

# MULTI-RESOLUTION ANALYSIS OF NON-UNIFORM DATA WITH JUMP DISCONTINUITIES AND IMPULSIVE NOISE USING ROBUST LOCAL POLYNOMIAL REGRESSION

Shing-Chow Chan and Zhiguo Zhang

Department of Electrical and Electronic Engineering  
The University of Hong Kong, Pokfulam Road, Hong Kong  
{schan, zgzhang}@eee.hku.hk

## ABSTRACT

This paper proposes a new method for performing multiresolution analysis (MRA) of non-uniform data with jump discontinuities and impulsive noise using robust M-estimator-based local polynomial regression (LPR). The basic idea is to interpolate the smoothed estimate, after performing the robust LPR, on a uniform grid in order to perform the MRA using the ordinary wavelet transform. Simulation results show that the new approach performs better than traditional LS-based LPR in preserving jump discontinuities and suppressing isolated impulses when the intersection confident intervals (ICI) bandwidth selection is employed.

## 1. INTRODUCTION

The classic wavelet transform for data smoothing and decomposition are mostly applied to uniform data which is

$$\text{modeled as: } Y^{(i)} = m(i \cdot \Delta t) + \varepsilon^{(i)}, \quad i = 1, \dots, N, \quad (1)$$

where  $m(X)$  is a smooth function specifying the conditional mean of  $Y^{(i)}$  given  $X^{(i)}$ , and  $\varepsilon^{(i)}$  are the additive noise. For non-uniformly sampled data, which are frequently encountered in modeling of large-scale systems, environmental studies, probability density estimation, etc, appropriately modifications have to be carried out. Generally, existing wavelet-based regression of non-uniform data can be broadly classified into two different classes of techniques. One is to interpolate the original data at equidistant points [3,14], and the other is to project the equispaced result onto the irregular grid [2,4]. A technique based on lifting [1] has also been reported. In this paper, we consider the problem of multi-resolution (MR) decomposition/analysis of non-uniform data with jump discontinuities and is corrupted by Gaussian as well as impulsive noises. As pointed out in [10], wavelet regression is closely related to another efficient nonparametric regression method called local polynomial regression (LPR) [7-12]. In LPR, the function to be estimated is assumed to be continuous locally so that the noisy observations can be fitted locally by a polynomial using a least-squares (LS) fit with a kernel function having a certain bandwidth. The bandwidth parameter  $h$  in LPR is closely related to the concept of scale in wavelet analysis. In wavelets, the scale parameters are chosen as powers of two and fixed basis functions are employed. Important advantages of LPR are that the data points can be non-uniformly spaced and the bandwidth can be varied locally to achieve a better bias-variance tradeoff. Motivated by these inherent and important advantages, we propose to perform the smoothing by a robust LPR and compute the samples of the data on a regular grid for performing MRA. The robust LPR employed in this paper is

based on M-estimation instead of the conventional LS fit. M-estimator-based LPR was studied in [9] for image smoothing with jump discontinuity. However, the bandwidths for the M-estimator and the kernel function are fixed and their selection usually requires human intervention. The proposed robust LPR employs the ICI rule introduced in [9] and [10] for choosing this local bandwidth. Simulation results show that it helps to preserve the jump discontinuities, while the M-estimator is capable of suppressing the outliers as well as stabilizing the estimate around the jump discontinuities. It should be noted the jump discontinuity problem in wavelet estimators was also studied in [3] using different interpolation methods considered here. Also, the problem of impulsive noise is not considered. As shown in the simulation section, direct interpolation of the observations can be substantially affected by the impulsive noises. The problem of robust wavelet analysis with impulsive outliers was also considered in [6] using a uniform design. Our paper is organized as follows: in Section 2, the basic principle of LPR is introduced. Sections 3 and 4 are devoted to the proposed M-estimation-based LPR algorithm with ICI rule. Interpolation of the nonuniform data to the uniform grid is considered in Section 5. Simulation results and comparisons are described in Section 6.

