

An effective scheme for estimating a smoother parameter in the method of regularization

Hongmei Bao¹, Kaoru Fueda²

¹Graduate School of Environmental Science, Okayama University, Okayama, Japan

²Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan

Email address:

gev421256@s.okayama-u.ac.jp (H. Bao), fueda@ems.okayama-u.ac.jp (K. Fueda)

To cite this article:

Hongmei Bao, Kaoru Fueda. An Effective Scheme for Estimating a Smoother Parameter in the Method of Regularization. *Applied and Computational Mathematica*. Vol. 2, No. 6, 2013, pp. 118-123. doi: 10.11648/j.acm.20130206.11

Abstract: We had proposed a scheme for the surface approximation which consists of the estimation by the regularization method and the evaluation by generalized CV with an influence function [1]. We have to decide the value of the optimal smoother parameter which can minimize the value of the evaluation function. Among the models which have suitable parameters, we have to choose the best model using information criteria such as CV or generalized CV with an influence function (GCVIF). However, the method of GCVIF is not practical, because it requires the calculation of the inverse matrix of the hat matrix and the influence function [2]. Those calculations take a large amount of time when n increases. An efficient scheme which will take a small amount of time is required. On the other hand, there are many parameters which we have to decide. Those are the coefficients of the spline functions and the total number of knots, and positions of the parameters and a smoother parameter of the penalized term. The range of the total number of knots is decided by the total number of sample points. The range of the positions of the knots is decided by the area of the surface. However, it is difficult to estimate the range of the value of the smoother parameter. Therefore, we have to estimate it quite roughly. In this paper, we propose an effective method to estimate the range of the smoother parameter and consequently obtain the parameter precisely. We can reduce the calculation time which does not contribute to the selection of the optimal model and we can determine a more accurate and smoother parameter in a small amount of time.

Keywords: Spline Interpolation, Penalized Coefficient, Smoother Parameter, Method of Regularization, Cross-Validation

1. Introduction

The smoothing spline surfaces are often used to estimate a three-dimensional shape of the surface. When we use the regularization method with the penalized term, the smoothing parameter is most important. Too large of a parameter produces a surface that is too flat and too small of a one causes over-fitting and creates a surface without fluency. The value of the appropriate parameter varies according to the shape of the surface to be estimated. We have to search for the appropriate range of the value of the smoothing parameter. When we have decided the set of the parameter and coefficients, we will evaluate them by the information criterion. However, it will take a long time to calculate when we have chosen CV. For the many sets of knots and many values of parameters we have to calculate the values of CV. However most of them don't contribute to the determination of the optimal model. We introduce a scheme which uses sets that are as small as possible.

Moreover, we can obtain a more accurate value of the smoother parameter.

At first, we start with small sets of knots and a set of rough values of the smoother parameters. After calculating the values of CV for these values, we can obtain a value of the smoother parameter by using spline interpolation. This value is only estimated by interpolation, so it is not necessarily accurate. Based on these parameters we estimate the values of CV for the many sets of knots. The model which has the smallest value of CV is the optimal set of knots. We can obtain the identical optimal set of knots from various estimated values of the smoother parameter. This has been verified by numerical calculation. Finally, we can determine the optimal value of the smoother parameter for the optimal set of knots.

2. Surface Approximation by B-Splines

2.1. B-Splines

A B-spline of degree n is a function composed of a linear combination of basis B-splines $B_{i,n}$ of degree n [3-5]. The $m-n-1$ basis B-splines of degree n can be defined, for $n=0,1,\dots,m-2$, using the Cox-de Boor recursion formula as follows:

$$B_{j,0}(x) = \begin{cases} 1 & (\text{if } x_j \leq x < x_{j+1}) \\ 0 & (\text{otherwise}) \end{cases} \quad (j=0,1,\dots,m-2)$$

$$B_{j,0}(x) = \frac{x - x_j}{x_{j+n} - x_j} B_{j,n-1}(x) + \frac{x_{j+n+1} - x}{x_{j+n+1} - x_{j+1}} B_{j+1,n-1}(x),$$

$$(j = 0, 1, \dots, m - n - 2)$$

where m is the total number of the knots and $t_0 \leq t_1 \leq \dots \leq t_{m-1}$. We set the approximation for the three dimensional surface:

