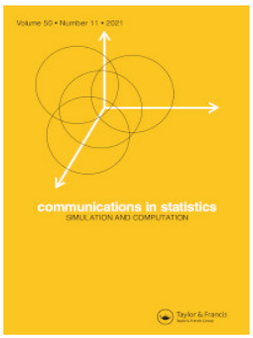

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Penalized I-spline monotone regression estimation

Junsouk Choi, JungJun Lee, Jae-Hwan Jhong & Ja-Yong Koo

To cite this article: Junsouk Choi, JungJun Lee, Jae-Hwan Jhong & Ja-Yong Koo (2021) Penalized I-spline monotone regression estimation, Communications in Statistics - Simulation and Computation, 50:11, 3714-3732, DOI: 10.1080/03610918.2019.1630433

To link to this article: https://doi.org/10.1080/03610918.2019.1630433

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2021, VOL. 50, NO. 11, 3714–3732
https://doi.org/10.1080/03610918.2019.1630433

Penalized I-spline monotone regression estimation

Junsouk Choi, JungJun Lee, Jae-Hwan Jhong, and Ja-Yong Koo

Department of Statistics, Korea University, Seoul, South Korea

ABSTRACT

We propose a penalized regression spline estimator for monotone regression. To construct the estimator, we adopt the I-splines with the total variation penalty. The I-splines lend themselves to the monotonicity because of the simpler form of restrictions, and the total variation penalty induces a data-driven knot selection scheme. A coordinate descent algorithm is developed for the estimator. If the number of complexity parameter candidates sufficiently increases, the algorithm considers all possible monotone linear spline fits to the given data. The pruning process of the algorithm not only provides numerical stability, but also implements the data-driven knot selection. We also compute the maximum candidate of the complexity parameter to facilitate complexity parameter selection. Extensive numerical studies show that the proposed estimator captures spatially inhomogeneous behaviors of data, such as sudden jumps.

ARTICLE HISTORY
Received 14 August 2018
Accepted 5 June 2019

KEYWORDS
Coordinate descent algorithm; I-splines; Knot selection; Maximum complexity parameter; Monotone regression; Total variation penalty

1. Introduction

In many disciplines, the imposition of the monotonicity constraint on the shape of the regression function is often necessary. The human growth curve estimation addresses a problem of the identification of monotonically non-decreasing growth functions (Cole, Freeman, and Preece 1998; Durbán et al. 2005). The dose-response curve estimation problem is a case where the monotonicity assumption plays a role in the explanation of real data (Kelly and Rice 1990). Ramsay (1988, 1998) showed that the monotone regression is closely related to the generalized Box-Cox transformation as well as the distribution and density estimation.

A vast literature exists for the nonparametric estimation of monotone regression functions. Brunk (1955) and Barlow, Bartholomew, and Bremner (1972) presented isotonic least squares estimators that are implemented using the pool-adjacent-violator algorithm of Robertson, Wright, and Dykstra (1988). These estimates are in the shapes of monotone step functions. Wu, Meyer, and Opsomer (2015) imposed a penalty on the range of the classical isotonic least squares estimator to resolve the inconsistency at the boundaries. In addition, various smoothing techniques with the monotonicity constraint have been studied (Friedman and Tibshirani 1984; Mammen 1991; Mammen and Thomas-Agnan 1999; Hall and Huang 2001; Mammen et al. 2001; Wang and Li 2008).

CONTACT Ja-Yong Koo jykoo@korea.ac.kr Department of Statistics, Korea University, Seoul 02841, South Korea.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

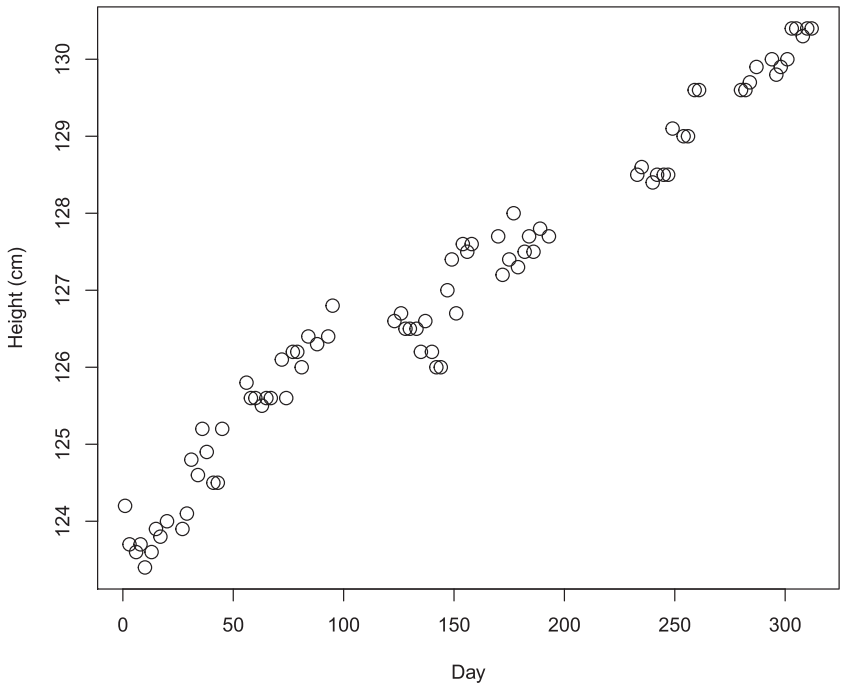

Figure 1. The plot of the growth data. Heights of a child are recorded over a 312-day period with $N = 83$ observations.

Many researchers have studied regression spline methods under the monotonicity constraint. Ramsay (1988) constructed a monotone regression spline estimator based on the I-splines by integrating the M-splines whose definition can be found in Ramsay (1988). Meyer (2008) extended the idea of Ramsay (1988) to other shape constraints, asserting that the imposition of the monotonicity or convexity assumption results in robust regression spline estimators with respect to the knot selection. A monotone smoother that is robust to the knot selection was also developed by Meyer (2012) based on the penalized spline of Eilers and Marx (1996).

Procedures for the adaptive knot selection are extensively discussed in the literature on regression splines (Mammen and van de Geer 1997; Osborne, Presnell, and Turlach 1998). However, the consideration of the knot selection methods is less evident in monotone regression splines. In various cases when monotone regression functions are inhomogeneous, knots need to be selected carefully to make the estimator flexible enough to capture the inhomogeneity. Proper knots should include more knots located at places where a regression function varies rapidly and fewer knots at the regions where the function seems stable. An example is the growth of a child data investigated in Thalange et al. (1996) and Ramsay (1998). The dataset involves the growth of a single child over a 312-day period with 83 observations. Human height is assumed to be non-decreasing when day passes and the prepubertal growth is modeled as periodic growth cycles including abrupt growth spurts with intervening growth arrest (Thalange et al. 1996). Datasets with such characteristics are often dealt in the human biology literature (Tillmann et al. 2002). The abrupt growth spurts appear as rapid jumps (near 150 Day) in Figure 1. Therefore a technique for choosing proper knots adaptive to data is

necessary in order to capture such growth, while maintaining the global fit. This article suggests a data-driven knot selection technique under the monotonicity constraint and its performance is tested in simulation studies.

The primary goal of this article is to propose a penalized regression spline method including an adaptive knot selection for monotone regression estimation. A modified version of the I-splines and the total variation penalty is used to construct the proposed estimator. The I-splines play a key role in simplifying the monotonicity constraint and the total variation penalty enable the estimator to get proper knots based on the given data. Being motivated by Jhong, Koo, and Lee (2017) that studied a penalized regression spline method using the B-splines and the total variation penalty, we devise a coordinate descent algorithm to implement the estimator. Through the pruning process of the algorithm, proper knots are automatically chosen, which produces an estimate that is smooth but flexible enough to capture abrupt changes in the data without compromising the overall fit. If the number of the complexity parameter candidates is sufficiently large, the algorithm deals with every monotone linear spline estimate. A formula for the upper bound of the complexity parameter is also derived to facilitate complexity parameter selection, subseqently reducing computational burden of the algorithm. Application to the simulated data and growth data reveals that the finite sample performance of the proposed method is satisfactory, particularly in the estimation of monotone regression functions having sudden jumps.

The rest of this article is organized as follow: Section 2 defines a monotone regression estimator using the modified I-splines with the total variation penalty. In Sec. 3, its implementation method based on a coordinate descent algorithm is described. Section 4 presents simulation studies and real data analysis to measure the finite sample performance of the proposed method; and lastly, this article is concluded in Sec. 5 with a brief discussion of future research topics. The technical details of the proposed implementation method are collected in the Appendix.

