TABLE I
SQUARE ROOT MEAN SQUARED ERRORS (SRMSE) OF
ESTIMATION USING THE LPA AND VARIOUS WAVELET METHODS

N	$\Gamma = var$	$\Gamma = 4.4$	$Wavelets$
$Blocks$			
256	0.61	1.56	(0.68-1.20)
512	0.46	0.92	(0.59-1.12)
1024	0.33	0.74	(0.47-1.03)
2048	0.25	0.49	(0.41-0.85)
$Heavy$			
256	0.47	0.79	(0.49-0.62)
512	0.37	0.67	(0.40-0.55)
1024	0.31	0.49	(0.32-0.46)
2048	0.24	0.34	(0.38-0.61)

A New Method for Varying Adaptive Bandwidth Selection

Vladimir Katkovnik

Abstract—A novel approach is developed to solve a problem of varying bandwidth selection for filtering a signal given with an additive noise. The approach is based on the intersection of confidence intervals (ICI) rule and gives the algorithm, which is simple to implement and adaptive to unknown smoothness of the signal.

Index Terms—Adaptive filtering, adaptive varying bandwidth, adaptive varying window length, segmentation.

I. INTRODUCTION

In this correspondence, we introduce an adaptive filter that produces piecewise smooth curves with a small number of discontinuities in the signal or its derivatives. It allows certain desirable features such as jumps or instantaneous slope changes to be preserved in the smooth curves. The algorithm is adaptive to unknown smoothness of the signal. The local polynomial approximation (LPA) is used as a tool for filter design as well as for a presentation of the developed general method of the bandwidth selection. This method can be applied for a variety of quite different linear and nonlinear problems where the bandwidth selection involves the bias-variance compromise usual for nonparametric estimation.

The author is with the Department of Statistics, University of South Africa, Pretoria, South Africa (e-mail: vlkatkov@hotmail.com).
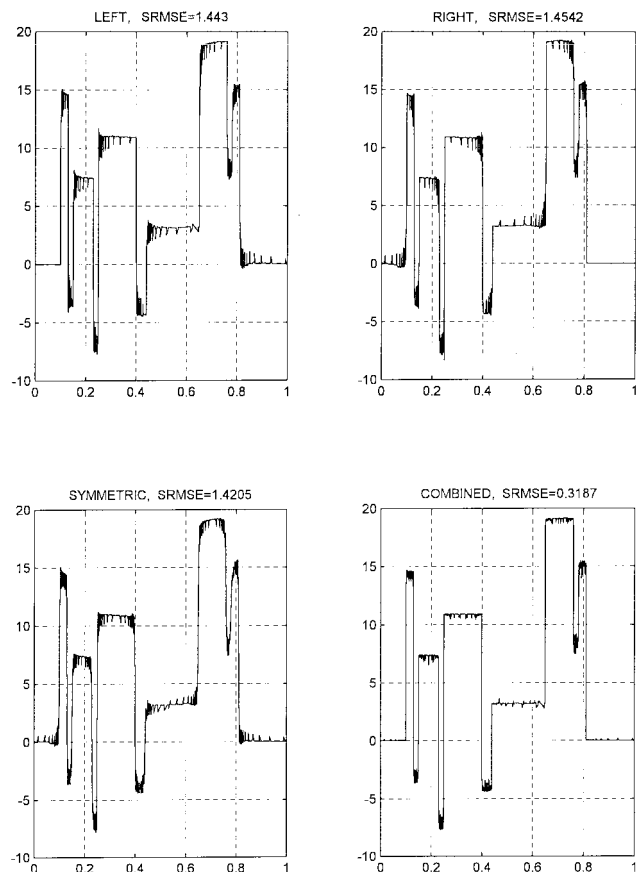
Fig. 1. Adaptive estimates, $m = 2$, $\Gamma = 4.4$.

