Korean J. Math. **22** (2014), No. 2, pp. 289–305
http://dx.doi.org/10.11568/kjm.2014.22.2.289

NEW SELECTION APPROACH FOR RESOLUTION AND BASIS FUNCTIONS IN WAVELET REGRESSION

Chun Gun Park

Abstract. In this paper we propose a new approach to the variable selection problem for a primary resolution and wavelet basis functions in wavelet regression. Most wavelet shrinkage methods focus on thresholding the wavelet coefficients, given a primary resolution which is usually determined by the sample size. However, both a primary resolution and the basis functions are affected by the shape of an unknown function rather than the sample size. Unlike existing methods, our method does not depend on the sample size and also takes into account the shape of the unknown function.

1. Introduction

In wavelet representation-based nonparametric regression, the focus has been on thresholding the wavelet coefficients, given a primary resolution which is usually determined by a sample size. By intuition, a proper primary resolution might be affected by the shape of an unknown function rather than the the sample size [10, 11].

Usually a shrinkage procedure involves determining a thresholding value and then applying a shrinkage function to the coefficients using the threshold. Several approaches to thresholding have been studied [1, 3–5, 9, 11, 14]. Donoho and Johnstone [3, 4] proposed VisuShrink

Received April 8, 2014. Revised May 28, 2014. Accepted May 28, 2014.
2010 Mathematics Subject Classification: 60A05, 62C10, 62J05, 62G08.
Key words and phrases: Bayes factor, Positerior model probability, Primary resolution, Wavelet basis functions, Wavelet regression.
This research was supported by Kyonggi University Research Grant 2010.

and SureShrink as minimize approaches. Nason [9] considered a cross-validation method, and Antoniadis and Fan [1] introduced nonlinear regularized wavelet estimators for estimating nonparametric regression functions when the sampling points are not uniformly spaced. Their method was developed using the smoothing clipped absolute deviation penalty. Vidakovic and Ruggeri [14] proposed a Bayesian Adaptive Multiresolution Shrinker method to address the problem of model-induced wavelet shrinkage. They showed that their approach resulted in simple optimal shrinkage rules to be used in fast wavelet denoising to address the problem of model-induced wavelet shrinkage. Johnstone and Silverman [5] investigated an empirical Bayesian thresholding rule. Park *et al.* [11] also proposed a Bayesian method to select a primary resolution and the wavelet basis functions with the thresholding value determined by the Bayes factor. Although Park *et al.*'s approach is able to select a primary resolution, which does not depend on the sample size, their selection method based on the Bayes factor cannot reflect the characteristic of the unknown function due to the criterion based on the Bayes factor.

To the best of our knowledge, none of these existing thresholding-based approaches can reflect the shape of the unknown function to select a primary resolution. If the unknown function is a smooth function, we have, by intuition, a high chance to select a low resolution level regardless of the a sample size. On the other hand, if an unknown function is a wiggled function, we can select a high resolution level [10, 11]. Hence, the primary resolution is not affected by the sample size, but the shape of the unknown function.

Therefore, in this paper, we propose a new approach to determine a primary resolution and wavelet basis functions. Our approach takes into account the pattern of posterior probabilities for nested models which depend on the shape of the unknown function. From Figure 1, our idea is motivated by the fact that ideal nested models have the special structure of the posterior probabilities. For an example, if the true nested model contains only one independent variable, then the posterior probabilities for any nested models are one. If the true model has only three independent variables for the nested models, the posterior probability corresponding to a nested model is less than one until including the three variables. Hence, it can be applied to any sample size, unlike typical discrete wavelet transformations with the sample size of the form $n = 2^J$ for some positive integer J. Furthermore, our approach selects a primary resolution and wavelet basis functions based on the true underlying function estimated by computing the corresponding wavelet coefficients of the selected wavelet basis functions. Therefore, if the chosen primary resolution based on the sample size is too high, our approach can reduce the computational cost. In contrast, if the chosen primary resolution based on the sample size is too low, the accuracy of fitting the unknown function could be increasing.

This article is organized as follows. In Section 2, we briefly review wavelet regression. Section 3 describes the motivation of our problem and our new approach to selecting a resolution and wavelet basis functions. In Section 4, we conduct a simulation to report some results for test functions with several sample sizes. Section 5 contains conclusion and further works.

2. Wavelet regression

Suppose that we have n paris of observations (x_i, y_i) and consider the following nonparametric regression

$$y_i = f(x_i) + \epsilon_i,\ \ i = 1, \ldots, n \tag{1}$$

where $f(\cdot)$ is an unknown function and ϵ_i's are independent and identically distributed random errors. We assume that x_i's are equally spaced points in the unit interval. The goal is to estimate the underlying function f. We can estimate $f(x)$ using a wavelet series. An orthogonal wavelet basis in $L_2(R)$ is a collection of functions obtained from translations and dilations of a scaling function ϕ and a mother wavelet ψ [2]. From (1) any function $f \in L_2(R)$, for any integer J_0, can therefore be represented by a wavelet series as

$$f(x) = \sum_{k\in Z} s_{(J_0,k)}\phi_{(J_0,k)}(x) + \sum_{j\geq J_0}\sum_{k\in Z} d_{(j,k)}\psi_{(j,k)}(x). \tag{2}$$