2. I-spline monotone regression estimator

Consider the problem of estimating the underlying regression function based on sampled data from the following model:

$$y_i = f(x_i) + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, N, \tag{1}$$

where $x_i \in \mathcal{I} \subset \mathbb{R}$ and ε_i are independent errors with the mean $\mathbb{E}(\varepsilon_i) = 0$ and the variance $\mathrm{Var}(\varepsilon_i) = \sigma^2 > 0$. It is assumed that the domain $\mathcal{I}$ is a compact interval $[a, b]$ with $-\infty < a < b < \infty$ and that f is a monotone function on $\mathcal{I}$. Our goal is to estimate f, based on the data $(x_1, y_1), \ldots, (x_N, y_N)$ from the model (1).

We slightly modify the I-splines by integrating the constant B-splines rather than the constant M-splines used by Ramsay (1988). The resulting spline basis is more appropriate for the adopted penalty (2). Let a sequence of knots $t_1 = a < t_2 < \cdots < t_K < b = t_{K+1}$ be given. In this article, for $k = 1, \ldots, K$, the I-splines are defined by

$$I_k(x) = \int_{t_1}^{x} B_k(z)dz = \begin{cases} 0 & \text{if} \quad x < t_k \\ x - t_k & \text{if} \quad t_k \le x < t_{k+1} \\ t_{k+1} - t_k & \text{if} \quad x \ge t_{k+1}, \end{cases}$$

where B_k's indicate the constant B-splines. Then, a linear spline is defined by a linear combination of $I_1, ..., I_K$, and a constant function as follows:

$$f(\cdot\ ;\beta) = \beta_0 + \sum_{k=1}^{K} \beta_k I_k(\cdot),$$

for $\beta = (\beta_0, \beta_1, ..., \beta_K)^\top \in \mathbb{R}^{K+1}$. Necessary and sufficient condition for $f(\cdot\ ;\beta)$ to be non-decreasing is $\beta_k \geq 0$ for all $k = 1, ..., K$. $f(\cdot\ ;\beta)$ is also non-increasing if and only if $\beta_k \leq 0$ for all $k = 1, ..., K$.

Without loss of generality, assume that f is non-decreasing and consider $f(\cdot\ ;\beta)$ under $\beta_k \geq 0$ for all $k = 1, ..., K$. This non-negativity of the basis coefficients supports the use of the I-splines rather than the linear truncated-power basis or the linear B-splines, as the latter require the non-negativity of the cumulative sum of basis coefficients or the monotonicity of basis coefficients to generate a monotonically non-decreasing function. Therefore, it is possible to construct a simpler optimization problem for the monotone regression estimation by utilizing the I-splines.

Consider the sum of squares of the residuals given by

$$L(\beta) = \frac{1}{2}\sum_{i=1}^{N} \big(y_i - f(x_i;\beta)\big)^2 = \frac{1}{2}\|Y - G\beta\|^2.$$

Here, $Y = [y_i] \in \mathbb{R}^N$ and G denotes the $N \times (K+1)$ design matrix filled with 1 in the first column and the (i, k)-th element $I_{k-1}(x_i)$ for $1 \leq i \leq N$ and $2 \leq k \leq K+1$. $\|\ \cdot\ \|$ denotes the Euclidean norm in $\mathbb{R}^N$. For the penalty of the penalized regression estimator, we adopt the total variation norm of the first derivative of $f(\ \cdot\ ;\beta)$ given by

$$\mathrm{pen}(\beta) = \sum_{k=1}^{K-1} |\beta_{k+1} - \beta_k|. \tag{2}$$

For $\lambda \geq 0, \hat{\beta}^\lambda$ is defined by the solution of the following optimization problem:

$$\min_{\beta \in \mathbb{R}^{K+1}} L^\lambda(\beta) \quad \text{subject to} \quad \beta_1, ..., \beta_K \geq 0, \tag{3}$$

where $L^\lambda(\beta) = L(\beta) + \lambda \mathrm{pen}(\beta)$. Then, we define the I-spline Monotone Regression (IMR) estimator by

$$\hat{f}^\lambda = f\Big(\ \cdot\ ;\hat{\beta}^\lambda\Big).$$

Note that the total variation penalty (2) controls the smoothness of $f(\ \cdot\ ;\beta)$, as it measures the total jump size of the I-spline coefficients. Moreover, if a component term $|\beta_{k+1} - \beta_k|$ of $\mathrm{pen}(\beta)$ equals zero, the corresponding knot t_{k+1} can be ruled out since it is no longer active. The knot selection problem is thus simplified to the choice of the complexity parameter λ, which also controls the smoothness of the proposed estimator. Unlike the fused lasso proposed by Tibshirani et al. (2005), the total variation penalty (2) does not include the lasso penalty. In our case the lasso penalty may help us make the I-spline coefficients being zero in order to estimate the zero slope for the region where data seems to be flat. However, since the monotonicity restriction is already a

strong condition, including the lasso penalty is not attractive in the monotone regression setting.

The penalty (2) and its knot selection property are satisfactory to capture flat trends without the lasso penalty. If knots are properly chosen so that they are placed at the locations where rapid changes occur, the entire domain will be partitioned into several intervals where the regression function has different behaviors. The monotonicity constraint makes the coefficients zero in the regions where data do not show strictly increasing trends, as the sample size increases. If the sample size is small, the coefficient estimates may not be exact zero and introducing the lasso penalty can be helpful. However, it is not appealing because its computational cost is much bigger than its benefit. If we insert the lasso term into our penalty, the computational burden of searching the optimal complexity parameter will drastically increase because this will double the number of complexity parameters. By the way of contrast, since the coefficient estimates are already very close to zero in those regions, the IMR estimate might not change substantially.

3. Implementation of the IMR estimator

A coordinate descent algorithm is proposed to minimize $L^{\lambda}(\beta)$ under the constraints $\beta_k \geq 0$ for all $k = 1, \ldots, K$. Let $\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_K$ be the current values of the intercept and the I-spline coefficients. Denote

$$L_k^{\lambda}(\beta_k) = L^{\lambda}\big(\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \beta_k, \ldots, \tilde{\beta}_K\big), \tag{4}$$

for $k = 0, 1, \ldots, K$. The coordinate descent algorithm iteratively updates the intercept and the I-spline coefficients according to the minimizer of the univariate penalized quadratic functions (4) until the convergence of the objective values. The existence of the unique minimizer of $L_k^{\lambda}(\,\cdot\,)$ over the corresponding domain is presented in Sec. 3.1. The form of coordinate-wise update is given by

$$\tilde{\beta}_0 \leftarrow \underset{\beta_0 \in \mathbb{R}}{\operatorname{argmin}}\, L_0^{\lambda}(\beta_0) \quad \text{and} \quad \tilde{\beta}_k \leftarrow \underset{\beta_k \in [0,\infty)}{\operatorname{argmin}}\, L_k^{\lambda}(\beta_k),$$

for $k = 1, \ldots, K$. The updates will be repeated until convergence is reached.

The overall strategy is as follows: First, compute the upper bound λ_M of the complexity parameter, for which no interior knot is involved in the computation of the IMR estimate. A method to compute λ_M is derived in Sec. 3.3. Second, generate a sequence of λ values that increase from $\lambda_1 = \epsilon\lambda_M$ to λ_M in the log scale; here, the values of $\epsilon = 10^{-6}$ and $M = 200$ are typical. Third, place the initial knots at distinct input values, and set the initial coefficients for λ_1 as the ordinary least squares estimates. Finally, sequentially compute the IMR estimates for the increasing sequence $\lambda_1 < \lambda_2 < \ldots < \lambda_M$, and select the optimal IMR estimate based on the Bayesian information criterion (BIC) that is defined in Sec. 3.2. In the procedure of the iterative computations of the IMR estimates, the results from λ_m are considered as the initial values for λ_{m+1}.