## 2. LOCAL POLYNOMIAL REGRESSION

In LPR, we are given noisy samples of a signal:

$$Y^{(i)} = m(X^{(i)}) + \varepsilon^{(i)}, \quad i=1, \dots, n, \quad (2)$$

where  $m(X)$  is a smooth function specifying the conditional mean of  $Y^{(i)}$  given  $X^{(i)}$ , and  $\varepsilon^{(i)}$  are independent identically distributed additive noise with zero mean and variance  $\sigma^2$ . We need to estimate the original signal  $m(X)$  and its derivatives  $m^k(X)$  from the noisy samples  $Y^{(i)}$  at location  $\mathbf{x} = (x_1, \dots, x_d)^T$ . One flexibility of LPR is that the data points  $X^{(i)}$  can be non-uniformly spaced and the observations around a point  $\mathbf{x}$  is approximated locally by the following polynomial:

$$\begin{aligned} P_p(\mathbf{X} : \mathbf{x}, \boldsymbol{\beta}) &= \sum_{K=0}^p \sum_{k_1+\dots+k_d=K} \beta_{k_1, \dots, k_d} \prod_{j=1}^d (X_j - x_j)^{k_j}, \\ &= \boldsymbol{\beta}^T \cdot \mathbf{P}_p(\mathbf{X} - \mathbf{x}), \quad i=1, \dots, n, \end{aligned} \quad (3)$$

where  $\mathbf{P}_p(\mathbf{X} - \mathbf{x}) = \{\prod_{j=1}^d (X_j - x_j)^{k_j} : k_1 + k_2 + \dots + k_d = K \text{ and } K=0, \dots, p\}$  and  $\boldsymbol{\beta} = \{\beta_{k_1, \dots, k_d} : k_1 + k_2 + \dots + k_d = K \text{ and } K=0, \dots, p\}$  is a vector containing the coefficients of the polynomials. We can estimate  $\boldsymbol{\beta}$  from  $(Y^{(i)}, X^{(i)})$  by the weighted least square method. Let

$w_h(\mathbf{x}; \mathbf{X}) = K_h(\mathbf{X} - \mathbf{x})$  be the weighting function for a sample at  $\mathbf{X}$  for estimating  $\boldsymbol{\beta}$  at  $\mathbf{x}$ . To allow an efficient tradeoff between bias and variance, the weighting function or kernel  $w_h(\mathbf{x}; \mathbf{X})$  for the LS fit is usually chosen as  $K_h(\mathbf{X} - \mathbf{x}) = |\mathbf{h}|^{-1} \cdot K(\mathbf{h}^{-1}(\mathbf{X} - \mathbf{x}))$ , where  $h$  is a bandwidth matrix and  $K(\cdot)$  is a non-negative function such as the Gaussian function or the Epanechnikov kernel  $K(u) = (1 - |u|^2)_+$ . Selecting a proper local bandwidth is very critical to achieve the best bias-variance tradeoff in estimating non-stationary signals. For slow varying parts of a signal, we would like the window size or bandwidth to be large so that more accurate estimates can be obtained by averaging out the additive noise as much as possible. At fast varying parts of a signal, however, we would like to have a smaller window size so that excessive bias errors due to the limited order of the fitting polynomial will not occur. The determination of local adaptive bandwidth has been a subject of intensive research in the statistics community. For a survey of this topic, see [8] and the references therein. The LS solution of  $\boldsymbol{\beta}$  is

$$\hat{\boldsymbol{\beta}}_{LS}(\mathbf{x}, \mathbf{h}) = \arg \min_{\boldsymbol{\beta}} E_{LS}(\mathbf{x}, \mathbf{h}), \quad (4)$$

and  $E_{LS}(\mathbf{x}, \mathbf{h}) = \sum_{i=1}^n K_h(\mathbf{x} - \mathbf{X}^{(i)}) \{Y^{(i)} - \boldsymbol{\beta}^T \cdot \mathbf{P}_p(\mathbf{X}^{(i)})\}^2$ . Differentiating  $E_{LS}(\mathbf{x}, \mathbf{h})$  with respect to  $\boldsymbol{\beta}$  and setting the derivative to zero, one gets

$$\hat{\boldsymbol{\beta}}_{LS}(\mathbf{x}, \mathbf{h}) = (\mathbf{P}_{XX})^{-1} \mathbf{P}_{YX} \quad (5)$$

where  $\mathbf{P}_{YX} = \sum_{i=1}^n K_h(\mathbf{x} - \mathbf{X}^{(i)}) Y^{(i)} \mathbf{P}_p(\mathbf{X}^{(i)})$ ,

and  $\mathbf{P}_{XX} = \sum_{i=1}^n K_h(\mathbf{x} - \mathbf{X}^{(i)}) \mathbf{P}_p(\mathbf{X}^{(i)}) \mathbf{P}_p^T(\mathbf{X}^{(i)})$ . (6)

