Comparison of Significant Approaches of Penalized Spline Regression (P-splines)

Saira Sharif

College of Statistical and Actuarial Sciences - University of the Punjab, Pakistan saira.sharif@ucp.edu.pk

Shahid Kamal

College of Statistical and Actuarial Sciences - University of the Punjab, Pakistan kamal_shahid@yahoo.com

Abstract

Over the last two decades P-Splines have become a popular modeling tool in a wide class of statistical contexts. Fundamentally, semiparametric regression methods combine the leads of parametric and nonparametric approaches to regression analysis, while in precise, penalized spline regression uses the knowledge of nonparametric spline smoothing as a generalization of smoothing splines that let more suppleness in a choice of model with respect to the basis functions and the penalty. The present article compares two significant approaches of penalized spline regression models named as p-splines based on different basis functions with numerous knot selections and various types of penalites. These model fits have been applied on Wood Strength data to compare by calculating nonlinear least square method; also approaches are compared on several aspects: numerical immovability, quality of fit, derivative estimation and smoothing. This comparison will help us to fit best suitable model for conforming best suitable conditions and scenarios.

Keywords: Penalized Splines, B-splines Basis, Truncated Power Basis, Ridge Penalty and Difference Penalty.

1. Introduction

Up until the mid-1990s most literature on spline-based nonparametric regression was concerned with smoothing splines and their multivariate extension, where the penalty takes a particular form and the number of basis functions roughly equals the sample size. P-splines are keen methods for modeling nonlinear smooth effects of covariates within the additive models framework. It's a great approach of smoothing in handling a wide range of nonparametric and semi-parametric modeling situations. These are the regressions which uses splines, or piecewise continuous polynomials, paired with a mathematical penalization to find the best fit to the data. P-splines are generally regression splines fit by least-squares with a roughness penalty. It is very flexible idea that can be seen as a generalization of spline smoothing with a more flexible choice of bases, penalties and knots.

Namely, one chooses a spline basis based on some sufficiently large set of knots and penalizes unnecessary structure. Different basis functions, form of the penalties, amount and location of knots all provide a wide spectrum of smoothers. In husk, penalized regression splines as simple smoothing model is fitted by usurping

$$\beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{j=1}^{\kappa} b_j (x - t_j)_+^2 = B(x)\theta$$

Pak.j.stat.oper.res. Vol.XIV No.2 2018 pp289-303

as an average function for the knots $t_1, t_2, ..., t_j$ with large k. A combine recipe is to take high dimensional basis B(x) by choosing open-handedly large number of knots. Then to minimize the criterion $(y - B\theta)'(y - B\theta) + \lambda\theta'D\theta$ with sufficiently chosen D as penalty matrix and data driven penalty parameter λ . However, in recent years, there has been a great deal of research on more general spline/penalty strategies. Besides for selecting the number of knots, Eilers & Marx (1996) suggest to use a generous numbers, while Rupert (2002) suggests min (40, n/4) knots and Wood (2006) proposes k = 10 *3(m-1) with m as dimension of covariates. Although practical significance has two selection flairs; one to choose adequate number of knots to fit the model and increase the numbers until the likelihood does not further increase and secondly choose a large k and decrease it slightly to check that the likelihood does not decrease. Whereas for Model selection Wager, Vaida & Kauermann (2007) and Kneib & Greven (2010) suggested automatic correction of degree of freedom.

In 1986 O'Sullivan introduced a class of penalized splines that is a direct generalization of smoothing splines based on B-spline basis functions provided by de Boor (1978). After that Eilers & Marx (1992; 1996) again raised the issue. In 1996, they introduced a modified version of O'Sullivan's estimator and coined the term P-splines. This P-spline combines regression on B-spline basis, with an equidistant grid of knots and a discrete roughness penalty, i.e., higher-order differences in the penalty to smooth scatterplots. They proposed to use a relatively large number of knots and a difference penalty on the coefficients of adjacent B-splines. While Claeskens, Krivobokova & Opsomer (2009) and Li & Ruppert (2008) presented that if generous no. of knots are chosen then penalty asymptotically dominates and Kauermann, Krivobokova & Fahrmeir (2009) revealed that if moderate no. of knots are chosen then there is no asymptotically influence at penalty. Setup presented by Eilers & Marx (1996) is easy to use, and allows great flexibility, that any order of penalty can be combined with any order of the B-spline basis.

Subsequently Ruppert & Carroll (1997) used the truncated power basis (TPB) to represent the components as penalized splines with smoothness controlled by a ridge type penalty over the coefficients of the parametrization. Rather than placing the knots on an equidistant grid, they chose to use quantiles of the observations for the variable X as knot locations. Later Wand (1999) derived an asymptotic approximation of P-splines mean squared error, when penalties are used in combination with fixed basis. Ruppert & Carroll (2000) and Ruppert (2002) provide a strong support to Penalized splines.