The algorithm is motivated by Friedman et al. (2007) and based on Jhong, Koo, and Lee (2017), but it is further specialized for the IMR estimator. The method investigated by Friedman et al. (2007) is equivalent to the method using constant spline basis with the initial knots placed at all of the input values. However, our method is based on the I-spline basis and thus the exact update formula is derived in the simplest form which

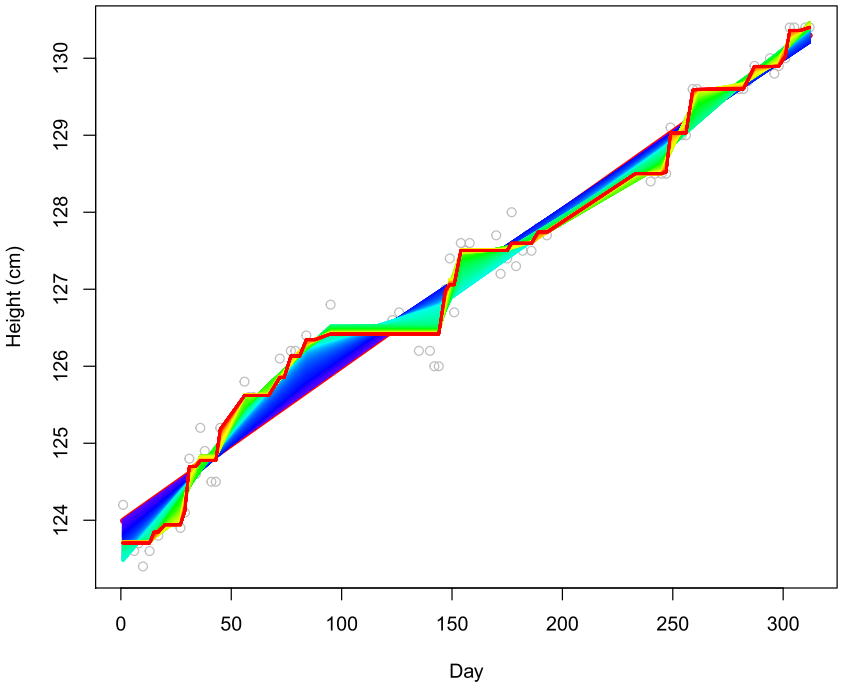

Figure 2. The IMR estimates for the growth data with $M = 1000$. The rainbow-colored lines indicate the IMR estimates for $\{\lambda_1, ..., \lambda_{1000}\}$ respectively. The red line (the isotonic least squares estimate) corresponds to λ_1, while the purple one (the monotone linear regression estimate) corresponds to λ_{1000}.

deals with the sign restriction of the I-splines. Moreover, in our algorithm, initial knots either can be any subset of the data points or can be values other than the input values. Thus more flexible choice can be made for the initial knot placement.

If M is sufficiently large, the algorithm searches all possible monotone piecewise linear spline estimates defined by the specified initial knots. When the initial knots are placed at the distinct input values, the IMR estimate corresponding to λ_1 is the isotonic least squares estimate that is the most wiggly piecewise linear fit to the given data, while λ_M leads to the monotone linear regression estimate that is completely linear. Figure 2 shows an example using the growth data. The algorithm is applied with $M = 1000$ and the rainbow-colored lines represent the piecewise linear fits that are achievable between the isotonic least squares estimate and the monotone linear regression estimate. The red line corresponding to λ_1 indicates the isotonic least squares estimate, while the purple one corresponding to λ_{1000} shows the monotone linear regression estimate. The algorithm chooses the optimal one by BIC among all monotone linear spline estimates between the isotonic least squares estimate and the monotone linear regression estimate.

3.1. Coordinate-wise updates

Observe that the gradient and the Hessian of L are respectively given by

$$\nabla L(\beta) = G^\top G\beta - G^\top Y \quad \text{and} \quad H = G^\top G.$$

Since

$$L_k''(\beta_k) = \frac{\partial^2}{\partial \beta_k^2} L\left(\tilde{\beta}_0, ..., \beta_k, ..., \tilde{\beta}_K\right) = H_{kk},$$

we have

$$L_k'(\beta_k) = \nabla L\left(\tilde{\beta}\right)_k + \mathrm{H}_{kk}\left(\beta_k - \tilde{\beta}_k\right),$$

where H_{kk} represents the (k, k)-th diagonal element of H and $\nabla L(\tilde{\beta})_k$ indicates the k-th element of $\nabla L(\tilde{\beta})$. Solving the linear equation $L_k'(\beta_k) = 0$ gives the solution $\tilde{\beta}_k - \nabla L(\tilde{\beta})_k / \mathrm{H}_{kk}$. Since L_k is a univariate quadratic function, we obtain

$$L_k(\beta_k) = \frac{\mathrm{H}_{kk}}{2}\left(\beta_k - \left(\tilde{\beta}_k - \frac{\nabla L\left(\tilde{\beta}\right)_k}{\mathrm{H}_{kk}}\right)\right)^2, \tag{5}$$

ignoring the terms independent of β_k.

Define

$$p(z \ ; \ c_1, c_2) = |z - c_1| + |z - c_2|, \tag{6}$$

where z, c_1, and c_2 are real numbers. Then, the penalty term can be written as

$$\mathrm{pen}\left(\tilde{\beta}_0, \tilde{\beta}_1, ..., \beta_k, ..., \tilde{\beta}_K\right) = \begin{cases} 0 & \text{if} \quad k = 0 \\ p\left(\beta_1 \ ; \ \tilde{\beta}_2, \tilde{\beta}_2\right)/2 & \text{if} \quad k = 1 \\ p\left(\beta_k \ ; \ \tilde{\beta}_{k-1}, \tilde{\beta}_{k+1}\right) & \text{if} \quad k = 2, ..., K-1 \\ p\left(\beta_K \ ; \ \tilde{\beta}_{K-1}, \tilde{\beta}_{K-1}\right)/2 & \text{if} \quad k = K. \end{cases} \tag{7}$$

ignoring the terms independent of β_k. Proposition 3.1 implies the explicit form of the minimizer of L_k^{λ} over $[0, \infty)$ with its existence. The proof is given in Appendix A.

Proposition 3.1. *Define an univariate penalized quadratic function q^{γ} by*

$$q^{\gamma}(z) = \frac{1}{2}(z - m)^2 + \gamma p(z \ ; \ c_1, c_2) \quad \text{if} \quad z \in \mathbb{R},$$

where $m \in \mathbb{R}, \gamma \geq 0$ and p is defined as (6). Then, the unique global minimizer $Q(m, c_1, c_2, \gamma)$ of q^{γ} over $[0, \infty)$ is given by

$$Q(m, c_1, c_2, \gamma) = \left[\mathrm{DST}\left(m - \frac{c_1 + c_2}{2} \quad ; \quad \frac{|c_1 - c_2|}{2}, \quad 2\gamma\right) + \frac{c_1 + c_2}{2}\right]_+,$$

where $[z]_+ = \max\{z, 0\}$ and $\mathrm{DST}(z \ , \ u, v)$ called the double soft-thresholding function is given by

$$\mathrm{DST}(z \ ; \ u, v) = \begin{cases} z - v & \text{if} \quad z \geq u + v \\ u & \text{if} \quad u \leq z < u + v \\ z & \text{if} \quad -u \leq z < u \\ -u & \text{if} \quad -u - v \leq z < -u \\ z + v & \text{if} \quad z < -u - v, \end{cases}$$

for $z \in \mathbb{R}$ and $u, v \geq 0$.

We finally obtain the coordinate-wise update formula given by

$$\tilde{\beta}_k \leftarrow \begin{cases} \tilde{\beta}_0 - \nabla L(\tilde{\beta})_0 / \mathrm{H}_{00} & \text{if} \quad k = 0 \\ Q\left(\tilde{\beta}_1 - \nabla L(\tilde{\beta})_1 / \mathrm{H}_{11}, \quad \tilde{\beta}_2, \quad \tilde{\beta}_2, \quad \dfrac{\lambda}{\mathrm{H}_{11}}\right) & \text{if} \quad k = 1 \\ Q\left(\tilde{\beta}_k - \nabla L(\tilde{\beta})_k / \mathrm{H}_{kk}, \quad \tilde{\beta}_{k-1}, \quad \tilde{\beta}_{k+1}, \quad \dfrac{2\lambda}{\mathrm{H}_{kk}}\right) & \text{if} \quad k = 2, \ldots, K-1 \\ Q\left(\tilde{\beta}_K - \nabla L(\tilde{\beta})_K / \mathrm{H}_{KK}, \quad \tilde{\beta}_{K-1}, \quad \tilde{\beta}_{K-1}, \quad \dfrac{\lambda}{\mathrm{H}_{KK}}\right) & \text{if} \quad k = K. \end{cases}$$

3.2. Pruning process and complexity parameter selection

The algorithm starts at λ_1 near zero, stopping at λ_M for which no interior knot is involved in the computation of the IMR estimate. When the complexity parameter λ increases from λ_m to λ_{m+1}, we eliminate or *prune* the knots which are no longer active. The pruning process provides the fusion of the neighboring pair of the corresponding I-splines, and Friedman et al. (2007) established that it is a way to get around the coordinate descent algorithm getting stuck. Proper knots are selected based on the given data via the pruning process and its performance will be examined in simulation studies of Section 4. As λ increases, the number of both knots and the basis decreases. The IMR estimate for $\lambda = \lambda_M$ appears to be a non-decreasing linear function.