From  $\hat{\boldsymbol{\beta}}_{LS}(\mathbf{x}, \mathbf{h})$ , we can also estimate the derivatives of  $m(\mathbf{x})$ :

$$m^{(k)}(\mathbf{x}, \mathbf{h}) = \frac{\partial^k}{\partial x^{k_1} \dots \partial x^{k_d}} m(\mathbf{x}), \quad (7)$$

from the polynomial  $\hat{m}(\mathbf{x})$  as follows:

$$\hat{m}^{(k)}(\mathbf{x}, \mathbf{h}) = \frac{\partial^k}{\partial x^{k_1} \dots \partial x^{k_d}} \hat{m}(\mathbf{x}, \mathbf{h}) = \prod_{j=1}^d (k_j!) \cdot \hat{\boldsymbol{\beta}}_{k_1, \dots, k_d}(\mathbf{x}, \mathbf{h}) \quad (8)$$

$$\hat{\boldsymbol{\beta}}_{k_1, \dots, k_d}(\mathbf{x}, \mathbf{h}) = \mathbf{1}_k^T \cdot \hat{\boldsymbol{\beta}}_{LS}(\mathbf{x}, \mathbf{h}). \quad (9)$$

where  $\mathbf{1}_k^T = [0, \dots, 0, 1, 0, \dots, 0]$  is a vector with a one in the  $k$ -location.

### 3. M-ESTIMATION AND IRLS

“M-estimation” refers to “generalized maximum likelihood estimation”, which is a formal approach to robust estimation developed by Huber in 1964. Later, Härdle & Gasser [11] combined M-estimation with nonparametric function fitting. More recently, Chu et al. [12] employed M-smoother with local linear fit to address the problem of smoothing with jump discontinuities. They have also been employed in robust adaptive filtering under impulsive noise [13]. Unlike LS-estimation, M-estimation minimizes a different objective function, which effectively down-weigh those data points with abnormally large errors. By so doing, the bandwidth is less sensitive to the impulsive noise and jump discontinuities. More precisely, the M-estimate  $\boldsymbol{\beta}(\mathbf{x}, \mathbf{h})$  is obtained by minimizing a score function:

$$\hat{\boldsymbol{\beta}}_M(\mathbf{x}, \mathbf{h}) = \arg \min_{\boldsymbol{\beta}} E_M(\mathbf{x}, \mathbf{h}) \quad (10)$$

$$E_M(\mathbf{x}, \mathbf{h}) = \sum_{i=1}^n K_h(\mathbf{x} - \mathbf{X}^{(i)}) \rho_{\xi}(Y^{(i)} - \boldsymbol{\beta}^T \cdot \mathbf{P}_p(\mathbf{X}^{(i)}))$$

where  $\rho_{\xi}(x) = \rho(x/\xi)$  with  $\rho(x)$  an M-estimate function (Huber function, 3-parts re-descending functions, etc [5]). The function  $\rho(x)$  usually levels off when the magnitude of  $x$  is large so that the estimation error  $e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \mathbf{P}_p(\mathbf{X}^{(i)})$  is de-emphasized when  $|e_i| > \xi$ , a certain threshold to be determined.

Since the main purpose of the scale parameter is to “reject the outlier”, its exact value is not that sensitive, provided it is not chosen too large or too small. Based on the results of the estimators we proposed in [13], we let  $\xi = 2.576 \cdot \sigma(\mathbf{X}^{(i)})$ , where  $\sigma^2(\mathbf{X}^{(i)})$  is the robust variance estimator

$$\sigma^2(\mathbf{X}^{(i)}) = c_1 \text{med}(A_e(i)) \quad (11)$$

where  $A_e(i) = \{(Y^{(i)} - Y^{(i-1)})^2, \dots, (Y^{(i-N_w-1)} - Y^{(i-N_w-2)})^2\}$ ,  $N_w$  is the length of the estimation window and  $c_1 = 1.483 \cdot (1 + \frac{5}{N_w-1})$  is a finite sample correction factor.