Afterwards, Ruppert, Wand & Carroll (2003) present an excellent overview of theory and applications of semi-parametric models based on penalized splines and named their approach as P-splines too. They proposed the use of P-splines as a low rank smoother that can be seen as a compromise between smoothing and regression splines. Ruppert, Wand & Carroll (2003) used truncated power functions in the basis with quantile based knots of the independent variable and a ridge penalty. P-splines are becoming more and more popular and achieved general recognition since Eilers & Marx (1996) and Ruppert, Wand & Carroll (2003) introduced it.

2. Methodology

This article is presented to compare and document strengths and weaknesses of the two approaches of P-splines, one of Eilers and Marx (1996) which uses B-spline basis, equally-spaced knots and difference penalties and second is Ruppert, Wand and Carroll (2003), which uses a truncated power basis, knots based on quantiles of the independent variable and a ridge penalty. Here is a great collaboration to make well-informed choices for what type of splines should be used, how knots should be spaced and what penalty should be chosen. This article is organized as follows: section 1 give details about spline models of both approaches and compares their components. Section 2 compares both approaches of P-splines by example using wood strength data.

Table 3.1: Wood Strength data

Concentration	1.0	1.5	2	3	4.0	4.5	5.0	5.5	6.0	6.5	7	8.0	9.0	10	11	12	13	14	15
Strength	6.3	11.1	20	24	26.1	30.0	33.8	34.0	38.1	39.9	42	46.1	53.1	52	52.5	48	42.8	27.8	21.9

Here variable "Concentration" is used as explanatory variable while variable "Strength" as dependent variable.

And in Section 3 in order to gain more insight into the practicability and the limitations of approaches both approaches are discussed in some aspects, including: numerical immovability, quality of the fit & visualization, derivative estimation and smoothing. In present article, we will call Eilers and Marx (1996) approach as E&M P-Spline, while Ruppert, Wand and Carroll (2003) approach as RWC P-spline. The methodology of this article is implemented in R-language.

3. Spline Models

3.1 Model 1: (E&M P-Spline)

As discussed above E&M P-Splines combine regression based on B-splines basis with equally spaced grid of knots and used discrete roughness higher-order finite differences as penalty to smooth scatter plots. Mathematically, Eilers & Marx (1996) presented penalized spline model is:

$$S(X) = \sum_{i=1}^{m} \left\{ y_i - \sum_{j=1}^{n} \hat{\alpha}_j B_j(X) \right\}^2 + \gamma \sum_{j=k+1}^{n} \left(\Delta^k \hat{\alpha}_j \right)^2$$
 Model 1

Whereas penalty proposal based on finite differences i.e., $\Delta^k \hat{\alpha}_j$

If suppose $D_k = \Delta^k \alpha$, then the matrix D_k is the kth differences of α , and $\lambda > 0$. Where

$$\hat{\alpha} = \mathbf{B}(\mathbf{B}'\mathbf{B} + \lambda\mathbf{D}_k)^{-1}\mathbf{B}'\mathbf{y}$$

In other words, If X_B is the X-matrix corresponding to the t *B*-spline basis i.e., $X_B = X_T L_{p_t}$

(4.1)

where \mathbf{L}_p is a square invertible matrix. Then fitted values for a penalized spline can be writing as

$$\hat{y} = X(X'X + \lambda^2 D)^{-1}X'y$$

Substitution of eq (4.1) allow us to express a penalized spline fit of degree p in terms of the *B*-spline basis as

$$\hat{y} = X_{B} (X'_{B}X_{B} + \lambda^{2p} L'_{p} D_{p} L_{p})^{-1} X'_{B} y$$
(4.2)

An important component of P-splines is number and position of knots as it is crucial problem especially with regression splines. Small number of knots results in a not flexible function space to capture the enough variability of the data. The position of the knots may possibly have a strong influence on estimation and large number of knots may lead to serious over fitting. For this Eilers & Marx (1996) proposed a remedy based on roughness penalty. Also they chose a moderate number of equally spaced knots within the domain of x to ensure enough exibility. Sufficient smoothness of the fitted curve is achieved through a difference penalty on adjacent B-spline coefficients.