Suppose that we are given noisy observations of a signal $y(x)$ with a sampling period Δ, i.e., $x_s = s\Delta$, $s = \cdots -1, 0, 1 \cdots$: $z_s = y(x_s) + \varepsilon_s$, where ε_s are independent and identically distributed Gaussian random errors $E(\varepsilon_s) = 0$, $E(\varepsilon_s^2) = \sigma^2$. It is assumed that
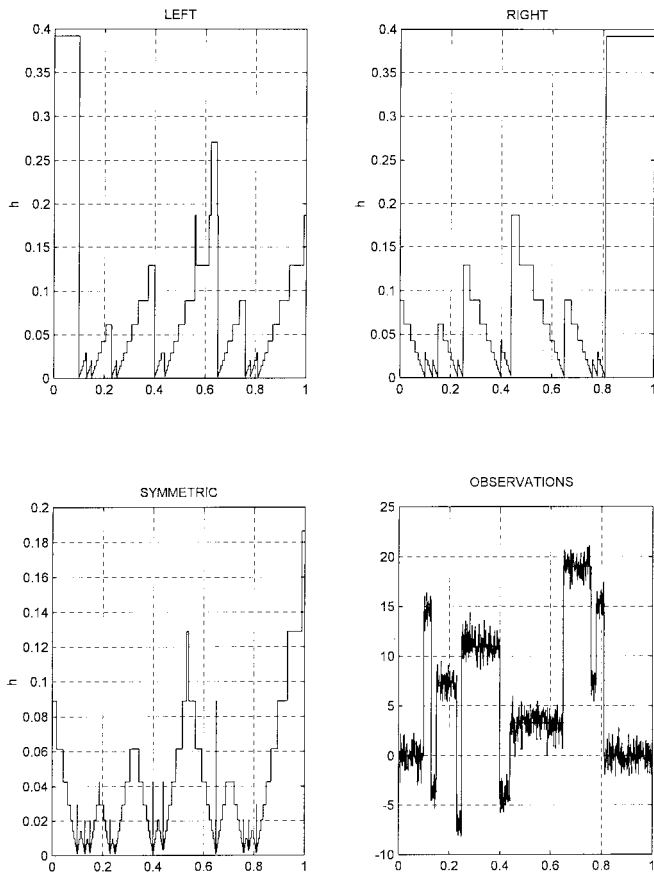
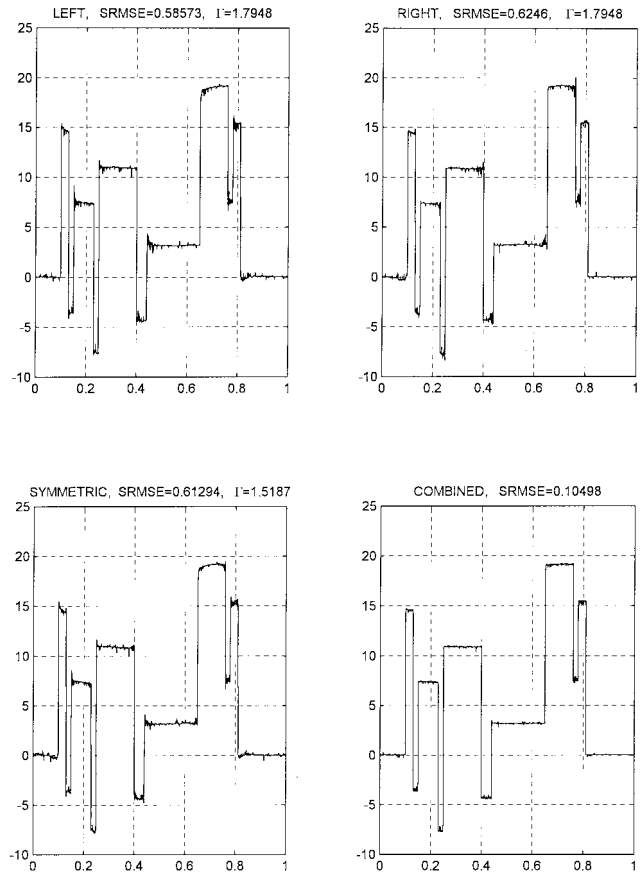Fig. 2.   Adaptive bandwidths, $m = 2$, $\Gamma = 4.4$ and observations.

Fig. 3.   Adaptive estimates, $m = 2$, $\Gamma = $ var.

$y(x)$ belongs to a nonparametric class of piecewise m-differentiable functions $\mathcal{F}_m = \{|y^{(m)}(x)| \leq L_m\}$. Our goal is to estimate the signal $y(x_s)$ and its derivatives $y^{(k)}(x_s)$, $1 \leq k \leq m - 1$, depending on the observations $(z_s)_{s=1}^N$ with the point-wise mean squared error (MSE) risks, which are as small as possible.