The scaling coefficients $s_{(J_0,k)}$ and the wavelet coefficients $d_{(j,k)}$ are defined to be $s_{(J_0,k)} = \int f(x)\phi_{(J_0,k)}(x)dx$ and $d_{(j,k)} = \int f(x)\psi_{(j,k)}(x)dx$, respectively, with smoothing function $\phi_{(J_0,k)}(x) = 2^{J_0/2}\phi(2^{J_0}x - k)$ and detailed function $\psi_{(j,k)}(x) = 2^{j/2}\psi(2^j x - k)$. For multi-resolution analysis, wavelet expansion with a primary resolution m is the orthogonal projection $P_s f$ of f onto a V-subspace which is a sequence of subspaces of functions in $L_2(R)$ [7]. It can be expressed in terms of the scaling function only

$$P_m f(x) = \sum_{k\in Z} s_{(J_0,k)}\phi_{(J_0,k)}(x) + \sum_{j\geq J_0}^{m}\sum_{k\in Z} d_{(j,k)}\phi_{(j,k)}(x) = \sum_{k\in Z} c_{(m,k)}\phi_{(m,k)}(x)$$

which implies that $\lim_{s\to\infty} P_s f(x) = f(x)$. Wavelet regression has in the following three steps:

- Step 1: Apply the discrete wavelet transformation (DWT) to y; DWT calculates the coefficients of the wavelet series expansion for a discrete signal of final extent. In other words, given a vector of function values $\mathbf{f} = (f(x_1), \ldots, f(x_n))^T$ at equally spaced points x_i, the DWT of $\mathbf{f}$ is given by

$$\mathbf{f} = \mathbf{W}\boldsymbol{\theta}$$

  where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ is a vector consisting of both discrete scaling and wavelet coefficients and $\mathbf{W}$ is an orthogonal $n \times n$ matrix associated with the chosen orthogonal wavelet basis. When x_i's are equally spaced in the interval [0,1], W consists of the values of the scaled wavelet basis functions $\tilde{\phi}_{m,k}(x_i) = n^{-1/2}\phi_{m,k}(x_i)$. We use Daubechies wavelets because of their nice properties such as orthogonality, compact support and different degrees of smoothness.
- Step 2: Threshold the empirical wavelet coefficients; this step is crucial as it is well known that shrinking the coefficients determines the properties of $\hat{f}$. Usually a shrinkage procedure involves determining a thresholding value and then applying a shrinkage function to the coefficients using the threshold.
- Step 3: Perform the the inverse DWT (IDWT) to get $\hat{\mathbf{f}}$; Because of the orthogonality of $\mathbf{W}$, the coefficients are obtained simply by IDWT

$$\boldsymbol{\theta} = \mathbf{W}^T\mathbf{f}.$$

  The corresponding wavelet coefficients are set to be $\theta_k = n^{1/2}c_{s,k}$. We note that in the DWT, it is required that the sample size is $n = 2^J$ for some positive integer J. However, our approach, which will be explained in the next section, does not depend on the sample size.

For a predetermined primary resolution m, these three steps for wavelet regression are typically performed. The primary resolution is usually determined solely by the sample size. However, a proper primary resolution might be affected by the shape of the underlying function rather than by the sample size. Therefore, in the next section, we propose a new method to select a primary resolution as well as the basis functions. Under a Bayesian framework, we propose a method of selecting a subset of wavelet basis functions, which may be viewed as a version of hard thresholding whose a thresholding value is determined by the pattern of posterior probabilities of the subsets of the wavelet basis functions. The true underlying function is estimated by the corresponding wavelet coefficients of the selected wavelet functions.

3. New selection method for resolution and basis functions

Given a primary resolution m, the model (2) can be defined as

$$Y = \mathbf{W}\mathbf{\Theta} + \epsilon \tag{3}$$

where $Y = (y_1, ..., y_n)^T$, $\mathbf{\Theta} = [\theta_{-1}, \theta_0, ..., \theta_m]$, $\mathbf{W} = [W_{-1}, W_0, ..., W_m]$, and $\boldsymbol{\epsilon}$ follows $N(0, \sigma^2 I_n)$.

Here θ_{-1} is the $n \times q(0)$ column vector of the coefficients corresponding to the wavelet basis functions in the smoothing part, $\{\phi_{0,k}\}$, and θ_i, $i = 0, 1, ..., m$, are the $n \times q(i)$ column vector of the coefficients corresponding to the wavelet basis functions in the detailed part, $\{\psi_{i,k}\}$, where $q(i)$ denotes the number of wavelet basis functions at the resolution i. Further define $W_{-1} = \{\phi_{(0,k)}\}$ and $W_i = \{\psi_{(i,k)}\}$, $i = 0, 1, ..., m$.

For the choice of a primary resolution, we propose a Bayesian model selection method based on the posterior model probability. However, direct application of a Bayesian model selection method has several disadvantages: (i) it tends to select a low primary resolution regardless of the unknown functions as the sample size increases, (ii) it is computationally expensive because the number of candidate models is huge and furthermore, the number of wavelet coefficients increase rapidly as the level of the primary resolution becomes high, and (iii) its selection criterion is based on the Bayes factor and cannot reflect the characteristic of the unknown function.

