Performance Analysis of the Adaptive Algorithm for Bias-to-Variance Trade-off

LJubiša Stanković

Abstract— **An algorithm for the mean squared error minimization, through the bias-to-variance ratio optimization, has been recently proposed and used in the literature. This algorithm is based on the analysis of the intersection of confidence intervals (ICI). The algorithm does not require explicit knowledge of the estimation bias for a "near to optimal" parameter estimation. This paper presents a detailed analysis of the algorithm performances, including procedures and relations that can be used for a fine adjustment of the algorithm parameters. Reliability of the algorithm is studied for various kinds of the estimation noise. Results are confirmed on a simulated example with uniform, Gaussian and Laplacian noise. An illustration of the algorithm application on a simple filtering example is given.**

I. Introduction

In numerous signal processing methods and applications of noisy signals, the result is a biased random variable. This is the case in filtering, smoothing, Fourier transform calculation, instantaneous frequency estimation, time-frequency distributions calculations, LMS adaptive algorithms, direction of arrival estimation, image and multidimensional signal processing, and many other applications, not restricted to signal processing [1]-[18]. The variance and bias in most of these cases are functions of one parameter (smoothing interval, number of samples, lag window, step-size value, number of sensors,...). Behavior of bias and variance is usually opposite with respect to this parameter. When the parameter increases then the variance increases (decreases) and the bias decreases (increases). The optimal parameter value can be determined by minimizing the estimation mean squared error (MSE), provided that some signal and noise parameters are explicitly known. However, these parameters are not available in advance. This is especially true for the estimation bias which is determined by the signal changes. The adaptive algorithm for determination of the parameter value close to the optimal one is recently proposed and intensively used [1]-[18]. The algorithm is based on the intersection of confidence intervals (ICI) [14], [15]. This algorithm does not require knowledge of the estimation bias value.

In the first part of this paper, after a review of the algorithm, a method for fine adjustment of the algorithm parameters is proposed. The second part of the paper introduces reliability analysis of the algorithm for various kinds of estimation noise. The paper is completed with a statistical study and numerical confirmation of the presented results.

II. Model and Optimal Parameter Value

Consider a noisy signal:

$$x(k) = f(k) + \epsilon(k), \tag{1}$$

with $f(k)$ being a signal and $\epsilon(k)$ being a stationary noise. Suppose that we want to estimate a quantity $Q(k)$ from this noisy signal. In general, this quantity (signal value, transform value, distribution value, instantaneous frequency, adaptive coefficient, direction of arrival...) is time-dependent. Also assume that its estimate $\hat{Q}(k)$ depends on a parameter h (smoothing interval, number of samples, lag window width, adaptive step-size value, number of sensors,...). Let the estimation bias be

$$bias(k,h) = \sqrt{B(k)h^n},$$

IEEE Transactions on Signal Processing, Vol.52, No. 5, May 2004

and the variance be

$$\sigma^2(h) = V/h^m. \quad (2)$$

Here, the parameter $B(k)$ depends on the unknown signal $f(k)$, and it is not known in advance. The variance of the estimate is assumed to be time-invariant (as it is true in all of the above mentioned applications), but it is dependent on the parameter h. Variance and squared bias dependence on h is of the $m-th$ and $n-th$ power, respectively. Note that, depending on the nature of parameter h, in some cases the bias and variance expressions can assume the form $bias(k,h) = \sqrt{B(k)/h^n}$ and $\sigma^2(h) = Vh^m$. It does not influence the algorithm applicability, since a simple substitution $h \to 1/h$ produces (2).

The MSE is of the form

$$E\left\{(Q(k) - \hat{Q}(k))^2\right\} = \frac{V}{h^m} + B(k)h^n. \quad (3)$$

The MSE in (3) has a minimum with respect to h. This minimum occurs for the optimal value of $h = h_{opt}(k)$. Thus, the optimal value of h follows from

$$\frac{\partial E\left\{(Q(k) - \hat{Q}(k))^2\right\}}{\partial h} =$$

$$-m\frac{V}{h^{m+1}} + nB(k)h^{n-1} = 0_{|h=h_{opt}}, \quad (4)$$

in the form

$$h_{opt}(k) = [mV/(nB(k))]^{1/(m+n)}. \quad (5)$$

Note that this relation is not useful in practice, because its right-hand side contains $B(k)$ which depends on the unknown signal $f(k)$.

For the optimal value of h relation (4) holds. Multiplying (4) by h, we get the relationship between the bias and standard deviation (2) for $h = h_{opt}(k)$,

$$bias(k, h_{opt}) = \sqrt{\frac{m}{n}}\sigma(h_{opt}). \quad (6)$$

Thus, the bias-to-standard deviation ratio is signal independent at $h = h_{opt}(k)$, $bias(k, h_{opt})/\sigma(h_{opt}) = \sqrt{m/n}$.