After determining a rough estimate of  $\sigma^2(\mathbf{X}^{(i)})$ , it can be scaled appropriately to obtain the scale parameter of the M-estimate function. Differentiating  $E_M(\mathbf{x}, \mathbf{h})$  with respect to  $\boldsymbol{\beta}$  and setting the derivative to zero, we get

$$\hat{\boldsymbol{\beta}}_M(\mathbf{x}, \mathbf{h}) = (\mathbf{P}_{\rho_{-XX}})^{-1} \mathbf{P}_{\rho_{-YX}}, \quad (12)$$

$\mathbf{P}_{\rho_{-YX}} = \sum_{i=1}^n w_{\rho}(\mathbf{x}; \mathbf{X}) Y^{(i)} \mathbf{P}_p(\mathbf{X}^{(i)})$ ,  $w_{\rho}(\mathbf{x}; \mathbf{X}) = \rho'(e_i) \cdot K_h(\mathbf{x} - \mathbf{X})$ ,

$e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \mathbf{P}_p(\mathbf{X}^{(i)})$  and  $\mathbf{P}_{\rho_{-XX}} = \sum_{i=1}^n w_{\rho}(\mathbf{x}; \mathbf{X}) \mathbf{P}_p(\mathbf{X}^{(i)}) \mathbf{P}_p^T(\mathbf{X}^{(i)})$ .

Note that this is a nonlinear equation, because the entries of  $\mathbf{P}_{\rho_{-YX}}$  and  $\mathbf{P}_{\rho_{-XX}}$  depend on  $e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \mathbf{P}_p(\mathbf{X}^{(i)})$ , which in turns depend on the parameter to be estimated. We can solve  $\hat{\boldsymbol{\beta}}_M(\mathbf{x}, \mathbf{h})$  using the iterative reweighted least squares (IRLS or IWLS) or other Newton-based methods. In the IRLS, one starts with an initial estimate of  $\hat{\boldsymbol{\beta}}_M^{(0)}(\mathbf{x}, \mathbf{h})$  and repeatedly solves (12) by replacing  $\boldsymbol{\beta}^T$  in  $e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \mathbf{P}_p(\mathbf{X}^{(i)})$  by  $\hat{\boldsymbol{\beta}}^{(l-1)}$ , where  $l$  is the iteration number:

$$\hat{\boldsymbol{\beta}}_M^{(l)}(\mathbf{x}, \mathbf{h}) = (\mathbf{P}_{\rho_{-XX}}^{(l-1)})^{-1} \mathbf{P}_{\rho_{-YX}}^{(l-1)}, \quad (13)$$

$l=1$  to  $l_{\max}$  or change of  $\hat{\boldsymbol{\beta}}_M^{(l)}(\mathbf{x}, \mathbf{h})$  is small enough,

where  $\mathbf{P}_{\rho_{-XX}}^{(l-1)} = \sum_{i=1}^n w_{\rho}^{(l-1)}(\mathbf{x}; \mathbf{X}) \mathbf{P}_p(\mathbf{X}^{(i)}) \mathbf{P}_p^T(\mathbf{X}^{(i)})$  and

$\mathbf{P}_{\rho_{-YX}}^{(l-1)} = \sum_{i=1}^n w_{\rho}^{(l-1)}(\mathbf{x}; \mathbf{X}) Y^{(i)} \mathbf{P}_p(\mathbf{X}^{(i)})$ . Let us use a 1-D scenario

to explain why M-estimation works even for impulsive noises and jump discontinuities. When the LPR is performed using LS fit and the ICI rule, the adaptive bandwidths at the locations of the impulses and jump discontinuities, are normally very small to limit the bias errors. Therefore, not only the edges, but also the impulsive noises are preserved. On the other hand, the M-estimate function,  $\rho_{\xi}(\cdot)$ , with an appropriate scale  $\xi$ , help to de-emphasize the effects of these outliers by assigning them a smaller weights.

#### 4. BANDWIDTH SELECTION USING ICI

Unlike plug-in bandwidth method, where parameters in certain analytical optimal bandwidth formulae are estimated and “plugged” into the formulae, empirical method usually starts with a finite set of window sizes:

$$\mathbf{H} = \{\mathbf{h}_1 < \mathbf{h}_2 < \dots < \mathbf{h}_J\}, \quad (14)$$

and determines the optimal bandwidth by evaluating the fitting results (note, in multivariate data, windows can be ordered according to the volume of their support). Let  $\hat{m}(\mathbf{x}, \mathbf{h}_j)$  be the estimate for the window  $\mathbf{h}_j$ . The variance and the bias of these estimators at  $\mathbf{x}$  are functions of the filter bandwidth  $\mathbf{h}$ , so is the mean square error (MSE). In fact, we have:

$$MSE(\mathbf{x}, \mathbf{h}) = Var[\hat{m}(\mathbf{x}, \mathbf{h})] + bias^2[\hat{m}(\mathbf{x}, \mathbf{h})]. \quad (15)$$

