Journal of Statistical Planning and Inference 184 (2017) 77–93

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Penalized B-spline estimator for regression functions using total variation penalty

Jae-Hwan Jhong [a], Ja-Yong Koo [a,*], Seong-Whan Lee [b]

[a] Department of Statistics, Korea University, Seoul 136-701, Republic of Korea

[b] Department of Brain and Cognitive Engineering, Korea University, Seoul 136-701, Republic of Korea

ARTICLE INFO

Article history:
Received 5 May 2016
Received in revised form 23 September 2016
Accepted 17 December 2016
Available online 29 December 2016

Keywords:
Adaptive estimation
Coordinate descent algorithm
LASSO
Oracle inequalities
Penalized least squares

ABSTRACT

We carry out a study on a penalized regression spline estimator with total variation penalty. In order to provide a spatially adaptive method, we consider total variation penalty for the estimating regression function. This paper adopts B-splines for both numerical implementation and asymptotic analysis because they have small supports, so the information matrices are sparse and banded. Once we express the estimator with a linear combination of B-splines, the coefficients are estimated by minimizing a penalized residual sum of squares. A new coordinate descent algorithm is introduced to handle total variation penalty determined by the B-spline coefficients. For large-sample inference, a nonasymptotic oracle inequality for penalized B-spline estimators is obtained. The oracle inequality is then used to show that the estimator is an optimal adaptive for the estimation of the regression function up to a logarithm factor.

1. Introduction

A common problem in nonparametric regression is providing an estimator of a regression function based on a sample from the distribution of a noisy response variable without assuming a particular functional form for the regression function. The literature on nonparametric regression is rich and diverse, and there are many methods for estimating regression functions given observations from regression models. Popular methods include local polynomials, kernels, splines, sieves, and wavelets; see Green and Silverman (1993), Fan and Gijbels (1996), Efromovich (1999), and Tsybakov (2009).

A closely related problem is high dimensional regression. To analyze high-dimensional data, various selection methods for variables have been considered by many researchers. Chen and Donoho (1994) propose Basis Pursuit, and Tibshirani (1996) introduces LASSO (Least Absolute Shrinkage and Selection Operator). Zou and Hastie (2005) consider Elastic Net, Fan and Li (2001) adopt non-convex penalties such as SCAD (Smoothly Clipped Absolute Deviation), and Zhang (2010) uses MCP (Minimax Concave Penalty). Zou (2006) shows that the adaptive LASSO enjoys oracle properties.

For regression spline problems, knot selection involves the choice of proper knots to estimate an underlying function. Knot selection is not unlike variable selection in linear regression. Osborne et al. (1998) propose an algorithm for knot selection for regression splines via the LASSO. Mammen and Geer (1997) consider least squares penalized regression estimators with total variation penalties and show that locally adaptive regression splines converge to the underlying function at the minimax rate. Rosset and Zhu (2007) consider a generic regularized optimization problem with LASSO penalty and apply the method to the locally adaptive regression splines of Mammen and Geer (1997). They use truncated power

* Corresponding author.
E-mail address: jykoo@korea.ac.kr (J.-Y. Koo).

