*)^+\|_\infty\} + \omega_4\{\|(B''\hat{\beta})^-\|_\infty - \|(B''\beta^*)^-\|_\infty\}] + \lambda_{1n}\omega_2\|(B'\hat{\beta})^-\|_\infty.$$

For the M_I term, by some routine calculation, with β^c from Lemma 5, we get

$$M_I = (\hat{\beta} - \beta^*)^T B^T B(\hat{\beta} - \beta^*) - 2(Y - B\beta^c)^T B(\hat{\beta} - \beta^*) - 2(\beta^c - \beta^*)^T B^T B(\hat{\beta} - \beta^*). \quad (\text{A.6})$$

According to the Cauchy-Schwartz inequality and Lemma A.1 of Tang et al. (2012), we have for some constant $C_1 > 0$, $(\beta^c - \beta^*)^T B^T B(\hat{\beta} - \beta^*) \leq C_1 n k_n^{-1} \|\hat{\beta} - \beta^*\|_2 \|\beta^c - \beta^*\|_2$. From the triangle inequality, we get $\|\beta^c - \beta^*\|_2 \leq \|\beta^c - \hat{\beta}\|_2 + \|\hat{\beta} - \beta^*\|_2$. Combining with Lemma 5, we get

$$(\beta^c - \beta^*)^T B^T B(\hat{\beta} - \beta^*) \leq O_p(n^{1/2} + \lambda_{2n} n^{\frac{2}{2r+1}}) \|\hat{\beta} - \beta^*\|_2 + n^{\frac{2r}{2r+1}} \|\hat{\beta} - \beta^*\|_2^2. \quad (\text{A.7})$$

Define $e = (f(X_1), \dots, f(X_n))^T - B\beta^c$. By Theorem 1 of Shi and Li (1995), we have $\|e\|_2^2 = n^{\frac{1}{2r+1}}$. According to Lemma A.1 of Tang et al. (2012) and the Cauchy-Schwartz inequality, we have

$$|(Y - B\beta^c)^T B(\hat{\beta} - \beta^*)| \leq |e^T B(\hat{\beta} - \beta^*)| + |\varepsilon^T B(\hat{\beta} - \beta^*)| = O_p(n^{1/2}) \|\hat{\beta} - \beta^*\|_2. \quad (\text{A.8})$$

Combining (A.6), (A.7) and (A.8) together, we have

$$M_I \geq -O_p(n^{1/2} + \lambda_{2n} n^{\frac{2}{2r+1}}) \|\hat{\beta} - \beta^*\|_2. \quad (\text{A.9})$$

For the M_p term, combining Lemmas 1 and 4 (with $r \leq 3$), we have the convergence rate of unpenalized derivative estimator as $|\|B'(x_1^{**})\beta^u\|_\infty - \|f'(x_1^{**})\|_\infty| = O_p(n^{-\frac{r-1}{2r+1}} \log n)$ and $|\|B''(x_2^{**})\beta^u\|_\infty - \|f''(x_2^{**})\|_\infty| = O_p(n^{-\frac{r-2}{2r+1}} \log n)$. For $f(\cdot)$ being strictly increasing and convex, we know that $\|f'(x_1^{**})\|_\infty > 0$, $\|f'(x_1^{**})\|_\infty = 0$, $\|f''(x_2^{**})\|_\infty > 0$, and $\|f''(x_2^{**})\|_\infty = 0$. By the definition of adaptive weight, with probability approaching one, we have, for some positive constants a_1, a_2, a_3, a_4 ,

$$\omega_1 \leq \{\|f'(x_1^{**})\|_\infty/2\}^{-1} \leq a_1, \quad \omega_2 \geq a_2(\log n)^{-1} n^{\frac{r-1}{2r+1}}, \quad (\text{A.10})$$

$$\omega_3 \leq \{\|f''(x_2^{**})\|_\infty/2\}^{-1} \leq a_3, \quad \omega_4 \geq a_4(\log n)^{-1} n^{\frac{r-2}{2r+1}}.$$

Hence by Lemma 4, (A.5), (A.10), Theorem 1 conditions $c_4 \lesssim \frac{\lambda_{1n}}{\lambda_{2n}} \lesssim k_n$ and $k_n \sim n^{\frac{1}{2r+1}}$, there exists some positive constants C_2 and C_3 such that

$$M_p \geq \lambda_{1n}\omega_2\|(B'\hat{\beta})^-\|_\infty \geq c_4\lambda_{2n}\omega_2\|B'(\beta^* - \hat{\beta})\|_\infty \quad (\text{A.11})$$

$$\geq \lambda_{2n}C_2(\log n)^{-1} n^{\frac{r-1}{2r+1}} k_n^{1/2} \|\beta^* - \hat{\beta}\|_2 \doteq C_3\lambda_{2n}(\log n)^{-1} n^{\frac{r-1/2}{2r+1}} \|\hat{\beta} - \beta^*\|_2.$$

Combining (A.9) and (A.11), we have, with probability approaching one,

$$Q(\hat{\beta}) - Q(\beta^*) \geq -C_4(n^{1/2} + \lambda_{2n} n^{\frac{2}{2r+1}}) \|\hat{\beta} - \beta^*\|_2 + C_3\lambda_{2n}(\log n)^{-1} n^{\frac{r-1/2}{2r+1}} \|\hat{\beta} - \beta^*\|_2,$$

for some positive constant C_4 . With $2.5 < r \leq 3$ and $\lambda_{2n}/(n^{\frac{1}{2r+1}} \log n) \rightarrow \infty$, we have $Q(\hat{\beta}) - Q(\beta^*) > 0$, which is a contradiction with $\hat{\beta}$ being the solution minimizing the optimization. Thus (A.4) is denied.