Note 1: The bias-to-standard deviation ratio is signal independent when $h = h_{opt}(k)$ for a more general form of the bias and variance functions: $bias(k,h) = \sqrt{B(k)z(h)}$, and $\sigma^2(h) = Vw(h)$, where $z(h)$ and $w(h)$ are positive monotonous functions satisfying the relation $z(h) = C_1/w^{C_2}(h)$, and C_1, C_2 are arbitrary positive constants. Proof is the same as for (4),(6).

Note 2: In many applications we can assume that the bias and standard deviation are of the same order of magnitude for a parameter h close to the optimal one, $bias(k, h_{opt})/\sigma(h_{opt}) \cong 1$. Then the presented analysis can be used in quite general cases, when the bias and standard deviation are just monotonous functions with opposite behavior.

III. Review of the Adaptive Algorithm

Here, we will review the adaptive algorithm [1]-[4], and introduce the parameters that will be analyzed in the sections that follow. The algorithm can produce $h_{opt}(k)$ or, due to discrete nature of h, inherent to this algorithm, a value close to $h_{opt}(k)$, **without having to know** $B(k)$.

Let us introduce a set $\mathbb{H}$ of discrete values of parameter h,

$$\mathbb{H} = \{h_s \mid\ h_s = ah_{s-1},\ s = 1, 2, ..., J\}, \quad (7)$$

with $a > 1$ and $h_0 > 0$. The algorithm can be generalized for any set $\mathbb{H} = \{h_s \mid\ h_s > h_{s-1},\ s = 1, 2, ..., J,\ h_0 > 0\}$.

In general, the exact optimal parameter h_{opt} is not equal to any of the values from the set $\mathbb{H}$. In order to relate h_{opt} with the values h_s from the set $\mathbb{H}$, assume that h_{opt} is close to a parameter h_{s^+} belonging to $\mathbb{H}$, $h_{s^+} \in \mathbb{H}$, i.e., $h_{s^+} \approx h_{opt}$. Then, we can write $h_{s^+} = a^p h_{opt}$, where p is a constant close to 0. According to (7) all other parameter values can be written as a function of h_{s^+} or h_{opt} as

$$\begin{aligned} h_s &= h_{s^+}a^{(s-s^+)} = h_{opt}a^{s-s^+ +p}, \\ (s - s^+) &= ..., -2, -1, 0, 1, 2, ... \end{aligned}$$

With this notation, having in mind (6), the bias and the standard deviation from (2) can be expressed, for any $h_s \in \mathbb{H}$, as

$$\begin{aligned} \sigma(h_s) &= \sqrt{V/h_s^m} \\ &= \sigma(h_{opt})a^{-(s-s^+ +p)m/2}, \end{aligned}$$

$$bias(k,h_s)=\sqrt{B(k)h_s^n}$$
$$=\sqrt{\frac{m}{n}}\sigma(h_{opt})a^{(s-s^+ +p)n/2}. \quad (8)$$

After we have defined the bias and the standard deviation of the estimate $\hat{Q}_s(k)$, obtained by using h_s, we can introduce the confidence intervals of the random variable $\hat{Q}(k)$. The confidence intervals play a crucial role in the algorithm. The estimate $\hat{Q}_s(k)$ is a random variable distributed around $Q(k)$ with the bias $bias(k,h_s)$ and the standard deviation $\sigma(h_s)$. The unbiased estimate $\left(\hat{Q}_s(k)-bias(k,h_s)\right)$ is centered around the true value $Q(k)$. Thus, we may write the relation:

$$\left|Q(k)-\left(\hat{Q}_s(k)-bias(k,h_s)\right)\right|\leq\kappa\sigma(h_s), \quad (9)$$

where the inequality holds with probability $P(\kappa)$ depending on parameter κ.[1] We will assume that κ is such that $P(\kappa)\to 1$.

The confidence intervals of the estimate, obtained by using a parameter $h_s\in\mathbb{H}$, are defined by $D_s=[L_s,U_s]$, where the lower and upper bound read

$$\begin{aligned} L_s &= \hat{Q}_s(k)-(\kappa+\Delta\kappa)\,\sigma(h_s),\\ U_s &= \hat{Q}_s(k)+(\kappa+\Delta\kappa)\,\sigma(h_s). \end{aligned} \quad (10)$$

Here, $\hat{Q}_s(k)$ is an estimate of $Q(k)$ obtained by parameter $h=h_s$, and $\sigma(h_s)$ is the standard deviation of $\hat{Q}_s(k)$.