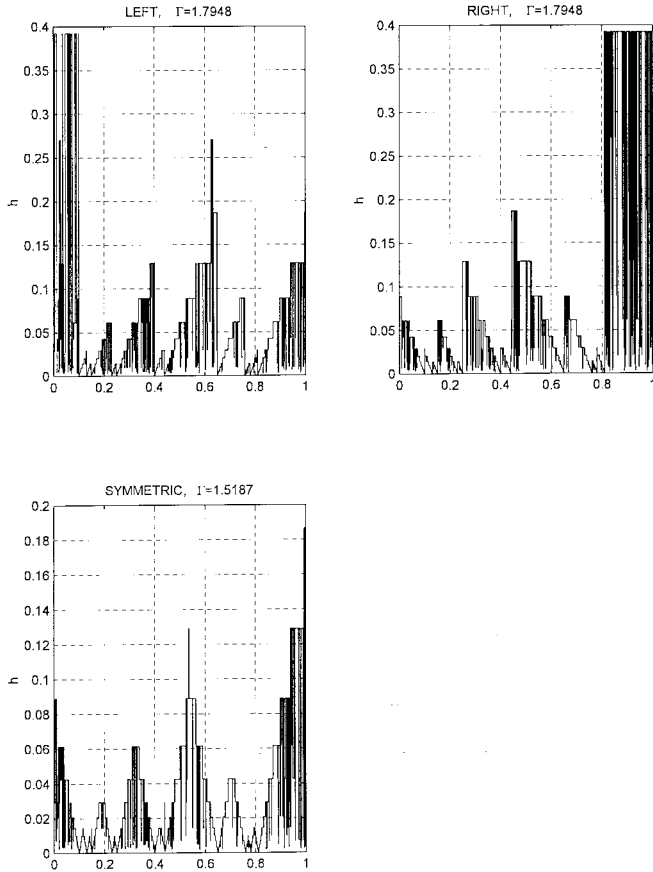The following loss function is applied in the standard LPA (e.g., [4] and [6])

$$J_h(x) = \frac{1}{N} \sum_{s=1}^{N} \rho_h(x_s - x)(z_s - C^T \phi(x_s - x))^2$$

$$\phi(x) = (1, x, x^2/2, \cdots, x^{m-1}/(m-1)!)'$$

$$C = (C_0, C_1, \cdots, C_{m-1})' \qquad (1)$$

where x is a "center," and m is the order of the LPA. The window $\rho_h(x) = \rho(x/h)/h$ is a function satisfying conventional properties of the "kernel" estimates, in particular, $\rho(x) \geq 0$, $\rho(0) = \max_x \rho(x)$, $\rho(x) \to 0$, as $|x| \to \infty$, and $\int_{-\infty}^{\infty} \rho(u)\,du = 1$. Here, h is a window "size" or a bandwidth. Minimizing $J_h(x)$ with respect to C, $\hat{C}(x, h) = \arg\min_{C \in R^m} J_h$ gives $\hat{y}_0(x, h) \triangleq \hat{C}_0(x, h)$ as an estimate of $y(x)$ and $\hat{y}_k(x, h) \triangleq \hat{C}_k(x, h)$, $k = 1, \cdots, m-1$, as estimates of the derivatives $y^{(k)}(x)$.

A crucial step in the nonparametric estimating is choosing the bandwidth h that controls the smoothness of the resultant estimate. Automatic bandwidth selectors generally fall into two broad classes of methods: plug-in and quality-of-fit approaches. The plug-in methods are based on the asymptotic formulas giving the optimal bandwidth as a function of the bias and variance of estimation. Main philosophical drawbacks of this approach are caused by using estimates of the higher order derivatives $y^{(m)}(x)$ in calculation of the bias, which is obviously a more difficult problem than the original estimation of the signal itself. The algorithms are quite complex and require

some additional parameters, in particular, determining a localization of the derivative estimates (e.g., [4], [10]). The second alternative approach does not have to deal with estimation of the bias. This group of methods uses the so-called quality-of-fit statistics, such as the cross-validation (CV), generalized cross-validation (GCV), C_p, Akaike criteria, etc. Most of these statistics can be represented in the form $Q = \log \hat{\sigma}^2 + \psi(h)$, where $\hat{\sigma}^2 = 1/N \sum_s (z_s - \hat{y}_0(x_s, h))^2$, and $\psi(h)$ is a penalty function designed to increase with decreasing the bandwidth h. The bandwidth h is selected by minimizing Q, and the penalty $\psi(h)$ preserves against taking too small values of h. Thus, minimizing Q is not linked directly to accuracy optimization. Most publications concerning this approach are related to a data-based global (constant) bandwidth (e.g., [5] and [6]). The LPA estimates with the varying bandwidth selection based on modified quality-of-fit statistics have been developed in [1] and [9]. This correspondence is inspired by the novel approach developed in [3], where a new bandwidth selection rule, which we name the intersection of confidence intervals (ICI), is proposed and substantiated. The ICI approach does not require estimation of the bias and differs from the quality-of-fit statistics mentioned above by its targeting to accuracy optimization. It is proved in [3] that the LPA equipped with the ICI rule for the bandwidth selection possesses simultaneously many attractive asymptotic properties, namely, 1) it is nearly optimal within the $\ln(1/\Delta)$ factor in the pointwise MSE risk for estimating both the function and its derivatives, and 2) it is spatially adaptive over a wide range of the classes of $y(x)$ in the sense that its quality is close to that which we could achieve if smoothness of $y(x)$ was known in advance.