If there exists multiple knots at which the signs of the first derivatives are incorrectly estimated, we can construct corresponding γ^* to show the contradiction similarly.

(I.2) Assume $\lim_{n \rightarrow \infty} P(\|(B''(x_2^{**})\hat{\beta})^-\|_\infty > 0) = \eta_2$ for some constant $0 < \eta_2 < 1$. By Lemma 6, we have the derivative transformation as

$$\delta = \tilde{B}''(\tilde{B}')^{-1}\gamma \doteq H\gamma = \begin{pmatrix} H_{(1)} & \tilde{H} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \tilde{\gamma} \end{pmatrix} = H_{(1)}\gamma_1 + \tilde{H}\tilde{\gamma},$$

$$\gamma = \begin{pmatrix} \gamma_1 \\ \tilde{\gamma} \end{pmatrix} = \begin{pmatrix} 1 & \tilde{0}^T \\ -\tilde{H}^{-1}H_{(1)} & \tilde{H}^{-1} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \delta \end{pmatrix}, \quad (\text{A.12})$$

where $\tilde{\gamma}$ is the sub vector of γ excluding the first component, $H_{(1)}$ is the first column of H , and $\tilde{H}$ contains the rest columns of H . By Lemma 6, we have

$$H = \tilde{B}''(\tilde{B}')^{-1} = \begin{pmatrix} -\kappa_{1,1} & \kappa_{1,2} & 0 & \cdots & 0 & 0 \\ 0 & -\kappa_{2,2} & \kappa_{2,3} & \cdots & 0 & 0 \\ 0 & 0 & -\kappa_{3,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\kappa_{k_n+1, k_n+1} & \kappa_{k_n+1, k_n+2} \end{pmatrix}_{(k_n+1) \times (k_n+2)}.$$

Thus $\tilde{H}$ has the dimension $(k_n + 1) \times (k_n + 1)$ and it is of full rank. We can solve out that

$$\tilde{H}^{-1} = \begin{pmatrix} \frac{1}{\kappa_{1,1}} & 0 & \cdots & 0 & 0 \\ \frac{\kappa_{1,2}}{\kappa_{1,1}\kappa_{2,2}} & \frac{1}{\kappa_{2,2}} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\prod_{i=1}^{k_n} \kappa_{i,i+1}}{\prod_{i=1}^{k_n} \kappa_{i,i}} & \frac{\prod_{i=2}^{k_n} \kappa_{i,i+1}}{\prod_{i=2}^{k_n} \kappa_{i,i}} & \cdots & \frac{1}{\kappa_{k_n,k_n}} & 0 \\ \frac{\prod_{i=1}^{k_n+1} \kappa_{i,i+1}}{\prod_{i=1}^{k_n+1} \kappa_{i,i}} & \frac{\prod_{i=2}^{k_n+1} \kappa_{i,i+1}}{\prod_{i=2}^{k_n+1} \kappa_{i,i}} & \cdots & \frac{\kappa_{k_n,k_n+1}}{\prod_{i=k_n}^{k_n+1} \kappa_{i,i}} & \frac{1}{\kappa_{k_n+1,k_n+1}} \end{pmatrix}_{(k_n+1) \times (k_n+1)}.$$

By the form of $\tilde{H}^{-1}$, if γ_1 is given, we can recover γ from δ by (A.12) as

$$\gamma = \begin{pmatrix} \gamma_1 \\ \tilde{\gamma} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \tilde{H}^{-1}\delta - \gamma_1\tilde{H}^{-1}H_{(1)} \end{pmatrix} = \begin{pmatrix} 1 & \vec{0}^T \\ -\tilde{H}^{-1}H_{(1)} & \tilde{H}^{-1} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \delta \end{pmatrix},$$

where $-\tilde{H}^{-1}H_{(1)} = \left(1, \frac{\kappa_{2,1}}{\kappa_{2,2}}, \dots, \frac{\prod_{i=1}^{k_n} \kappa_{i,i+1}}{\prod_{i=1}^{k_n} \kappa_{i+1,i+1}}\right)^T$, and for equally spaced knots, $-\tilde{H}^{-1}H_{(1)} = \vec{1}$. Combining with (A.2), we have

$$\beta = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{1} & (\tilde{B}')^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 & \vec{0}^T \\ \vec{1} & 1 & \vec{0}^T \\ \vec{1} & -\tilde{H}^{-1}H_{(1)} & \tilde{H}^{-1} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \gamma_1 \\ \delta \end{pmatrix}. \quad (\text{A.13})$$

That is, given β_1 and γ_1 , we can recover β from δ .