$$u(x, y) = \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} w_{ij} M_i(x) N_j(y), \quad (1)$$

where p_1, p_2 is the total number of basis B-splines $\{M_i(x)\}, \{N_j(y)\}$ respectively, and these functions have the support $[\xi_{i-r}, \xi_i]$, $[\eta_{j-r}, \eta_j]$ for the x, y direction respectively. We have to satisfy the Schoenberg-Whitney condition [6] because if there is no sample point in the domain $\{(x, y) | \xi_{i-r} \leq x < \xi_i, \eta_{j-r} \leq y < \eta_j\}$, then we cannot determine the parameter w_{ij} . Usually, B-splines with order four (degree three) are used in the calculation. Along the x direction, we set the knots x_1, x_2, \dots, x_p and the knots at both ends were four-folded. Therefore, the total number of basis B-splines should be $p-4$. At every interval of $[x_{n-1}, x_n], n = 5, 6, \dots, p-3$ there exists four basis B-splines, as below.

$$B_{n,1}(x) = \frac{-(x-x_n)^3}{(x_n-x_{n-3})(x_n-x_{n-2})(x_n-x_{n-1})}, \quad (2)$$

$$B_{n,2}(x) = \frac{(x-x_{n-3})(x-x_n)(x-x_{n-1})}{(x_n-x_{n-3})(x_n-x_{n-2})(x_n-x_{n-1})} + \frac{(x-x_{n-2})(x-x_n)(x-x_{n+1})}{(x_{n+1}-x_{n-2})(x_n-x_{n-2})(x_n-x_{n-1})} - \frac{(x-x_{n-1})(x-x_{n+1})(x-x_{n+1})}{(x_{n+1}-x_{n-1})(x_{n+1}-x_{n-2})(x_n-x_{n-1})}, \quad (3)$$

$$B_{n,3}(x) = \frac{(x-x_{n-2})(x-x_{n-2})(x-x_n)}{(x_{n+1}-x_{n-2})(x_n-x_{n-2})(x_n-x_{n-1})} - \frac{(x-x_{n-2})(x-x_{n-1})(x-x_{n+1})}{(x_{n+1}-x_{n-2})(x_n-x_{n-1})(x_{n+1}-x_{n-1})} + \frac{(x-x_{n-1})(x-x_{n-1})(x-x_{n+2})}{(x_{n+1}-x_{n-1})(x_n-x_{n-1})(x_{n+2}-x_{n-1})}, \quad (4)$$

$$B_{n,4}(x) = \frac{(x-x_{n-1})^3}{(x_{n+2}-x_{n-1})(x_{n+1}-x_{n-1})(x_n-x_{n-1})}. \quad (5)$$

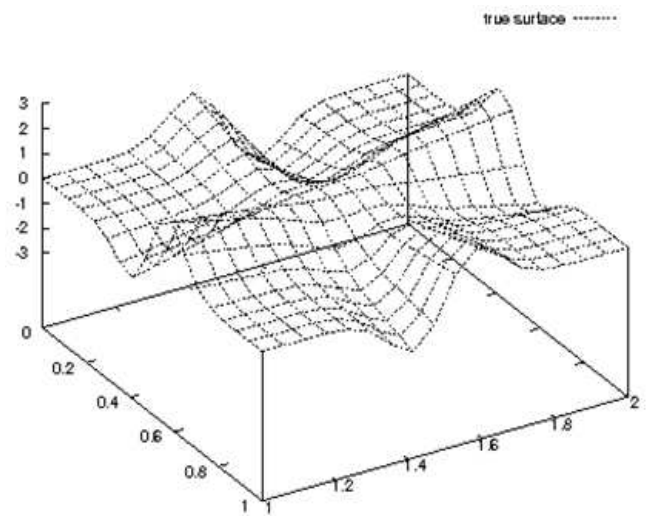
We denote the total number of knots (n_x, n_y) where n_x and n_y are the total number of knots along the x and y directions respectively.