This correspondence presents a development of the results obtained in [3] and further in [8]. Mainly, the modification concerns optimization of the threshold of the ICI according to the pointwise MSE risk.

Fig. 4.   Adaptive bandwidths, $m = 2$, $\Gamma = $ var.

Fig. 5.   Adaptive estimates, $m = 2$, $\Gamma = 3.0$.

It is shown that this threshold is an important design parameter of the algorithm, which influences the accuracy in a crucial way. Further, the cross-validation proves to be a good criteria for selection of a data-driven threshold. This data-driven adjustment of the threshold results in a very valuable accuracy improvement of the adaptive estimation.

II. THE IDEA OF THE ICI AND THE ALGORITHM

The estimation error of the LPA can be represented in the form

$$|e_k(x, h)| = |y^{(k)}(x) - \hat{y}_k(x, h)| \leq \omega_k(x, h) + |\zeta_k(x, h)| \quad (2)$$

where $\omega_k(x, h)$ is the estimation bias, and $\zeta_k(x, h)$ is a random error with the probability density $N(0, std_k(x, h))$. Then, $|\zeta_k(x, h)| \leq \chi_{1-\alpha/2} \cdot std_k(x, h)$ holds with the probability $p = 1 - \alpha$, where $\chi_{1-\alpha/2}$ is the $(1 - \alpha/2)$th quantile of the standard Gaussian distribution and with the same probability

$$|e_k(x, h)| \leq \omega_k(x, h) + \chi_{1-\alpha/2} std_k(x, h). \quad (3)$$
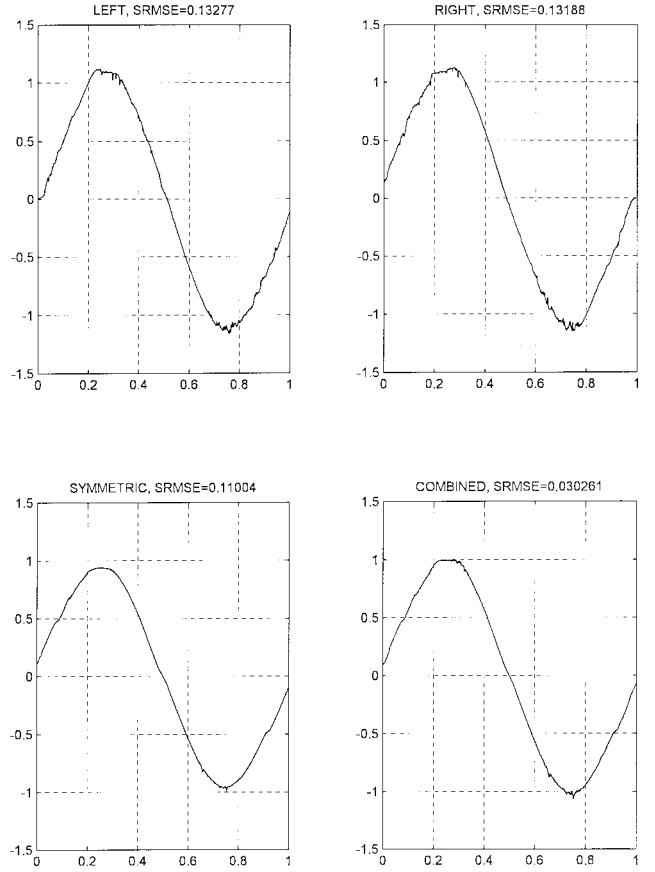
Let h_0 be the bandwidth corresponding to a desirable proportion between the bias and the random error given in the form

$$h_0 = \max\{h: \omega_k(x, h) \leq X \cdot \chi_{1-\alpha/2} std_k(x, h)\} \quad (4)$$

where $X > 0$ is a parameter determining the proportion (selection of X is discussed in [8]). Then, according to the properties of the bias and standard deviation, $\omega_k(x, h) \leq X \cdot \chi_{1-\alpha/2} std_k(x, h)$ for $h \leq h_0$, and the (3) can be weakened for $h \leq h_0$ to

$$|e_k(x, h)| \leq (1 + X)\chi_{1-\alpha/2} std_k(x, h). \quad (5)$$

In what follows, we use (5) corresponding to different bandwidth values in order to test the hypotheses $h \leq h_0$ and, in this way, to find h close to h_0. Let us introduce a finite set of bandwidth