The pruning process looks similar to the fusion cycle of Friedman et al. (2007). However the reason why we do pruning is not simply because the neighboring basis coefficients are equal as the fusion cycle does. It is because the identical coefficients means that the knot in the middle is being inactive and the dimension of the design matrix can be reduced. This implies that pruning process should contain not only a step to find and combine the identical neighboring coefficients but also a step to remove knots and re-calculate the corresponding design matrix whose dimension is reduced.

For the criterion to choose the optimal complexity parameter $\hat{\lambda}$ among $\lambda_1, \ldots, \lambda_M$, the BIC is defined by

$$\mathrm{BIC}(\lambda) = N \log\left(L\left(\hat{\beta}^{\lambda}\right)/N\right) + K^{\lambda} \log N,$$

where K^{λ} is the dimension of $\hat{\beta}^{\lambda}$. The optimal complexity parameter is chosen to minimize $\mathrm{BIC}(\lambda)$. The BIC tends to produce a smooth regression spline estimate, compared with other methods for complexity parameter selection such as the Akaike information criterion (AIC) or cross-validation. The resulting smoothness from the BIC contributes to the robustness of the IMR estimate with respect to outliers.

3.3. Upper bound of the complexity parameter

In order to make the search for the optimal complexity parameter conveniently restricted, we want to compute an upper bound of the complexity parameter beyond which every interior knot is out of work. Osborne, Presnell, and Turlach (2000) has shown that there exists an upper bound for the complexity parameter beyond which the

solution to the lasso problem is the null vector. In the setting of the IMR estimator, the way of Osborne, Presnell, and Turlach (2000) does not guarantee the closed form of the bound when the I-spline coefficients are more than two. It is because the total variation penalty (2) is defined based on the difference of the coefficients. Thus we need another approach to calculate such upper bound in the case of the IMR estimator and Jhong, Koo, and Lee (2017) motivates us to get one way. As the algorithm we proposed works, the pruning process keeps to reduce the number of the I-splines until only two remains. In this case it is possible to get the closed form of an upper bound in the similar way of Osborne, Presnell, and Turlach (2000).

Based on the pruning process described in Sec. 3.2, the remaining consideration for the computation of λ_M is regarding the situation where only one interior knot remains. With the assumption that only one interior knot $\xi \in (a, b)$ exists, the I-splines are written as I_1^ξ and I_2^ξ in this subsection. Let $\hat{\beta}^{\lambda,\xi} = (\hat{\beta}_0^{\lambda,\xi}, \hat{\beta}_1^{\lambda,\xi}, \hat{\beta}_2^{\lambda,\xi})$ denote the solution to (3) with only the single knot ξ. The centered design matrix is denoted by $G_c^\xi \in \mathbb{R}^{N\times 2}$, whose $(i,\ k)$-th element is $I_k^\xi(x_i) - \frac{1}{N}\sum_{i=1}^N I_k^\xi(x_i)$ and $Y_c \in \mathbb{R}^N$ represents the centered response vector. The proposition 3.2 provides the upper bound beyond which the knot ξ is ineffective and the I-splines I_1^ξ and I_2^ξ merge. It is a specialized version of Theorem 2 of Jhong, Koo, and Lee (2017) which we adapt to the setting of the I-splines. It provides a method to compute λ_M that remove every interior knot in the end of the pruning process. The proof is given in Appendix A.

Proposition 3.2. *Suppose that G_c^ξ has full-rank and*

$$\lambda \geq \lambda^\xi \triangleq \frac{|d^\top \left(\mathrm{G}_c^{\xi\top}\mathrm{G}_c^\xi\right)^{-1}\mathrm{G}_c^{\xi\top} Y_c|}{d^\top \left(\mathrm{G}_c^{\xi\top}\mathrm{G}_c^\xi\right)^{-1} d}, \tag{8}$$

where $d = (1, -1)^\top$. Then, $\mathrm{f}(\cdot\ ; \hat{\beta}^{\lambda,\xi})$ is a non-decreasing linear function.

If we define $\lambda_M = \max_{\xi\in\{x_1,\ldots,x_N\}}\lambda^\xi$ and choose $\lambda \geq \lambda_M$, the pruning process will delete even the last interior knot and the IMR estimate will be a linear function. Note that our approach is essentially based on the knot selection property of the IMR method. The use of λ_M facilitates the search of the optimal complexity parameter, because it works as an upper bound beyond which the complexity parameter candidates do not need to be considered.

4. Numerical studies

4.1. Simulations

We conduct simulations to examine the finite sample performance of the IMR estimator by comparing with other monotone regression estimators. For the comparison, the penalized isotonic regression (PIR) method of Wu, Meyer, and Opsomer (2015) and the shape restricted regression splines (SRRS) method of Meyer (2008) are chosen. The PIR method fits the isotonic regression with a penalty on the range of the regression function and is available in the `isotonic.pen` package in `R`. The tools for the SRRS method can be obtained from the `ConSpline` package in `R` and it utilizes quadratic monotone

regression splines. The results of the PIR and SRRS estimates that are presented in this section are obtained using these packages in R with the given default options.

In order to compare our method with another penalized monotone linear splines method, we take into account the constrained penalized splines (CPS) method of Meyer (2012). The linear B-splines are used with a lot of knots placed at every five observation. We specify the second order difference matrix for its penalty and the BIC described in Meyer (2012) is adopted to select the optimal complexity parameter. The CPS method were coded using the R package coneproj for cone projection and quadratic programing.

We examine the monotone regression model with the regular design on $[0, 1]$. The following two different distributions for ε_i are considered: normal distribution $N(0, 0.1^2)$ and mixture distribution $0.9N(0, 0.1^2) + 0.1N(0, 0.3^2)$. The latter is utilized to generate outliers. Samples of sizes 100 and 500 are generated, and the following four monotone regression functions on $[0, 1]$ are considered:

$$\begin{aligned} f_1(x) &= x - [x-0.25]_+ + 4[x-0.5]_+ - 2[x-0.75]_+, \\ f_2(x) &= \exp(x), \\ f_3(x) &= 1/\left(1 + \exp(-40(x-0.5))\right), \end{aligned}$$

and

$$f_4(x) = 10(x-0.5)^3 + 20[x-0.49]_+ - 20[x-5.1]_+.$$

f_1 is a monotone linear spline that the IMR method is expected to fit well, while f_2 is a simple smooth monotone curve. Jumps, smooth or non-smooth rapid increases, are positioned on f_3 and f_4, thereby making the knot selection an important issue for estimating them. f_3 has a smooth jump at 0.5 with flat areas before and after the jump. f_4 is constructed to produce data sets imitating the growth data. It reveals as a smooth monotone curve with a non-smooth jump at 0.5. The shapes of these functions are displayed in Figure 3.

The generation of 100 data sets is completed for each combination of the underlying regression functions, the error distributions, and the sample sizes. As a discrepancy measure between the true regression function f and a function g, we adopt the mean squared error (MSE) defined by

$$\mathrm{MSE}(g) = \frac{1}{N}\sum_{i=1}^{N}\left(g(x_i) - f(x_i)\right)^2.$$

The maximum deviation (MXDV) defined in Breiman and Peters (1992) is also considered as a discrepancy measure:

$$\mathrm{MXDV}(g) = \max_{1\le i\le N} |g(x_i) - f(x_i)|.$$

The MSE is generally used to measure the global performance of an estimator. The MXDV is more adequate than the MSE to examine the local performance over areas where drastic changes occur in the underlying regression function.