2.2. Sample Surface

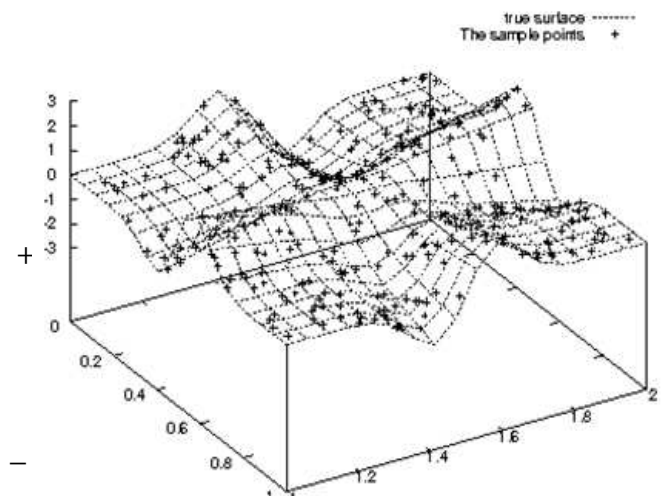
To evaluate our scheme, we have made the next numerical experiment. We have used the sample surface with the equation as follows:

$$z = (1-x)\exp(-x^2) + xy\exp(-x^2) + xy\exp(-y^2).$$

We have taken the randomized samples with Gaussian noise. We set the region of the samples $[0,1] \times [1,2]$ shown in Fig.1 and set $n=300$ as the total number of samples. Using these samples, we estimated the coefficients of the models based on the various conditions. For the determination of the optimal model we evaluated the models by CV.



(a) True surface



(b) Samples

Figure 1. Experimental surface.

3. Estimation of Coefficients of B-Spline for Regularization Method

For the nonlinear statistical modeling, the maximum penalized likelihood methods are often used [7-9]. Suppose that we have n observations $\{(z_\alpha, x_\alpha); \alpha = 1, \dots, n\}$, where z_α are the response variables generated from unknown true distribution $G(z|x)$ having a probability density of $g(z|x)$ and x_α are the vectors of explanatory variables. We estimate w , which is a vector consisting of the unknown parameters, and determine the model $z = u(x|w)$. Let $f(z_\alpha|x_\alpha; \theta)$ be a specified parametric model, where θ is a vector of unknown parameters included in the model. The regression model with Gaussian noise is denoted as:

$$z_\alpha = u(x_\alpha|w) + \varepsilon_\alpha, \quad \varepsilon_\alpha \sim N(0, \sigma^2), \quad \alpha = 1, \dots, n$$

$$f(z_\alpha|x_\alpha; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{\{z_\alpha - u(x_\alpha; w)\}^2}{2\sigma^2} \right],$$

where $\theta = (w' \sigma^2)'$. The parameter will be determined by the maximization of the penalized log-likelihood function, expressed as:

$$\ell_\lambda(\theta) = \sum_{\alpha=1}^n \log f(z_\alpha|x_\alpha; \theta) - \frac{n}{2} \lambda H(w) \quad (6)$$

As the regularized term or penalized terms $H(w)$ with an m -dimensional parameter vector w , various types are used depending on the dimension of explanatory variables or the purpose of the analysis. For the three-dimensional approximation we use [10]:

$$H(w) = \iint \left\{ \left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right\} dx dy, \quad (7)$$

and it is represented in the quadratic form:

$$H(w) = w' K w \quad (8)$$

Therefore, (6) will be

$$\ell_\lambda(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (z - Bw)'(z - Bw) - \frac{n}{2} \lambda w' K w,$$

where $z = (z_1, \dots, z_n)'$, $u(x_\alpha|w) = w' b(x_\alpha)$ and B is $n \times m$ matrix composed of the basis functions as:

$$B = (b(x_1)', \dots, b(x_n'))'.$$

With respect to θ , differentiating $\ell_\lambda(\theta)$ and setting the result equal to zero obtains the solution. As a result, the estimations of the parameters are:

$$\begin{aligned} \hat{w} &= (B' B + n \lambda \hat{\sigma}^2 K)^{-1} B' z, \\ \hat{\sigma}^2 &= \frac{1}{n} (z - B \hat{w})' (z - B \hat{w}). \end{aligned} \quad (9)$$

At first, we set the constant value of $\beta = \lambda \hat{\sigma}^2$ and determine \hat{w} for a given value of β . After we obtain the variance estimator $\hat{\sigma}^2$ we can then obtain the smoothing parameter $\lambda = \beta / \hat{\sigma}^2$.