Therefore, we consider a new method under nested models [11]

$$\text{Model } -1, \; y = W_{-1}\theta_{-1} + \epsilon$$

$$\begin{aligned}
&\text{Model 0,} \quad y = W_{-1}\theta_{-1} + W_0\theta_0 + \epsilon \\
&\qquad\qquad \vdots \\
&\text{Model m,} \quad y = W_{-1}\theta_{-1} + W_0\theta_0 + \ldots + W_m\theta_m + \epsilon,
\end{aligned}$$

and use noninformative prior distributions of θ_m and m in order to obtain a closed form of the posterior probability so that the computational cost is reduced. Selecting the primary resolution is equivalent to finding the best model. For each model on a given resolution level, we calculate the posterior model probability using the following priors distributions;

- Given σ^2 and m, the prior for θ_m is

$$p(\theta_m|\sigma^2, m) \propto \text{constant}$$

- The prior for each model M is

$$p(M) = \frac{1}{m+2}, \; M = -1, 0, 1, \ldots, m.$$

Using these priors and the orthogonal property $(W_m^T W_m = I_{q(m)})$, a posterior probability of the primary resolution parameter m is then given by

$$\begin{aligned}
p(m|Y, \sigma^2) &\propto \int p(m)p(y|\theta_m, \sigma^2, m)p(\theta_m|\sigma^2, m)d\theta \\
&\propto \exp(\frac{1}{2}Y^T W_m W_m^T Y). \qquad (4)
\end{aligned}$$

To select a primary level of resolution, the Bayes factor might be used. However, we note that the selection based on the Bayes factor does not take into account the characteristics of the unknown function. If the unknown function is smooth, we might select the low resolution level because the posterior probability already reaches 1 at the low resolution level. If the unknown function is a wiggled function, we can select the high resolution level because the posterior probability reaches 1 at the high resolution level. Hence the change pattern of posterior probability rapidly reaches 1 when the unknown function is smooth, while slowly goes to 1 when the unknown function is wiggled. However, the selection based on the Bayes factor is not able to incorporate these important facts. Hence we propose a new selection procedure for a primary resolution based on the pattern of the posterior probability as well as the characteristic of the unknown function.

In convenience of notation, for a selected resolution m_0, we rewrite the wavelet regression (2) and (3) in order to select the basis functions as follows:

$$(5) \qquad y_i = \varphi_1(x_i)\theta_1 + \cdots + \varphi_N(x_i)\theta_N + \epsilon_i, \ i = 1, ..., n,$$

where until the primary resolution, m_0, θ_i is the ith wavelet coefficient corresponding to φ_i being the ith wavelet basis function, and N is the number of the wavelet basis functions which are defined such that

$$|\sum_{i=1}^{n} \varphi_1(x_i)y_i| \geq |\sum_{i=1}^{n} \varphi_2(x_i)y_i| \geq \cdots \geq \sum_{i=1}^{n} |\varphi_N(x_i)y_i|.$$

From (5) we can obtain the same form of the posterior model probability except for the order of the wavelet basis functions in the design matrix and the posterior model probability for the model selection is

$$(6) \qquad p(k|Y, X, \sigma^2) \propto \exp(\frac{1}{2\sigma^2} Y^T B_k B_k^T Y), \ k = 1, 2, \ldots, N$$

where $B_k = [\{\varphi_1(X)\}, \ldots, \{\varphi_k(X)\}]$ is a $n \times k$ matrix and $X = (x_1, ..., x_n)^T$.

3.1. Motivation for the selection of nested models.

The motivation of our new selection procedure for a primary resolution and wavelet basis functions is to use the pattern of posterior probabilities for the concave nested models. When the shape of the unknown function is very smoothing and the sample size is very large, i.e. more than 2^{10}, the fitted functions may be overestimated and the computational cost using MCMC is very expensive.

To overcome these problems for the wavelet regression model selection, we employ ideal nested models from Figure 1. If the optimal model is Model 1, then the posterior probabilities for all other nested models, except for Model -1 and Model 0, are all 1. However the pattern of actual posterior probabilities might be the same as that of the ideal nested models. Therefore, we want to find the proper resolution level such that the pattern of actual posterior probabilities might be similar to that of ideal nested models.

3.2. The procedure for the selection of nested models.

We want to find the resolution (or basis functions) whose posterior probability is the most close to the ideal posterior probability. To do this, we proceed using the following procedure for selection of resolution using