For small values of h_s, when $s\ll s^+$, the bias is negligible, (8), thus $Q(k)\in D_s$ (with probability $P(\kappa+\Delta\kappa)\to 1$). Then, obviously, $D_{s-1}\cap D_s\neq\emptyset$, since at least the true value $Q(k)$ belongs to both confidence intervals. For $s\gg s^+$ the variance is small, but the bias is large (8). It is clear that there always exists such a large s that $D_s\cap D_{s+1}=\emptyset$ for a finite $(\kappa+\Delta\kappa)$.

The parameter $\Delta\kappa$ in D_s can be determined so that the largest s, for which the sequence of pairs of the confidence intervals D_{s-1} and D_s intersect, is $s=s^+$. Then, the intersection of confidence intervals D_{s-1} and D_s, which occurs when

$$\left|\hat{Q}_{s-1}(k)-\hat{Q}_s(k)\right|\leq(\kappa+\Delta\kappa)[\sigma(h_{s-1})+\sigma(h_s)], \quad (11)$$

works as an indicator of the event $s=s^+$, i.e., the event $h_s\approx h_{opt}$. Illustration of the probability density functions (pdf) of the estimate $\hat{Q}_s(k)$ for various values of h_s is shown in Fig.1.

IV. Parameters in the Adaptive Algorithm

There are three possible ways of choosing algorithm parameters κ, $\Delta\kappa$, and p.

A. Heuristic Approach

When our knowledge about the variance and bias behavior, given by (3), is not quite reliable, an approximative approach for determination of κ, $\Delta\kappa$, and p can be used. Then, we can use $a=2$ and assume, for example, a value of $\kappa\cong 2.5$ such that $P(\kappa)\cong 0.99$ for Gaussian distribution of estimation error. The value of $\Delta\kappa$ should take into account the bias for the expected optimal parameter value (6). It is common to assume that, for the optimal value of h, the bias and standard deviation are of the same order (Note 2), resulting in $\Delta\kappa\cong 1$. Then we can expect that the obtained value h_{s^+} is close to h_{opt}, thus $p\cong 0$, and the algorithm is completely defined, since all parameters for **the key algorithm equation (11)** are defined. This simple heuristic procedure has been successfully used in [1],[2].

B. "Analytic" Approach

When the knowledge about the variance and bias behavior is reliable, i.e., when (3) accurately describes estimation error, then we can calculate all algorithm parameters. According to the algorithm basic idea and monoto-

[1] If we assume, for example, that the random variable $\hat{Q}_s(k)$ is Gaussian, with the mean value $M=Q(k)+bias(k,h_s)$ and the standard deviation $\sigma(h_s)$, then the probability that $\hat{Q}_s(k)$ takes a value within the interval $[M-\kappa\sigma(h_s),M+\kappa\sigma(h_s)]$, i.e., $\left|\hat{Q}_s(k)-M\right|\leq\kappa\sigma(h_s)$, is $P(\kappa)=0.95$ for $\kappa=2$, and $P(\kappa)=0.997$ for $\kappa=3$. If the random variable $\hat{Q}_s(k)$ is uniformly distributed then the probability that $\hat{Q}_s(k)$ takes a value within $[M-\kappa\sigma(h_s),M+\kappa\sigma(h_s)]$ is $P(\kappa)=1$ for $\kappa>\sqrt{3}$. Therefore, a value of κ which guaranties $P(\kappa)$ close to 1 is the only algorithm condition. The algorithm applicability does not depend on the particular distribution of random variable $\hat{Q}_s(k)$. Detailed analysis of the algorithm reliability will be presented in Section V.

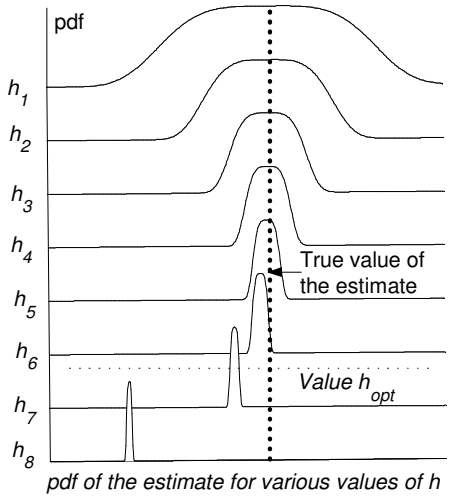

Fig. 1. Illustration of the probability density functions (pdf) of the estimate $\hat{Q}_s$ for various values of h_s

TABLE I
PARAMETERS IN THE ADAPTIVE ALGORITHM FOR $a = 2$ AND VARIOUS m, n, AND κ.

