A Data-Adaptive Knot Selection Scheme for Fitting Splines

Xuming He , Lixin Shen, and Zuowei Shen

Abstract— **A critical component of spline smoothing is the choice of knots, especially for curves with varying shapes and frequencies in its domain. We consider a two-stage knot selection scheme for adaptively fitting splines to data subject to noise. A potential set of knots is chosen based on information from certain wavelet decompositions with the intention to place more points where the curve shows rapid changes. The final knot selection is then made based on statistical model selection ideas. We show that the proposed method is well suited for a variety of smoothing and noise filtering needs.**

Index Terms: Least squares; Model selection; Knot; Smoothing; Spline; Wavelet decomposition.

I. Introduction

The task of fitting a spline to noisy data is often performed in diverse domains of applications such as computer aided design, pattern recognition, data smoothing and denoising in engineering. Suppose that n pairs of observations $\{(x_i, y_i),\ \ i = 1, 2, \ldots, n\}$ are available from $y_i = f(x_i) + e_i$ $(i = 1, \ldots, n)$, where f is a smooth function to be approximated by a spline of order $k+1$, and e_i are random noise with mean 0. The function $f(x)$ is the conditional mean of y given x. In this paper we consider regression splines with free knots with the purpose of fitting curves that can better adapt to locality than smoothing splines ([4]) and yet retain the desirable properties of splines.

Assume that x_i's are contained in a finite interval $[a, b]$. If the set of knots for the spline of order $k+1$ is given as $T_m = \{a = t_0 \le t_1 \le \cdots \le t_m \le t_{m+1} = b\}$, then the splines form a linear space with $m+k+1$ dimensions. Let $B_j(x)$ $(j = 1, 2, \cdots, m+k+1)$ be any basis function so that we seek $f(x) = \sum_{j=1}^{m+k+1} c_j B_j(x)$ such that $Obj(T_m) = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{m+k+1} c_j B_j(x_i))^2$ is minimized. This is the well-known least squares principle. The coefficients c_j can be obtained easily by solving a linear system. The main question we consider is how to find T_m, the set of knots.

We consider a data-adaptive scheme to construct a set of potential knots based on a very simple idea that more knots are needed when there are faster changes in the curve. We use wavelets decomposition to find a subset of observed x_i's where faster changes in th curve are likely to occur. This is followed by a stepwise knot deletion scheme to select knots for final use. Through the two-stage process, we aim to overcome the problems that could result from the convenient choice of "uniform" pool of knots. We include a comparison with the cubic smoothing splines and with the wavelet thresholding and reconstruction method ([3]). For the types of problems we consider, the wavelet thresholding and reconstruction method tends to underperform the splines.

The rest of the paper is organized as follows. In Section 2, we explore the idea of wavelets decomposition for preselection of potential knots and a model selection method for final automated knot selection. A specific implementation of the proposed method is given in Section 3 with a simulation study to compare the proposed method with several well-known existing methods of curve fitting. We conclude the paper with some additional comments on spline fitting in Section 4.

Research is partially supported by the Wavelets Strategic Research Program at the National University of Singapore. Xuming He (Corresponding author) is with the Department of Statistics, University of Illinois, USA (email:he@stat.uiuc.edu). Lixin Shen and Zuowei Shen are with the Department of Mathematics, National University of Singapore, Singapore. (matslx@nus.edu.sg, matzuows@math.nus.edu.sg). The authors thank referees for helpful comments.

II. Knot Selection

A. Pre-selection of knots

Our purpose is to find a set of potential knots that mark the major trend as well as detailed features of the curve. The basic idea of the wavelet decomposition algorithm is to pass the data through a pair of high and low pass filters. The output from the low pass filter will pass through the next pair of filters. The low pass filter is designed to de-noise and the high pass filter retains local details in the data. The information passing through the high pass filter provides certain order of differences of the original data values and can be used to help determine the positions where there are jumps in the curve or its higher-order derivatives. The order of differences obtained through the high pass filter depends on the choice of the filter itself.