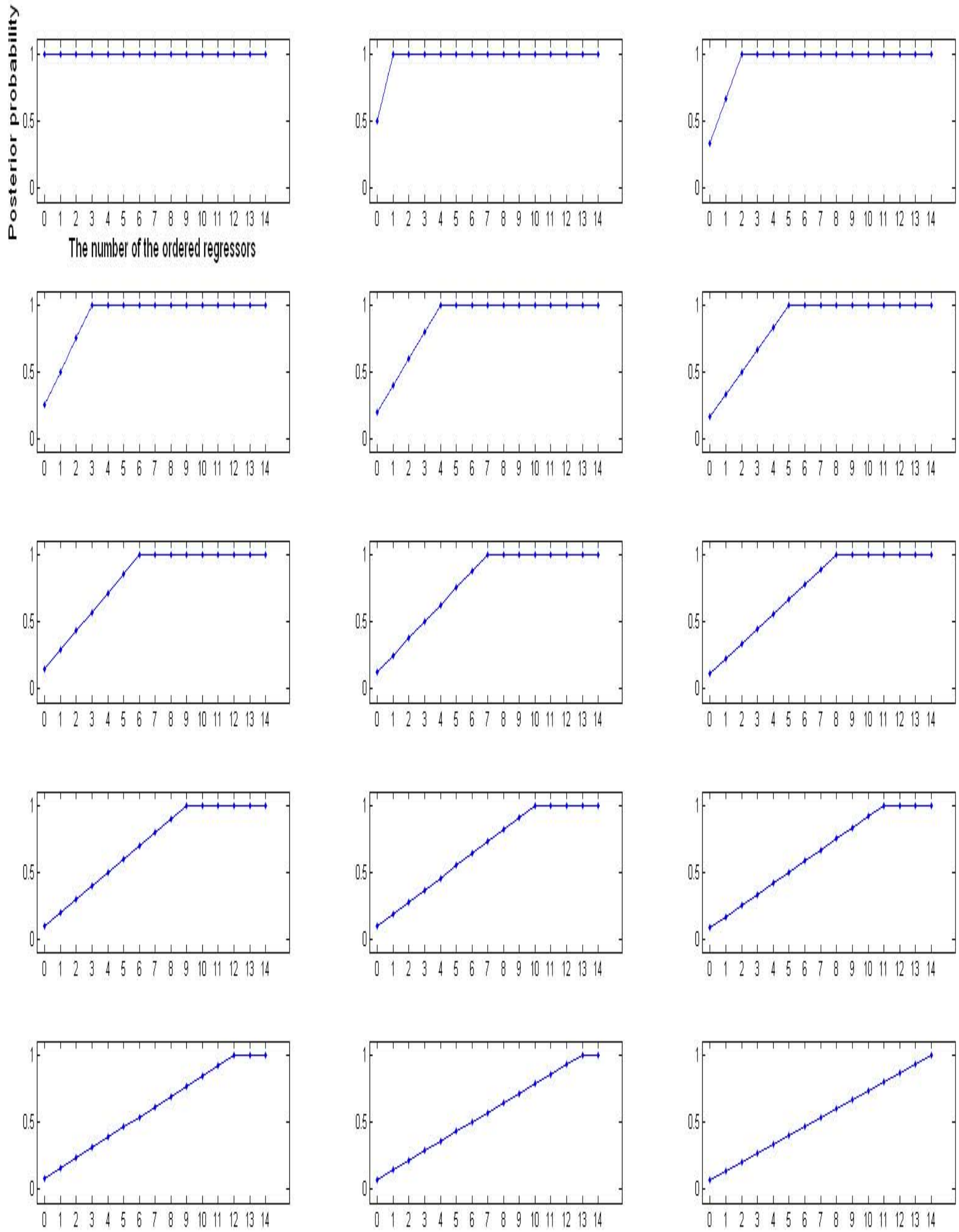

FIGURE 1. Posterior probabilities under ideal nested models

the relative ratio between the ideal posterior probability and the observed posterior probability. Consider the plot (Figure 2) between the levels of resolution and posterior probabilities.

- Step 1: Calculate posterior probabilities from (4) or (6).
- Step 2: Construct the rectangle with four points $(i,\ p(i|Y,\sigma^2))$, $(i+1,\ p(i|Y,\sigma^2))$, $(i,\ 1)$, $(i+1,\ 1)$ and obtain the area of this rectangle.
- Step 3: Within this rectangle, obtain the linear line passing through the two points $(i,\ p(i|Y,\sigma^2))$ and $(i+1{,}1)$.
- Step 4: Within this rectangle, obtain the area under the linear line to pass two points $(i,\ p(i|Y,\sigma^2))$ and $(i+1{,}1)$; define this area as A.
- Step 5: Obtain the area under the linear line passing through the two points $(i,\ p(i|Y,\sigma^2))$ and $(i+1,\ p(i+1|Y,\sigma^2))$; define this area as A_i.
- Step 6: Obtain $B_i = A - A_i$.
- Step 7: Repeat Step 2-6 for all i and obtain A_i and B_i.
- Step 8: Calculate ratios R_i

$$R_i = \frac{B_i}{A_i + B_i},\ \ i = 1, 2, \ldots, m.$$

- Step 9: Find a change resolution point such that

$$m_0 = \{j|\min\ R_j\ j = 2, \ldots, m\}$$

provides a dramatic change of the posterior probability. If there is no such change point, then $m_0 = m$, which is the maximum of resolutions obtained from

$$m_0 = \text{argmin}_m abs(Y^TY - \hat{Y}_m^T\hat{Y}_m),\ \ m = -1, 0, 1, ...$$

where $\hat{Y}_m$ is the fitted Y values on the resolution m using the wavelet coefficients estimated by quadrature type estimator [11]. Here m_0 is called a maximum resolution.

4. Simulation study

In simulation study we fucus on determining a primary resolution. To do this, we consider the following four nonlinear functions $f(x)$ (see Figure 3),

- *Cosine* function: $f(x) = 0.5\cos(2.2\pi/3 + 8x) + 0.5$;