m	1	1	3	3	3	3	3	3
n	4	4	4	4	4	8	8	8
κ	2	3	2	3	5	2	3	5
$\Delta\kappa$	0.86	1.29	0.39	0.58	0.97	0.09	0.14	0.23
p	0.97	1.22	0.34	0.51	0.72	-0.13	-0.03	0.11
p_1	1.18	1.41	0.59	0.76	0.97	0.19	0.30	0.43

nous nature of the bias and standard deviation with respect to h, only three confidence intervals, D_{s^+-1}, D_{s^+}, and D_{s^++1}, should be considered. The confidence intervals D_{s^+-1} and D_{s^+} **should intersect,** while D_{s^+} and D_{s^++1} **should not intersect**. Assuming that relation (9) holds, and that the bias is positive, this condition means that the minimal possible value of upper D_{s^+-1} bound, (10), denoted by $\min\{U_{s^+-1}\}$, is always greater than or equal to the maximal possible value of the lower D_{s^+} bound, denoted by $\max\{L_{s^+}\}$, i.e.,

$$\min\{U_{s^+-1}\} \geq \max\{L_{s^+}\}.$$

The condition that D_{s^+} and D_{s^++1} do not intersect is given by

$$\max\{U_{s^+}\} < \min\{L_{s^++1}\}.$$

The maximal and minimal values of $\hat{Q}_s(k)$ follow from (9), as $Q(k) + bias(k, h_s) - \kappa\sigma(h_s) \leq \hat{Q}_s(k) \leq Q(k) + bias(k, h_s) + \kappa\sigma(h_s)$. By substituting these values into (10), the above two inequalities result in

$$bias(h_{s^+-1}) + \Delta\kappa\sigma(h_{s^+-1}) \geq bias(h_{s^+}) - \Delta\kappa\sigma(h_{s^+}),$$

and

$$bias(h_{s^+}) + (2\kappa + \Delta\kappa)\sigma(h_{s^+}) < bias(h_{s^++1}) - (2\kappa + \Delta\kappa)\sigma(h_{s^++1}). \quad (12)$$

Since the inequalities are written for the worst case, we can calculate the algorithm parameters by using the corresponding equalities. By using (8) we get the parameters

$$\Delta\kappa = \frac{2\kappa}{a^{(n+m)/2} - 1},$$

$$a^p = \left(\Delta\kappa\sqrt{\frac{n}{m}}\frac{a^{m/2}+1}{1-a^{-n/2}}\right)^{2/(m+n)}. \quad (13)$$

Values of parameters $\Delta\kappa$ and p for various values of m and n, and the most commonly used case $a = 2$, are given in Table I.

Note 3: "Linear or logarithmic scale"? For further, and very fine, tuning of the algorithm parameters, one may wish that the adaptive parameter is unbiased in the logarithmic, rather than in the linear scale (due to definition (7)). The additional logarithmic shift, due to a difference in arithmetic and geometric mean, is denoted by Δp.[2] Logarithmic shift $p_1 = p + \Delta p$ is presented in Table I.

Therefore, the adaptive value (as an estimate of the optimal parameter h value) should be

$$h_a = \hat{h}_{opt} = h_{s^+}/a^{p_1}$$

Note that the set $\mathbb{H}$ of parameter h values is **a priori assumed and fixed**. Therefore, as long as we can calculate the logarithmic shift p_1, we can use it in the following ways:

a) To calculate value of $Q(k)$ with the new value $h_a = h_{s^+}/a^{p_1}$ as the best estimate of h_{opt},

b) To remain within the assumed set of $h_s \in \mathbb{H}$, and to decide only whether to correct the obtained h_{s^+} or not. If $|p_1| \leq 1/2$, the correction of h_{s^+} is a^{p_1}. In this case it is smaller than the parameter h discretization step. Thus, if we remain within the assumed set $\mathbb{H}$ we can use $h_a = h_{s^+}$. For parameters m, n and κ (Table I) which produce $1/2 < p_1 \leq (1+1/2)$ it is better to use $h_a = h_{s^+}/a = h_{s^+-1}$, as the adaptive parameter value. Fortunately, the loss of accuracy for the adaptive h_a, as long as they are of h_{opt} order, is not significant since the MSE varies slowly around its stationary point, (4). Thus, in numerical implementations we can use only the values of h from the given set $\mathbb{H}$.

[2]The estimation bias and variance are exponential functions with respect to m and n, (8). Thus, the confidence interval limits change as $2^{(s-s^+)(m+n)/2}$, for the case when $a = 2$. The mean value for this exponential function, for two successive confidence intervals, for example $(s - s^+) = 0$ and $(s - s^+) = 1$, is $(1 + 2^{(m+n)/2})/2$. With respect to the geometrical mean $\sqrt{2^{(m+n)/2}}$ of these two intervals the mean value is shifted by approximately $\Delta p \cong 2\left[\log_2\left((1 + 2^{(m+n)/2})/2\right)\right]/(m + n) - 1/2$, resulting in logarithmic shift $p_1 = p + \Delta p$, presented in Table I.

C. Statistical Approach

The third approach for the parameter $(\kappa + \Delta\kappa)$ estimation is based on the statistical nature of confidence intervals, and on the a posteriori check of the fitting quality [2]. This approach is beyond the scope of this article.