$H = \{h_1 < h_2 < \cdots < h_J\}$ starting with quite a small h_1 and, according to (5), determine a sequence of the confidence intervals $\mathcal{D}_k(i)$ of the biased estimates as

$$\mathcal{D}_k(i) = [L_i, U_i]$$
$$U_i = \hat{y}_k(x, h_i) + \Gamma \cdot std_k(x, h_i)$$
$$L_i = \hat{y}_k(x, h_i) - \Gamma \cdot std_k(x, h_i) \quad (6)$$

where $\Gamma = (1 + X)\chi_{1-\alpha/2}$ is a threshold of the confidence interval. Then, for $h = h_j$, (5) is of the form $y^{(k)}(x) \in \mathcal{D}_k(j)$, and we can conclude from (3) that while the inequality $\omega_k(x, h) \leq X \cdot \chi_{1-\alpha/2} std_k(x, h)$ holds for $h = h_j$, $1 \leq j \leq i$, all of the intervals $\mathcal{D}_k(j)$, $1 \leq j \leq i$ have a point in common, namely, $y^{(k)}(x)$.

The following ICI statistic tests the very existence of this common point and gives the adaptive bandwidth value. Consider the intersection of the intervals $\mathcal{D}_k(j)$, $1 \leq j \leq i$ with increasing i, and let i^+ be the largest of those i for which the intervals $\mathcal{D}_k(j)$, $1 \leq j \leq i$ have a point in common. This i^+ defines the adaptive bandwidth and the adaptive LPA estimate as

$$\hat{y}_k^+(x) = \hat{y}_k(x, h^+(x, k)), \qquad h^+(x, k) = h_{i^+}. \quad (7)$$

The following algorithm implements the procedure (7):

$$\overline{L}_{i+1} = \max[\overline{L}_i, \quad L_{i+1}], \underline{U}_{i+1} = \min[\underline{U}_i, U_{i+1}]$$
$$i = 1, 2, \cdots, J, \overline{L}_1 = L_1, \underline{U}_1 = U_1. \quad (8)$$

Then, the optimal window length h_i^+ is the largest i when $\overline{L}_i \leq \underline{U}_i$ is still satisfied. We wish to emphasize that this bandwidth ICI selection procedure requires a knowledge of the estimate and its variance only and, in general, gives different bandwidths for estimates of the signal and derivatives.
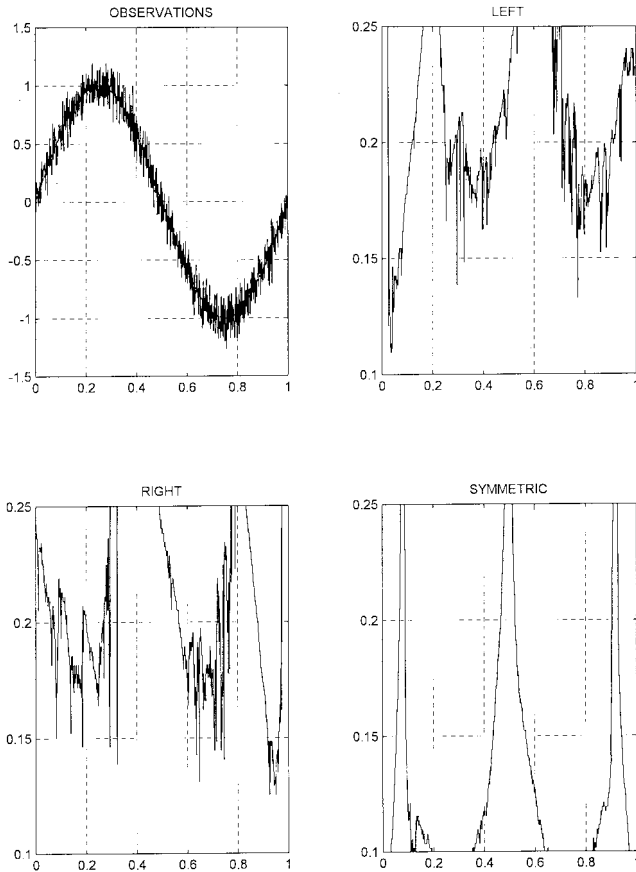
Fig. 6.   Observations and adaptive bandwidths, $m = 2$, $\Gamma = 3.0$.