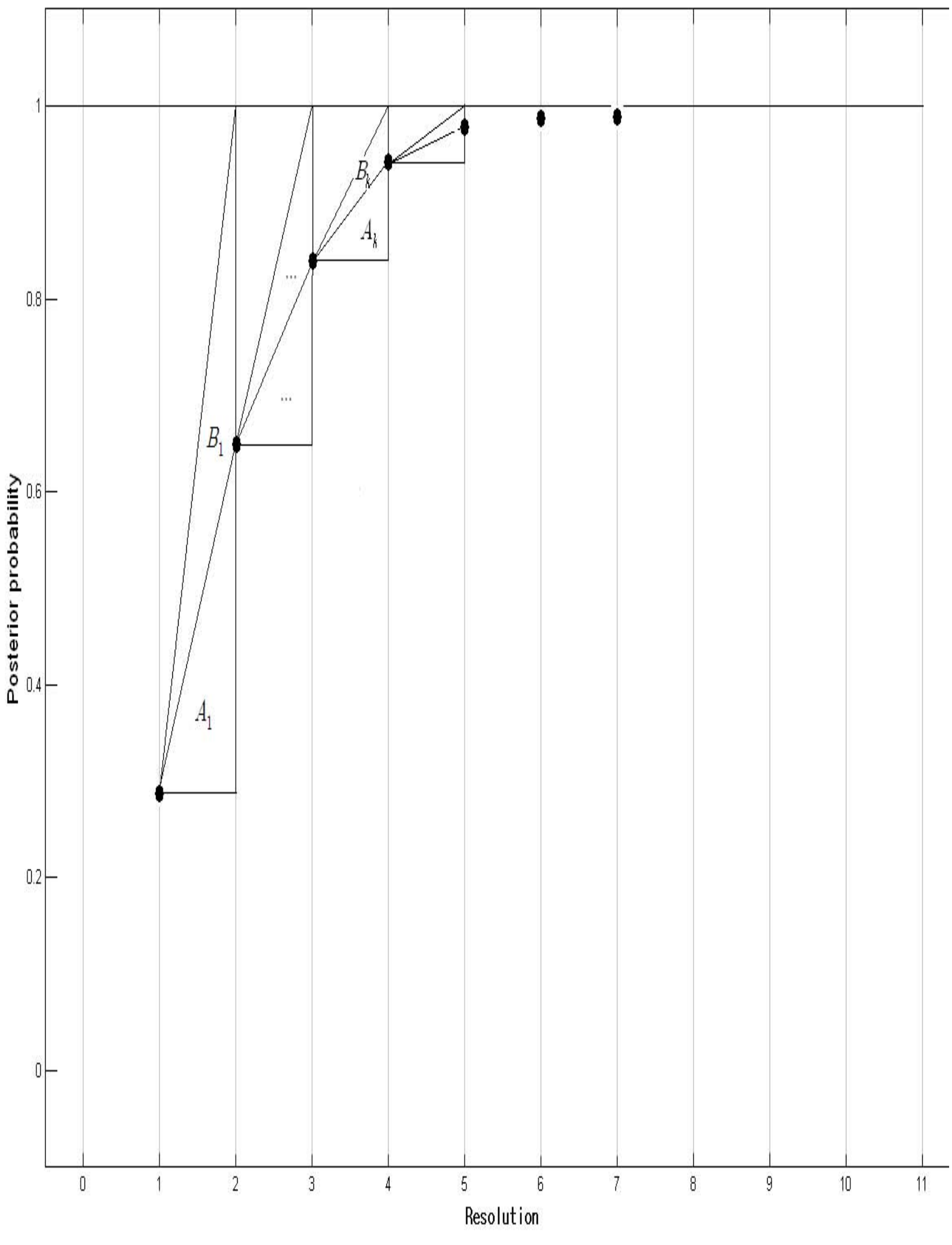

Figure 2. Resolution selection method

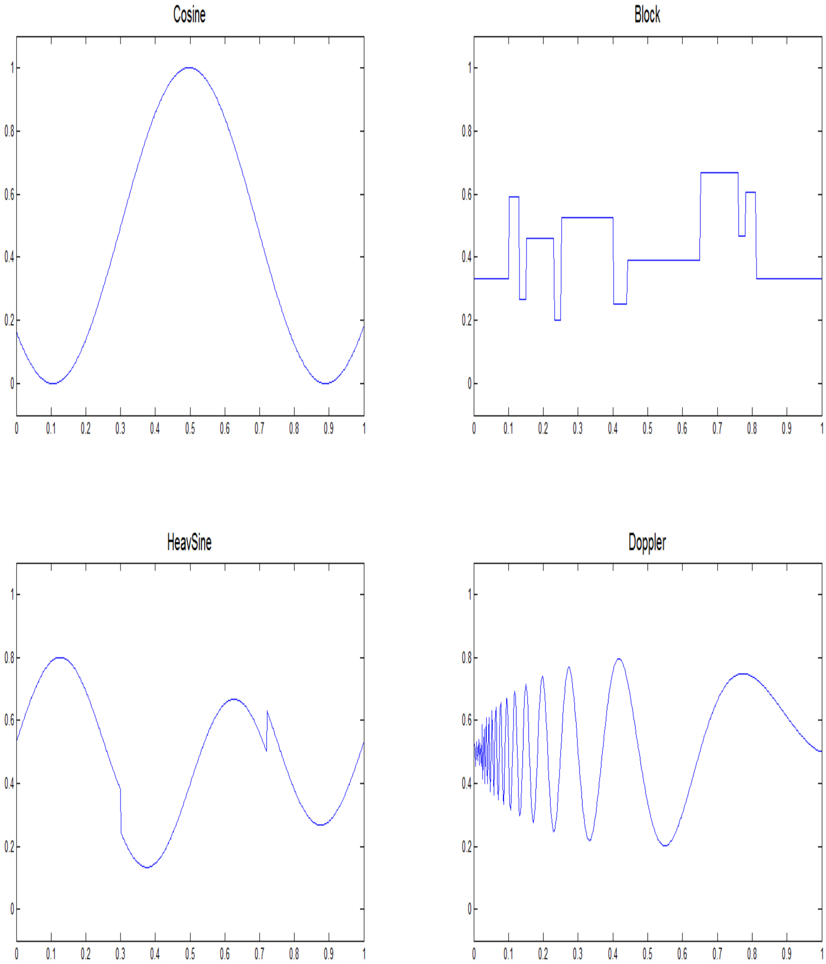

FIGURE 3. Four True functions used for simulation studies

- *Block* function: $f(x) = \sum_{j=1}^{11} h_j K(x - t_j)$,
  where $K(t) = \{1 + sign(t)\}/2$, $t_j = (.1, .13, .15, .23, .25, .40, .44, .65, .76, .78, .81)$ and $h_j = (4, -5, 3, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2)$;
- *Heavy Sine* function: $f(x) = 4\sin(4\pi x) - sign(.72 - x)$;
- *Doppler* function: $f(x) = 0.6(\sqrt{x(1-x)})\sin\{2.1\pi/(x + 0.05)\} + 0.2$;

with two cases of error variances, σ^2=(0.05^2, 0.1^2), and eight cases of sample sizes, $n = (30, 32, 100, 128, 500, 512, 2000, 2048)$. Note that the

Cosine function has a relatively smooth compared to other functions, while *Doppler* function is the most wiggled function among them. Neither the *Block* function nor the *Heavy Sine* function are continuous functions. We expect that the *Cosine* function has a lower primary resolution, while the other functions might have higher primary resolutions.

