ROBUST LOCAL POLYNOMIAL REGRESSION USING M-ESTIMATOR WITH ADAPTIVE BANDWIDTH

Shing-Chow Chan and Zhiguo Zhang

Department of Electrical and Electronic Engineering, The University of Hong Kong

ABSTRACT

In this paper, a new method for robust local polynomial regression (LPR) using M-estimator with adaptive bandwidth is proposed. This is motivated by the limitation of traditional LPR in detecting and removing impulsive noise or outlies. By using M-estimation technique and the intersection of confidence intervals (ICI) rule for choosing an adaptive local bandwidth, a robust LPR algorithm is developed. Simulation results show that the new M-estimation-based LPR performs considerably better than the traditional LS-based method in removing the impulsive noise as well as preserving the jump discontinuities, which are frequently found in image and video processing.

1. INTRODUCTION

Local polynomial regression [1, 2] is a very flexible and efficient nonparametric method for data smoothing and density estimation. Given a set of noisy samples of a signal, the data points are fitted locally by a polynomial using a least-square (LS) fit with a kernel function having a certain bandwidth. 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 estimations 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 [3] and references therein. One very useful method is called the empirical-bias bandwidths selection (EBBS) proposed by Ruppert [4]. Unfortunately, the complexity of the EBBS is rather high. More recently, Goldenshluger and Nemirovski [5] and Katkovinik [6] studied a new bandwidth selection rule called the intersection confidence intervals (ICI), which is very simple to use and yield reasonable and efficient results.

When it comes to nonparametric regression, LPR with ICI performs very well in removing Gaussian noise, while preserving signal edges, especially when the observed signals exhibit so called "jump discontinuities". The main reason behind this nice property is that the filter bandwidth (effective window size) can automatically be adjusted to a small value in the neighborhood of these jumping points, so that excessive smoothing is avoided to preserve the jump discontinuities. This is also a reminiscence of the adaptive nature of local bandwidth selector in striking for the best bias-variance tradeoff. Data smoothing under jump discontinuities has also been studied in [10] for image smoothing. To avoid the jump discontinuities from affecting the signal estimates, robust statistics using M-estimator is introduced. Basically, the M-estimator de-emphasizes or ignores those samples on the other side of the discontinuity to stabilize the estimate. However, the bandwidths for the M-estimator (the scale parameters) and the kernel function are fixed and their selection usually requires human intervention.

Although the performance of LPR with ICI efficiently handles jump discontinuities, they are still sensitive to impulsive noise, because the algorithm is based on least square estimation, which implicitly assumes that the additive noise is Gaussian. For noises with long tail in their probability distribution functions, we observed that the estimates are substantially affected and they cannot be removed satisfactorily. This motivates us to replace the LS estimation with M-estimation in the LPR with ICI rule. It can also be viewed as incorporating an ICI adaptive bandwidth selector into Chu's M-estimator-based LPR. The M-estimator serves two major roles here: to combat the impulsive noise, and to stabilize the estimates across jump discontinuities. After that, the ICI rule will try to strike for the best bias-variance tradeoff.

Our paper is organized as follows: in Section 2, the basic principle of LPR is introduced. The ICI rule and its application to LPR are introduced in Section 3. Section 4 is devoted to the proposed M-estimation-based LPR algorithm with ICI rule. Simulation results and comparisons are described in Section 5. Finally, conclusions are drawn in Section 6.

2. LOCAL POLYNOMIAL REGRESSION

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

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

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

$$P_p(\boldsymbol{X}:\boldsymbol{x},\boldsymbol{\beta}) = \sum_{K=0}^{p} \sum_{k_1+\cdots k_d=K} \beta_{k_1,\dots,k_d} \prod_{j=1}^{d} (X_j - x_j)^{k_j},$$
$$= \boldsymbol{\beta}^T \cdot \boldsymbol{P}_p(\boldsymbol{X}-\boldsymbol{x}), \ i=1,\dots,n, \tag{2}$$

where $\boldsymbol{\beta} = \{\beta_{k_1,\dots,k_d} : k_1+k_2+\dots+k_d=K \text{ and } K=0,\dots,p\}$ is a vector containing the coefficients of the polynomials, and

$$P_p(X-x) = \{[\prod_{j=1}^{d}(X_j - x_j)^{k_j} : k_1 + k_2 + ... + k_d = K \text{ and } K = 0,...,p\} .$$