4.2 Model 2: (RWC P-Spline)

Another approach of p-spline models was offered by Ruppert et al., (2003). They used truncated power functions, F, as basis with quantile based knots and have a ridge penalty on the truncated power function (TPF), whatever their degree. For a given degree p, column j of F is given by

$$f_{ij} = \begin{cases} 0 & x < t_j \\ (x_i - t_j)^p & x \ge t_j \end{cases}$$

The vector *t* contains the knots chosen as quantiles of the x_s . The model for $E(y) = \mu$ is given by

$$\mu = \beta_0 + \beta_1 X + \dots + \beta_p X^p + \sum_{j=1}^{K} b_j (X - t_j)_+^p$$

or

$$\mu_j = \sum_{k=0}^p \beta_k X_i^k + \sum_{j=1}^{n-1} f_{ij} b_j$$

or
$$\mu = X\beta + Fb$$
,

where X is a m by p+1 matrix with to in row i. So mathematically, Ruppert, Wand and Carroll (2003) presented p-spline model as:

$$S(X) = \sum_{i=1}^{m} \left\{ y_i - \left(\beta_0 + \beta_1 x + \dots + \beta_p x^p + \sum_{j=1}^{k} b_j \left(x - t_j \right)_+^p \right) \right\}^2 + \lambda |b|^2$$
 Model 2

i.e. minimization function as

$$Q_F = |y - X\beta - Fb|^2 + \lambda |b|^2,$$

in which we recognize a ridge penalty on b. Minimization of Q_{F} leads to the system of equations

$$\begin{bmatrix} X'X & X'F \\ F'X & F'F + \lambda I \end{bmatrix} \begin{bmatrix} \beta \\ b \end{bmatrix} = \begin{bmatrix} X'y \\ F'y \end{bmatrix}$$

For the number of basis functions in B chosen "too large", means very small value of λ , the fitted curve over-fitted the data giving too many fluctuations. Increasing λ increases the smoothness and in the limit a polynomial fit of degree p is obtained. Quantiles of x are chosen for the positions of the knots. By increasing λ the smoothness can be tuned. Also, we can express the eq (4.2) in form of penalty used by B-Splines basis as

$$\hat{y} = X_{p} (X'_{p}X_{p} + \lambda \widetilde{D})^{-1} X'_{p} y$$

That is a type of ridge regression.

While a ridge penalty on TPF is equivalent to a difference penalty of (fixed order) on B-splines; B-splines allow a flexible choice of the order of the penalty, such as B-splines can be computed almost trivially from TPF.

As to apply p-splines, different spline bases will be chosen for a balance of numerical stability and ease of implementation. Thus in next section we look at basis functions used for both P-spline models and their mutual relationships.

Truncated Power Function basis vs. B-Spline basis

Truncated polynomials as basis are simple and useful for understanding spline regression but not numerically stable when the number of knots is large and the smoothing parameter λ close to zero. In this case the computation has to be organized carefully, involving QR or Demmler-Reinsch decomposition While B-Spline basis are easy to calculate and numerically superior alternatives.

We begin with TPF and equally-spaced knots since they are a little easier to explain, as TPFs Basis are 1, x, K, x^{p} , $(x - t_{1})^{p}_{+}$, K, $(x - t_{k})^{p}_{+}$ and B-splines can be derived from them as differences of TPF.

$$\begin{split} B_{j}(x; 0) &= f_{j-1}(x; 0) - f_{j}(x; 0) = -\Delta f_{j}(x; 0) \\ B_{j}(x; 1) &= f_{j-2}(x; 1) - 2 f_{j-1}(x; 1) + f_{j}(x; 1) = \Delta^{2} f_{j}(x; 1) \\ N \\ B_{j}(x; p) &= -1^{p+1} \Delta^{p+1} f_{j}(x; p) / (h^{p} p!) \end{split}$$

Where B-splines based on a K+1 non decreasing set of knots, $t_0, t_1, t_2, ..., t_k$ are defined by de Boor (1978), as the *i*th B-splines function of degree *p* is obtained by recurrence from first-order B-splines

$$B_{i,p,t}(X) = \frac{X - t_i}{t_{i+p} - t_i} B_{i,p-1,t}(X) + \frac{t_{i+1+p} - X}{t_{i+1+p} - t_{i+1}} B_{i+1,p-1,t}(X)$$

for all real numbers x, with

$$B_{i,0,t}(X) = \begin{cases} 1, & \text{if } t_i \leq X < t_{i+1,} \\ 0, & \text{otherwise} \end{cases}$$

Pak.j.stat.oper.res. Vol.XIV No.2 2018 pp289-303

Saira Sharif, Shahid Kamal

Now by using example of wood strength data, we derive B-splines from differences of TPFs. Thus the matrix of TPF basis and matrix of B-spline basis derived from TPFs are given below.

From wood strength data presented in section 3, we have:

min(x) is 1.0 i.e. minimum concentration

and

max(x) is 15.0 i.e. maximum concentration.

If we use 3 breaks and 3rd degree polynomial for this data, then 10 equally spaced knots are found as:

-13.000,	-8.333,	-3.667,	1.000,	5.667,
0.333,	15.000,	19.667,	24.333,	29.000.

