Wavelet Estimator in Nonparametric Regression to improve Least Squares Estimation

E.M. de Souza - emsouza@uem.br - UEM, Maringá, Paraná, Brasil.
J.F.G. Monico - galera@fct.unesp.br - UNESP, Presidente Prudente, São Paulo, Brasil.

Abstract. A nonparametric regression with a wavelet estimator is proposed to correct undesired effects (bias) that, in general, can not be handle in functional or stochastical models of Least Squares method. An application was carried out with real data from Global Positioning System to correct the multipath effect from signal reflection. Improviments of up to 99% was reached in root mean squared error of the residuals.

1. Introduction

Nonparametric regression has been a fundamental tool in data analysis over the past three decades and is still an expanding area of ongoing research. The goal is to recover an unknown function f, based on sampled data that are contaminated with additive Gaussian noise [14]. Only very general assumptions about f are made such as that it belongs to a certain class of smooth functions. Nonparametric regression (or denoising) techiques provide a very effective and simple way of finding structure in data sets without the imposition of a parametric regression model (as in linear or polynomial regression, for example) [1].

Various nonlinear estimators in nonparametric regression have been proposed. The most popular are variable-bandwidth kernel methods and adaptive regression splines [1]. Although some of these methods achieve the optimal asymptotic rates, they can be computationally intensive and they are usually designed to denoise regular functions.

[5] have introduced nonlinear wavelet estimators in nonparametric regression through thresholding. These function classes model the notion of different amounts of smoothness in different locations more effectively than the usual smooth classes. The performance of these fuction classes is, in general, described in terms of convergence rates that are achieved over large function classes, however, concise accounts of mean squared error for single functions also exist and have been discussed in detail by [8]. It has also been shown that nonlinear wavelet thresholding estimators are asymptotically near optimal or optimal while traditional linear estimators are suboptimal for estimation over some particular classes of spaces [6]. Indeed, the good behaviour of wavelet estimators for diferent sizes of samples was described by [10] and [1].

A variety of different wavelet families exist that combine compact support with various degrees of smoothness and numbers of vanishing moments [4], and these have been most intensively used in practical applications in statistics. Many types of functions encountered in practice can be sparsely (i.e. parsimoniously) and uniquely represented in terms of a wavelet bases, due to their special structure.

Wavelets have been a real revolution in science and technology due to the high number, quality, efficiency and diversity of their applications in different areas. Especially, the wavelet theory represents a tool for signal analysis in Phisics, voice and pattern recognition, image codification and segmentation, denoising, density estimation and regression. Moreover, wavelet shrinkage and wavelet thresholding estimators can be implemented through fast algorithms.

In this paper it is discussed about the wavelet estimator in nonparametric regression to estimate effects that can not be handled in functional or stocastical models of Least Squares (LS) method. If the function model is adequate, the residuals obtained from the LS solution should be randomly distributed. However, the observations are sometimes contaminated by several types of biases and not all can be correctely taken into acount. Consequently, the obtained residuals are not randomly distributed. This is the case of observations from Global Positioning System (GPS), where a double-differencing (DD) technique is commonly used for constructing the functional model as it can eliminate or reduce many of the GPS biases (atmospheric, orbital, receiver and satellite clock biases) for short baselines [11]. However, the multipath effect, due to the signal reflections, is not eliminated because it depends of the geometry and environment of each GPS receiver station. Therefore, multipath is a major residual error source in the DD GPS observables.

The multipath effect distorts the signal because it is always delayed compared to line-of-sight signal due to the longer travel paths caused by the reflection. If the multipath effect is from long delays, it is characterized by high frequencies; otherwise, it is of low frequencies. The low-frequency multipath from short delays causes the largest errors, but the complety effect is a sum of different multipath that change during the time because of the satellite moviment. This makes this effect very difficult to be reduced or modeled by classical methods. [15] used the wavelets multiresolution analysis to reduce the high-frequency multipath from the DD GPS temporal series. However, the low-frequency multipath is very difficult to be reduced directly from the DD time series.

Thus it was proposed the wavelet regression (WR) to estimate the multipath effect from the residual DD temporal series, which is obtained from the LS solution. Then, the multipath components are directly corrected in the functional model.

Section 2 recalls the wavelet theory and the discrete wavelet transform. Section 3 briefly discusses the nonlinear wavelet approach to nonparametric regression of the residual time series. The experiment with real data and results are present in section 4. Finally, the conclusions are given in section 5.

2. Wavelets Family

Wavelets are building block functions and localized in time and frequency. They are obtained from a single function ψ, called the mother wavelet, by translations and dilations [4]

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right),\ a,b \in R,\ a \neq 0, \tag{2.1}$$

where a represents the dilation parameter and b the translation parameter.