4. Evaluation of the Model

For the B -spline, the total number and the locations of the knots are important. When the knots at both ends are four-folded, the least total number of knots is ten along the x and y axis respectively. The total number of parameters is $(n_x - 4) \times (n_y - 4) + 1$ which consists of the coefficients of the basis and the variance, and this value should be less than the total number of the sample points, at least. We have set 300 sample points, and the adequate total number of knots along every axis is less than or equal to 20. Also, for every (n_x, n_y) , we prepared 100 sets of randomized knots generated uniformly. However, if some of them do not satisfy the Schoenberg-Whitney condition [5], then we must generate another set of knots again. Furthermore, if some equations of matrices made from ill-conditioned sets cannot be solved properly, then we also must generate another set of knots again. We prepared 100 solvable sets for every (n_x, n_y) . We denote a set of knots as $S_{j,k}$ ($j = 1, 2, \dots, 121$, $k = 1, 2, \dots, 100$) where $j = 11(n_x - 10) + (n_y - 10) + 1$ represents the number of knots and k represents the serial number of the sets which has the same total number of knots. On the other hand, we tried eight values of the smoother parameter which is the coefficient of the penalized term. We set the values of β from 10^{-1} to 10^{-8} , so we considered 96,800 models and determined the parameters of those by the regularization method. The evaluations of the models are done by the value of CV. We use the log-likelihood for Cross-Validation (CV) as:

$$\begin{aligned} CV &= -2 \sum_{\alpha=1}^n \log (f(x_\alpha, \theta^{(-\alpha)})) \\ &= \sum_{\alpha=1}^n \left\{ \log(2\pi \hat{\sigma}^{2(-\alpha)}) + \frac{(z_\alpha - \hat{u}^{(-\alpha)})^2}{\hat{\sigma}^{2(-\alpha)}} \right\}. \end{aligned} \quad (10)$$

where $\hat{\sigma}^{2(-\alpha)}$, $\theta^{(-\alpha)}$, $\hat{u}^{(-\alpha)}$ are determined by the data without an α -th sample. We only tried very rough values of β . The results are shown in Table 1-2.

In order to determine a better model, we need a more accurate value of β . However, it will take much time to calculate the values of CV's, so we studied to obtain a more efficient scheme. In these models, there are many sets of knots and β 's which are not useful to determine the best model. We saved time for calculation and tried to obtain a more correct value of β and verify the validity of our scheme.

5. Optimal Smoother Parameter

5.1. Interpolation by Spline Function

For every set of knots $S_{j,k}$, we have only eight values of CV for $\beta = 10^{-1}, 10^{-2}, \dots, 10^{-8}$. Fig. 2 shows the interpolation of the value of CV by the spline function.

Let $x_i = -\log_{10} \beta_i$ and y_i is the value of CV for x_i . Spline interpolation functions $f_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i$, ($i = 1, 2, \dots, N - 1$) are defined over the interval $[x_i, x_{i+1}]$ ($i = 1, 2, \dots, N -$

1). We determine the coefficients of these functions as follows.

$$\begin{aligned} f_i(x_{i+1}) &= f_{i+1}(x_{i+1}) = y_{i+1} \\ f'_i(x_{i+1}) &= f'_{i+1}(x_{i+1}) \\ f''_i(x_{i+1}) &= f''_{i+1}(x_{i+1}) \text{ for } i = 1, 2, \dots, N-2 \end{aligned}$$

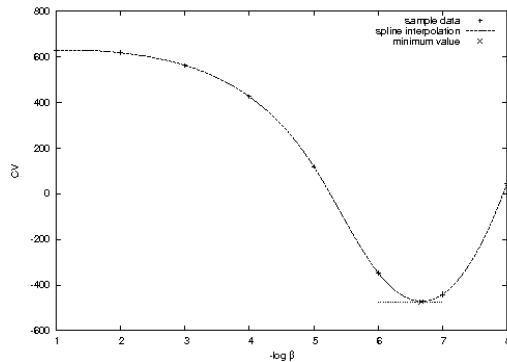


Figure 2. Spline Interpolation.