Table 4.1:Thus the matrix of truncated power basis for wood strength data is as
follows:

2744.000	813.0370	101.6296	0.000	0.00000000	0.0000000	0	0	0	0
3048.625	950.8287	137.9213	0.125	0.00000000	0.00000000	0	0	0	0
3375.000	1103.3704	181.9630	1.000	0.00000000	0.00000000	0	0	0	0
4096.000	1455.7037	296.2963	8.000	0.00000000	0.00000000	0	0	0	0
4913.000	1876.0370	450.6296	27.000	0.00000000	0.00000000	0	0	0	0
5359.375	2113.5787	544.6713	42.875	0.00000000	0.00000000	0	0	0	0
5832.000	2370.3704	650.9630	64.000	0.00000000	0.00000000	0	0	0	0
6331.625	2647.1620	770.2546	91.125	0.00000000	0.00000000	0	0	0	0
6859.000	2944.7037	903.2963	125.000	0.00000000	0.00000000	0	0	0	0
7414.875	3263.7454	1050.8380	166.375	0.57870370	0.00000000	0	0	0	0
8000.000	3605.0370	1213.6296	216.000	2.37037037	0.00000000	0	0	0	0
9261.000	4357.3704	1587.9630	343.000	12.70370370	0.00000000	0	0	0	0
10648.000	5207.7037	2032.2963	512.000	37.03703704	0.00000000	0	0	0	0
12167.000	6162.0370	2552.6296	729.000	81.37037037	0.00000000	0	0	0	0
13824.000	7226.3704	3154.9630	1000.000	151.7037037	0.2962963	0	0	0	0
15625.000	8406.7037	3845.2963	1331.000	254.0370370	4.6296296	0	0	0	0
17576.000	9709.0370	4629.6296	1728.000	394.3703703	18.9629630	0	0	0	0
19683.000	11139.3704	5513.9630	2197.000	578.7037037	49.2962963	0	0	0	0
21952.000	12703.7037	6504.2963	2744.000	813.037037	101.6296296	0	0	0	0

0.00164	-0.00656	0.00984	-0.00656	0.00164	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00164	-0.00656	0.00984	-0.00656	0.00164	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00164	-0.00656	0.00984	-0.00656	0.00164	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00164	-0.00656	0.00984	-0.00656	0.00164	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00164	-0.00656	0.00984	-0.00656	0.00164	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00164	-0.00656	0.00984	-0.00656	0.00164

 Table 4.2: And the required difference matrix is:

T.LL. 4.2	TTI • 1	D I I I	• • •	4	1
1 able 4.3:	I hus our reamred	B-SDIIDE DASIS	matrix lising	truncated no	wer basis is:
I dole liet	Indo our required	2 spine subis	marin asing	n ancatea po	

1.66667e-01	1.66667e-01	1.66667e-01	0	0	0
1.18630e-01	6.55802e-01	2.25363e-01	0.000205	0	0
8.08431e-02	6.25668e-01	2.91849e-01	0.001640	0	0
3.10982e-02	5.22352e-01	4.33431e-01	0.013120	0	0
7.59232e-03	3.86237e-01	5.61893e-01	0.044278	0	0
2.60417e-03	3.15104e-01	6.11979e-01	0.070313	0	0
4.85909e-04	2.46842e-01	6.47716e-01	0.104956	0	0
7.59232e-06	1.85139e-01	6.65413e-01	0.149440	0	0
2.43978e-15	1.33443e-01	6.61747e-01	0.204750	6.07386e-05	0
4.25213e-15	9.23758e-02	6.37626e-01	0.269050	9.49040e-04	0
2.96663e-15	6.07386e-02	5.96696e-01	0.338678	3.88727e-03	0
3.59338e-15	2.08333e-02	4.79167e-01	0.479167	2.08333e-02	0
5.42536e-15	3.88727e-03	3.38678e-01	0.596696	6.07386e-02	0
6.15294e-15	6.07386e-05	2.047450e-01	0.661747	1.33443e-01	0
8.30050e-15	5.27105e-16	1.049563e-01	0.647716	2.46842e-01	0.000486
5.98387e-15	-7.98833e-16	4.427843e-02	0.561893	3.86237e-01	0.007592
-4.30970e-17	9.69566e-15	1.31195e-02	0.433431	5.22352e-01	0.031098
-1.24162e-14	1.35755e-14	1.63994e-03	0.291849	6.25668e-01	0.080843
-1.18928e-14	1.50358e-14	-6.96356e-15	0.166667	6.66667e-01	0.166667

The main advantage of the truncated power function basis is the simplicity of its construction and the ease of interpreting the parameters in a model that corresponds to these basis functions. However, there are two weaknesses when to use this basis for regression. These functions grow rapidly without bound as increases, resulting in numerical precision problems when the data span a wide range. Furthermore, many or even all of these basis functions can be nonzero when evaluated at some value, resulting in a design matrix with few zeros that precludes the use of sparse matrix technology to speed up computation. This weakness can be addressed by using a B-spline basis.