For $\lim_{n \rightarrow \infty} P(\|(B''(x_2^{**})\hat{\beta})^-\|_\infty > 0) = \eta_2$, without loss of generality, we can assume

$$\hat{\delta}_\ell = (\tilde{B}'\hat{\beta})_\ell < 0, \quad \hat{\delta}_i \geq 0, \quad \forall i \neq \ell, \quad (\text{A.14})$$

an incorrectly estimation at the ℓ th knot. Define $\delta^{**} = (\delta_1^{**}, \dots, \delta_{k_n+1}^{**})^T$, with $\delta_i^{**} = \hat{\delta}_i$, $\forall i \leq \ell - 1$, and $\delta_\ell^{**} = \dots = \delta_{k_n+1}^{**} = 0$. Corresponding γ^{**} can be constructed by (A.12). Since $-\tilde{H}^{-1}H_{(1)}$ is determinant for given knots, we have $\gamma_i^{**} = \hat{\gamma}_i$, $1 \leq i \leq \ell - 1$ and $\gamma_{\ell+1}^{**} = \gamma_{\ell+1}^{**} = \dots = \gamma_{k_n+1}^{**} = \hat{\gamma}_{\ell-1}$. It is easy to derive that $\|(\gamma^{**})^-\|_\infty = 0$ and $\|(\gamma^{**})^+\|_\infty \leq \|\hat{\gamma}^+\|_\infty$. Then, for the corresponding β^{**} by (A.13) from γ^{**} , we get

$$\begin{aligned} \beta_1^{**} &= \hat{\beta}_1, \quad \gamma_1^{**} = \hat{\gamma}_1, \quad (B''\beta^{**})^+ = (B''\hat{\beta})^+, \quad (B''\beta^{**})^- = \vec{0}, \\ \|(B'\beta^{**})^+\|_\infty &\leq \|(B'\hat{\beta})^+\|_\infty, \quad \|(B'\beta^{**})^-\|_\infty \leq \|(B'\hat{\beta})^-\|_\infty. \end{aligned} \quad (\text{A.15})$$

In addition, γ_1^{**} can be replaced by any other $\hat{\gamma}_i$, and the inequalities in (A.15) still hold.

Similar to (I.1), we can decompose $Q(\hat{\beta}) - Q(\beta^{**})$ as M'_p and M_p , where

$$\begin{aligned} M'_p &= (Y - B\hat{\beta})^T(Y - B\hat{\beta}) - (Y - B\beta^{**})^T(Y - B\beta^{**}), \\ M_p &= \lambda_{1n}[\omega_1\{ \|(B'\hat{\beta})^+\|_\infty - \|(B'\beta^{**})^+\|_\infty \} + \omega_2\{ \|(B'\hat{\beta})^-\|_\infty - \|(B'\beta^{**})^-\|_\infty \}] + \lambda_{2n}\omega_4\|(B''\hat{\beta})^-\|_\infty. \end{aligned}$$

The bound of M'_p is similar to M_l in (I.1). For M'_p , by (A.15) and $k_n \sim n^{\frac{1}{2r+1}}$, we have

$$\begin{aligned} M'_p &\geq \lambda_{2n}\omega_4\|(B''\hat{\beta})^-\|_\infty = \lambda_{2n}\omega_4\|B''(\beta^{**} - \hat{\beta})\|_\infty \\ &= \lambda_{2n}C_5(\log n)^{-1}n^{\frac{r-2}{2r+1}}k_n^{\frac{3}{2}}\|\beta^{**} - \hat{\beta}\|_2 \doteq C_6\lambda_{2n}(\log n)^{-1}n^{\frac{r-1/2}{2r+1}}\|\beta^{**} - \hat{\beta}\|_2, \end{aligned} \quad (\text{A.16})$$

for some positive constants C_5 and C_6 . Similar to (I.1), we can prove $Q(\hat{\beta}) - Q(\beta^{**}) > 0$, which is a contradiction and deny (A.14).

Also, similar arguments can be derived when the signs of second derivatives are incorrectly estimated at multiple knots.

(I.3) We discuss the simultaneous sign identification of derivatives. We use A_s to denote the event that “on all knots, sth derivative is correct in sign”, $s = 1, 2$, and $A_3 = A_1 \cap A_2$. From results in (I.1) and (I.2), we have $\lim_{n \rightarrow \infty} P(A_1^c) = \lim_{n \rightarrow \infty} P(\|B'\hat{\beta}^-\|_\infty \neq 0) = 0$ and $\lim_{n \rightarrow \infty} P(A_2^c) = \lim_{n \rightarrow \infty} P(\|B''\hat{\beta}^-\|_\infty \neq 0) = 0$, and therefore $\lim_{n \rightarrow \infty} P(A_3^c) = 0$. Thus, the proposed estimator can provide correct sign identification at knots for both first and second derivatives simultaneously. Extending from the knots to meshes, the proposed estimator based on quadratic B-splines is consistent in shape detection in the whole domain.