V. Algorithm Reliability Analysis

Here we will analyze the probability that the algorithm produces "a false result" when the algorithm parameters are chosen according to Section IV.B. Consider values of parameter h from the discrete set $\mathbb{H}$. In general, the optimal parameter h_{opt} does not correspond to any particular value from $\mathbb{H}$. Let the optimal value h_{opt} lie between two adjacent values from the set $\mathbb{H}$, denoted by h_k and h_{k+1}. A false result is the value h_a obtained by the algorithm, which does not correspond to either h_k or h_{k+1}. A false result may be produced when the probability of (9) being satisfied is not $P(\kappa) = 1$. Since we start the analysis with the lower biases and larger variances, **a false result will appear if two confidence intervals do not intersect, despite a small bias**. Now, we will find that probability.

Assume that the error $\hat{Q}_{s-1}(k) - Q(k)$ takes a value x, $x > 0$, for a parameter $h = h_{s-1}$. Probability of this event is $\mathbf{p}_{s-1}(x)dx$, where $\mathbf{p}_{s-1}(x)$ is a pdf of the error $\hat{Q}_{s-1}(k) - Q(k)$. A false result will be produced if the error $\hat{Q}_s(k) - Q(k)$, with the next parameter h value $(h = h_s)$, is such that two confidence intervals, for h_{s-1} and h_s, do not intersect despite a very small (assume zero) bias, i.e.,

$$\left|\left(\hat{Q}_{s-1}(k) - Q(k)\right) - \left(\hat{Q}_s(k) - Q(k)\right)\right|$$

$$> (\kappa + \Delta\kappa)[\sigma(h_{s-1}) + \sigma(h_s)],$$

see (11). Since we assumed $\hat{Q}_{s-1}(k) - Q(k) = x$, this event occurs when the estimation error $\hat{Q}_s(k) - Q(k)$ produced by using h_s is greater than $x + (\kappa + \Delta\kappa)[\sigma(h_{s-1}) + \sigma(h_s)]$ or lower than $-(\kappa + \Delta\kappa)[\sigma(h_{s-1}) + \sigma(h_s)] + x$. Probability of this event is

$$\wp(x) = \int_{x+(\kappa+\Delta\kappa)[\sigma(h_{s-1})+\sigma(h_s)]}^{\infty} \mathbf{p}_s(y)dy$$

$$+\int_{-\infty}^{-(\kappa+\Delta\kappa)[\sigma(h_{s-1})+\sigma(h_s)]+x} \mathbf{p}_s(y)dy.$$

Therefore, the false result probability is:

$$P_F = 2\int_0^{\infty}\int_{x+(\kappa+\Delta\kappa)[\sigma(h_{s-1})+\sigma(h_s)]}^{\infty} \mathbf{p}_{s-1}(x)\mathbf{p}_s(y)dydx$$

$$+2\int_0^{\infty}\int_{-\infty}^{-(\kappa+\Delta\kappa)[\sigma(h_{s-1})+\sigma(h_s)]+x} \mathbf{p}_{s-1}(x)\mathbf{p}_s(y)dydx. \quad (14)$$

Special cases:

1. **Limited distribution of error**, $\mathbf{p}_s(x) = 0$ for $|x| > (\kappa + \Delta\kappa)\sigma(h_s)$: Then we have $P_F = 0$, i.e., it is impossible to get a false result. For example, for a uniformly distributed error $(\kappa + \Delta\kappa) > \sqrt{3}$ guaranties $P_F = 0$.

2. **Gaussian distributed error** $\hat{Q}(k) - Q(k)$, with $\mathbf{p}_s(x) = \exp\left(-x^2/(2\sigma_s^2)\right)/(\sigma_s\sqrt{2\pi})$: The false result probability is:

$$P_F = \mathrm{erfc}\left(\frac{(\kappa+\Delta\kappa)}{\sqrt{2}}\frac{\alpha+1}{\sqrt{\alpha^2+1}}\right), \quad (15)$$

where $\alpha = \sigma(h_{s-1})/\sigma(h_s) = a^{m/2}$ and $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty}\exp(-t^2)dt$. The last expression is derived by considering integration domain of (14) in the joint two-dimensional space (x, y).

For $a = 2$, $n = 4$, $m = 1$ and other parameters from Table I, we get that for $\kappa = 2$ the false result probability is $P_F < 0.0001$. For $m = 3$ we get $P_F < 0.003$ with $\kappa = 2$. For $\kappa = 3$ these probabilities are $P_F < 10^{-8}$ and $P_F < 10^{-5}$ for $m = 1$ and $m = 3$, respectively.