The simplest system of TPF uses p = 0; it consists of step functions with jumps of size 1 at the knots. The right branch of a TPF of degree p looks like $(x - t_j)^p$; the left branch is zero. Following figure shows quadratic TPF bases vs quadratic B-Spline basis, with equally-spaced knots.





Figure 1: B-spline bases with equally spaced knots vs

Truncated Power bases with equally spaced knots

5. Comparisons of Models by Example:

For Model 1:

$$S(x) = \sum_{i=1}^{m} \left\{ y_i - \sum_{j=1}^{n} \hat{\alpha}_j B_j(X) \right\}^2 + \lambda \sum_{j=k+1}^{n} (\Delta^k \hat{\alpha}_j)^2$$

Cubic B-spline basis with equally spaced knots and squared difference penalty are used. For calculating smoothing parameter GCV criterion is customized.

By setting 10 equally spaced knots as:

1,	2.556,	4.112,	5.668,	7.224,
8.778,	10.336,	11.889,	13.448,	15

0	0	0	0	0	0	0	0	0	0	0
0.557472	0.124527	0.005534	0	0	0	0	0	0	0	0
0.533812	0.37635	0.04427	0	0	0	0	0	0	0	0
0.091122	0.566338	0.338653	0.003886	0	0	0	0	0	0	0
0.000091	0.20475	0.661739	0.133412	0	0	0	0	0	0	0
0	0.070331	0.612005	0.31506	0.002602	0	0	0	0	0	0
0	0.013126	0.433478	0.522309	0.031086	0	0	0	0	0	0
0	0.000205	0.225412	0.65579	0.118597	0	0	0	0	0	0
0	0	0.080871	0.625701	0.291789	0.001638	0	0	0	0	0
0	0	0.016691	0.456614	0.501085	0.0256096	0	0	0	0	0
0	0	0.000487	0.246909	0.647688	0.104916	0	0	0	0	0
0	0	0	0.020849	0.479247	0.479086	0.0208173	0	0	0	0
0	0	0	0	0.105010	0.647754	0.2467516	0.0004844	0	0	0
0	0	0	0	0.001644	0.291955	0.6256086	0.0807920	0	0	0
0	0	0	0	0	0.031128	0.5224586	0.4333094	0.013104	0	0
0	0	0	0	0	0	0.1335298	0.6617565	0.204623	0.000090	0
0	0	0	0	0	0	0.0038963	0.3387756	0.566281	0.0910469	0
0	0	0	0	0	0	0	0.0442955	0.376451	0.533756	0.045497
0	0	0	0	0	0	0	0	0	0	1

 Table 4.4:
 We construct cubic b-spline basis at wood strength data as shown below

1	-3	-3	-1	0	0
-3	10	-12	6	-1	0
3	-12	19	-15	6	-1
-1	6	-15	19	-12	3
0	-1	6	-12	10	-3
0	0	-1	3	-3	1

 Table: 4.5: And we found difference penalty matrix of order 3 as

For fitting Model1 at wood strength data we use 10 equally spaced knots, 3^{rd} order difference penalty, and calculated by using GCV criterion with GCV = 83.3827 with 6 df. Plot of the model1 fit is shown below:



Figure 2: Model1 Fit (GCV =83.3827, df= 6)

Table 4.6: We found the residual table by above p-spline (model 1) fit as follow:

Concentration	Residuals
1	0.60452391
1.5	-2.25444024
2	2.18956401
3	0.99529131
4	-1.75211456
4.5	-0.27581749
5	1.10047958
5.5	-1.12322335
6	0.55307371
6.5	-0.07062922

Concentration	Residuals
7	-0.39433215
8	-1.04115275
9	2.20970088
10	-0.73938135
11	0.03696717
2	-0.67702568
13	2.31202487
14	-2.83762416
15	1.16411551

Here $SS_e = 39.04372$ and $SS_t = 3416.885$ and R^2 is 0.9885733. Similarly for Model 2:

$$S(x) = \sum_{i=1}^{m} \left\{ y_{i} - \left(\beta_{0} + \beta_{1}x + ... + \beta_{p}x^{p} + \sum_{j=1}^{K} b_{j}(x - t_{j})_{+}^{p} \right) \right\}^{2} + \lambda |b|^{2}$$

Truncated power basis with 10 quantile based knots and ridge penalty are used. For calculating smoothing parameter GCV criterion is customized.

Table 4.7: 10 quantile based knots are calculated as:

9.090909%	18.18182%	27.27273%	36.36364%	45.45455%	54.54545%	63.63636%	72.72727%	81.81818%	90.90909%
1.818182	3.272727	4.454545	5.272727	6.090909	6.909091	8.454545	10.090909	11.727273	13.363636

By using above code 'tbs' applied on wood strength data with 10th degree and 10 quantile based knots, we construct the following truncated power basis matrix in table 4.9 (shown on next page).