Figure 4 summarizes the results of the simulation studies for the regression function f_1. If the sample size is small, the IMR, CPS and SRRS estimates show similar results

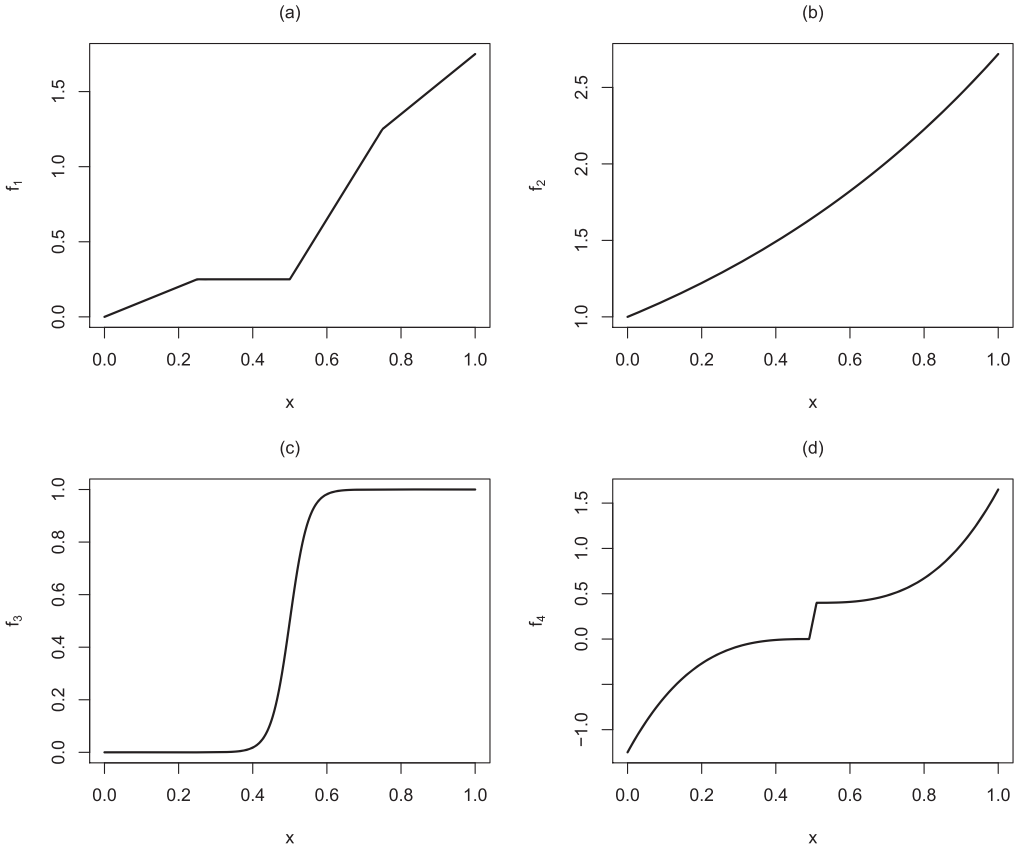

Figure 3. Plots of the monotone regression functions for simulations: f_1 (a), f_2 (b), f_3 (c), and f_4 (d).

under both the normal errors and the mixture errors, while the PIR method does not seem to work well in this case. However, as the sample size increases, the accuracy of the IMR estimate is improved more rapidly than those of the others. The IMR method can be a good choice when data show linear trends in some regions.

The simulation results for the regression function f_2 are demonstrated by Figure 5. In the cases of f_2, the IMR, CPS and SRRS methods seem to adapt well to the smooth regression function. Since the PIR method is not able to get a smooth fit, it shows poor performance in every case. Even though the SRRS method utilizes the cubic splines that are much smoother than the linear splines, the IMR estimator shows the similar results with the SRRS estimator. It validates that the knot selection technique we proposed is also satisfactory to make the IMR estimator smooth enough when the underlying regression curve has a certain degree of smoothness.

The results of each case for the regression function f_3 are shown by Figure 6. If we consider f_3 with the small sample size, the IMR, CPS and PIR estimates are better than the SRRS method in the sense of both local and global performance. With the enough sample size, the realized values of the MSE and MXDV of the IMR and CPS methods are lower than those of other methods. The function f_3 has flat regions and a rapid change which increases smoothly. In this situation, the IMR and CPS methods estimate both the stable trend and the rapid change quite well.

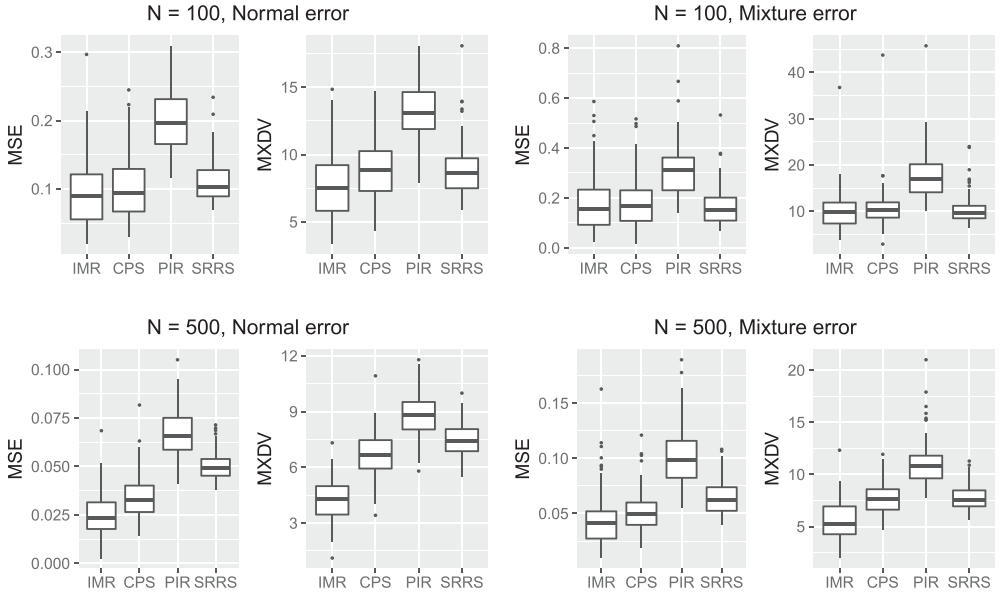

Figure 4. Boxplots of the realized values ($\times 10^2$) of the MSE and MXDV for the simulation studies on the regression function f_1.

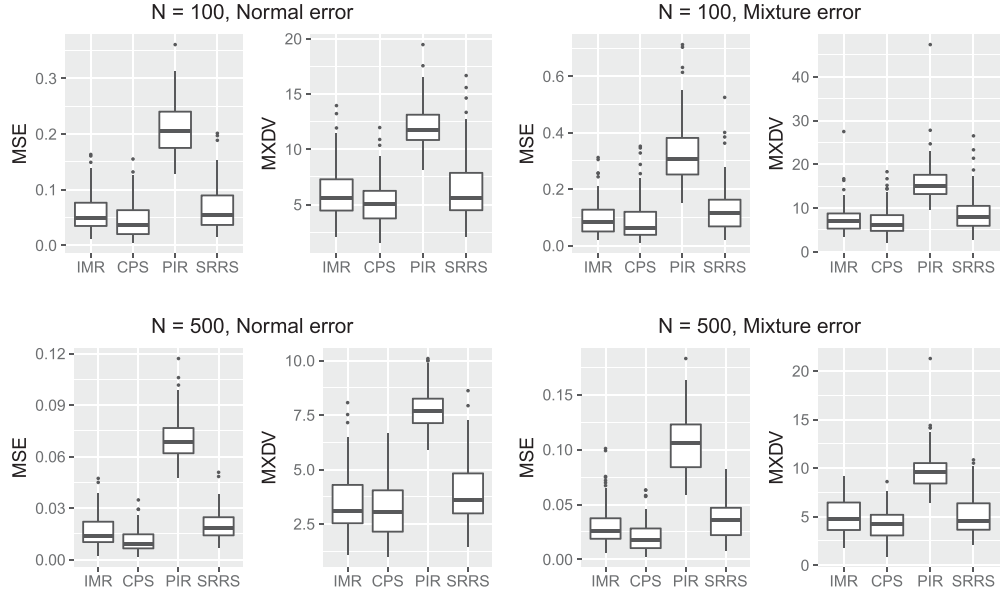

Figure 5. Boxplots of the realized values ($\times 10^2$) of the MSE and MXDV for the simulation studies on the regression function f_2.