Figure 4-11 show the results of selecting primary resolutions for each function with 100 repetitions. For the *Cosine* function, our selected resolutions are 2 under all combinations. For the *Block* function, our selected resolution is 4 when the sample size is 30 or 32 and it is 5 when the sample size is 100 or 128. However, it is 7 when the sample size is 500 or 512, while it is 9 when the sample size is 2000 or 2048. For the *Doppler* function, our selected resolutions are similar to the maximum resolution. For the *HeaviSine* function, our selected resolutions are most 3, while the maximum resolution increases as the sample size increases.

Therefore, the proposed method estimates lower primary resolutions for the *Cosine* function and the *HeaviSine* function, and higher resolutions for other functions.

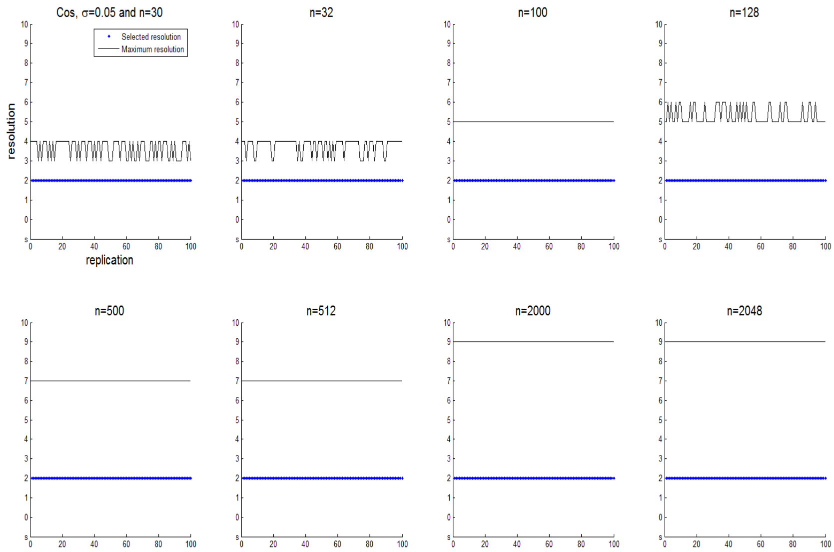

FIGURE 4. Resolutions selected for the cosine function with $\sigma = 0.05$

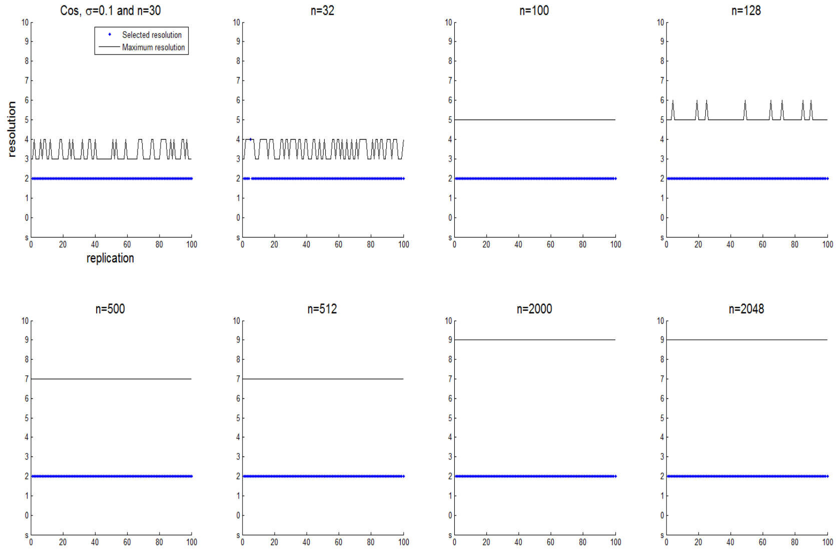

FIGURE 5. Resolutions selected for the cosine function with $\sigma = 0.1$

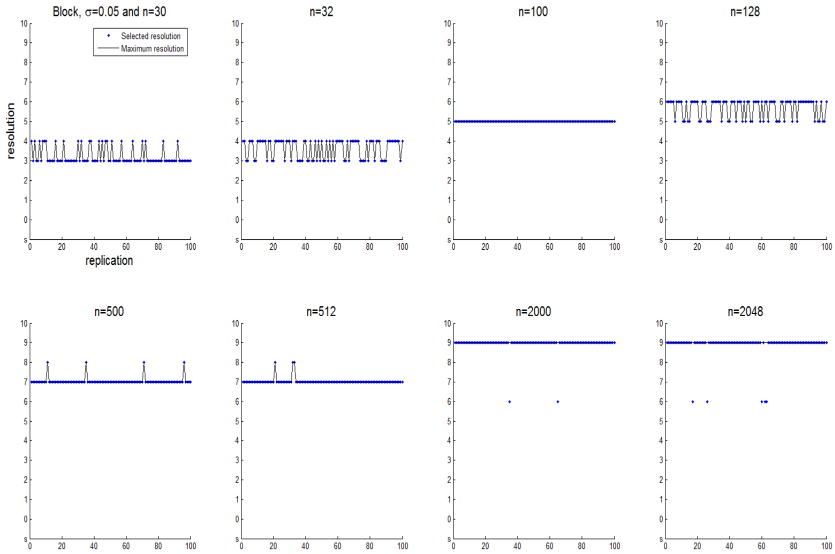

FIGURE 6. Resolutions selected for the block function with $\sigma = 0.05$

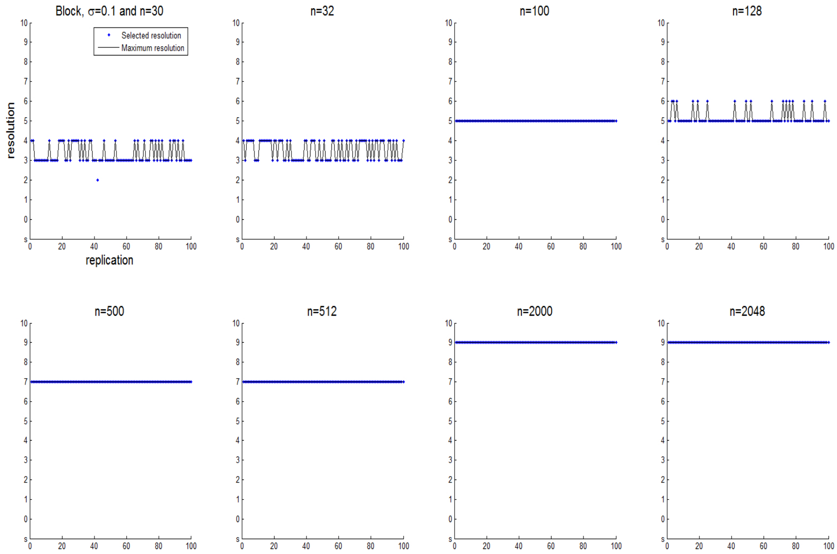

FIGURE 7. Resolutions selected for the block function with $\sigma = 0.1$

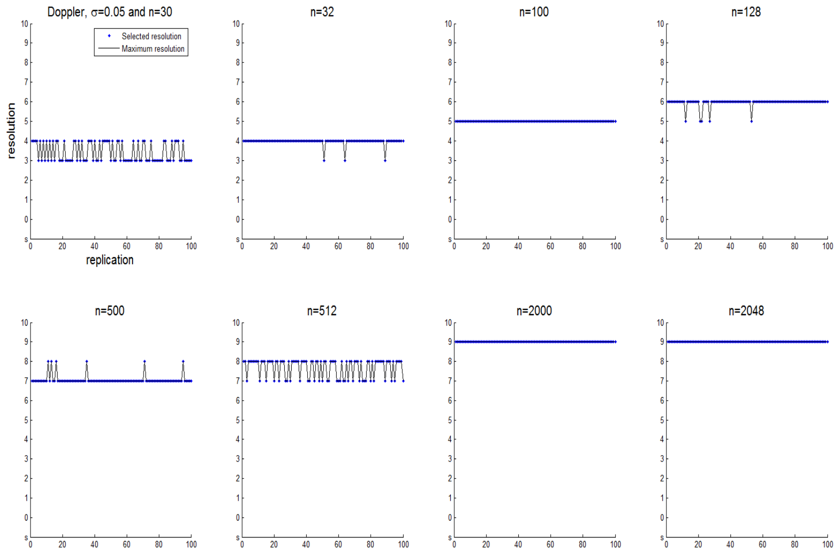

FIGURE 8. Resolutions selected for the doppler function with $\sigma = 0.05$

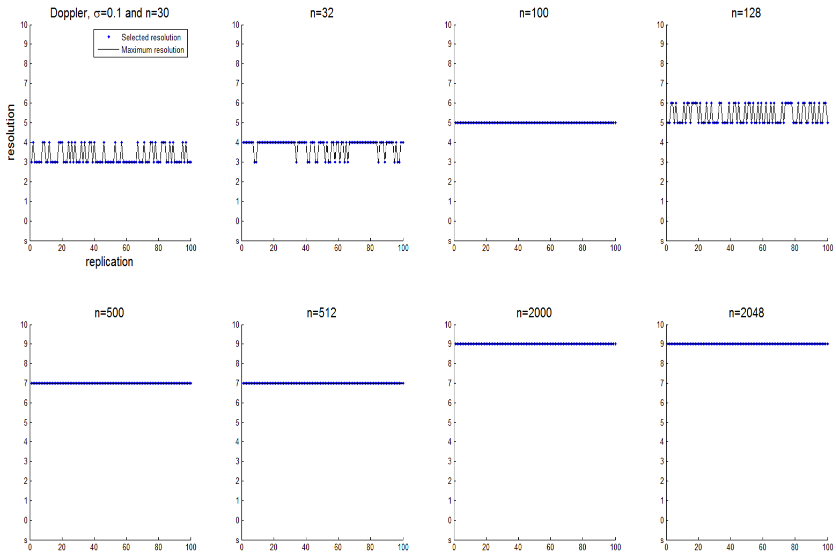

FIGURE 9. Resolutions selected for the doppler function with $\sigma = 0.1$

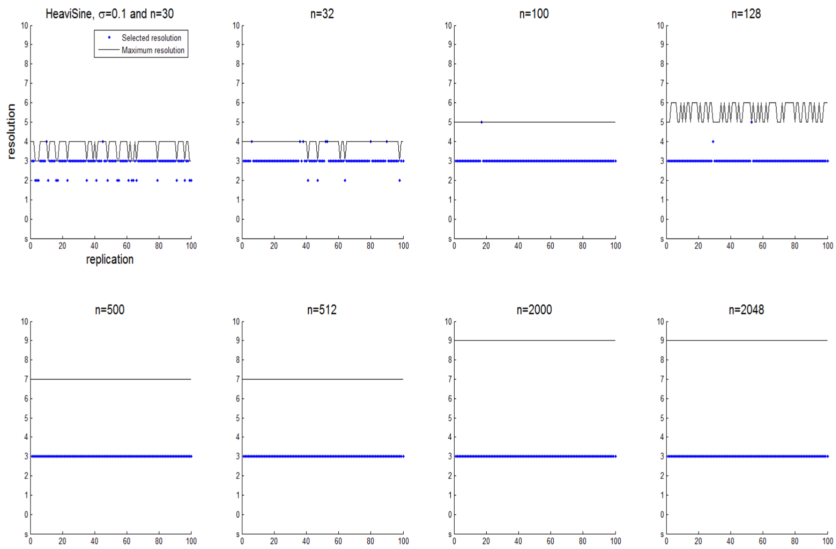