For some very special choices of ψ, a e b, $\psi_{a,b}$ constitute an orthonormal basis for $L_2(R)$. In particular, with the choice $a = 2^j$ and $b = k2^j$, with $j,k \in Z$, then there is ψ, such that $\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$ is an orthonormal basis [2]. But a remaining question is how to obtain the wavelet ψ. An approach was developed by [4] to construct this wavelet using the scaling function (or father wavelet) that is the solution to the equation:

$$\phi(t) = \sqrt{2}\sum_k h_k \phi(2t-k), \tag{2.2}$$

where h_k is given by

$$h_k = \sqrt{2}\int_{-\infty}^{\infty} \phi(t)\,\phi(2t-k)\,dt. \tag{2.3}$$

The equation (2.3) creates an orthonormal family in $L_2(R)$, $\phi_{j,k}(t) = 2^{-j/2}\phi\left(2^{-j}t-k\right),\ \ j,k \in Z$. In these conditions, ψ can be obtained by

$$\psi(t) = \sqrt{2}\sum_k g_k \phi(2t-k), \tag{2.4}$$

where g_k is

$$g_k = \sqrt{2}\int_{-\infty}^{\infty} \psi(t)\,\phi(2t-k)\,dt. \tag{2.5}$$

Thus, considering the orthonormal system $\{\phi_{j,k}(t), \psi_{j,k}(t), j,k \in Z\}$, $f(t) \in L_2(R)$ can be written by [12]

$$f(t) = \sum_k c_{J,k}\phi_{J,k}(t) + \sum_{j \leq J}\sum_k d_{j,k}\,\psi_{j,k}(t),$$

where J is the coarse scale and

$$d_{j,k} = \langle f, \psi_{j,k}\rangle = \sum_{n \in Z} g_n \langle f, \phi_{j-1,2k+n}\rangle = \sum_{n \in Z} g_{n-2k}c_{j-1,n} \tag{2.6}$$

$$c_{j,k} = \langle f, \phi_{j,k}\rangle = \sum_{n \in Z} h_n \langle f, \phi_{j-1,2k+n}\rangle = \sum_{n \in Z} h_{n-2k}c_{j-1,n} \tag{2.7}$$

with h_n and g_n given by (2.3) and (2.5).

The equations (2.6) and (2.7) represent the Mallat or pyramidal algorithm basis [9] and can be interpreted as low-pass and high-pass filters, respectively. This means that the high frequencies components (details or noise) of the signal are separated from the low frequencies ones in many resolution levels. The name multiresolution analysis comes from this idea. Furthermore, the equations (2.6) and (2.7) can be seem as a convolution process followed by a downsampling by two that applied to a sequence implies that the even or odd samples are eliminated. Consequently, the level j has half of the coefficients of the level $j-1$, indicating the pyramidal name.

Actually, the pyramidal algorithm performs the Discrete Wavelet Transform (DWT), which can be seem in matricial form as

$$d = Wf, \tag{2.8}$$

where d is an $n \times 1$ vector comprising both discrete scaling coefficients, $c_{j,k}$, and discrete wavelet coefficients, $d_{j,k}$, and W is an orthogonal $n \times n$ matrix associated with the orthonormal wavelet basis chosen. Several wavelet bases are available, such as Daubechies, Symmlets, Coiflets, among others. Details can be found in [4] and [17].

The Inverse DWT (IDWT) essentially performs the same operations associated with the DWT in opposite direction. Instead of downsampling, the signal is interpoted: zeros are add among the coefficients (upsampling). More details of pyramidal algorithm can be found in [2] and [9].

The great advantage of pyramidal algorithm is that for the wavelet transformation only O(N) operations are necessary.

3. Wavelet Regression Approach

Considering LS residuals (v_i) that can be modeled by standard univariate nonparametric regression:

$$v_i = f(t_i) - \sigma\varepsilon_i,\ i = 1, \ldots, n, \tag{3.1}$$

where $t_i = i/n$, ε_i are independent N(0, 1) random variables and the noise level σ may be known or unknown. The function f is desirable to be estimated without assuming any particular parametric structure for f, that is in this case, the multipath effect. One of the basic approaches to nonparametric regression is to consider the unknown function f expanded as a generalised Fourier series and then to estimate the generalised Fourier coefficients from the data. Thus, the original nonparametric problem is transformed to a parametric one, although the potential number of parameters is infinite. An appropriate choice of basis for the expansion is therefore a key point in relation to the efficiency of such an approach. A good basis should be parsimonious in the sense that a large set of possible response functions can be approximated well by only few terms of the generalized Fourier expansion employed. Wavelet series allow a parsimonious expansion for a wide variety of functions, including inhomogeneous cases. It is therefore natural to consider applying the generalized Fourier series approach using a wavelet series [1].