Figure 7 illustrates the simulation results on the regression function f_4 which has a sudden jump in the middle of the smooth mean trend. When the small sample is generated, the IMR and CPS method appears to be slightly better than the other methods in

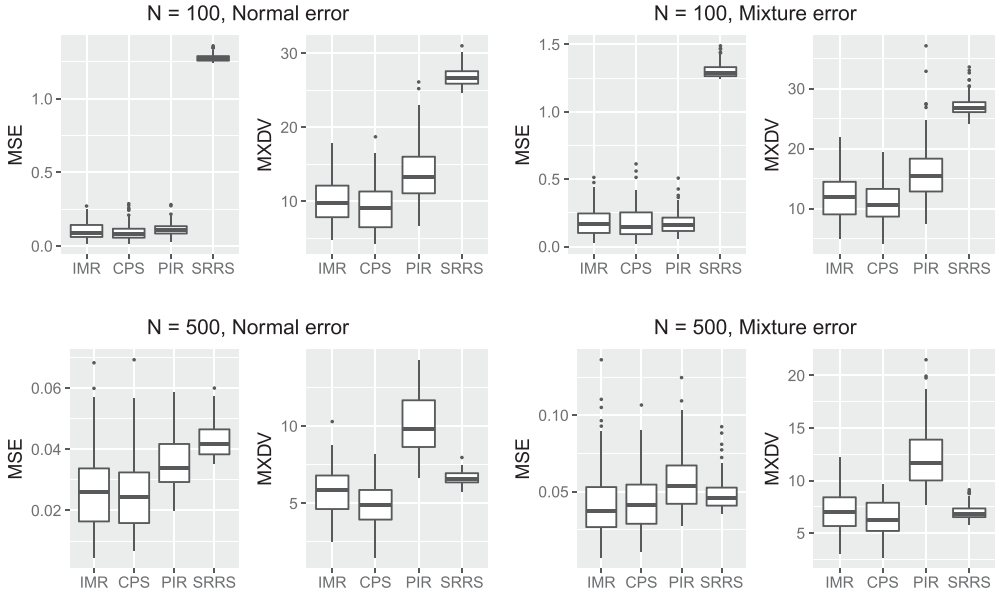

Figure 6. Boxplots of the realized values ($\times 10^2$) of the MSE and MXDV for the simulation studies on the regression function f_3.

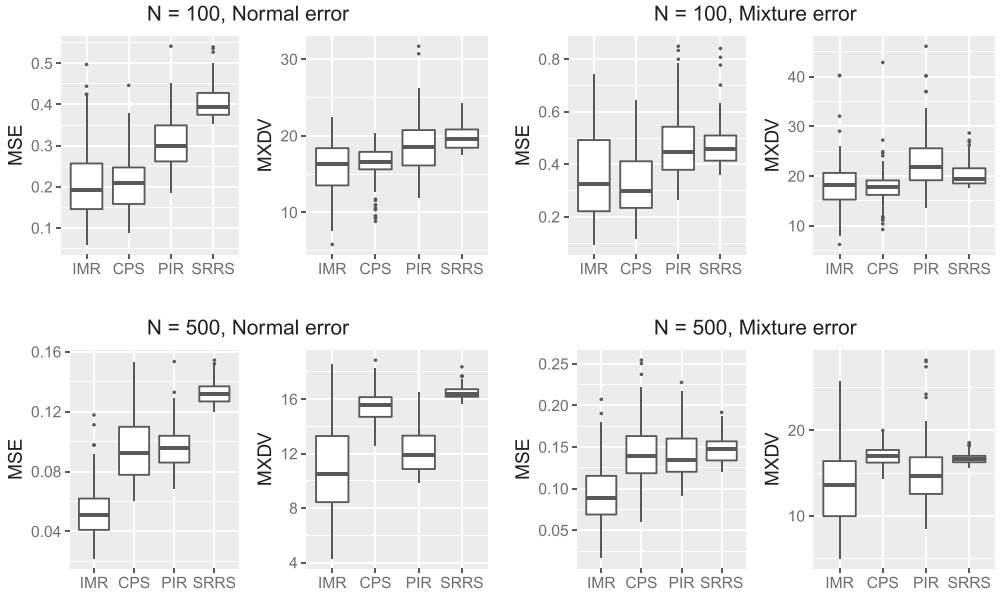

Figure 7. Boxplots of the realized values ($\times 10^2$) of the MSE and MXDV for the simulation studies on the regression function f_4.

the estimation of the global smoothness. As the sample size increases, however, the IMR method demonstrates much better global and local performance. The CPS and SRRS methods have large MXDV values, implying that they are not good at capturing

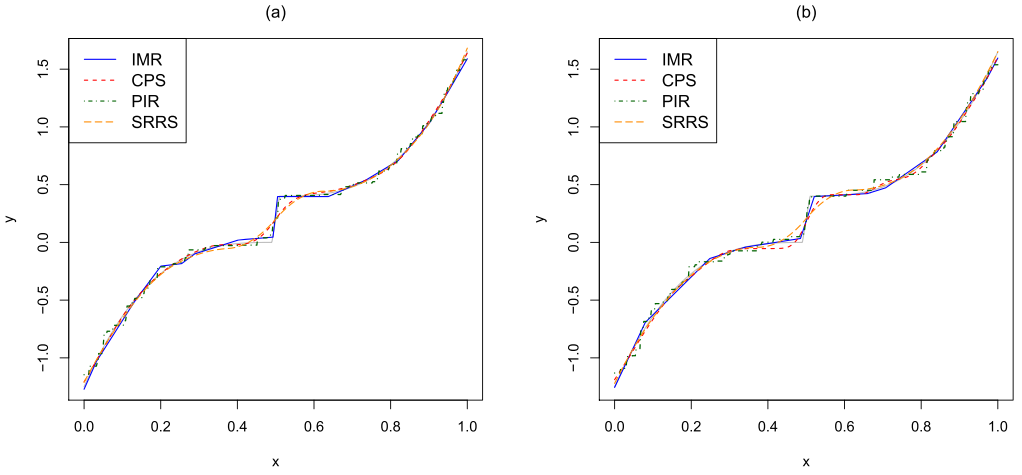

Figure 8. The IMR, CPS, PIR and SRRS estimates for f_4 corresponding to the median of the realized values of the (a) MSE and (b) MXDV from all 100 simulated datasets. The case of $N = 500$ and the normal errors is considered.

the local behavior of f_4. The PIR estimator appears to be fine in catching the local change, but the global performance is not as good as that of the IMR estimator. On the other hand, the IMR method estimates both global and local changes in the regression function. The simulation studies demonstrates that the finite sample performance of the IMR method is satisfactory for each setting and numerical results corresponding to Figures 4–7 are presented in Tables 1 and 2 of Appendix B.

In order to highlight the difference among the four methods, we include Figure 8 that displays each estimate of the four methods for f_4 with $N = 500$ and the normal errors. We pick up the IMR (blue solid), CPS (red dashed), PIR (green dot-dashed) and SRRS (orange long-dashed) estimates corresponding to the median of the realized values of the (a) MSE and (b) MXDV from 100 simulated datasets. The gray solid lines indicate the regression function f_4 which involves both smooth global trend and rapid sharp jump. We can see that the IMR estimates adapt well to the local level of smoothness without compromising the overall fits. This adaptabity demonstrates that the proposed knot selection technique is effective to capture the inhomogeneous behavior of the regression function, meanwhile, makes the estimate smooth enough to avoid overfitting. However, the CPS and SRRS method shows the over-smooth fits, which means they do not capture the sudden jump at 0.5. Even though the CPS method uses many knots so that enough knots are placed near the jump, the method has a trouble in capturing the sharp jump. The CPS method seems to lose the flexibility to catch sudden jumps in order to make its estimate smooth. Thus, if a true monotone regression functions has abrupt jumps at some points, the CPS method will not identify the changes and show a smooth fit only. One can see that the jump is captured by the PIR estimates, but the PIR method makes its estimate too wiggly to fit well to the smooth regions of the regression function.

4.2. The growth data

The growth data are analyzed in depth with the IMR method. Since it is typically assumed that human height is monotonically increasing, it is reasonable to apply the IMR method to

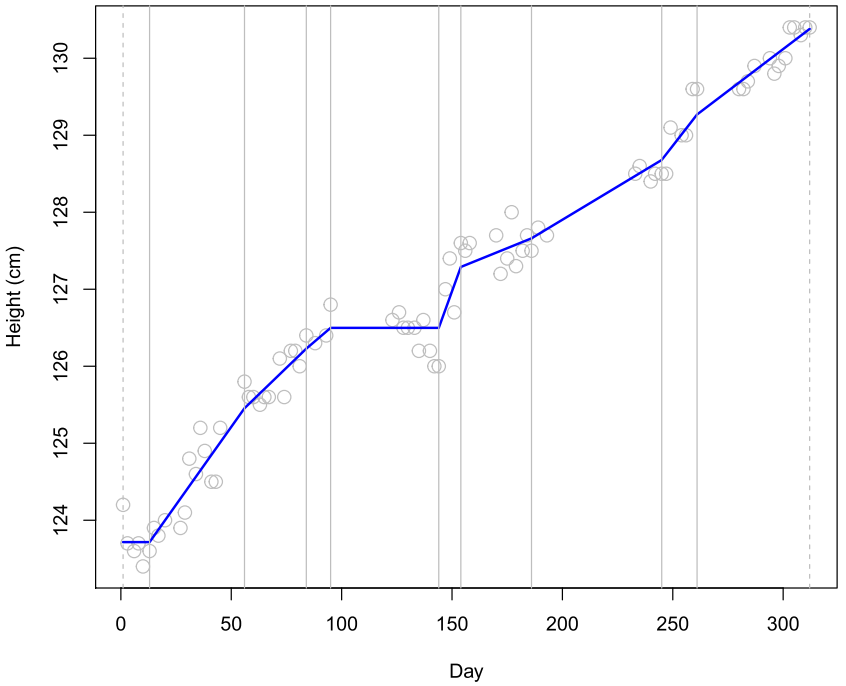

Figure 9. The optimal IMR estimate for the growth data. The blue line represents the optimal IMR estimate with $\hat{\lambda} = 3.36$ and the gray vertical lines indicate the location of active interior knots for the estimate. The dotted vertical lines show the boundary knots.

the growth data. Figure 9 displays the IMR estimate with the optimal complexity parameter $\hat{\lambda} = 3.36$. The gray vertical lines indicates 9 active interior knots of the IMR estimate. The growth of a single child seems to occur in short bursts rather than continuously, so the knot selection becomes crucial for the IMR method to fit the data well.