We can estimate $\boldsymbol{\beta}$ from $(Y^{(i)}, X^{(i)})$ by the weighted least square method. Let $w_{\boldsymbol{h}}(\boldsymbol{x}:X) = K_{\boldsymbol{h}}(X - x)$ be the weighting function for a sample at X for estimating $\boldsymbol{\beta}$ at $\boldsymbol{x}$. To allow efficient tradeoff between bias and variance, $w_{\boldsymbol{h}}(\boldsymbol{x}:X)$ is usually chosen as a d-dimensional non-negative function $K_{\boldsymbol{h}}(X_i - x)$, where $\boldsymbol{h}$ is a bandwidth matrix and $K_{\boldsymbol{h}}(X-x) = |\boldsymbol{h}|^{-1} \cdot K_{\boldsymbol{h}}(\boldsymbol{h}^{-1}(X-x))$. For separable windows, we have

$$K_{\boldsymbol{h}}(X-x) = \frac{1}{h_1...h_d}\prod_{\alpha=1}^{d}K_{h_\alpha}(\frac{X_\alpha - x_\alpha}{h_\alpha})) ,$$
$$\boldsymbol{h} = diag\{h_1,...,h_d\} . \tag{3}$$

The LS solution of $\boldsymbol{\beta}$ is

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

where $E_{LS}(\boldsymbol{x},\boldsymbol{h}) = \sum_{i=1}^{n}K_{\boldsymbol{h}}(\boldsymbol{x} - X^{(i)})\{Y^{(i)} - \boldsymbol{\beta}^T \cdot P_p(X^{(i)})\}^2$
$$= \overline{Y} - 2\boldsymbol{\beta}^T P_{YX} + \boldsymbol{\beta}^T P_{XX} \boldsymbol{\beta} ,$$

$\overline{Y} = \sum_{i=1}^{n}w_{\boldsymbol{h}}(\boldsymbol{x}:X^{(i)})(Y^{(i)})^2$, $P_{YX} = \sum_{i=1}^{n}K_{\boldsymbol{h}}(\boldsymbol{x} - X^{(i)})Y^{(i)}P_p(X^{(i)})$,

and $P_{XX} = \sum_{i=1}^{n}K_{\boldsymbol{h}}(\boldsymbol{x} - X^{(i)})P_p(X^{(i)})P_p^T(X^{(i)})$.

Differentiating $E_{LS}(\boldsymbol{x},\boldsymbol{h})$ with respect to $\boldsymbol{\beta}$ and setting the derivative to zero, we can get

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

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

$$m^{(k)}(\boldsymbol{xh}) = \frac{\partial^k}{\partial x^{k_1}...\partial x^{k_d}}m(\boldsymbol{x}) , \tag{6}$$

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

$$\hat{m}^{(k)}(\boldsymbol{x},\boldsymbol{h}) = \frac{\partial^k}{\partial x^{k_1}...\partial x^{k_d}}\hat{m}(\boldsymbol{x},\boldsymbol{h})$$
$$= \prod_{j=1}^{d}(k_j!) \cdot \hat{\beta}_{k_1,...,k_d}(\boldsymbol{x},\boldsymbol{h}) \tag{7}$$

$$\hat{\beta}_{k_1,...,k_d}(\boldsymbol{x},\boldsymbol{h}) = 1_{\boldsymbol{k}}^T \cdot \hat{\boldsymbol{\beta}}_{LS}(\boldsymbol{x},\boldsymbol{h}) , \tag{8}$$

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

3. INTERSECTION OF CONFIDENCE INTERVALS

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

$$H = \{h_1 < h_2 < ... < h_J\} , \tag{9}$$

and determines the optimal bandwidth by evaluating the fitting results (note, in multivariate data, windows are order according

to the volume of their support). Let $\hat{m}(\boldsymbol{x},\boldsymbol{h}_j)$ be the estimate for the window $\boldsymbol{h}_j$. The variance and the bias of these estimators at $\boldsymbol{x}$ are functions of the filter bandwidth $\boldsymbol{h}$, so is the mean square error (MSE). In fact, we have:

$$MSE(\boldsymbol{x},\boldsymbol{h}) = Var[\hat{m}(\boldsymbol{x},\boldsymbol{h})] + bias^2[\hat{m}(\boldsymbol{x},\boldsymbol{h})] . \tag{10}$$