In the usual wavelet applications the decomposition algorithm is often bundled with compression and reconstruction. Donoho and Johnstone [3] use thresholding to dampen the noise effect in the reconstruction. In this paper the wavelet decomposition ideas are used only to seek out a set of potential knots. They are followed by statistical model selection which identifies the final set of the knots.

The specific wavelets used in the paper are Haar wavelet and cubic spline wavelet frame (see [1] and [2]). The former leads to simple averaging as low pass and the first-order difference as high pass. The latter leads to local smoothing as low pass and the fourth-order difference as high pass. After the data pass through such filters, we take the local extrema as potential knots; see Section 3 for a specific implementation.

B. Final Knot Selection

Given a set of potential knots $\{a = t_0 < t_1 < \cdots < t_m < t_{m+1} = b\}$ from the preceeding subsection, we now see the parallel between knot selection and model selection in the context of linear regression. One more knot corresponds to a larger model. Knot selection is more conveniently done with the use of the truncated power basis functions for the splines because each knot is represented by one basis function. For a fixed order of polynomials, the basis functions are

$$\pi = \{1,\ x,\ \cdots,\ x^k,\ (x-t_1)_+^k,\ \cdots,\ (x-t_m)_+^k\}$$

where $x_+^k = \max\{x^k, 0\}$. We then regress y on all the "variables" in the basis π and perform variable selection. We focus on backward elimination for its simplicity and computation efficiency.

When regressing a response variable y on a set of basis functions π, the well-known least squares estimate of the projection yields the fitted value $\hat{y}_i = \pi(x_i)^T(\sum_{j=1}^n \pi(x_j)\pi(x_j)^T)^{-1}\sum_{j=1}^n \pi(x_j)y_j$, where $\pi(x_i)$ is the column vector π evaluated at $x = x_i$, and the residual sum of squares is $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. Backward elimination starts from the full model, and at each step, removes one basis function that gives the smallest increase in the residual sum of squares. The usual practice in statistics is to stop the elimination process when all remaining basis functions are statistically significant, see [5]. Not to miss potentially smaller models, we continue the elimination process to get a collection of nested models from the largest (with m internal knot) to the smallest (no internal knots).

We consider using the Schwarz information criterion

$$SIC_q = \log(RSS_q) + \log(n)q/(2n)$$

to evaluate the competing models, where q is the dimension of the model and RSS_q is the residual sum of squares of the model being considered. For cubic splines, q is equal to four plus the number of internal knots. The algorithm for the final knot selection goes as follows.

Step 0: start from the full model with, say, m internal knots. Let $q = m + k + 1$ where $k + 1$ is the order of splines used.

Step 1: Sequentially find the best model with backward elimination:

(1a) Remove the i-th knot from the current model and re-evaluate the residual sum of squares RSS_{-i}, where $i = 1, 2, ...$

(1b) Choose the model with the smallest RSS_{-i} from all those considered in (1a) and then replace q by $q - 1$.

(1c) Compute SIC_q for the model chosen in (1b).

(1d) Go back to step (1a) until $q = m + 1$.

Step 2: Choose the final set of knots corresponding to the model chosen in step 1 with the smallest SIC_q.

III. Implementation and Simulation

The wavelet decomposition ideas discussed in Section 2 offer some flexibility in choosing the resolution level and the number of potential knots. In general, the number of decompositions and the number of knots should increase with the size of the data set at a slow rate.

To assess the performance of the proposed method for knot selection, we give a specific set of parameters in our implementation later referred to as KNOT. Our choices of the parameters can serve as a guideline for curve fitting problems with modest data size. The algorithm consists of four steps:

(1). The data pass through a high pass filter corresponding to the Haar wavelet. The locations for the N_1 largest local maxima and N_1 smallest local minima are selected. $N_1 = 10$ is used in study.