3. **Heavy-tailed Laplacian error** $\hat{Q}(k) - Q(k)$, with $\mathbf{p}_s(x) = \exp(-|x/\sigma_s|)/(2\sigma_s)$: In this case (14) assumes the form:

$$P_F = \frac{\exp(-(\kappa+\Delta\kappa)(1+1/\alpha))\alpha^2}{\alpha^2-1}$$

$$-\frac{\exp(-(\kappa+\Delta\kappa)(1+\alpha))}{\alpha^2-1}. \quad (16)$$

Heavy-tailed distributed error requires larger values of κ in order to satisfy the condition that $P(\kappa)$ is close to 1. For example, for $a = 2$, $n = 4$, $m = 3$ and $\kappa = 2$, we get $P_F < 0.05$, while for $\kappa = 3$ we get $P_F < 0.009$. For $\kappa = 5$ the false result probability is $P_F < 0.0004$. For $m = 1$ and the same other parameters we get $P_F < 0.015$, $P_F < 0.0013$, and $P_F < 0.000011$, for $\kappa = 2$, $\kappa = 3$, and $\kappa = 5$, respectively. Obviously, for heavy-tailed distributed errors larger values of κ are required in order to produce highly reliable results.

The false result probability P_F is calculated for two confidence intervals with h_{s-1} and h_s. For an optimal value h_{opt} close to h_{s^+} the overall false result probability is equal to the sum of probabilities that the false result occurred: in h_0 and h_1 confidence intervals check, that it did not occur during the first check but it occurred in h_1 and h_2 check, and so on until h_{s^+-1} and h_{s^+}. This probability is equal to $P_{FO} = P_F + (1-P_F)P_F + (1-P_F)^2P_F + ... + (1-P_F)^{s^+-1}P_F$. Since P_F is small we can write $P_{FO} \cong s^+P_F$. Probability that a false result will be obtained trough more confidence intervals checks, for large h_{opt}, is slightly higher than the overall false result probability for a small h_{opt}, when only a few checks should be done.

In this section we have considered false results due to nonintersection of confidence intervals intersection, when the bias is small and the standard deviation is large. The opposite event of confidence intervals intersection when the bias is large and the standard deviation is small is almost impossible, Fig.1.

VI. Illustrations with Statistical Study

Example 1 **(Gaussian error):** We have modeled a biased random variable as

$$\triangle Q = \hat{Q} - Q = \mathbf{w}\sqrt{V/h^m} + \sqrt{B(k)h^n}, \quad (17)$$

whose MSE is of form (3). Here, $\mathbf{w} = \mathcal{N}(0,1)$ is a Gaussian (zero-mean, unity-variance) random variable, $m = 3$, $n = 4$, and $V = 1$. The bias parameter $B(k)$ in $\triangle Q$ is logarithmically changed within $\frac{1}{7}\log_2(mV/nB(k)) \in [-4, 4]$, with the step 0.008 (in total 1000 values are considered).

For each of 1000 parameter $B(k)$ values, we have calculated optimal value for h according to (5), and plotted $\log_2 h_{opt}$ as a thick gray line in Fig.1.

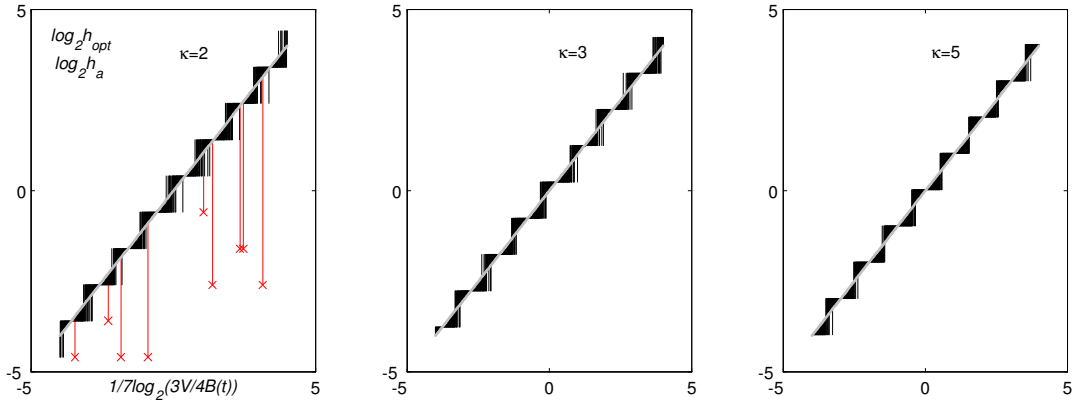

Fig. 2. Gaussian distribution of error: Optimal window width h_{opt} (straight gray line), and adaptive window widths h_a (end of the vertical lines, starting from the optimal window width line) for $a = 2$, $m = 3$, $n = 4$, $V = 1$. The variance to bias ratio $V/B(t)$ is logaritmicaly changed, in 1000 points. The adaptive width $h_a = h_{s+}/2^{p_1}$ is obtined by correcting h_{s+} by 2^{-p_1}, Table I. False results are indicated by "x".