(II) Proof of part (II) of Theorem 1. Now we establish the convergence rate of the proposed estimator. It is easy to show that

$$\|B\hat{\beta} - f(X)\|_2 \leq \|B\hat{\beta} - B\beta^c\|_2 + \|B\beta^c - f(X)\|_2 \leq \lambda_{\max}^{1/2}(B^TB)\|\hat{\beta} - \beta^c\|_2 + \|B\beta^c - f(X)\|_2.$$

Combining $\lambda_{\max}(B^T B) = O(n^{\frac{2r}{2r+1}})$ from Lemma 2, $\|B\beta^c - f(X)\|_2 = O_p\{n^{\frac{1}{2(2r+1)}}\}$ from Lemma 3 and $\|\hat{\beta} - \beta^c\|_2 = O_p(n^{\frac{-r+1/2}{2r+1}} + \lambda_{2n} n^{\frac{-2r+2}{2r+1}})$ from Lemma 5, we have

$$n^{-1} \sum_{i=1}^n \{B(X_i)\hat{\beta} - f(X_i)\}^2 = n^{-1} \|B\hat{\beta} - f(X)\|_2^2 = O_p(n^{\frac{-2r}{2r+1}} + \lambda_{2n}^2 n^{\frac{-4r-3}{2r+1}}).$$

With $2.5 < r \leq 3$ and $n^{-\frac{r-3/2}{2r+1}} \lambda_{2n} \rightarrow 0$, we have $n^{-1} \sum_{i=1}^n \{B(X_i)\hat{\beta} - f(X_i)\}^2 = O_p\{n^{-2r/(2r+1)}\}$.

A.4. Proof of Theorem 2

The proof of part (II) is similar to that of part (II) of Theorem 1, thus is omitted. We use the method of contradiction to prove part (I). We only prove Theorem 2 for $0.5 < \nu \leq 1$, so that $2.5 + \eta \leq r \leq 3$ for some positive constant η , and the proof for $\nu > 1$ is similar.

(I.1) Given that $\exists x_1, x_2 \in (0, 1)$, s.t. $f'(x_1) \cdot f'(x_2) < 0$, then $\exists x_0 \in (x_1, x_2)$, s.t. $f'(x_0) < 0$. If $\lim_{n \rightarrow \infty} P\{\|(B'(x_1^{**}))\hat{\beta}\|_{\infty} = 0\} \neq 0$, then with positive probability $\hat{\gamma}_i = (\hat{B}'\hat{\beta})_i > 0$ for all $i = 0, 1, \dots, k_n + 1$. Without loss of generality, we can assume that

$$\hat{\gamma}_j > 0, \text{ while } f'(x_j^*) < 0, \quad (\text{A.17})$$

a false derivative estimation at j th knot x_j^* . According to Lemma 5, we have that $\|\hat{\beta} - \beta^c\|_2 = O_p\{\max(n^{\frac{-r}{2r+1}}, \lambda_{2n} n^{\frac{-2r-2}{2r+1}})\}$. By Lemma 4 we have $\lim_{n \rightarrow \infty} \gamma_j^c = f'(x_j^*) < \eta_j < 0$, for some constant $\eta_j < 0$.

If assumption $\hat{\gamma}_j > 0$ is true, then $|\hat{\gamma}_j - \gamma_j^c| \geq |\eta_j|$, which implies that $\|\hat{\gamma} - \gamma^c\|_2 \geq |\eta_j|$. However, from Lemma 5, we have $\|\hat{\gamma} - \gamma^c\|_2 = \|B'(\hat{\beta} - \beta^c)\|_2 = O_p\{\max(n^{\frac{-r-1/2}{2r+1}}, \lambda_{2n} n^{\frac{-2r-5/2}{2r+1}})\}$. Under Theorem 1 conditions $n^{\frac{-2r-5/2}{2r+1}} \lambda_{2n} \rightarrow 0$ and $2.5 < r \leq 3$, we have $\|\hat{\gamma} - \gamma^c\|_2 \rightarrow 0$. Therefore, for $n^{\frac{-2r-5/2}{2r+1}} \lambda_{2n} \rightarrow 0$, we have $\|\hat{\gamma} - \gamma^c\|_2 \rightarrow 0$.

Thus, the assumption (A.17) is conflicted with Lemma 5, which makes the contradiction. Hence, we have $\lim_{n \rightarrow \infty} P\{\|(\hat{\gamma})^-\|_{\infty} = 0\} = 0$ when $f'(x_1) \cdot f'(x_2) < 0$. In a similar way, we can prove that $\lim_{n \rightarrow \infty} P\{\|(\hat{\gamma})^+\|_{\infty} = 0\} = 0$ when $f'(x_1) \cdot f'(x_2) < 0$.

(I.2) Given that $\exists x_1, x_2 \in (0, 1)$, s.t. $f''(x_1) \cdot f''(x_2) < 0$, using the similar strategy as in (I.1), we can prove that $P\{\|(\hat{\delta})^-\|_{\infty} = 0 \text{ or } \|(\hat{\delta})^+\|_{\infty} = 0\} \rightarrow 0$ if $n^{\frac{-2r-3}{2r+1}} \lambda_{2n} \rightarrow 0$.

A.5. Proof of Lemmas 5 and 6

Proof of Lemma 5. To simplify the notation, we set $B \doteq B(X)$. By the triangle inequality, we have $\|\hat{\beta} - \beta^c\|_2 \leq \|\hat{\beta} - \beta^u\|_2 + \|\beta^u - \beta^c\|_2$. According to the central limit theorem, it is easy to prove that $\|\beta^u - \beta^c\|_2 = \|(B^T B)^{-1} B^T \varepsilon\|_2 = O_p(n^{-1/2} k_n^{1/2})$. Next we derive the bound for $\|\hat{\beta} - \beta^u\|_2$.