Based on these conditions, we can obtain the coefficients by solving the next matrix equation.

$$\begin{bmatrix} 2(h_1 + h_2) & h_2 & 0 & \cdots & 0 \\ h_2 & 2(h_2 + h_3) & h_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{N-2}2(h_{N-2} + h_{N-1}) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N-2} \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{N-2} \end{bmatrix} \quad (11)$$

where $h_i = x_{i+1} - x_i$, $v_i = 3(\frac{y_{i+1}-y_i}{h_i} - \frac{y_i-y_{i-1}}{h_{i-1}})$. We can determine all variables from $\{b_i\}$ as follows:

$$a_i = \frac{b_{i+1} - b_i}{3h_i}, c_i = \frac{y_{i+1} - y_i - a_i h_i^3 - b_i h_i^2}{h_i}.$$

Our aim is to obtain the value of x which gives the minimum value of the interpolated CV. If the spline

functions are only three-dimensional polynomials, we can easily differentiate them and calculate the zero points.

$$f'_i(x) = 3a_i(x - x_i)^2 + 2b_i(x - x_i) + c_i = 0$$

Let \tilde{x}_j be the zero point for every j . Instead of calculating CV's for all sets $S_{j,k}$ ($j = 1, 2, \dots, 121$, $k = 1, 2, \dots, 100$), we select only one $k = k_j$ for every j , and calculate CV's for selected sets S_{j,k_j} ($j = 1, 2, \dots, 121$). In those sets, we make the spline interpolation and determine the minimum estimated values m_j ($j = 1, 2, \dots, 121$) and \tilde{x}_j which gives m_j for every set. Among them, we determine the minimum value of m_j and \tilde{x}_j which gives it. We denote those values as m_{\min} and x_{\min} . Then, we obtain $\beta_{\min} = 10^{-x_{\min}}$. Using this β_{\min} , we estimate the coefficients of the estimated surface and calculate CV for all sets $S_{j,k}$ ($j = 1, 2, \dots, 121$, $k = 1, 2, \dots, 100$). The value of β_{\min} , that is $10^{-x_{\min}}$ varies depending on the selection of k_j , and the statistical values of x_{\min} based on 100 experiments are as follows.

The result of the calculation based on the set of various values of β_{\min} which include a maximum one and minimum one shows that the model $S_{71,21}$ is always the best.

5.2. Estimation of the Optimal β

After the determination of the best set of knots, we have to determine the optimal value of smoother parameter β . We have only vaguely estimated the value of β . To obtain the best value of CV, we calculate it based on various values of β only on the best set, which was selected above. The range of β is set from $b - 3\sigma$ to $b + 3\sigma$, where b is the mean of the best ten values of \tilde{x}_j in Table 3 and σ is the standard deviation of these values. The results of these calculations are shown in Table 9.

Table 1. All values of CV by previous method

n_x	n_y	j	k	β							
				10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
10	10	1	1	628.70	619.04	572.46	484.71	409.37	385.67	424.62	446.15
10	10	1	2	628.79	619.96	581.09	506.56	464.77	470.70	502.74	514.65
		\vdots									
10	10	1	100	628.79	619.92	579.63	507.01	464.46	453.35	480.04	487.96
10	11	2	1	628.58	617.90	565.17	463.64	351.10	339.66	402.22	414.08
10	11	2	2	628.95	621.43	590.44	534.84	523.31	541.29	581.61	645.75
		\vdots									
10	11	2	100	628.85	620.57	585.10	509.99	469.65	471.24	482.55	486.20
		\vdots									
20	20	121	1	628.55	617.59	561.84	415.92	106.62	-145.01	367.30	3028.74
		\vdots									
20	20	121	99	628.57	617.81	564.13	438.16	216.48	-10.95	127.41	1393.35
20	20	121	100	628.56	617.74	563.23	425.62	152.27	-32.05	173.46	3192.88

Table2. Minimum value of CV for each β by previous method

n_x	n_y	j	k	β	σ^2	λ	CV
13	13	37	19	10^{-1}	0.003851	2.59673E+01	31.07
18	18	97	37	10^{-2}	0.053463	1.87045E-01	40.25
20	15	66	92	10^{-3}	0.322460	3.10116E-03	537.93
17	20	118	95	10^{-4}	0.165861	6.02913E-04	367.40
20	15	66	92	10^{-5}	0.048688	2.05391E-04	74.36
16	14	51	21	10^{-6}	0.006367	1.57050E-04	-346.11
16	14	51	21	10^{-7}	0.002303	4.34181E-05	-441.16
19	13	43	4	10^{-8}	0.002769	3.61129E-06	-244.25