Now, we have assumed that the bias parameter was not known, as it is the case in practical application. For a given unknown $B(k)$, the value of $\triangle\hat{Q}(k)$ was simulated for each $h_s \in \mathbb{H}$, according to (17). The assumed set of possible values of h was

$$\mathbb{H} = \{1/16, 1/8, 1/4, 1/2, 1, 2, 4, 8, 16, 32, 64\},$$

and $\kappa = 2$. The key algorithm relation (11) was tested for each of $B(k)$ values, with the known standard deviation $\sigma(h_s) = \sqrt{V/h_s^m}$. The largest value of h_s for which the equation (11) was still satisfied, was denoted by h_{s+}. Value $\Delta\kappa = 0.39$, corresponding to $m = 3$, $n = 4$, $\kappa = 2$, was used (Table I). The adaptive values $h_a = h_{s+}/2^{p_1}$, $p_1 = 0.59$ (Table I), obtained in this way, are connected with the optimal value line by thin vertical lines in Fig.1.

The same simulation is repeated with $\kappa = 3$ and $\kappa = 5$.

We can conclude that the presented algorithm almost always chooses the value h_s from $\mathbb{H}$ which is one of two the nearest values to the optimal one. However, for relatively small $\kappa = 2$ there are few complete misses of the optimal value. The number of these misses ("false results") is in full accordance with the algorithm reliability analysis from the previous section, eq.(15).

Example 2 (**Uniform error**)*:* When the error $Q(k)-\hat{Q}(k)$ is uniformly distributed within the interval $[-\Delta_h, \Delta_h]$ (for example, the error due to the quantization noise), then the variance is $\sigma(h) = \Delta_h^2/3$. It is obvious that (9) is satisfied with $P(\kappa) = 1$ for $\kappa > \sqrt{3}$. The previous experiment is repeated with a simulated random variable $\triangle Q = \hat{Q} - Q = \mathbf{w}\sqrt{V/h^m} + \sqrt{B(k)h^n}$ where $\mathbf{w} = \mathcal{U}(0,1)$ is a uniform (zero-mean, unity-variance) random variable, again with $m = 3$, $n = 4$, $\kappa = 2$ and $V = 1$. Values $\Delta\kappa = 0.39$ and $h_a = h_{s+}/2^{p_1}$, $p_1 = 0.59$, follow from Table I. The bias parameter $B(k)$ in $\triangle Q$ is logarithmically varied within the same interval, $\frac{1}{7}\log_2(mV/nB(k)) \in [-4, 4]$, with step 0.008. The adaptive values h_a obtained by the algorithm are again connected with the optimal thick line, by vertical lines, Fig.2. We can see that, opposite to the Gaussian distributed case, there is no false results, since $P(\kappa) = 1$ for all used κ.

Example 3 **(Laplacian error)***:* The same parameters as in the previous two examples are assumed here, but with the Laplacian distributed error $Q - \hat{Q}$. Laplacian random variable of unity variance $\mathbf{w} = \mathcal{L}(0,1)$ is formed by using $\mathbf{w} = (\mathbf{w}_1\mathbf{w}_2 + \mathbf{w}_3\mathbf{w}_4)/\sqrt{2}$, where $\mathbf{w}_i$ are Gaussian random variables $\mathbf{w}_i = \mathcal{N}(0,1)$. Values of $\kappa = 2$, $\kappa = 3$, and $\kappa = 5$, are considered. Since this noise is of a heavy-tailed type, the lowest $\kappa = 2$ here produces quite low $P(\kappa)$,