As mentioned earlier, the bias of the estimation will increase rapidly if the bandwidth  $\mathbf{h}$  becomes so large that the underlying data at  $\mathbf{x}$  cannot be modeled by the local polynomial of a given order. On the other hand, the larger the window size, the smaller will be the variance of the estimator. So there exists an optimal bandwidth  $\mathbf{h}_{opt}(\mathbf{x})$  where the  $MSE(\mathbf{x}, \mathbf{h})$  is minimized. To determine this optimal bandwidth, the ICI rule examines a sequence of confidence intervals of the estimates  $\hat{m}(\mathbf{x}, \mathbf{h}_j)$ :

$$D_j = [L_j, U_j], \quad (16)$$

$$U_j = \hat{m}(\mathbf{x}, \mathbf{h}_j) + \Gamma \cdot std_k(\mathbf{x}, \mathbf{h}_j), \quad L_j = \hat{m}(\mathbf{x}, \mathbf{h}_j) - \Gamma \cdot std_k(\mathbf{x}, \mathbf{h}_j),$$

where  $std_k(\mathbf{x}, \mathbf{h}_j)$  is the standard deviation of the estimate and  $\Gamma > 0$  is a threshold parameter of the confidence interval. Define the following quantities from the confident intervals

$$\bar{L}_j = \max[\bar{L}_{j-1}, L_j], \quad \underline{U}_j = \min[\underline{U}_{j-1}, U_j], \quad (17)$$

$j=1, 2, \dots, J$ ,  $\bar{L}_0 = \underline{U}_0 = 0$ . The largest value of these  $j$  for which  $\underline{U}_j \geq \bar{L}_j$  gives  $j^+$  and it yields a bandwidth  $\mathbf{h}_j^+$ , which is the required optimal ICI bandwidth. In other words, the optimal bandwidth  $\mathbf{h}_j^+$  is the largest  $j$  when  $\underline{U}_j \geq \bar{L}_j$  is still satisfied.

Note, the ICI window sizes are different for different position of  $\mathbf{x}$ . Because the optimal bandwidth is decided by  $\Gamma$ ,  $\Gamma$  plays a crucial part in the performance of the algorithm. When  $\Gamma$  is large, the segment  $D_j$  becomes wide, and it will cause the value of  $\mathbf{h}_j^+$  to be bigger. This will result in over-smoothing. On the contrary, when  $\Gamma$  is small, the segment  $D_j$  would become narrow, and it will yield a small value of  $\mathbf{h}_j^+$  so that the noise cannot be removed effectively. In [10], Katkovnik used Cross-Validation to determine a reasonable threshold  $\Gamma$ .

#### 5. INTERPOLATION AND MRA

After removing the impulsive and Gaussian noises from the original non-uniform data, it can be interpolated to a uniform grid for performing the MR using the wavelet transform. In this section, we use local regression with Gaussian filter to interpolate the data and the filter bandwidth can be obtained automatically by ICI rule. There are several methods for carrying out the interpolation. Suppose that we have a local polynomial representation at two adjacent points  $x_0$  and  $x_1$ :  $p(x : x_0)$  and  $p(x : x_1)$ . Then, the value of a point  $x \in (x_0 : x_1)$  can be obtained by linear interpolation from the estimate of  $p(x : x_0)$  and  $p(x : x_1)$ . Another simple method is to perform a

linear interpolation using LPR and ICI again, which has the advantage of preserving the jump-discontinuities.

#### 6. SIMULATIONS

We now evaluate the proposed algorithm using a 1-D with jump-discontinuities and impulsive noises. First of all, we shall demonstrate the effectiveness of the M-estimator-based LPR over the LS-based LPR using ICI bandwidth selection and uniform data. Figures 1 to 3 show the original noisy signal and the estimates obtained from the two approaches. It can be seen that the M-estimator-based LPR is able to preserve the jump discontinuities while suppressing the isolated impulses. Therefore, we would expect a direct interpolation of this noisy signal onto a uniform grid be substantially affected by the jump discontinuities and impulsive noise, and it is not suitable for performing multi-resolution analysis. We now proceed to the nonuniform data in Fig. 4. The additive noise is Gaussian noise with zero mean and variance 0.01. The amplitude of the impulsive noises is generated randomly with a variance of 1.5. To better visualize the effects, their locations are fixed at  $x=0.1, 0.2, 0.55, 0.7$  and  $0.95$ . The two jump discontinuities locations are at  $x=0.35, 0.8$ . The original observation signal and the M-estimation result were shown in Fig.4. We can clearly see from Fig.1 that M-estimator-based LPR can preserve those jump discontinuities or edges, while removing the isolated impulses. The M-estimator we used is the Huber function