Fig. 7.   Adaptive estimates, $m = 2$, $\Gamma = $ var.

The threshold Γ in $\mathcal{D}_k(j)$ plays a crucial role in the performance of the adaptive algorithm. It can be shown that the MSE of estimation has a minimum with respect to Γ. Larger values of Γ give $h_{i+} > h_0$ and oversmooth the signal while smaller values of Γ give $h_{i+} < h_0$ and undersmooth the signal. The asymptotic analysis given in [3] provides the convergence orders of the estimates as $\Delta \to 0$. However, it can not be used in order to find the optimal value of the threshold Γ. There are great difficulties in obtaining the optimal Γ from theoretical analysis. It seems that the data-driven adjustment of Γ is the only reasonable way, at least for now. It follows from our simulation experience that the cross-validation determines a reasonable and efficient selector for Γ. For the estimator $\hat{y}_0(x_s, h^+(x_s, 0))$, which is written as $\hat{y}_0(x, h) = \sum_s g(x, x_s, h)z_s$, the cross-validation loss function can be represented as the sum of the weighted squared residuals

$$I_{\text{CV}} = \sum_s \left(\frac{z_s - \hat{y}_0(x_s, h^+(x_s, 0))}{1 - g(x_s, x_s, h^+(x_s, 0))} \right)^2. \qquad (9)$$

Thus, the adaptive estimates are assumed to be calculated for every $\Gamma \in G$, $G = \{\Gamma_1, \Gamma_2, \cdots \Gamma_{N_G}\}$, and $\hat{\Gamma} = \arg\min_{\Gamma \in G} I_{\text{CV}}$ gives the adjusted threshold value used in the final estimate. Our attempts to use, instead of the cross-validation, another quality-of-fit statistic, in particular, the C_P, Akaike criteria, and its modifications (e.g., [5]), have not shown any improvement in the accuracy.

III. SIMULATION

Let ρ_L, ρ_R, and ρ_S be the left, right, and symmetric windows in (1), i.e., $\rho_L(u) = 0$ for $u > 0$, $\rho_R(u) = 0$ for $u < 0$, and $\rho_S(u) = \rho_S(-u)$. We consider combined estimates obtained by combining all three left, right, and symmetric estimates with
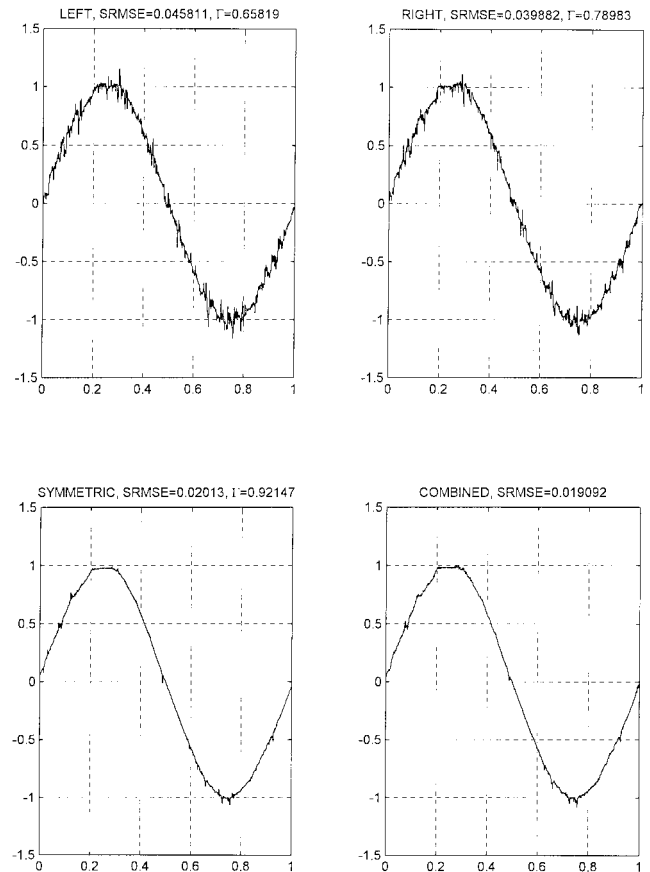
the inverse variances as the weights (see [3] and [8]). We use the linear LPA, $m = 2$, the observation interval [0, 1], $N = 1024$, and $\Delta = 1/N$. The signal outside of the segment [0, 1] is equal to 0. The standard deviation σ for $std_k(x, h)$ is estimated by $\hat{\sigma} = \{\text{median}(|z_s - z_{s-1}|: s = 2, \cdots, N)\}/(\sqrt{2} \cdot 0.6745)$.