The aim is to estimate f with the minimum squared error:

$$\frac{1}{n}E\left\|\widehat{f}-f\right\|_{L^2}=\frac{1}{n}\sum_{i=0}^{n-1}\left\{\left[\widehat{f}\left(i/n\right)-f\left(i/n\right)\right]^2\right\}, \tag{3.2}$$

under the condition that with high probability, $\widehat{f}$ is at least as smooth as f.

The transform (2.8) applied to (2.4) generates

$$Wy = Wf + W\varepsilon \tag{3.3}$$

and as W is orthogonal, it transforms white noise in white noise. Thus, if $\omega_{j,\,k}$ are the wavelet coeficients of $f\left(t_i\right)$, by (3.3) we have

$$y_{j,k} = \omega_{j,k} + \sigma\, z_{j,k}, \tag{3.4}$$

where $z_{j,k} \sim i.i.d.\, N(0,1)$. In other words, (3.4) says that wavelet coeficients of a time series with noise can be write as wavelet coeficients without noise added to white noise [12].

The sparseness of the wavelet expansion makes it reasonable to assume that essentially only a few large $d_{j,k}$ contain information about the underlying function f, while small d_{jk} can be attributed to the noise which uniformly contaminates all wavelet coefficients. If we can decide which are the significant large wavelet coefficients, then we can retain them and set all others equal to zero, thus obtaining an approximate wavelet representation of the underlying function f. It is also advisable to keep the scaling coefficients $c_{j,k}$, the coefficients on the lower coarse levels, intact because they represent low-frequency terms that usually contain important components about the underlying function f. This makes part of the wavelet thresholding process in a wavelet regression estimator, that works as follows:

1) Find the DWT of LS residuals v_i to obtain the wavelet coefficients using the pyramidal algorithm as explained in section 2. In relation to the mother wavelet, Symmlets with 8 vanishing moments was used. [15] showed that, for GPS time series, this mother wavelet is better than other Symmlets and Daubechies wavelets.

2) Modify the wavelet coefficients by thresholding.

3) Reconstruct the function f (multipath effect) with the IDWT, as described in section 2.

The crucial part of the regression is step 2 [13]. Nearly all the relevant works on thresholding rules and optimal thresholding estimators are contained within [7] and [5]. The thresholding estimators presented in this work were analyzed by [15]. So, in this paper, only the thresholding estimators that presented the best performance for GPS time series will be described.

Let d_i, $i = 1, ..., n$ the wavelet coefficients. The hard threshold function (T_λ^R) given by

$$T_\lambda^R = \begin{array}{c} 0,\ |d_i| < \lambda \\ d_i,\ |d_i| \geq \lambda \end{array}, \tag{3.5}$$

where λ is the universal threshold parameter

$$\lambda = \widehat{\sigma}\sqrt{2\log n},$$

that provides a fast and automatic thresholding. The σ is the observation noise level and should be estimated from each temporal series residuals in the first decomposition level (the finest scale). As the empirical wavelet coefficients at the finest scale are, with a small fraction of exceptions, essentially pure noise, [5] proposed the following estimator:

$$\widehat{\sigma} = med\left\{\left|d_{J-1,k}\right|,\, 0 \leq k < n/2\right\}/0,6745,$$

where med is the median, $J-1$ is the finest scale and the factor $0,6745$ is statisticaly determined in [7].

By (3.5), if an estimate of a coefficient is sufficiently large in absolute value that is, if it exceeds a predetermined threshold λ then the corresponding term in the empirical wavelet expansion is retained; otherwise it is omitted. Therefore, the undesired parts (noise) are removed by modifying the wavelet coefficients and the wavelet regression can estimate the multipath effect from the residual temporal series. Then, the multipath components are directly correct in functional model.

To perform the methodology briefly discussed here, the methods were implemented in Fortran and C languages linked by Delay Link Libraries (DLL).

4. Experiment with real data and results

In order to test the proposed method based on WR to estimate effects that can not be handled in functional or stocastical models of LS method, an experiment with real GPS data was carried out. To cause the multipath effect, refletor plates were placed near the GPS receiver.

WR was applied to each residual temporal series to estimate the multipath effect as described in sections 2 and 3. Then, the multipath components were directly correct in the LS functional model and the LS was performed again. The LS with the wavelet regression is denoted by LSWR.

The residual temporal series obtained from LS and LSWR were compared. To ilustrate the results, one of the residual time series obtained from LS and LSWR methods is plotted in Figure 1.

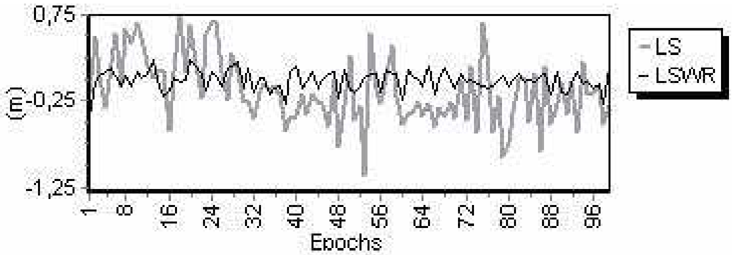

Figura 1: Residual time serie from LS estimation.