As mentioned earlier, the bias of the estimation will increase rapidly if the bandwidth $\boldsymbol{h}$ becomes so large that the underlying data at $\boldsymbol{x}$ 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 $\boldsymbol{h}_{opt}(x)$ where the $MSE(\boldsymbol{x},\boldsymbol{h})$ is minimized. To determine this optimal bandwidth, the ICI rule examines a sequence of confidence intervals of the estimates $\hat{m}(\boldsymbol{x},\boldsymbol{h}_j)$:

$$D_j = [L_j, U_j] ,$$
$$U_j = \hat{m}(\boldsymbol{x},\boldsymbol{h}_j) + \Gamma \cdot std_k(\boldsymbol{x},\boldsymbol{h}_j) , \tag{11}$$
$$L_j = \hat{m}(\boldsymbol{x},\boldsymbol{h}_j) - \Gamma \cdot std_k(\boldsymbol{x},\boldsymbol{h}_j) ,$$

where $std_k(\boldsymbol{x},\boldsymbol{h}_j)$ 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

$$\overline{L}_j = \max[\overline{L}_{j-1}, L_j] ,$$
$$\underline{U}_j = \min[\underline{U}_{j-1}, U_j] , \tag{12}$$
$$j=1,2,...,J, \ \overline{L}_0 = \underline{U}_0 = 0 .$$

The largest value of these j for which $\underline{U}_j \geq \overline{L}_j$ gives j^+ and it yields a bandwidth $\boldsymbol{h}_j^+$, which is the required optimal ICI bandwidth. In other words, the optimal bandwidth $\boldsymbol{h}_j^+$ is the largest j when $\underline{U}_j \geq \overline{L}_j$ is still satisfied. Note, the ICI window sizes are different for different position of $\boldsymbol{x}$.

Because the optimal bandwidth is decided by Γ, Γ plays a crucial part in the performance of the algorithm. When Γ is large, the segment D_j becomes wide, and it will cause the value of $\boldsymbol{h}_j^+$ to be bigger. This will result in over-smoothing. On the contrary, when Γ is small, the segment D_j would become narrow, and it will yield a small value of $\boldsymbol{h}_j^+$ so that the noise cannot be removed effectively. In [2], Katkovnik used Cross-Validation to determine a reasonable threshold Γ.

4. 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 [9] combined M-estimation with nonparametric function fitting. More recently, Chu et al. [10] 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 [11,12,13]

We apply this local polynomial regression combined with ICI using LS method algorithm to 1-D signal and obtain results that preserve edge well, jump discontinuities included. In a jump point, the algorithm can provide a small bandwidth that ignores other neighbor points that have large difference with the preference point. But when the signal contains many outlies, such as impulsive noise, the method cannot perform well.

In order to smooth these outliers, it is generally preferred to estimate $\boldsymbol{\beta}(\boldsymbol{x},\boldsymbol{h})$ by minimizing the M-estimate function.

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

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

(13)

where $\rho_\xi(.)$ is an M-estimate function, where the estimation error $e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \boldsymbol{P}_p(\boldsymbol{X}^{(i)})$ is de-emphasized when $|e_i| > \xi$. Since the main purpose of the scale parameter is to "reject the outliers", it exact value is not that sensitive, provided it is not chosen too large or too small. A simple estimate is $\xi = 2.576 \cdot \sigma(\boldsymbol{X}^{(i)})$. Using the robust variance estimator, the robust estimate of $\sigma^2(\boldsymbol{X}^{(i)})$ should be

$$\sigma^2(\boldsymbol{X}^{(i)}) = c_1 med(A_e(i)) \qquad (14)$$

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

After determining a rough estimate of $\sigma^2(\boldsymbol{X}^{(i)})$, it can be scaled appropriately to obtain the scale parameter of the M-estimate function.