1) We compare the adaptive LPA estimates with the results achieved by the wavelet filters on the test functions *Blocks* and *HeavySine.* These functions, noise, and conditions of the Monte Carlo statistical modeling are exactly as given in ([2, Tab. II, p. 1218]). It is assumed that $H = \{h_s | h_s = \Delta \cdot 1.45^{s+1}, s = 2, \cdots, 17\}$. The radix 1.45 for h_s is used instead of the standard dyadic 2 in order to get a more accurate fitting for both the optimal bandwidth and the estimated signal. The root of mean squared errors of estimation

$$\text{SRMSE} = \sqrt{\frac{1}{M} \sum_{j=1}^{M} \frac{1}{N} \sum_{s=1}^{N} \left(y(x_s) - \hat{y}_0^{[j]}(x_s, h^+(x_s, 0)) \right)^2}$$

$$(10)$$

are presented in Table I. The average over $M = 20$ simulation runs is calculated in SRMSE, where the superscript $[j]$ indicates the randomness of the estimate for every jth run. The second and third columns of the table give the SRMSE for the LPA estimator, respectively, with the adjusted threshold parameter $\Gamma = $ var and fixed $\Gamma = 4.4$, as it is used in the simulation given in [3]. The fourth column presents the interval of the SRMSE values obtained in [2] for the wavelet filters with different adaptive thresholds. It is evident from the table that the developed algorithm with the adjusted threshold parameter in all cases achieves a better accuracy than the wavelet estimators, whereas it is not true for the algorithm with a fixed value of the threshold parameter $\Gamma = 4.4$. The algorithm with the adjusted threshold parameter demonstrates the accuracy improvement around 1.5 to 2 times in comparison with the algorithm with $\Gamma = 4.4$.
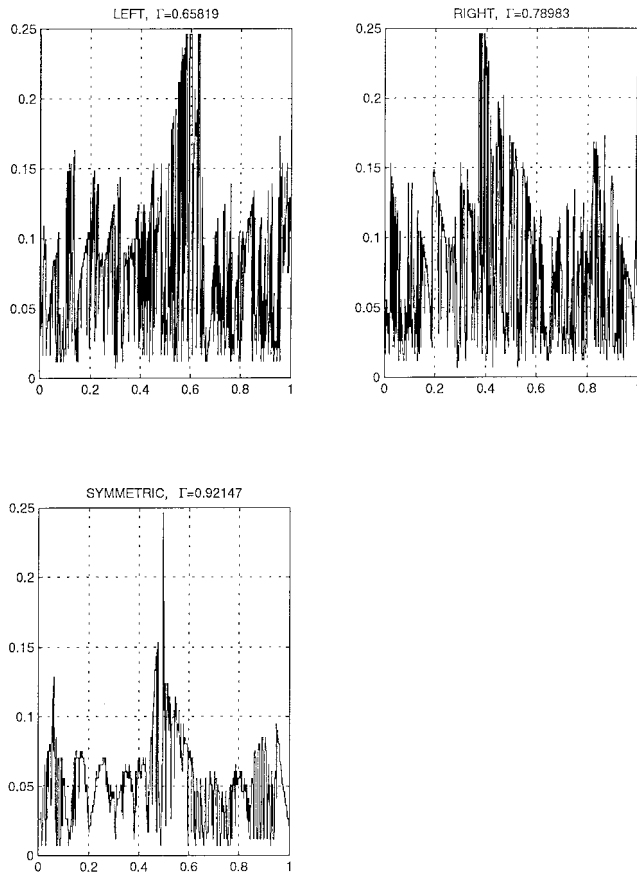
Fig. 8. Adaptive bandwidths, $m = 2$, $\Gamma = var$.

It is emphasized that in the Monte Carlo simulation, the variance of the adjusted $\Gamma = var$ between different runs is very small and does not exceed 10% of its mean value. The optimal accuracy is achieved for the comparatively small values of Γ from the interval (0.8, 1.5) that is different for the left, right, and symmetric windows. As $\Gamma = (1 + X)\chi_{1-\alpha/2}$, it means that in this optimization, $\chi_{1-\alpha/2}$, as well as the probability $p = 1 - \alpha$, become small. It follows that the accuracy optimization results in a high probability of the inequalities (3) to be violated.