For $\|\hat{\beta} - \beta^c\|_2$, we assume that $\hat{\beta} - \beta^u = d_n u$, where d_n is a scalar and u is a vector satisfying $\|u\|_2 = 1$. By Lemma 2, we have

$$(Y - B\hat{\beta})^T(Y - B\hat{\beta}) - (Y - B\beta^u)^T(Y - B\beta^u) = d_n^2 u^T B^T B u \geq d_n^2 n k_n^{-1} \lambda_{\min}(V). \quad (\text{A.18})$$

Since $\hat{\beta}$ is the minimizer of $Q(\beta)$, thus $Q(\hat{\beta}) - Q(\beta^u) \leq 0$. We write

$$\begin{aligned} Q(\hat{\beta}) - Q(\beta^u) &= d_n^2 u^T B^T B u + \left\{ \lambda_{1n}(\omega_1 \|\hat{\gamma}^+\|_{\infty} + \omega_2 \|\hat{\gamma}^-\|_{\infty} - \omega_1 \|\gamma^{u+}\|_{\infty} - \omega_2 \|\gamma^{u-}\|_{\infty}) \right. \\ &\quad \left. + \lambda_{2n}(\omega_3 \|\hat{\delta}^+\|_{\infty} + \omega_4 \|\hat{\delta}^-\|_{\infty} - \omega_3 \|\delta^{u+}\|_{\infty} - \omega_4 \|\delta^{u-}\|_{\infty}) \right\} \\ &\doteq d_n^2 u^T B^T B u + M_1. \end{aligned} \quad (\text{A.19})$$

Now we consider the penalty difference M_1 . In the case that $f(\cdot)$ is strictly increasing and strictly convex, we know $\|f'(x_1^{**})^+\|_{\infty} > 0$ and $\|f''(x_2^{**})^+\|_{\infty} > 0$. By Lemma 4, we have, with probability approaching one,

$$\omega_1 \leq \{\|f'(x_1^{**})^+\|_{\infty}/2\}^{-1} \leq a_1, \quad \omega_3 \leq \{\|f''(x_2^{**})^+\|_{\infty}/2\}^{-1} \leq a_3,$$

where a_1 and a_3 are some positive constants. By the definitions of ω_2 and ω_4 , we have

$$\begin{aligned} M_1 &= \lambda_{1n} \omega_1 (\|\hat{\gamma}^+\|_{\infty} - \|\gamma^{c+}\|_{\infty}) + \lambda_{1n} \omega_2 (\|\hat{\gamma}^-\|_{\infty} - \|\gamma^{u-}\|_{\infty}) \\ &\quad + \lambda_{2n} \omega_3 (\|\hat{\delta}^+\|_{\infty} - \|\delta^{c+}\|_{\infty}) + \lambda_{2n} \omega_4 (\|\hat{\delta}^-\|_{\infty} - \|\delta^{u-}\|_{\infty}) \\ &\geq -\lambda_{1n} a_1 \|\hat{\gamma}^+\|_{\infty} - \|\gamma^{u+}\|_{\infty} + \lambda_{1n} (\omega_2 \|\hat{\gamma}^-\|_{\infty} - 1) \\ &\quad - \lambda_{2n} a_3 \|\hat{\delta}^+\|_{\infty} - \|\delta^{u+}\|_{\infty} + \lambda_{2n} \omega_4 \|\hat{\delta}^-\|_{\infty} - \lambda_{2n} \\ &\geq -\lambda_{1n} a_1 \|\hat{\gamma}^+\|_{\infty} - \|\gamma^{u+}\|_{\infty} - \lambda_{2n} a_3 \|\hat{\delta}^+\|_{\infty} - \|\delta^{u+}\|_{\infty} - \lambda_{1n} - \lambda_{2n}. \end{aligned}$$

By Lemma 1 we have

$$|\|\hat{\gamma}^+\|_\infty - \|\gamma^{u+}\|_\infty| \leq \|d_n \tilde{B}' u\|_\infty, \quad |\|\hat{\delta}^+\|_\infty - \|\delta^{u+}\|_\infty| \leq \|d_n \tilde{B}'' u\|_\infty.$$

According to the norm inequality $\|\cdot\|_\infty \leq \|\cdot\|_2$ and the exact form of B', B'' , we have

$$\|d_n \tilde{B}' u\|_\infty \leq \|d_n \tilde{B}' u\|_2 \leq d_n C_7 k_n, \quad \|d_n \tilde{B}'' u\|_\infty \leq \|d_n \tilde{B}'' u\|_2 \leq d_n C_8 k_n^2,$$

for some positive constants C_7 and C_8 . Given $\frac{\lambda_{1n}}{\lambda_{2n}} \lesssim k_n$, we have

$$M_1/\lambda_{2,n} \geq -a_1 C_7 d_n k_n^2 - a_3 C_8 d_n k_n^2 - k_n - 1. \quad (\text{A.20})$$

Let $k_n = C_9 n^{\frac{1}{2r+1}}$ for some positive constant C_9 . Combining (A.18), (A.19) and (A.20), we have, with probability approaching one,

$$\begin{aligned} Q(\hat{\beta}) - Q(\beta^u) &= d_n^2 u^T B^T B u + M_1 \geq d_n^2 n k_n^{-1} \lambda_{\min}(V) - \lambda_{2,n} \{(a_1 C_7 + a_2 C_8) d_n k_n^2 + k_n + 1\} \\ &\geq C_9^{-1} d_n^2 n^{\frac{2r}{2r+1}} \lambda_{\min}(V) - 2 \lambda_{2,n} (a_1 C_7 + a_2 C_8) C_{04}^2 d_n n^{\frac{2}{2r+1}}. \end{aligned}$$

Thus, if the order of d_n is larger than $\lambda_{2n} n^{-(2r-2)/(2r+1)}$, we have $Q(\hat{\beta}) - Q(\beta^c) > 0$, which is a contradiction to the definition of $\hat{\beta}$, the minimizer of $Q(\cdot)$. Combining the fact that $\|\beta^u - \beta^c\|_2 = O_p(n^{-1/2} k_n^{1/2})$, we have $\|\hat{\beta} - \beta^c\|_2 = O_p[\max\{n^{-r/(2r+1)}, \lambda_{2n} n^{-(2r-2)/(2r+1)}\}]$.