FIGURE 10. Resolutions selected for the heavisine function with $\sigma = 0.05$

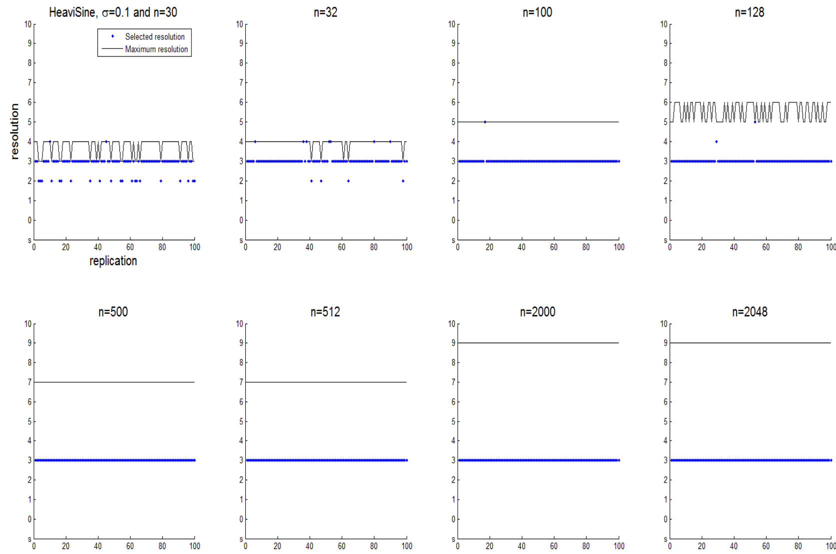

FIGURE 11. Resolutions selected for the heavisine function with $\sigma = 0.1$

5. Conclusion

In this paper, we proposed a new method for selecting the primary resolution and the wavelet basis functions. The main contribution of our proposed method is that it does not depend on the sample size, which is a major restriction of the classical wavelet shrinkage method and also provides new criterion selection based on the characteristics of the unknown function. To the best of our knowledge, none of the existing thresholding based approaches can reflect the characteristics of the unknown function to select a primary resolution and wavelet basis functions. The simulation study suggests that our approach can be an alternative to a classical nonlinear wavelet shrinkage. Our approach can be applicable to any type of functions including non-smoothing functions. However, our approach is based on equally spaced data. It is worthwhile future research to extend to the case of unequal spaced data.

References

[1] A. Antoniadi and J. Fan, *Regularization of Wavelet Approximations*, J. Amer. Statist. Assoc. **96** (2001), 939–955.

[2] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, (1992).

[3] D.L. Donoho and I.M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81** (1994), 425–455.

[4] D.L. Donoho and I.M. Johnstone, *Adapting to unknown smoothing via wavelet shrinkage*, J. Amer. Statist. Assoc. **90** (1995), 1200–1224.