$$\rho(e) = \begin{cases} e^2/2, & 0 < |e| < \xi \\ \xi^2/2, & \text{otherwise} \end{cases}. \quad \text{Other M-estimate function such as}$$

cauchy or Hampel three parts redescending function can also be used. The threshold  $\xi$  is computed from (11) in Section 3. Fig. 5 shows the signal after interpolating to a uniform grid using linear LPR with ICI and LS fit. It can be seen that the jump discontinuities are well preserved. The wavelet decomposition of this uniform data is shown in Fig. 6. The jump discontinuities at  $x=0.35$  and  $0.8$  show up as sharp changes of the wavelet coefficients. Note the sign of the coefficients, which indicate the polarity of the jumps.

#### 7. CONCLUSION

A new method for performing multiresolution analysis of non-uniform data using a robust M-estimator-based LPR and ICI bandwidth is presented. Simulation results show that the new approach is able to preserve jump discontinuities and suppressing isolated impulses.

#### REFERENCES

- [1] E. Vanraes, M. Jansen, and A. Bultheel, “Stabilized Wavelet Transforms for Nonequispaced Data Smoothing,” *Signal Processing*, Vol. 82, No. 12, pp. 1979-1990, 2002.
- [2] I. Daubechies, I. Guskov, P. Schr oder, and W. Sweldens, “Wavelets on Irregular Point Sets,” *Phil. Trans. R. Soc. Lond. A*, Vol. 357(1760), pp. 2397--2413, 1999.
- [3] P. Hall, and B. A. Turlach, “Interpolation Methods for Nonlinear Wavelet Regression with Irregularly Spaced Design,” *Ann. Statist.* Vol. 25, pp. 1912-1925, 1997.
- [4] T. Cai and L. D. Brown, “Wavelet Shrinkage For Nonequispaced Samples,” *Annals of statistics*, Vol. 26, No. 5, pp. 1783-1799, 1998.
- [5] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outliers detection*. New York: Wiley 1987.
- [6] A. G. Bruce, D. L. Donoho, H.-Y. Gao, and R. D. Martin, “Denoising and robust nonlinear wavelet analysis,” *SPIE Proceedings, Wavelet Applications*, Vol. 2242, 1994.
- [7] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London, 1996.
- [8] D. A. Burt, “Bandwidth Selection Concerns for Jump Point Discontinuity Preservation in the Regression Setting Using M-smoothers

and the Extension to Hypothesis Testing”, Phd thesis, Virginia Tech, 2000.

[9] A. Goldenshluger and A. Nemirovski, “On Spatial Adaptive Estimation of Nonparametric Regression,” *Math. Meth. Stat.*, Vol. 6, No. 2, pp. 135–170, 1997.

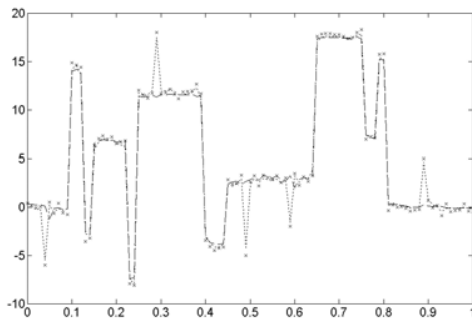
[10] V. Katkovnik, “A New Method for Varying Adaptive Bandwidth Selection,” *IEEE Trans. on SP*, Vol. 47, No. 9, pp.2567-2571, Sept. 1999.

[11] W. Härdle and T. Gasser, “Robust non-parametric function fitting”, *Journal of the Royal Statistical Society, Ser. B* 46, pp. 42-51, 1984.

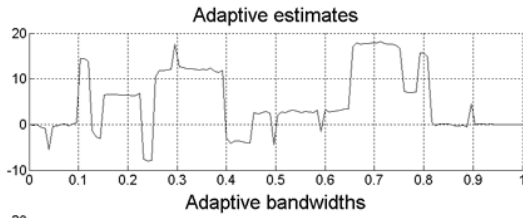
[12] C. K. Chu, I. K. Glad, F. Godtliebsen, F. and J. S. Marron, “Edge-preserving smoothers for image processing”, *JASA*, 93(442), pp. 526-541, 1998.