Differentiating $E_M(\boldsymbol{x},\boldsymbol{h})$ with respect to $\boldsymbol{\beta}$ and setting the derivative to zero, we can get

$$\hat{\boldsymbol{\beta}}_M(\boldsymbol{x},\boldsymbol{h}) = (\boldsymbol{P}_{\rho\_XX})^{-1}\boldsymbol{P}_{\rho\_YX} \qquad (15)$$

where $\boldsymbol{P}_{\rho\_YX} = \sum_{i=1}^{n} w_\rho(\boldsymbol{x}:\boldsymbol{X})Y^{(i)}\boldsymbol{P}_p(\boldsymbol{X}^{(i)})$,

$$\boldsymbol{P}_{\rho\_XX} = \sum_{i=1}^{n} w_\rho(\boldsymbol{x}:\boldsymbol{X})\boldsymbol{P}_p(\boldsymbol{X}^{(i)})\boldsymbol{P}_p^T(\boldsymbol{X}^{(i)}),$$

$w_\rho(\boldsymbol{x}:\boldsymbol{X}) = \rho'(e_i) \cdot K_{\boldsymbol{h}}(\boldsymbol{x}-\boldsymbol{X})$, and $e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \boldsymbol{P}_p(\boldsymbol{X}^{(i)})$.

Note, this is a nonlinear equation, because the entries of $\boldsymbol{P}_{\rho\_YX}$ and $\boldsymbol{P}_{\rho\_XX}$ depend on $e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \boldsymbol{P}_p(\boldsymbol{X}^{(i)})$, which in turns depend on the parameter to be estimated. We can solve $\hat{\boldsymbol{\beta}}_M(\boldsymbol{x},\boldsymbol{h})$ using iterative reweighed least squares (IRLS or IWLS). Here are the details:

Start with an initial estimate of $\hat{\boldsymbol{\beta}}_M^{(0)}(\boldsymbol{x},\boldsymbol{h})$, solve (15) as

$$\hat{\boldsymbol{\beta}}_M^{(l)}(\boldsymbol{x},\boldsymbol{h}) = (\boldsymbol{P}_{\rho\_XX}^{(l-1)})^{-1}\boldsymbol{P}_{\rho\_YX}^{(l-1)},$$

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

(16)

where $\boldsymbol{P}_{\rho\_YX}^{(l-1)} = \sum_{i=1}^{n} w_\rho^{(l-1)}(\boldsymbol{x}:\boldsymbol{X})Y^{(i)}\boldsymbol{P}_p(\boldsymbol{X}^{(i)})$,

$$\boldsymbol{P}_{\rho\_XX}^{(l-1)} = \sum_{i=1}^{n} w_\rho^{(l-1)}(\boldsymbol{x}:\boldsymbol{X})\boldsymbol{P}_p(\boldsymbol{X}^{(i)})\boldsymbol{P}_p^T(\boldsymbol{X}^{(i)}),$$

$w_\rho^{(l-1)}(\boldsymbol{x}:\boldsymbol{X}) = L_g(e_i^{(l-1)}) \cdot K_{\boldsymbol{h}}(\boldsymbol{x}-\boldsymbol{X})$,

$e_i^{(l-1)} = Y^{(i)} - (\hat{\boldsymbol{\beta}}^{(l-1)})^T \cdot \boldsymbol{P}_p(\boldsymbol{X}^{(i)})$, $L_g(u) = \begin{cases} \rho_\xi'(u)/u & if\ u \neq 0 \\ 0 & if\ u = 0 \end{cases}$.

We can relax this nonlinear equation, by replacing $\boldsymbol{\beta}^T$ in $e_i = Y^{(i)} - \boldsymbol{\beta}^T \cdot \boldsymbol{P}_p(\boldsymbol{X}^{(i)})$ by $\hat{\boldsymbol{\beta}}^{(l-1)}$. Therefore, the matrix $\boldsymbol{P}_{\rho\_YX}$ and vector $\boldsymbol{P}_{\rho\_YX}$ can now be approximated as

$$\boldsymbol{P}_{\rho\_XX}^{(l-1)} = \sum_{i=1}^{n} w_\rho^{(l-1)}(\boldsymbol{x}:\boldsymbol{X})\boldsymbol{P}_p(\boldsymbol{X}^{(i)})\boldsymbol{P}_p^T(\boldsymbol{X}^{(i)}) \quad \text{and}$$