Proof of Lemma 6. By the definition,

$$B_{j,3}(x) = (x_{j+2}^* - x_j^*)^{-1} (x - x_j^*) B_{j,2}(x) + (x_{j+3}^* - x_{j+1}^*)^{-1} (x_{j+3}^* - x) B_{j+1,2}(x), j = 1, \dots, k_n + 3,$$

where $x_{-2}^* = x_{-1}^* = x_0^* = 0$, $x_{k_n+1}^* = x_{k_n+2}^* = x_{k_n+3}^* = 1$. Then

$$\frac{d}{dx} B_{j,3}(x) = 2(x_{j+2}^* - x_j^*)^{-1} B_{j,2}(x) - 2(x_{j+3}^* - x_{j+1}^*)^{-1} B_{j+1,2}(x),$$

where at each knot x_i^* , $B_{j,2}(x_i^*)$ can only take value 1 or 0. We construct $B' \doteq \left(\frac{d}{dx} B_{\cdot,3}(x)|_{x=x_1^{**}}\right) = M_1^T$ and decompose B' as $(B'_{(1)} \tilde{B}')$, where $B'_{(1)} = (-2(x_1^* - x_{-1}^*)^{-1}, 0, \dots, 0)^T$ and

$$\tilde{B}' = \begin{pmatrix} 2(x_1^* - x_{-1}^*)^{-1} & 0 & \cdots & 0 & 0 \\ -2(x_2^* - x_0^*)^{-1} & 2(x_2^* - x_0^*)^{-1} & \cdots & 0 & 0 \\ 0 & -2(x_3^* - x_1^*)^{-1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2(x_{N-3}^* - x_{N-4}^*)^{-1} & 0 \\ 0 & 0 & \cdots & -2(x_{N-2}^* - x_{N-3}^*)^{-1} & 2(x_{N-2}^* - x_{N-3}^*)^{-1} \end{pmatrix},$$

where $\tilde{B}'$ is a square matrix and has full rank.

Similarly, we can derive the second derivative of the basis function as

$$\begin{aligned} \frac{d^2}{dx^2} B_{j,3}(x) &= 2\{(x_{j+2}^* - x_j^*)(x_{j+1}^* - x_j^*)\}^{-1} B_{j,1}(x) + 2\{(x_{j+3}^* - x_{j+1}^*)(x_{j+3}^* - x_{j+2}^*)\}^{-1} B_{j+2,1}(x) \\ &\quad - 2\{(x_{j+3}^* - x_{j+1}^*)(x_{j+2}^* - x_j^*)(x_{j+2}^* - x_{j+1}^*)\}^{-1} (x_{j+3}^* - x_{j+1}^* + x_{j+2}^* - x_j^*) B_{j+1,1}(x), \end{aligned}$$

where at each knot x_i^* , $B_{j,1}(x_i^*)$ can only take value 1 or 0. Notice that $\frac{d^2}{dx^2} B_{j,3}(x)$ is not continuous at knot, thus we consider the right-limit as

$$\frac{d^2}{dx^2} B_{j,3}(x)|_{x=x_i^*} \doteq \lim_{h \rightarrow 0^+} \frac{d^2}{dx^2} B_{j,3}(x+h)|_{x=x_i^*}, i = 0, \dots, k_n,$$

for $j = 1, \dots, k_n + 3$. Then, we can construct

$$\begin{aligned} B'' &\doteq \left(\frac{d^2}{dx^2} B_{\cdot,3}(x)|_{x=x_2^{**}}\right) = M_2^T M_1^T \\ &= \begin{pmatrix} \nabla_{0,1} & \nabla_{0,2} & \nabla_{0,3} & \cdots & 0 & 0 \\ 0 & \nabla_{1,2} & \nabla_{1,3} & \cdots & 0 & 0 \\ 0 & 0 & \nabla_{2,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \nabla_{N-4,N-1} & 0 \\ 0 & 0 & 0 & \cdots & \nabla_{N-3,N-1} & \nabla_{N-3,N} \end{pmatrix}, \end{aligned}$$

where $\nabla_{ij} \neq 0$. We decompose B'' as $(B'_{(1)} \tilde{B}'')$, where $B'_{(1)} = (\nabla_{0,1}, 0, \dots, 0)^T$ and

$$\tilde{B}'' = \begin{pmatrix} \nabla_{0,2} & \nabla_{0,3} & 0 & \cdots & 0 & 0 \\ \nabla_{1,2} & \nabla_{1,3} & \nabla_{1,4} & \cdots & 0 & 0 \\ 0 & \nabla_{2,3} & \nabla_{2,4} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \nabla_{N-4,N-1} & 0 \\ 0 & 0 & 0 & \cdots & \nabla_{N-3,N-1} & \nabla_{N-3,N} \end{pmatrix}.$$