For fitting Model 2 at wood strength data we use 10^{th} degree polynomial and 10 quantile knots. Here smoothing parameter is calculated by using GCV criterion with GCV = 10.83 with degree of freedom as 11 (It includes 1 df for intercept). So, plot of the above Fit of model 2 at original fit of given data is shown below:



Figure 3: Model2 Fit (GCV = 10.82, df = 11)

Table 4.8:	Calculated residua	als of model 2 fo	r above fit of]	P-spline are:

Concentration	Residuals
1	0.18278469
1.5	-0.79271766
2	0.98937231
3	-0.39981358
4	-0.84202914
4.5	0.59500398
5	1.43657058
5.5	-1.29464984
6	0.23541956
6.5	-0.14865309

Concentration	Residuals
7	-0.06268281
8	-0.46362547
9	1.64615185
10	-1.78414969
11	0.76135459
12	0.10433867
13	-0.24706469
14	0.09802063
15	-0.01363087

Table 4.9

1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
1	1.5	2.25	3.38	5.0625	7.5938	11.39063	17.08594	25.62891	38.44336	57.66504	0	0	0	0	0	0	0	0	0	0
1	2	4	8	16	32	64	128	256	512	1024	3.947963e ⁻⁸	0	0	0	0	0	0	0	0	0
1	3	9	27	81	243	729	2187	6561	19683	59049	5.315042	0	0	0	0	0	0	0	0	0
1	4	16	64	256	1024	4096	16384	65536	262144	1048576	2444.475	4.13974e ⁻²	0	0	0	0	0	0	0	0
1	4.5	20.25	91.13	410.06	1845.28	8303.766	37366.95	168151.3	756680.6	3405063	19243.91	7.751948	3.7651e ⁻¹⁴	0	0	0	0	0	0	0
1	5	25	125	625	3125	15625	78125	390625	1953125	9765625	106353.9	236.3791	2.33123e ⁻³	0	0	0	0	0	0	0
1	5.5	30.25	166.38	915.06	5032.84	27680.64	152243.5	837339.4	4605367	2.532952e ⁷	457744.8	3004.236	1.559738	3.67683e-7	0	0	0	0	0	0
1	6	36	216	1296	7776	46656	279936	1679616	1.00777e ⁷	6.046618e ⁷	1635503	22765.95	77.72529	4.13974e-2	0	0	0	0	0	0
1	6.5	42.25	274.63	1785.06	11602.91	75418.89	490222.8	3186448	2.071191e ⁷	1.346274e ⁸	5059941	122562.2	1282.030	7.751948	1.3128e4	0	0	0	0	0
1	7	49	343	2401	16807	117649	823543	5764801	4.035361e ⁷	2.824752e ⁸	1.395795e ⁷	517501.6	11419.67	236.3791	3.8554e ⁻¹	3.8554e ⁻¹¹	0	0	0	0
1	8	64	512	4096	32768	262144	2097152	1.677722e ⁷	1.342177e ⁸	1.073742e ⁹	8.150088e ⁷	5573225	313847.9	22765.95	643.0816	2.387182	0	0	0	00
1	9	81	729	6561	59049	531441	4782969	4.304672e ⁷	3.874205e ⁸	3.486784e°	3.650430e ⁸	3.797333e ⁷	3765071	517501.6	43408.32	1597.171	2331.233	0	0	0
1	10	100	1000	10000	100000	1000000	1.00e ⁷	1.00e ⁸	1.00e°	1.00e ¹⁰	1.344306e9	1.898411e ⁸	2.750246e ⁷	5573225	833216.2	79590.7	77.72529	0	0	0
1	11	121	1331	14641	161051	1771561	1.948717e ⁷	2.143589e ⁸	2.357948e ⁹	2.593742e ¹⁰	4.258796e ⁹	7.590361e ⁸	1.443438e ⁸	3.797333e7	8128506	1312799	11419.67	3.8554e-1	0	0
1	12	144	1728	20736	248832	2985984	3.583181e ⁷	4.299817e ⁸	5.159780e°	6.191736e ¹⁰	1.197439e ¹⁰	2.563218e°	5.982106e ⁸	1.898411e ⁸	5.190470e ⁷	1.169374e ⁷	313847.9	643.0816	2.2766e-6	0
1	13	169	2197	28561	371293	4826809	6.274852e ⁷	8.157307e ⁸	1.060450e ¹⁰	1_378585e ¹¹	3.055795e ¹⁰	7.584220e°	2.076594e ⁹	7.590361e ⁸	2.478615e ⁸	7.027829e ⁷	3765071	43408.32	11.1520	0
1	14	196	2744	38416	537824	7529536	1.054135e ⁸	1.475789e ⁹	2.066105e ¹⁰	2.892547e ¹¹	7.196497e ¹⁰	2.017870e ¹⁰	6.280094e°	2.563218e ⁹	9.577798e ⁸	3.213802e ^s	2.750246e ⁷	8.33216.2	3876.83	0.01089
1	15	225	3375	50625	759375	11390630	1.70859e ⁸	2.562891e ⁹	3.844336e ¹⁰	5.766504e ¹¹	1.583993e ¹¹	4.920058e ¹⁰	1.70079e ¹⁰	7.584220e ⁹	3.150169e ⁹	1.202190e°	1.443438e ⁸	8128506	140960.7	137.657