(2). The data pass through a low pass filter in the form of $\frac{1}{16}(y_{i-2} + 4y_{i-1} + 6y_i + 4y_{i+1} + y_{i+2})$ L times. Then the third order differences in the form of $\frac{1}{16}(y_{i-2} - 4y_{i-1} + 6y_i - 4y_{i+1} + y_{i+2})$ are computed from which the locations of the N_2 largest local maxima and N_2 smallest local minima are found. $N_2 = 20$ is used in our study here with $L = 2$.

(3). The locations found in steps (1) and (2) are combined. Then a total number of $N + 1$ locations taken as the i/N-th quantile ($i = 0, 1, ..., N$) are selected to produce the set of potential knots.

(4). Final knot selection is performed using the method described in Section 2B with $k = 3$.

The choices of N_1, N_2 and N are mainly subjective. An examination of the scatter plots are helpful. If you see frequent changes in the underlying curve, large values are called for.

In our simulation study, we generate data from the following regression model $y_i = f(x_i) + e_i$, where e_i's are normally distributed with mean 0 and variance σ^2. The main variants in our study reported here are the shape of f and the level of noise as measured by σ^2.

For each given function f on $[a, b]$ and the variance σ^2, we generate $M = 500$ samples of size n. For each sample, an estimate $\hat{f}_n$ is obtained and the following root mean squared error $RMSE = \{n^{-1}\sum_{i=1}^n (\hat{f}_n(x_i) - f(x_i))^2\}^{1/2}$ is computed. The average $RMSE$ and its standard error are then obtained from the M trials. The results are summarized in Table 1 for four different cases. The four test functions and the design points used in this section are

1. $f(x) = \cos(5x + \pi/8)$, $\quad x \in [0, \pi]$, $x_i = i\pi/128$, $1 \le i \le 128$.

2. $f(x) = \begin{cases} \frac{(a-2)(x+a)}{a} - a, & -a \le x < 0; \\ \frac{(a+2)(x-a)}{a} + a, & 0 \le x < a; \\ x, & \text{otherwise,} \end{cases}$

where $a = 0.1$, $x \in [-1, 1]$ and $x_i = i/64 - 1$, $1 \le i \le 128$.

3. $f(x) = 1/x$, $\quad x \in (0, 4]$, $x_i = i/32$, $1 \le i \le 128$.

4. $f(x) = \sin(3/x)$, $x_i = 0.15 + |z_i|$ $(i = 1, \cdots, 256)$, where z_i is a random sample from the standard normal distribution.

Case 1 is a curve with rather even changes. Case 2 is not everywhere differentiable with a deep and narrow valley. Case 3 has fast varying derivatives for x near 0. Case 4 has some high frequency signals for x near 0.

The four methods of investigation in our comparison are

UNIF: the cubic regression spline starting from a set of N uniform knots on $[a, b]$ followed by knot selection of Section 2B;

KNOT: our cubic regression spline as detailed in Section 3;

WAVE: the wavelet thresholding and reconstruction method of Donoho and Johnstone (with the symmetry option for boundaries);

SMOO: the cubic smoothing spline as implemented in Splus. The smoothing parameter is chosen by the generalized cross-validation method.

In both UNIF and KNOT methods, we used a set of potential knots with $N = 20$ and the same stepwise knot selection procedure follows so the comparison is purely on the pre-selection of knots. The average root-mean-squared-errors and their standard errors are given in Table 1 for two different noise levels for each test function.

Our empirical comparisons show that the advantage of pre-selection of knots over UNIF is more evident in cases 3 and 4 where the spatial variation of the curves is higher. There are some cases where the root mean squared errors from KNOT are slightly higher than from UNIF. This is because of somewhat higher variability in the pre-selection of knots using the KNOT method, especially when the noise level is high.

It is clear from the comparison that WAVE is not designed for the type of smoothing problems under consideration. SMOO is a main competitor for KNOT. These two methods show rather similar overall degree of accuracy for estimating the underlying curves in three cases 1,2 and 4. However, in case 3 our proposed method KNOT has a definite advantage over SMOO. This is because the smoothing spline has to undersmooth the curve for most values of x in order to catch the sharp spike for x near 0.