The IMR estimate seems to adapt quite well to the growth data over the entire range. It implies that the performance of the proposed knot selection technique is satisfactory. The IMR estimate shows evident jumps near days 150 and 250. Moreover, the monotone constraint on the shape of the regression function prevents the IMR estimate from overfitting. Although some data points fall off near days 10 and 150, the IMR estimate demonstrates the absence of meaningful changes before day 10, and between days 95 and 145.

5. Concluding remarks

This article has demonstrated that the I-splines are beneficial to develop the penalization methodology for monotone regression splines, since they simplify the optimization problem for the estimation of monotone regression functions.

As the implementation method, the coordinate descent algorithm has been developed to manage the non-negativity or non-positivity constraint on the I-spline coefficients and the non-seperable structure of the total variation penalty. In this method, the pruning process carries out the data-driven knot selection, guaranteeing convergence of the algorithm. The maximum candidate of the complexity parameter that removes all interior knots has been derived to alleviate the computational burden of complexity parameter selection. In

addition, if the number of the complexity parameter candidates is sufficiently large, all monotone linear spline estimates are considered in the algorithm. Numerical studies based on both simulated and real data have shown the satisfactory performance of the IMR method, especially when a regression function has sudden jumps.

The IMR estimator can be generalized by adopting the quadratic I-splines. The quadratic I-splines are defined by integrating the linear B-splines, and the imposition of non-negativity or non-positivity constraint on the coefficients produces a monotone quadratic spline. We also expect that the total variation penalty for the second derivative of the quadratic spline will lead to an effective data-driven knot selection. The monotone regression spline estimator using the quadratic I-splines and its implementation will be studied in our future work.

Funding

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2015R1D1A1A01057747).

Appendix A. Proofs of the propositions

Appendix A provides the proofs of Propositions 3.1 and 3.2. In the appendix, $\text{sign}(z)$ for $z \in \mathbb{R}$ is defined as the sub-differential $\partial|z|$ of the absolute function $|\cdot|$ at z, where

$$\text{sign}(z) \in \begin{cases} \{1\} & \text{if} \quad z > 0 \\ \{-1\} & \text{if} \quad z < 0 \\ [-1, 1] & \text{if} \quad z = 0. \end{cases}$$

A.1 Proof of Proposition 3.1

Without loss of generality, assume that c2 is greater than c_1. Since q^γ is the sum of the strictly convex function $\frac{1}{2}(z-m)^2$ and the convex function $\gamma p(z; c_1, c_2)$, q^γ is strictly convex. It follows that q^γ has the unique minimizer Q over $[0, \infty)$.

Observe that the sub-differential $\partial q^\gamma(z)$ of the function $q^\gamma(\cdot)$ at z is given by $(z-m) + \gamma\text{sign}(z-c_1) + \gamma\text{sign}(z-c_2)$. According to Rockafellar Rockafellar (1970), a point z^* satisfying $0 \in \partial q^\gamma(z^*)$ is the minimizer of q^γ over $\mathbb{R}$. Observe

$$z^* = \begin{cases} m-2\gamma & \text{if} \quad m \geq c_2 + 2\gamma \\ c_2 & \text{if} \quad c_2 \leq m < c_2 + 2\gamma \\ m & \text{if} \quad c_1 \leq m < c_2 \\ c_1 & \text{if} \quad c_1 - 2\gamma \leq m < c_1 \\ m + 2\gamma & \text{if} \quad m < c_1 - 2\gamma. \end{cases}$$

Since q^γ is minimized over $[0, \infty)$, the minimizer Q is given by

$$Q = \begin{cases} z^* & \text{if} \quad z^* \geq 0 \\ 0 & \text{if} \quad z^* < 0. \end{cases}$$

A.2. Proof of Proposition 3.2

The following proof is essentially identical to the proof of Theorem 2 in Jhong, Koo, and Lee (2017) except some notations, but we include this for the completeness.

Based on the centered I-spline basis, we can focus on estimating $\beta 1$ and $\beta 2$. Observe

$$\hat{\beta}_{-1}^{\lambda,\xi} = \left(\hat{\beta}_1^{\lambda,\xi}, \hat{\beta}_2^{\lambda,\xi}\right) = \underset{\beta_1, \beta_2 \geq 0}{\text{argmin}} \frac{1}{2} \|Y_c - G_c^\xi \beta_{-1}\|^2 + \lambda |d^\top \beta_{-1}|,$$

where $\beta_{-1} = (\beta_1, \beta_2)$. Referring to Luenberger and Ye (1984) for the definition of the Karush-Kuhn-Tucker (KKT) optimality condition, the KKT stationary condition is given by

$$0 = G_c^{\xi\top}\left(Y_c - G_c^{\xi}\hat{\beta}_{-1}^{\lambda,\xi}\right) - \lambda \operatorname{sign}\left(d^\top \hat{\beta}_{-1}^{\lambda,\xi}\right) d. \tag{9}$$

Choose any $\lambda \geq \lambda^{\xi}$. Suppose $d^\top \hat{\beta}_{-1}^{\lambda,\xi} \neq 0$. It follows from the KKT stationary condition (9) that

$$\hat{\beta}_{-1}^{\lambda,\xi} = \left(G_c^{\xi\top} G_c^{\xi}\right)^{-1} G_c^{\xi\top} Y_c - \lambda \operatorname{sign}\left(d^\top \hat{\beta}_{-1}^{\lambda,\xi}\right)\left(G_c^{\xi\top} G_c^{\xi}\right)^{-1} d.$$

Observe

$$\begin{aligned} 0 &< \operatorname{sign}\left(d^\top \hat{\beta}_{-1}^{\lambda,\xi}\right) d^\top \hat{\beta}_{-1}^{\lambda,\xi} \\ &= \operatorname{sign}\left(d^\top \hat{\beta}_{-1}^{\lambda,\xi}\right) d^\top \left(G_c^{\xi\top} G_c^{\xi}\right)^{-1} G_c^{\xi\top} Y_c - \lambda d^\top \left(G_c^{\xi\top} G_c^{\xi}\right)^{-1} d. \end{aligned}$$

The positive definiteness of $G_c^{\xi\top} G_c^{\xi}$ implies that

$$\lambda < \frac{|d^\top \left(G_c^{\xi\top} G_c^{\xi}\right)^{-1} G_c^{\xi\top} Y_c|}{d^\top \left(G_c^{\xi\top} G_c^{\xi}\right)^{-1} d} = \lambda^{\xi}.$$

This contradicts the assumption $\lambda \geq \lambda^{\xi}$, which implies the desired result.

Appendix B. Tables for the simulation results

Appendix B includes tables that show the numerical results of the simulation studies in Sec. 4.1.

Table 1. Average of each criterion ($\times 10^2$) over 100 runs of IMR, CPS, PIR and SRRS for f_1 and f_2 with its standard error in parentheses.