$\boldsymbol{P}_{\rho\_YX}^{(l-1)} = \sum_{i=1}^{n} w_\rho^{(l-1)}(\boldsymbol{x}:\boldsymbol{X})Y^{(i)}\boldsymbol{P}_p(\boldsymbol{X}^{(i)})$, respectively, by replacing $\boldsymbol{\beta}^T$ in $\hat{\boldsymbol{\beta}}^{(l-1)}$. Since $\hat{\boldsymbol{\beta}}^{(l-1)}$ is already known, the last estimation $\hat{\boldsymbol{\beta}}_M^{(l)}(\boldsymbol{x},\boldsymbol{h})$ can be computed by solving (16), where the right hand sides are all constant matrix and vector. Since each iteration progress requires the solution of a LS problem, it is called IRLS method. Alternatively, (15) can be solved using Newton type of algorithms. Let us use a 1-D scenario to explain why M-estimation works even for impulsive noise. 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 noise are preserved. On the other hand, the M-estimate function, $\rho_\xi(.)$, with an appropriate scale ξ, help to de-emphasis the effects of these outliers by assigning them a smaller weights.

5. SIMULATIONS

We now evaluate the proposed algorithm using a 1-D noisy signal with jump-discontinuities and impulsive noise. The additive noise is Gaussian with zero mean and variance 0.3. The amplitude of the impulsive noise is generated randomly with a variance of 1.8. To better visual effects, their locations are fixed at x=0.05, 0.3, 0.5, 0.6 and 0.9. The following are stimulated:

1:- LPR using LS and ICI;
2:- LPR using M-estimation and ICI.

The original observation signal and the two estimation results were shown in Fig.1. We can clearly see from Fig.1 that M-estimation can achieve better result than that of LS method, especially in the vicinity of the impulses. LS-based LPR can only preserve those jump discontinuities or edges, but it can't distinguish the isolated impulses and remove them. On the other hand, M-estimation not only preserves those jump discontinuities and edges but also removes effectively the impulsive noise. The M-estimator $\rho_\xi(.)$ we used is the Huber

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

such as cauchy or Hampel three parts redescending function can be used. Here we can use (14) in Section 4 to compute an appropriate value of the threshold ξ.
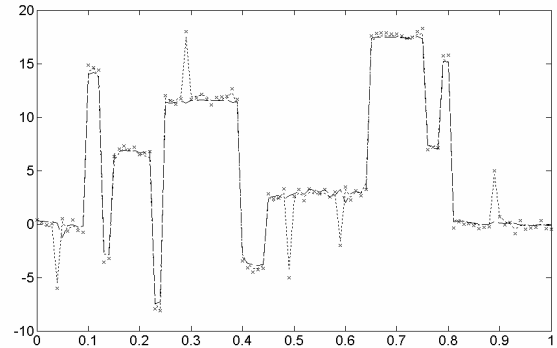

Fig.1 "x" is observation data, "..." is the estimate signal using LPR with LS method and "---" is the estimate signal using LPR with M-estimation function. Impulsive noise are added at positions x = 0.05, 0.3, 0.5, 0.6 and 0.9. The error distribution used to simulate the data still includes

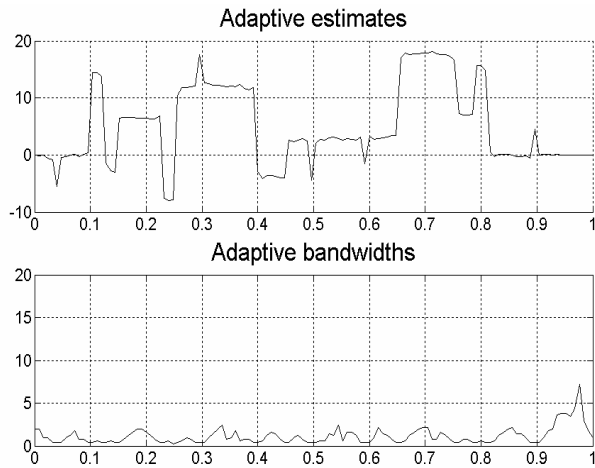Gaussian noise with mean 0 and variance 0.3. Iterative operation of IRLS is only once.