[13] Y. Zou, S. C. Chan, and T. S. Ng, “A Recursive Least M-estimate (RLM) Adaptive Filter For Robust Filtering in Impulse Noise,” *IEEE Signal Processing Letters*, IEEE, 7(11), pp. 324-6, 2000.

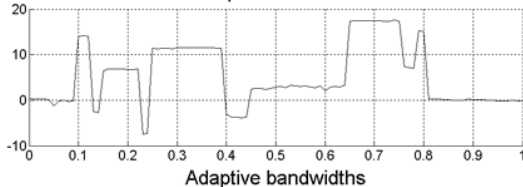
[14] A. Kovac and B. W. Silverman, “Extending the scope of wavelet regression methods by coefficient-dependent thresholding,” *JASA* 95, pp. 172-183, 1997.



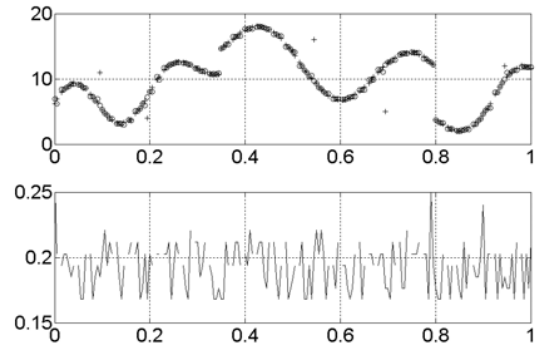
**Fig.1.** “x” - observation data, “...” - estimate using LS-based LPR with ICI, “---” - estimate using LPR with M-estimator-based LPR with ICI. Impulsive noise are added at positions  $x = 0.05, 0.3, 0.5, 0.6$  and  $0.9$ . Additive Gaussian noise has mean 0 and variance  $1/3$ .



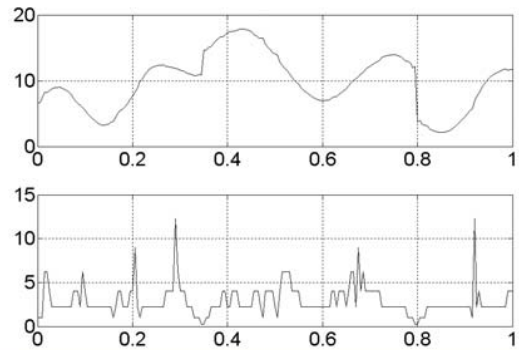
**Fig.2.** Adaptive bandwidths with Gaussian kernel.  $m=1, \Gamma=0.002$ .



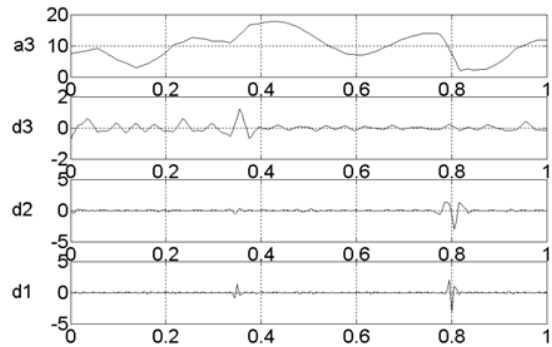
**Fig.3.** Adaptive bandwidths with Gaussian kernel.  $m=1, \Gamma=0.004$ .



**Fig.4.** Upper figure: “+” - observed non-uniform data, “o” - estimate signal using LPR with M-estimation function and ICI (Iterative operation of IRLS is only once). Lower figure: local adaptive bandwidth, which is represented by  $\sigma$  in the kernel Gaussian filter  $K_h(x) = \frac{1}{2\pi\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ .  $m=1, \Gamma=0.35$ .



**Fig.5.** Upper figure: interpolated uniform data using linear LPR with LS and ICI. Lower figure: adaptive bandwidth which is represented by  $\sigma$  in the kernel Gaussian filter  $K_h(x) = \frac{1}{2\pi\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ .  $m=1, \Gamma=0.2$ .



**Fig.6.** The wavelet we used is db3 in MATLAB wavelet toolbox, and three level of wavelet decomposition was performed.