Function	Error	N	Method	MSE	MXDV
f_1	Normal	100	IMR	0.09 (0.005)	7.65 (0.240)
			PIR	0.20 (0.005)	13.20 (0.213)
			SRRS	0.11 (0.003)	8.86 (0.197)
			MPLS	0.10 (0.005)	8.92 (0.225)
		500	IMR	0.03 (0.001)	4.23 (0.051)
			PIR	0.07 (0.001)	8.87 (0.055)
			SRRS	0.05 (0.000)	7.47 (0.038)
			MPLS	0.03 (0.001)	6.70 (0.055)
	Mixture	100	IMR	0.18 (0.012)	10.19 (0.424)
			PIR	0.32 (0.011)	17.64 (0.500)
			SRRS	0.17 (0.008)	10.37 (0.312)
			MPLS	0.18 (0.010)	10.77 (0.427)
		500	IMR	0.04 (0.001)	5.58 (0.081)
			PIR	0.10 (0.001)	11.08 (0.095)
			SRRS	0.06 (0.001)	7.80 (0.053)
			MPLS	0.05 (0.001)	7.71 (0.068)
f_2	Normal	100	IMR	0.06 (0.003)	6.05 (0.229)
			PIR	0.21 (0.005)	12.04 (0.191)
			SRRS	0.07 (0.004)	6.41 (0.287)
			MPLS	0.05 (0.003)	5.26 (0.218)
		500	IMR	0.02 (0.000)	3.52 (0.064)
			PIR	0.07 (0.001)	7.74 (0.042)
			SRRS	0.02 (0.000)	3.89 (0.063)
			MPLS	0.01 (0.000)	3.23 (0.059)
	Mixture	100	IMR	0.10 (0.006)	7.66 (0.337)
			PIR	0.33 (0.011)	15.85 (0.466)
			SRRS	0.13 (0.009)	8.69 (0.419)
			MPLS	0.10 (0.008)	7.09 (0.336)
		500	IMR	0.03 (0.001)	4.96 (0.083)
			PIR	0.11 (0.001)	9.75 (0.092)
			SRRS	0.04 (0.001)	5.15 (0.092)
			MPLS	0.02 (0.001)	4.36 (0.074)

Table 2. Average of each criterion ($\times 10^2$) over 100 runs of IMR, CPS, PIR and SRRS for f_3 and f_4 with its standard error in parentheses.

Function	Error	N	Method	MSE	MXDV
f_3	Normal	100	IMR	0.10 (0.006)	10.04 (0.291)
			PIR	0.11 (0.004)	13.88 (0.390)
			SRRS	1.28 (0.003)	26.80 (0.122)
			MPLS	0.10 (0.006)	9.20 (0.316)
		500	IMR	0.03 (0.001)	5.81 (0.071)
			PIR	0.04 (0.000)	10.15 (0.084)
			SRRS	0.04 (0.000)	6.65 (0.018)
			MPLS	0.03 (0.001)	4.94 (0.064)
	Mixture	100	IMR	0.18 (0.011)	12.36 (0.661)
			PIR	0.18 (0.012)	11.23 (0.487)
			SRRS	1.31 (0.006)	27.26 (0.182)
			MPLS	0.17 (0.012)	11.03 (0.484)
		500	IMR	0.04 (0.001)	7.47 (0.170)
			PIR	0.06 (0.001)	12.35 (0.016)
			SRRS	0.05 (0.000)	6.99 (0.031)
			MPLS	0.04 (0.001)	6.44 (0.076)
f_4	Normal	100	IMR	0.21 (0.010)	15.79 (0.341)
			PIR	0.31 (0.006)	18.62 (0.373)
			SRRS	0.41 (0.004)	19.71 (0.153)
			MPLS	0.21 (0.007)	16.19 (0.274)
		500	IMR	0.05 (0.001)	10.77 (0.149)
			PIR	0.10 (0.001)	12.18 (0.068)
			SRRS	0.13 (0.000)	16.49 (0.022)
			MPLS	0.10 (0.001)	15.49 (0.053)
	Mixture	100	IMR	0.36 (0.017)	18.23 (0.523)
			PIR	0.47 (0.013)	23.05 (0.599)
			SRRS	0.48 (0.009)	20.29 (0.237)
			MPLS	0.33 (0.013)	17.88 (0.401)
		500	IMR	0.09 (0.002)	13.52 (0.187)
			PIR	0.14 (0.001)	15.17 (0.168)
			SRRS	0.15 (0.001)	16.73 (0.027)
			MPLS	0.14 (0.002)	16.99 (0.056)

References

Barlow, R., D. Bartholomew, and J. Bremner. 1972. *Statistical inference under order restrictions*. New York: Wiley.

Breiman, L., and S. Peters. 1992. Comparing automatic smoothers (a public service enterprise). *International Statistical Review/Revue Internationale de Statistique* 60:271–90. doi:10.2307/1403679.

Brunk, H. D. 1955. Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics* 26 (4):607–16. doi:10.1214/aoms/1177728420.

Cole, T. J., J. V. Freeman, and M. A. Preece. 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine* 17 (4):407–29. doi:10.1002/(SICI)1097-0258(19980228)17:4<407::AID-SIM742>3.3.CO;2-C.

Durbán, M., J. Harezlak, M. Wand, and R. Carroll. 2005. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* 24 (8):1153–67. doi:10.1002/sim.1991.

Eilers, P. H., and B. D. Marx. 1996. Flexible smoothing with b-splines and penalties. *Statistical Science* 11 (2):89–102. doi:10.1214/ss/1038425655.

Friedman, J., T. Hastie, H. Höfling, R. Tibshirani., et al. 2007. Pathwise coordinate optimization. *The Annals of Applied Statistics* 1 (2):302–32. doi:10.1214/07-AOAS131.

Friedman, J., and R. Tibshirani. 1984. The monotone smoothing of scatterplots. *Technometrics* 26 (3):243–50. doi:10.1080/00401706.1984.10487961.

Hall, P., and L.-S. Huang. 2001. Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics* 29:624–47. doi:10.1214/aos/1009210683.

Jhong, J.-H., J.-Y. Koo, and S.-W. Lee. 2017. Penalized b-spline estimator for regression functions using total variation penalty. *Journal of Statistical Planning and Inference* 184 :77–93. doi:10.1016/j.jspi.2016.12.003.

Kelly, C., and J. Rice. 1990. Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics* 46 (4):1071–85. doi:10.2307/2532449.

Luenberger, D. G., and Y. Ye. 1984. *Linear and nonlinear programming*. vol. 2. New York: Springer.

Mammen, E. 1991. Estimating a smooth monotone regression function. *The Annals of Statistics* 19 (2):724–40. doi:10.1214/aos/1176348117.

Mammen, E., J. Marron, B. Turlach, and M. Wand. 2001. A general projection framework for constrained smoothing. *Statistical Science* 16 (3):232–48. doi:10.1214/ss/1009213727.

Mammen, E., and C. Thomas-Agnan. 1999. Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics* 26 (2):239–52. doi:10.1111/1467-9469.00147.

Mammen, E., and S. van de Geer. 1997. Locally adaptive regression splines. *The Annals of Statistics* 25 (1):387–413. doi:10.1214/aos/1034276635.

Meyer, M. C. 2008. Inference using shape-restricted regression splines. *The Annals of Applied Statistics* 2 (3):1013–33. doi:10.1214/08-AOAS167.

Meyer, M. C. 2012. Constrained penalized splines. *Canadian Journal of Statistics* 40 (1):190–206. doi:10.1002/cjs.10137.

Osborne, M. R., B. Presnell, and B. A. Turlach. 1998. Knot selection for regression splines via the LASSO. *Computing Science and Statistics* 30: 44–9.

Osborne, M. R., B. Presnell, and B. A. Turlach. 2000. On the lasso and its dual. *Journal of Computational and Graphical Statistics* 9 (2):319–37. doi:10.2307/1390657.

Ramsay, J. O. 1988. Monotone regression splines in action. *Statistical Science* 3 (4):425–41. doi:10.1214/ss/1177012761.

Ramsay, J. O. 1998. Estimating smooth monotone functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60 (2):365–75. doi:10.1111/1467-9868.00130.

Robertson, T., F. Wright, and R. Dykstra. 1988. Order restricted statistical inference. In *Probability and mathematical statistics*, eds. Robertson, T., F. T. Wright, and R. L. Dykstra. Chichester: John Wiley and Sons.

Rockafellar, R. T. 1970. *Convex analysis*. vol. 28. Princeton: Princeton University Press.

Thalange, N., P. Foster, M. Gill, D. Price, and P. Clayton. 1996. Model of normal prepubertal growth. *Archives of Disease in Childhood* 75 (5):427–31. doi:10.1136/adc.75.5.427.

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (1): 91–108. doi:10.1111/j.1467-9868.2005.00490.x.

Tillmann, V., P. Foster, M. Gill, D. P. Rice, and P. Clayton. 2002. Short-term growth in children with growth disorders. *Annals of Human Biology* 29 (1):89–104.

Wang, X., and F. Li. 2008. Isotonic smoothing spline regression. *Journal of Computational and Graphical Statistics* 17 (1):21–37. doi:10.1198/106186008X285627.

Wu, J., M. C. Meyer, and J. D. Opsomer. 2015. Penalized isotonic regression. *Journal of Statistical Planning and Inference* 161:12–24. doi:10.1016/j.jspi.2014.12.008.