Adaptive estimates

Adaptive bandwidths

Fig.2 Adaptive bandwidths here are represented by h in the kernel Gaussian function $K_h(x)$. m=1, Γ =0.002.

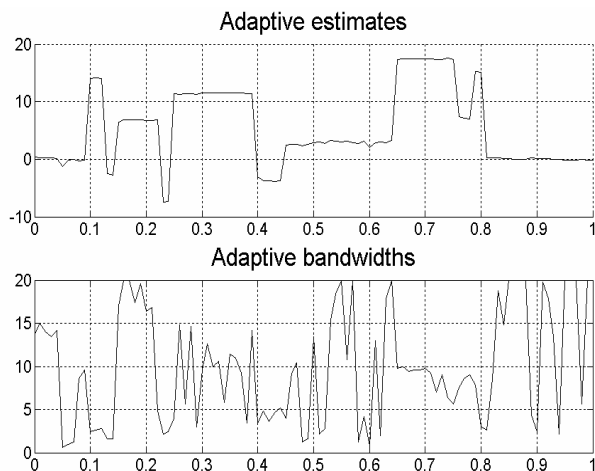### Adaptive estimates

Adaptive bandwidths

Fig.3 Adaptive bandwidths here are represented by h in the kernel Gaussian function $K_h(x)$. m=1, Γ =0.004.

We can see from Fig.2 and Fig.3 that in the discontinuities and impulses, the bandwidths of the LS-based method and M-estimation are both forced to a small value to reduce the bias in fitting them. We can still see that the bandwidths in Fig.3 are generally much bigger than the corresponding bandwidths in Fig.2. The reason is that in M-estimation, the Huber function can de-emphasize the effects of the impulses, which usually have an error much greater than ξ , and consequently the bandwidths can be bigger.

6. CONCLUSION

A new method for robust LPR under jump discontinuities and impulsive noise is presented in the paper. It is based on M-estimation, which can effectively combat the impulsive noise, while stabilizing the estimator around jump discontinuities. The ICI rule is used to select an adaptive local bandwidth for smoothing. Simulation results showed that the proposed method performs considerable better than traditional LS-based LPR

using ICI rule. These techniques are expected to have interesting applications in image and video processing.

7. REFERENCES

[1] M. P. Wand, M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, 1995.

[2] J. Fan, I. Gijbels, *Local Polynomial Modelling and Its Applications,* Chapman and Hall, London, 1996.

[3] 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 Dissertation*, Virginia Tech, 2000.

[4] D. Ruppert, "Empirical-bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation," *JASA*, 92, pp. 1049-1062, 1997.

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

[6] V. Katkovnik, "A New Method for Varying Adaptive Bandwidth Selection," *IEEE Transactions on signal processing*, Vol. 47, No. 9, pp.2567-2571, Sept. 1999.

[7] V. Katkovnik, "On Adaptive Local Polynomial Approximation with Varying Bandwidth," *IEEE International Conference of Acoustics, Speech and Signal Processing*, Seattle, Washington, United States of America, 1998.

[8] V. Katkovnik, K. Egiazarian, and J. Astola, "Application of the ICI principle to window size adaptive median filtering," Signal Processing, Vol. 83, N2, pp. 251-257, 2003.

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

[10] C. K. Chu, I. K. Glad, F. Godtliebsen, F. and J. S. Marron, "Edge-preserving smoothers for image processing", *Journal of the American Statistical Association*, 93(442), pp. 526-541, 1998.

[11] Y. Zou, S. C. Chan, T. S. Ng, "Least Mean M-estimate Algorithms for Robust Adaptive filtering in impulse noise", *IEEE Trans. on Circuits and Systems-II-analog and digital signal processing*, IEEE, 47, No. 12, pp. 1564-1569, 2000.

 [12] Y. Zou, S. C. Chan, 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.

[13] S. C. Chan and Y. Zou, "A Recursive Least M-estimate Algorithm for Robust Adaptive Filtering in Impulsive Noise: Fast Algorithm and Convergence Performance Analysis," to appear in *IEEE Trans. Signal Processing*.