http://dx.doi.org/10.1016/j.jspi.2016.12.003

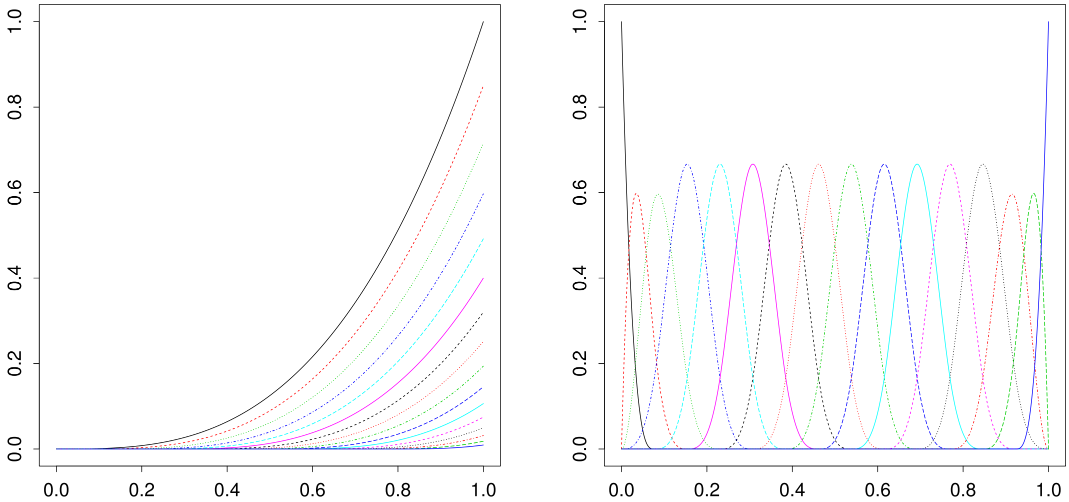

Fig. 1. For cubic splines, corresponding to order of splines $r = 4$, the left plot shows truncated power basis functions and the right plot displays B-splines.

splines to solve the penalized optimization problem and it can be represented as a LASSO problem. Harchaoui and Levy-Leduc (2010) prove nonasymptotic risk bounds for function estimator with total variation penalty for piecewise constant functions. More recently, Dalalyan et al. (2014) show nonasymptotic oracle prediction inequalities for the fused lasso (Tibshirani and Saunders, 2005) and piecewise constant functions.

There are mainly two kinds of basis function used to represent the splines for a given set of knots: B-splines and truncated power basis splines. The B-spline basis functions provide a numerically superior alternative basis to the truncated power basis (de Boor, 2001). Since the representation of truncated power basis splines allows for easier statistical interpretation of coefficients, many papers adopt the truncated power basis splines (Friedman, 1991; Hansen et al., 1997; Koo, 1997). Osborne et al. (1998) give an algorithm for the knot selection for regression splines via the LASSO based on truncated power basis functions. Trend filtering (Tibshirani, 2014) can be represented as a LASSO problem with basis functions, which are closed to truncated power basis. Fig. 1 shows the shape of cubic truncated power basis (left) and B-spline basis functions (right). Because predictor matrix for truncated power basis is dense, several computational problems may occur when the number of knots becomes large. B-splines have small supports so that the predictor matrix is sparse and the information matrix is banded.

In this paper, we carry out a study on both numerical implementation and nonasymptotic oracle property of penalized B-spline estimators of any order with total variation penalty. Once we express the estimator with a linear combination of B-splines, the coefficients are estimated by minimizing a penalized residual sum of squares. B-splines are adopted for both numerical implementation and asymptotic analysis because they have small supports, so the information matrices are sparse and banded, whereas the truncated power basis leads to dense information matrices that can lead to several computational problems for large sample sizes. A new coordinate descent algorithm is introduced to handle total variation penalty given by the B-spline coefficients. The proposed updating formula does not resort to Karush–Kuhn–Tucker (KKT) optimality conditions (Luenberger, 1984), which are mathematically demanding. For large-sample inference, a nonasymptotic oracle inequality for penalized B-spline estimators of any order is obtained. The oracle inequality is then used to show that the estimators are optimal adaptive in the minimax sense for the estimation of the regression function up to a logarithm factor. Using B-splines of arbitrary order for both implementation and theory appears to be new in penalized spline estimation of regression functions.

The outline of the paper is as follows. In Section 2, we define penalized B-spline regression estimators based total variation penalty. An updating method based on a coordinate descent algorithm and its implementation are described in Section 3. In Section 4, we illustrate the performance of the proposed method using both simulated and real example. An oracle inequality and adaptivity results are given in Section 5. All proofs are provided in Section 6. Section 7 summarizes the conclusions of the paper.

2. Total variation regression spline estimator

Consider the observations $(x_1, y_1), \ldots, (x_N, y_N)$ from the regression model

$$y_i = f(x_i) + \varepsilon_i, \quad \text{for } i = 1, \ldots, N, \tag{2.1}$$

where $y_i \in \mathbb{R}$ are responses, $x_i \in \mathcal{I}$ are fixed input points, and ε_i are independent errors with mean $\mathbb{E}(\varepsilon_i) = 0$ and variance $\mathrm{Var}(\varepsilon_i) = \sigma^2 > 0$ for $i = 1, \ldots, N$. It is assumed that the domain $\mathcal{I}$ is the compact unit interval $[0, 1]$. Given data $(x_1, y_1), \ldots, (x_N, y_N)$, our goal is to estimate f based on regression splines with a penalty.