Since the spline knots are non-stochastic and nearly uniformly spaced in domain $[0, 1]$, combining the above discussion and (A.2) in the main file, we can show that

$$\tilde{\beta} = (\tilde{B}')^{-1} \gamma - (\tilde{B}')^{-1} B'_{(1)} \beta_1 = (\tilde{B}')^{-1} \gamma + \tilde{1} \cdot \beta_1,$$

where one can easily check $-(\tilde{B}')^{-1} B'_{(1)} = \tilde{1}$ by the definitions. Hence we have

$$\delta = B'_{(1)} \beta_1 + \tilde{B}'' \{(\tilde{B}')^{-1} \gamma + \tilde{1} \cdot \beta_1\} = \tilde{B}'' (\tilde{B}')^{-1} \gamma,$$

where $B'_{(1)} + \tilde{B}'' \tilde{1} = \tilde{0}$ by definition. If the knots are equally spaced, denoting $\kappa = k_n + 1$,

$$\begin{aligned} \tilde{B}'' (\tilde{B}')^{-1} &= \begin{pmatrix} -\kappa_{1,1} & \kappa_{1,2} & 0 & \cdots & 0 & 0 \\ 0 & -\kappa_{2,2} & \kappa_{2,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \kappa_{k_n, k_n+1} & 0 \\ 0 & 0 & 0 & \cdots & -\kappa_{k_n+1, k_n+1} & \kappa_{k_n+1, k_n+2} \end{pmatrix} \\ &= \begin{pmatrix} -\kappa & \kappa & 0 & \cdots & 0 & 0 \\ 0 & -\kappa & \kappa & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \kappa & 0 \\ 0 & 0 & 0 & \cdots & -\kappa & \kappa \end{pmatrix}_{(k_n+1) \times (k_n+2)}. \end{aligned}$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.csda.2020.106956>. The online supporting information contains additional simulation results. The developed R package is available at <http://blogs.gwu.edu/judywang/software/>.

References

- Abrevaya, J., Jiang, W., 2005. A nonparametric approach to measuring and testing curvature. *J. Bus. Econom. Statist.* 23, 1–19.
- Ahkim, M., Gijbels, I., Verhasselt, A., 2017. Shape testing in varying coefficient models. *Test* 26, 429–450.
- Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., Zoghbi, H.Y., 1999. Rett syndrome is caused by mutations in x-linked *mecp2*, encoding methyl-cpg-binding protein 2. *Nature Genet.* 23 (2).
- Boente, G., Rodriguez, D., Vena, P., 2020. Robust estimators in a generalized partly linear regression model under monotony constraints. *Test* 29 (1), 50–89.
- de Boor, C., 1972. On calculating with b-splines. *J. Approx. Theory* 6, 50–62.
- Carroll, R.J., Delaigle, A., Hall, P., 2011. Testing and estimating shape-constrained nonparametric density and regression in the presence of measurement error. *J. Amer. Statist. Assoc.* 106 (493), 191–202.
- Chahrour, M., Zoghbi, H.Y., 2007. The story of rett syndrome: From clinic to neurobiology. *Neuron* 56 (3), 422–437.
- Chang, I., Chien, L., Hsiung, C.A., Wen, C., Wu, Y., 2007. Shape restricted regression with random bernstein polynomials. In: *Complex Datasets and Inverse Problems*. Institute of Mathematical Statistics, pp. 187–202.
- Collins, A.L., Levenson, J.M., Vilaythong, A.P., Richman, R., Armstrong, D.L., Noebels, J.L., Sweatt, J.D., Zoghbi, H.Y., 2004. Mild overexpression of *mecp2* causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* 13 (21), 2679–2689.
- Dette, H., Neumeyer, N., Pilz, K.F., 2006. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 12, 469–490.
- Du, P., Parmeter, C.F., Racine, J.S., 2013. Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statist. Sinica* 23 (3), 1347–1371.
- Efron, B., 2004. The estimation of prediction error. *J. Amer. Statist. Assoc.* 99 (467), 619–632.
- Feng, X., Sedransk, N., Xia, J.Q., 2014. Calibration using constrained smoothing with applications to mass spectrometry data. *Biometrics* 70, 398–408.
- Gallant, A.R., Golub, G.H., 1984. Imposing curvature restrictions on flexible functional forms. *J. Econometrics* 26 (3), 295–321.
- Ghosal, S., Sen, A., van der Vaart, A.W., 2000. Testing monotonicity of regression. *Ann. Statist.* 28, 1054–1082.
- Goffin, D., Allen, M., Zhang, L., Amorim, M., Wang, I.J., Reyes, A.S., Mercado-Berton, A., Ong, C., Cohen, S., Hu, L., 2012. Rett syndrome mutation *mecp2* t158a disrupts dna binding, protein stability and erp responses. *Nature Neurosci.* 15 (2), 274–283.
- Green, P.J., Silverman, B.W., 1994. *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman and Hall, CRC, New York.
- Grimsley, J.M.S., Monaghan, J.M., Wenstrup, J.J., 2011. Development of social vocalizations in mice. *PLOS ONE* 6 (3), 1–15.
- Guy, J., Cheval, H., Selfridge, J., Bird, A., 2011. The role of *mecp2* in the brain. *Annu. Rev. Cell Dev. Biol.* 27 (1), 631–652, PMID: 21721946.