Table3. Minimum value of CV for each j

n_x	n_y	j	\tilde{x}_j	m_j
10	10	1	5.824489	384.30
10	11	2	5.585633	330.07
		\vdots		
20	19	120	6.232998	-150.10
20	20	121	6.200994	-159.55

Table4. Statistical values of x_{\min}

average	6.56285
median	6.55971
standard deviation	0.10246
maximum	6.85326
minimum	6.32370

Table5. Model evaluation for $\beta_{\min} = 10^{-6.85326}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00240204	-462.82
15	14	60	80	0.00369185	-429.77
17	14	82	79	0.00271836	-425.07
		\vdots			

Table6. Model evaluation for $\beta_{\min} = 10^{-6.67209}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00261375	-473.44
15	14	60	80	0.00391828	-457.14
16	15	72	99	0.00325389	-429.55
		\vdots			

Table7. Model evaluation for $\beta_{\min} = 10^{-6.53701}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00288071	-469.07
15	14	60	80	0.00420413	-457.27
16	15	72	99	0.00356967	-441.77
		\vdots			

Table8. Model evaluation for $\beta_{\min} = 10^{-6.32370}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00364017	-439.82
16	15	72	99	0.00438565	-429.78
15	14	60	80	0.00502962	-421.37
		\vdots			

Table9. Values of CV for randomized

$-\log\beta$	β	λ	CV
6.66160	2.1796×10^{-7}	0.00263041	-473.493
6.65547	2.2106×10^{-7}	0.00264043	-473.490
6.67365	2.1200×10^{-7}	0.00261131	-473.435
6.64192	2.2807×10^{-7}	0.00266333	-473.407
6.64123	2.2843×10^{-7}	0.00266452	-473.400
	\vdots		

6. Conclusion

We can reduce the amount of calculations by almost one eighth. At first, we only used one set of each number of knots. Although the estimated value of β_{\min} depends on the selection of sets, we need not mind the difference because the standard deviation of x_{\min} is quite small. Based on β_{\min} , we can determine the best set of knots. Also, the determination does not depend on the value of the estimated β_{\min} .

Concerning the selected optimal set of knots, we finally obtained the optimal smoother parameter β by randomization, and its effective digit became higher than the previous method.

It is our future study how to find out set a more appropriate range of β in the final step.

References

- [1] Bao, H. and Fueda, K.(2013). "A New Method for the Model Selection in B-spline Surface Approximation with an Influence Function"*Science Journal of Applied Mathematics and Statistics*(submitted to).
- [2] Bao, H. and Fueda, K.(2013). "A Method for Topographical Estimation of Lake Bottoms by B-spline Surface"*American Journal of Theoretical and Applied Statistics Mathematics and Statistics*2(4): 102-109.

- [3] Cox, M.G.(1972). "The numerical evaluation of B -splines", *J. Inst. Math. Appl.*, 10, pp.134-149.
- [4] Cox, M.G.(1975). "An algorithm for spline interpolation", *J. Inst. Math. Appl.*, 15, pp.95-108.
- [5] de Boor, C.(1972). "On calculation with B -splines", *J. Approx. Theory*, 6, pp.50-62.
- [6] Schoenberg, I. J., Whitney, A.(1953). "On Pólya frequency functions III", *Trans. Amer. Math. Soc*, Vol. 74. pp. 246-259, pp. 246-259.
- [7] Good, I. J. and Gaskins, R.A.(1971). "Non parametric roughness penalties for probability densities", *Biometrika*, Vol. 58. pp. 255-277.
- [8] Good, I. J. and Gaskins, R.A.(1980). "Density estimation and bump hunting by the penalized likelihood method exemplified by scattering and meteorite data", *Journal of American Standard Association*, Vol. 75. pp. 42-56.
- [9] Green, P. J., Silverman, B. W.(1994). "*Nonparametric Regression and Generalized Linear Models*", Chapman and Hall, London.
- [10] Umeyama, S. (1996). "Discontinuity extraction in regularization using robust statistics", *Technical report of IEICE*, PRU95-217 (1996). pp. 9-16.