Figs. 1–4 for the *Blocks* are presented in order to illustrate a qualitative difference between the adaptive estimates and bandwidths obtained with the fixed and adjusted threshold Γ. The estimates for the left, right, and symmetric windows, as well as the combined estimates, are given in Fig. 1 for the fixed $\Gamma = 4.4$. The corresponding adaptive varying bandwidths are shown in Fig. 2. The similar results for the adjusted $\Gamma = var$ are presented in Figs. 3 and 4. The values of the SRMSE, which are calculated according to (10) with $M = 1$, make clear how the combined estimator improves the accuracy. The comparison shows also that the values of SRMSE in Figs. 1 and 3 are always in favor of the estimates with the adjusted threshold Γ. The bandwidths of Fig. 2 obtained with $\Gamma = 4.4$ have a smoother appearance than those in Fig. 4, corresponding to $\Gamma = var$, but the SRMSE values definitely show better results for the estimates with the adjusted $\Gamma = var$.

It is clear from Fig. 2 that the ICI rule is quite sensitive with respect to the jumps of the signal and, for all left, right, and symmetric windows, gives the nearly ideal bandwidth selection exactly covering corresponding pieces of the segment [0, 1], where the signal is constant. Similar but "noisy" curves are seen in Fig. 4, corresponding to the smaller values of the adjusted Γ.

2) Let $y(x) = \sin(2\pi x)$ and $H = \{h_s: h_{s+1} = h_s + 2, s = 1, 2, \cdots, 125, h_1 = 2\}$. We use this fine uniform bandwidth grid to demonstrate in a clear form how varying adaptive bandwidths can be if the signal has a varying curvature. The adaptive estimates, as well as the corresponding values of SRMSE for the fixed $\Gamma = 3.0$ and the adjusted Γ, are shown, respectively, in Figs. 5 and 7. The adaptive bandwidths for the fixed and adjusted Γ are given in Figs. 6 and 8, respectively. It can be seen from these curves that at the peaks of $y(x)$, where the linear polynomial approximation is not able to be good enough in a large window, the adaptive bandwidths are decreasing.

Note that the values of SRMSE in Fig. 7 for the left, right, and symmetric window filters with the adjusted Γ are always better than those in Fig. 5 given for the fixed Γ. Thus, again, the adjustment of the threshold Γ improves the accuracy.

IV. CONCLUDING REMARKS

A novel approach to solve a problem of bandwidth selection for filtering a signal with additive noise is presented. The linear LPA is used in order to demonstrate the efficiency of the approach, whereas a development to nonlinear observations and robust M-estimates is possible. The algorithm is simple to implement and requires calculation of the estimates and their standard deviations for a set of the bandwidth values. For the data given on the regular grid, the fast implementation of the algorithm is done in MATLAB. The adaptive estimator is built as J parallel filters, which differ in the bandwidth h_j, $j = 1, 2, \cdots, J$ and the selector, which determine the best $h^+(x_s)$ and the corresponding estimate $\hat{y}_0(x_s, h^+(x_s, 0))$ for every x_s. In simulation, the adaptive LPA demonstrates a better accuracy as compared with the adaptive wavelet estimates. We wish to note that this advantage of the LPA is not accidental and reflects deep similarities and links that exist between the LPA and wavelets [7].

REFERENCES

[1] W. S. Cleveland and C. Loader, "Smoothing by local regression: Principles and methods," in *Statistical Theory and Computational Aspects of Smoothing*, W. Hardel and M. Schimek, Eds. Heidelberg, Germany: Physica-Verlag, 1996, pp. 10–49.
[2] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *JASA*, vol. 90, no. 432, pp. 1200–1224, 1995.
[3] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Stat.*, vol. 6, no. 2, pp. 135–170, 1997.
[4] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Application*. London, U.K.: Chapman & Hall, 1996.
[5] C. M. Hurvich and J. S. Simonoff, "Smoothing parameter selection in nonparametric regression using an improved AIC criterion," *J. R. Stat. Soc.*, ser. B, vol. 60, pp. 271–293, 1998.
[6] V. Katkovnik, *Nonparametric Identification and Smoothing of Data (Local Approximation Methods)*. Moscow, Russia: Nauka, 1985 (in Russian).
[7] ——, "Wavelets and local polynomial approximation," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, Paris, France, 1996, pp. 237–240.
[8] ——, "On multiple window local polynomial approximation with varying adaptive bandwidths," in *Proc. Comput. Stat. 13th Symp.*, Bristol, U.K., 1998, R. Payne and P. Green, Ed. Heidelberg, Germany: Physica-Verlag, pp. 353–358.
[9] J. A. McDonald and A. B. Owen, "Smoothing with split linear fits," *Technometrics*, vol. 28, no. 3, pp. 195–208, 1986.
[10] D. Ruppert, "Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation, *JASA*, vol. 92, no. 439, pp. 1049–1062, 1997.