We further note that the root-mean-squared-error is only one summary measure of performance. It does not tell the whole story. We find that both WAVE and SMOO tend to have small zigzags in the fitted curves. Figure 1 compares the fitted curves with one simulated data under Case 4.

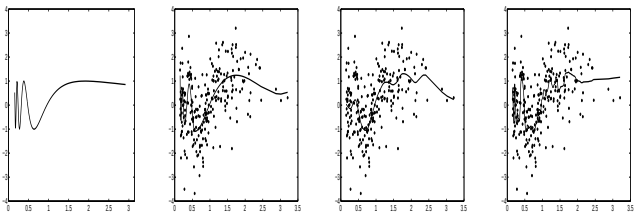

Fig. 1. Comparison of three different methods under Case 4 (from left to right: true curve, KNOT, SMOO, and WAVE)

IV. Concluding Remarks

We propose a two-step knot selection method for fitting spline to noisy data: a pre-selection of knots by identifying potential change points using wavelet decomposition ideas followed by basis selection via a statistical model selection criterion. The combination of the two provides a means to better adapt to the location-dependent shapes and curvatures of the underlying curve than regression splines with uniform knots and the globally adjusted smoothing splines. As with other smoothing methods, some tuning constants such as the set size of potential knots need to be chosen. The procedure, however, is computationally simple and stable. Since the fitted curve is a spline with typically a small number of knots, a compact global model representation is available and convenient for some further analyses.

We recommend, based on our study, that the proposed method KNOT be used for curve fitting in cases where one expects faster changes of the target curve in certain areas than in others with a relatively high signal-to-noise ratio.

Comparisons are made with the wavelet thresholding and reconstruction methods. Splines generally have better performance for curve smoothing, but this by no means implies inferiority of the wavelet method in statistical applications. Each method has its distinctive features that are useful for specific purposes. We hope that the present paper would support this line of our reasoning.

		UNIF		KNOT	
Cases	σ	RMSE	STD	RMSE	STD
1	0.1	0.0400	0.0003	0.0403	0.0003
	0.2	0.0766	0.0006	0.0775	0.0006
2	0.1	0.0647	0.0002	0.0519	0.0005
	0.2	0.0919	0.0005	0.1014	0.0012
3	1	0.5074	0.0024	0.3458	0.0040
	2	0.7261	0.0052	0.6478	0.0083
4	0.05	0.1142	0.0003	0.0296	0.0007
	1	0.2684	0.0024	0.2892	0.0025
		WAVE		SMOO	
Cases	σ	RMSE	STD	RMSE	STD
1	0.1	0.0597	0.0004	0.0360	0.0003
	0.2	0.1028	0.0006	0.0651	0.0006
2	0.1	0.0891	0.0004	0.0559	0.0003
	0.2	0.1517	0.0008	0.0955	0.0005
3	1	1.1615	0.0064	0.6518	0.0026
	2	2.0269	0.0109	1.1111	0.0052
4	0.05	0.0533	0.0002	0.0292	0.0001
	1	0.3773	0.0019	0.3616	0.0016

TABLE I
Estimated RMSE and standard errors for the four test cases.

References

[1] I. Daubechies, *Ten Lectures on Wavelets.* CBMS-NSF Reg. Conf. Series on App. Math., vol. 61, SIAM, Philadelphia, PA. 1992.

[2] I. Daubechies, B. Han, A. Ron and Z. Shen, Framelets: MRA-based constructions of wavelet frames, preprint (2000).

[3] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425-455, 1994.

[4] G. Wahba, *Spline models for observational data.* Society for Industrial and Applied Mathematics, 1990.

[5] A. J. Miller, *Subset Selection in Regression.* Monographs on Statistics and Applied Probability 40, London: Chapman and Hall, 1990.