- Hagberg, B., Aicardi, J., Dias, K., Ramos, O., 1983. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: Report of 35 cases. *Ann. Neurol.* 14 (4), 471–479.
- Hall, P., Huang, L., 2001. Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* 29 (3), 624–647.
- He, X., Shi, P., 1998. Monotone B-spline smoothing. *J. Amer. Statist. Assoc.* 93 (442), 643–650.
- Huang, J.Z., Wu, C.O., Zhou, L., 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* 14, 763–788.
- Huang, J.Z., Yang, L., 2004. Identification of non-linear additive autoregressive models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 (2), 463–477.
- Jacques, J., Preda, C., 2013. Funclust: A curves clustering method using functional random variables density approximation. *Neurocomputing* 112, 164–171.
- Ma, S., Racine, J.S., 2013. Additive regression splines with irrelevant categorical and continuous regressors. *Statist. Sinica* 23, 515–541.
- Ma, S., Song, P.X., 2015. Varying index coefficient models. *J. Amer. Statist. Assoc.* 110, 341–356.
- Mahrt, E.J., Perkel, D.J., Tong, L., Rubel, E.W., Portfors, C.V., 2013. Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. *J. Neurosci.* 33 (13), 5573–5583.
- Marschik, P.B., Einspieler, C., Sigafos, J., 2012. Contributing to the early detection of rett syndrome: The potential role of auditory gestalt perception. *Res. Dev. Disabil.* 33 (2), 461–466.
- Meyer, M.C., 2008. Inference using shape-restricted regression splines. *Ann. Appl. Stat.* 2 (3), 1013–1033.
- Molitor, J., Sun, D., 2002. Bayesian analysis under ordered functions of parameters. *Environ. Ecol. Stat.* 9 (2), 179–193.
- Papp, D., Alizadeh, F., 2014. Shape-constrained estimation using nonnegative splines. *J. Comput. Graph. Statist.* 23 (1), 211–231.
- Santos, M., Silva-Fernandes, A., Oliveira, P., Sousa, N., Maciel, P., 2007. Evidence for abnormal early development in a mouse model of rett syndrome. *Genes Brain Behav.* 6 (3), 277–286.
- Scattoni, M.L., Crawley, J.N., Ricceri, L., 2009. Ultrasonic vocalizations: A tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. *Neurosci. Biobehav. Rev.* 33 (4), 508–515. Risk Factors for Mental Health: Translational Models from Behavioral Neuroscience.
- Scattoni, M.L., Gandhi, S.U., Ricceri, L., Crawley, J.N., 2008. Unusual repertoire of vocalizations in the btbr t/tf/j mouse model of autism. *PLOS ONE* 3 (8), 1–15.
- Seijo, E., Sen, B., 2011. Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.* 39 (3), 1633–1657.
- Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R., Zoghbi, H., 2002. Mice with truncated mecp2 recapitulate many rett syndrome features and display hyperacetylation of histone h3. *Neuron* 35 (2), 243–254.
- Shen, X., Huang, H., 2006. Optimal model assessment, selection, and combination. *J. Amer. Statist. Assoc.* 101 (474), 554–568.
- Shi, P., Li, G., 1995. Global convergence rates of b-spline m-estimators in nonparametric regression. *Statist. Sinica* 303–318.
- Sklar, J.C., Wu, J., Meiring, W., Wang, Y., 2013. Nonparametric regression with basis selection from multiple libraries. *Technometrics* 55 (2), 189–201.
- Stone, C.J., 1985. Additive regression and other nonparametric models. *Ann. Statist.* 13 (2), 689–705.
- Tang, Y., Wang, H.J., Zhu, Z., Song, X., 2012. A unified variable selection approach for varying coefficient models. *Statist. Sinica* 22 (2), 601–628.
- Terrell, D., 1996. Incorporating monotonicity and concavity conditions in flexible functional forms. *J. Appl. Econometrics* 179–194.
- Tibshirani, R., Knight, K., 1999. The covariance inflation criterion for adaptive model selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 61 (3), 529–546.
- Wang, J., Ghosh, S.K., 2012. Shape-restricted nonparametric regression with bernstein polynomials. *Comput. Statist. Data Anal.* 56 (9), 2729–2741.
- Wang, J., Meyer, M., 2011. Testing the monotonicity or convexity of a function using regression splines. *Canad. J. Statist.* 39, 89–107.
- Wang, X., Shen, J., 2013. Uniform convergence and rate adaptive estimation of convex functions via constrained optimization. *SIAM J. Control Optim.* 51, 2753–2787.
- Wang, L., Yang, L., 2009. Spline estimation of single-index models. *Statist. Sinica* 19, 765–783.
- Yatchew, A., Härdle, W., 2006. Nonparametric state price density estimation using constrained least squares and the bootstrap. *J. Econometrics* 133 (2), 579–599.
- Ye, J., 1998. On measuring and correcting the effects of data mining and model selection. *J. Amer. Statist. Assoc.* 93 (441), 120–131.
- Zeng, P., Shi, J., Kim, W., 2019. Simultaneous registration and clustering for multi-dimensional functional data. *J. Comput. Graph. Statist.* 28 (4), 943–953.
- Zhou, S., Wolfe, D.A., 2000. On derivative estimation in spline regression. *Statist. Sinica* 10, 93–108.