[5] I.M. Johnstone and B.W. Silverman, *Empirical Bayes selection of wavelet thresholds*, Ann. Statist. **33** (2005), 1700–1752.

[6] J.D. Hart, Nonparametric Smoothing and Lack-of-Fit Tests, Berlin: Springer Verlag, (1997).

[7] S.G. Mallat, *A theory for multiresolution image denoising schemes using generalized Gaussian and complexity priors*, IEEE Transactions on Pattern Analysis and Machine Intelligence **11** (1989), 674–693.

[8] M. Misiti, Y. Misiti, G. Oppenheim and J.M. Poggi, Wavelet toolbox for use with MATLAB, The Math Works Incorporation, (1994).

[9] G.P. Nason, *Wavelet shrinkage by cross-validation*, J. R. Stat. Soc. Ser. B Stat. Methodol. **58** (1996), 463–479.

[10] C.G. Park, M. Vannucci and J.D. Hart, *Bayesian Methods for Wavelet Series in Single-Index Models*, J. Comput. Graph. Statist. **14** (4) (2005), 770–794.

[11] C.G. Park, H.S. Oh and H. Lee, *Bayesian selection of primary resolution and wavelet basis functions for wavelet regression*, Comput. Statist. **23** (2008), 291–302.

[12] M. Smith and R. Kohn, *Nonparametric regression using Bayesian variable selection*, J. Econometrics **75** (1996), 317–343.

[13] M. Smith and R. Kohn, *A Bayesian approach to nonparametric bivariate regression*, J. Amer. Statist. Assoc. **92** (1997), 1522–1535.

[14] B. Vidakovic and F. Ruggeri, *BAMS method: theory and simulations*, The Indian Journal of Statistics, Series B, **63** (2001), 234–249.

[15] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, NY, (1999).

Chun Gun Park
Department of Mathematics
Kyonggi University
Gyeonggi-Do 443-760, Korea
E-mail: cgpark@kgu.ac.kr