Here from table 4.8, $SS_{e=}$ 13.4542, SS_{t} = 3416.885 and R^{2} is 0.9960624. To test the statistical significance of difference we use the approximate *F*-test, which is based on the following statistic

$$F = \frac{R_{larger}^2 - R_{smaller}^2}{\left(1 - R_{larger}^2\right) \left(df_{res, smaller} - df_{res, larger}\right) / df_{res, larger}}$$

Under the null hypothesis this statistic will have an approximate F-distribution with dfres, smaller – dfres, larger and dfres, larger degrees of freedom.

For the given dataset the value of the corresponding F-statistic is 4.184 and the corresponding p-value is 0.165, which means that the difference between both P-spline fits is neither statistically nor practically significant.

6. Comparison on different aspects:

a. Numerical Immovability

Technically TPF can be used directly as a basis for regression. But it is not to be recommended especially for the large-scale data or when the number of knots is large and $p \ge 1$ their numerical condition can be poor. If smoothing is the only goal, TPF of degree 1 might be workable, but higher degrees, which are needed when one wishes to smoothly estimate (second) derivatives are essentially unusable.

In very large problems extra care is also needed for using B-spline basis. For example in an engineering application of long series (e.g. 15000 observations) of echo sounding measurements had to be smoothed and reduced to a uniform grid, we can use bases with 2000 B-splines, but not possible for TPF basis. Another technique is available, i.e. to first fill a matrix of this size with a TPF basis and then compute differences to reduce it to a sparse matrix. But it is not a good idea. A nice property of B-splines with equi-spaced knots is that we get the same set of values, but shifted by k columns. Hence compute the basis in a matrix with p + 1 columns and transfer the results to the right columns of a sparse matrix.

b. Quality of Fit

The details of a P-spline model can be envisaged in an attractive way. Figure below on left illustrates fit with a rich B-spline basis and a second-order penalty. The coefficients of the individual B-splines are shown, scaled by their coefficients at the positions of their maxima. These points are close to the fitted curve and present the skeleton of the fit. And Figure below on right illustrates fit with a Truncated Power function basis and a ridge penalty. Here the estimated curves (bold) are shown together with the used cubic truncated polynomial basis functions.



Figure 4: cubic B-splines & second order penalty (λ =0.01) vs cubic Truncated Powers & a ridge penalty (λ =0.002)

This Presentation shows a significant role of Penalized spline using B-splines and difference penalties that it forces the skeleton, i.e., the coefficients, to follow a smooth pattern. Consequently the full curve that follows from them will also be smooth while penalized spline using TPF and ridge penalty do not lend such an insightful presentation. Also shows that the larger the values of the smoothing parameter λ , the more the fit shrinks towards a polynomial fit, while smaller values of λ result in a wiggly "overfitted" estimate. This is clearly visible in figures where different smoothing parameters are selected.

c. Smoothing

Like in the case of time series and spectra where the number of large observations in the discrete data series, equidistant sampled sometimes only a smoothed discrete series is needed. In such problems, B-spline basis of degree zero, with a knot at every observation, may be an attractive choice. The basis then is the identity matrix, the system of equations becomes $(I + \lambda \Delta'_d \Delta_d) \alpha = y$ and the smooth series μ coincide. Because discrete smoother is very attractive for long data series in case, we have to access the sparse matrix software. So where the sparseness is vital, TPF will not work in such settings. Eilers and Marx (2010) expressed that a smaller-scale application in which the discrete approach may be appropriate is histogram smoothing. And to estimate a density one constructs a histogram with many, say 150, narrow bins. Such a plotted histogram will look unattractive and ambiguous, but generalized linear smoothing with penalized splines completely changes performance. Also Eilers and Marx presented histogram smoothing with Penalized B-splines in a Poisson regression setting. But did not found any possibility by Penalized splines with TPF basis especially with quantile based knots.