From all the time series analyzed, the improvement in the root mean squared error (RMSE) reached up 99%, which represented about 0.8 m of error reduction in the residuals.

To compare the quality of the observations, before (LS) and after (LSWR) the low-frequency multipath mitigation, the Global Overall Model (GOM) test statistic [18] was implemented. This test statistic has a Chi-squared distribution and allows a global detection of errors. It could be verified by GOM statistic that the observations quality improved after using the LSWR to correct the multipath.

Now, in order to evaluate the estimative accuracy using the LS and LSWR methods, the estimated coordinates (location of the GPS receiver) were compared with the known ones. After multipath mitigation using the LSWR, discrepancies between the coordinates were the smallest, improving 10 mm on average. The standard deviations were also significantly improved using the LSWR. Therefore, one can conclude that multipath was the main error affecting the accuracy of the coordinates and that the LSWR reduced significantly this error.

5. Conclusions

The wavelet estimator theory for nonparametric regression was briefly discussed and it seems to be a powerful method to correct effects that can not be handled in functional or stocastical models of LS method. Real data were used to estimate multipath effects and reduce them from GPS time series. In the results, the RMSE of the residuals improved up to 99% using the LSWR method. Furthermore, after multipath estimation and reduction using LSWR, the accuracy of the estimatives (coordinates) improved up to 16mm.

Referências

[1] Antoniadis, A.; Bigot, J. and Sapatinas, T., "Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study", *Journal of Statistical Software,* v. 6, i6, 2001.

[2] Chui, C.K., "An introduction to wavelets", Boston: Academic Press, 1992.

[3] Cohen, A.; Daubechies, I and Feauveau, J.C., Biorthogonal Bases of Compactly Supported Wavelets, *Communications on Pure and Applied Mathematics*, v. 55, pp. 458-560,1992.

[4] Daubechies, I., "Ten Lectures on Wavelets", SIAM, Philadelphia, 1992.

[5] Donoho, D.L. and Johnstone, I.M., "Ideal Spatial Adaptation by Wavelet Shrinkage", *Biometrika*, v. 81, 425-455, 1994.

[6] Donoho, D.L. and Johnstone, I.M. "Minimax estimation via wavelet shrinkage". *Ann.Statist.*, 26, 1998.

[7] Donoho D.L. et al, "Wavelet Shrinkage: Asymptopia? (with discussion)". *Journal Royal Statistical Society*, v.57, pp. 301-369, 1995.

[8] Hall, P. and Patil, P., "On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non-linear wavelet methods". *J. R. Statist. Soc.* B, v. 58, 1996.

[9] Mallat, S. "A wavelet tour of signal processing ". San Diego, CA: Academic Press, 1998.

[10] Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M.H. and Patil, P., "Exact risk analysis of wavelet regression". *J. Comp. Graph. Statist.*, 7, 1998.

[11] Monico, J.F.G., "Posicionamento pelo GNSS: descrição, fundamentos e aplicações", 2ed, 480p, São Paulo: UNESP, 2008.

[12] Morettin, P.A., "Ondas e Ondaletas: Da análise de Fourier à Análise de Ondaletas". São Paulo: EDUSP, 1999.

[13] Nason, G.P., "Wavelet Regression by Cross-Validation". TR 447, Dept. Statistics, Stanford University, 1994.

[14] Nason, G.P., "Wavelet Methods in Statistics with RSeries: Use R", Softcover, 260 p. 2008.

[15] E.M. Souza and Monico,J .F.G., Wavelet Shrinkage: High Frequency Multipath Reduction from GPS Relative Positioning. *GPS Solutions*, **8,** pp. 152-159, 2004.

[16] E.M. Souza, A. Pagamisse, M. Meneguette Jr e J.F.G. Monico, Processamento de Sinais com Wavelets: Redução do Efeito do Multicaminho do Sinal GPS, *Tendências em Matemática Aplicada e Computacional*, v. 6, pp. 177-186, 2005.

[17] Souza, E.M.; Pagamisse, A.; Monico, J.F.G. and Polezel, W.G.C., Comparação das Bases de Wavelets Ortonormais e Biortogonais: Implementação,Vantagens e Desvantagens no Posicionamento com GPS. TEMA. Tendências em Matemática Aplicada e Computacional, v. 8, p. 149-158, 2007.

[18] Teunissen, P.J.G., Quality Control and GPS, em "GPS for Geodesy" (P.J.G. Teunissen e A. Kleusberg, eds.), 2 ed., pp. 271-318, Berlin, Springer Verlag, 1998.