Similarly for Multidimensional smoothing applications like image smoothing and for many others of large real-world problem, Eilers and Marx (2010) has shown that tensor products of B-splines and difference penalties are a practical and effective tool. But by contrast, such problems with tensor products are impossible by Penalized Spline smoothing using TPF basis. And when smoothing periodic data on a linear axis, it may happen that both ends do not join smoothly. In such case again B-spline basis with difference penalty seems quite easy.

d. Derivative Estimation

Frequently one is not only interested in a fitted curve, but also in its derivatives. The derivative is also quite an intuitive concept. Basically derivatives are used to measure how a system changes with time. Like in a chemical reaction of say, oxygen + hydrogen goes to water, we can use the derivative of the rate equation to measure how long it will take for a certain percentage of the reactants to be turned into products. Also for example, in the business world, if we figure out the equation of a curve we can find the values for all of the minimums and maximums (that would be where the derivative in equal to zero) so that we can determine profit and loss points, etc.

Penalized splines allow easy calculation of derivatives. For penalized splines of Ruppert, Wand and Carroll (2003) using TPF basis only need is to differentiate the polynomial branches weighted by the estimated coefficients and sum them. But the situation of Eilers and Marx (1996) using B-spline basis is little more complicated. But in the case of equally-spaced knots, the derivative of a weighted sum of B-splines can be computed by:

$$\frac{d}{dt}\sum_{k}B_{k}(t;p)\alpha_{k}=\sum_{k}(p-1)B_{k}(t;p-1)\Delta\alpha_{k}/h$$

If the second derivative curve is required then both TPF and B-spline basis has to be cubic. But we know that cubic TPF have a very poor condition number, so required great attention with the accomplishment of computations for derivative estimation. While cubic B-splines do not comprise such problems.

6. Conclusion

By above comparison, we got that there are two main approaches of Penalized spline smoothing (both named as P-Spline), first approach as Model1with B-splines and difference penalty and 2nd approach as Model2 with truncated power function (TPF) basis besides a ridge penalty. Both Models of P-splines (With B-Spline basis and TPF basis) have become popular in statistics and in applied fields, as can be judged from citation counts.

So by comparing these two main approaches, we found:

- Ridge Penalty of Model2 has great worth especially in small set of data.
- Equally-spaced knots presented in Model1 are always to be preferred as they allow easy smooth interpolation with B-splines as well as TPF.
- As the basis of Model1 (B-spline basis) can be computed almost trivially from basis of Model2 (TPF basis). So Model1 allow a flexible choice of order of the penalty.
- Both Models allow mixed model approach.
- Model1 have excellent numerical properties over Model2.
- Basis of Model1 allow informative conception.
- Model2 is not applied for large-scale sets, while Model1 is sparse and can lend for such problems.
- Both basis and penalties of Model1 are easily adjusted in the smoothing agenda.

As Ruppert, Wand and Carroll (2003), even not anyone after that did not provide any practical advantage of Model2 having TPF with non-uniform knots and ridge penalty. So there can be conclusion that if truncated power functions are chosen, use equally-spaced knots. Otherwise simply use Model1 having B-spline bases with difference penalties.

References

- 1. Claeskens, G.; Krivobokova, T. and Opsomer, J. D. (2009). Asymptotic Properties of Penalized Spline Estimators. *Biometrika* 96 (3), 529-544.
- 2. DeBoor, C. (1978). *A Practical Guide to Splines* (Revised edition). New York, Springer.
- 3. O'Sullivan, F. (1986). A Statistical Perspective on ill-posed Inverse Problems. *Statistical Society*. 1 (4), 502-527.
- 4. Eilers, P. H. C. and Marx, B. D. (1992) Generalized Linear Models with P-Splines. *In Advance in GLIM and Statistical Modelling (Fahrmeir et al., eds.).* New York, Springer.
- 5. Eilers, P. H. C. and Marx, B. D. (1996) Flexible Smoothing with B-splines and Penalties. J stat 11: 89-102.
- 6. Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots and penalties. Wiley Interdisciplinary Reviews: Computational Statistics, 2, 637–653.
- 7. Greven, S. and Kneib, T. (2010) On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika* 97(4), 773-789.
- 8. Kauermann, G., Krivobokova, T. & Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society, Series B* 71 (2), 487-503.
- 9. Li, Y. & Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* 95, 415-436.
- 10. Ruppert, D. and Carroll, R.J. (1997). Penalized regression splines, manuscript.
- 11. Ruppert, D. (2002). Selecting the Number of Knots for Penalized Splines. *Journal* of Computational and Graphical Statistics 11 (4), 735-757.
- 12. Ruppert, D. & Carroll, R. (2000). Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics* 42, 205–253.
- 13. Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge.
- 14. Wager, C., Vaida, F. & Kauermann, G. (2007). Model Selection for P-spline Smoothing using Akaike information criteria. *Australian and New Zealand Journal of Statistics* 49 (2), 173-190.
- 15. Wand, M. P. (1999). On the optimal amount of smoothing in penalized spline regression. *Biometrika* 86, 936–940.
- 16. Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC Press.