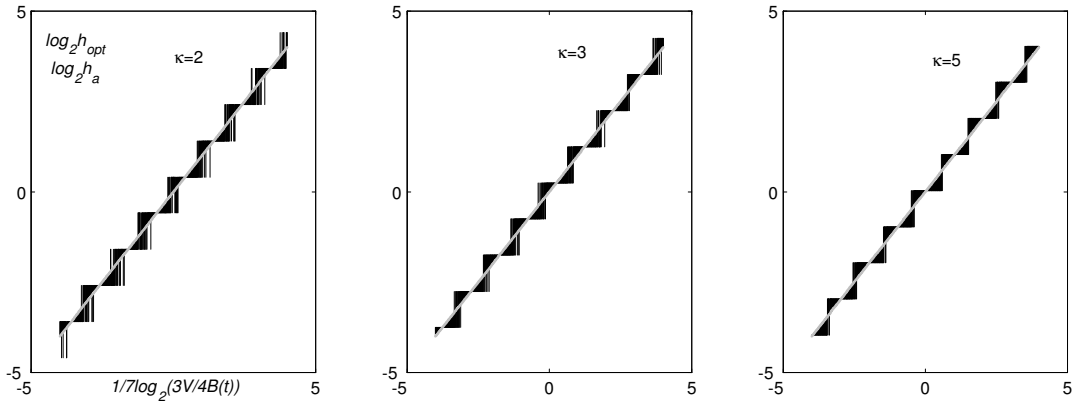

Fig. 3. Uniformly distributed error: Optimal window width h_{opt} (straight gray line), and adaptive window widths h_a (end of the vertical lines, starting from the optimal window width line) for $a = 2$, $m = 3$, $n = 4$, $V = 1$. The variance-to-bias ratio $V/B(t)$ is logaritmicaly changed, in 1000 points. The adaptive width $h_a = h_{s+}/2^{p_1}$ is obtined by correcting h_{s+} by 2^{-p_1}, Table I.

with a small reliability of the algorithm, Fig.3. Number of false result points is in agreement with (16). Thus, in order to improve the algorithm performance, higher values of κ ($\kappa = 3$, $\kappa = 5$) should be used.

Example 4: Application of the proposed algorithm on the smoothing of signal $f(t)$, corrupted by a zero mean stationary Gaussian additive noise $\epsilon(t)$ of variance σ_ϵ^2, will be the topic of this example. The aim of this example is not to show the efficiency of the proposed algorithm in signal denoising, or to compare it with other methods for the same application. The aim is only to illustrate the presented algorithm on one specific, very simple example, already used in literature, for example [11].

Consider estimation of $f(t)$ from

$$x(t) = f(t) + \epsilon(t)$$

based on a simple smoothing:

$$\hat{f}(t) = \frac{1}{h}\int_{-h/2}^{h/2} x(t+\tau)d\tau.$$

Bias of this estimator is:

$$bias(t,h) = E\{\hat{f}(t)\} - f(t)$$

$$= \frac{1}{h}\int_{-h/2}^{h/2} f(t+\tau)d\tau - f(t) \cong f''(t)\frac{h^2}{24},$$

where the Taylor expansion $f(t+\tau) = f(t) + f'(t)\tau + f''(t)\tau^2/2 + ...$ is used.

Note 4: The bias given by the last equation is only an approximation up to the second order term. The bias increases as parameter h increases. However, the bias cannot increase infinitely as $h \to \infty$. In this example, it is obvious that there is a limit for the bias. The maximal possible bias is equal to the maximal possible difference in the signal, i.e., $|bias(t,h)| \leq 1$. We should be aware of this fact, especially in order to avoid using extremely large values for κ.

Variance of the estimator is

$$\sigma^2(h) = \frac{\sigma_\epsilon^2}{h}.$$

Therefore, this case approximately corresponds to the described model, with $m = 1$ and $n = 4$.

For illustration we will consider

$$f(t) = \frac{1}{1 + (t/7.5)^{40}}$$

within the interval $|t| < 25$, with the step $\Delta t = 1/25$. The nonnoisy signal is shown in Fig.4a, while the signal with additive Gaussian white noise, with standard deviation $\sigma_\epsilon = 0.1$, is shown in Fig.4b. The noise standard deviation value σ_ϵ, needed for the algorithm, is

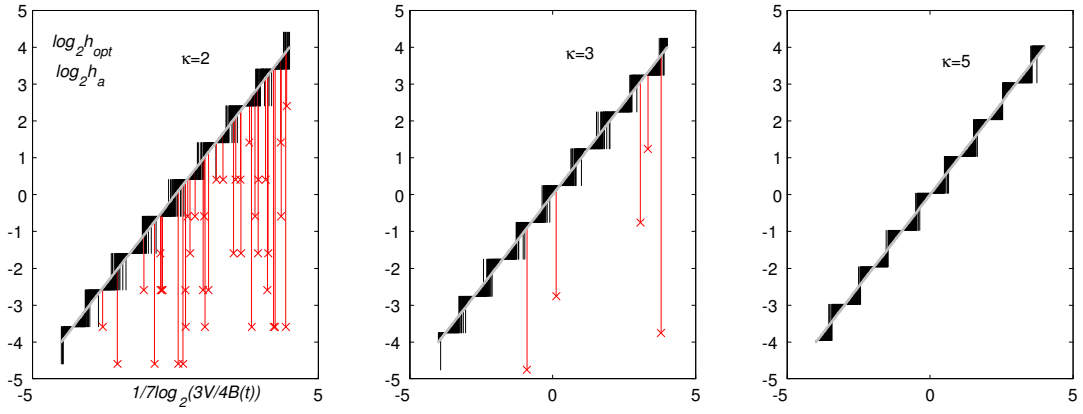

Fig. 4. Laplacian (heavy-tailed) distributed error: Optimal window width h_{opt} (straight gray line), and adaptive window widths h_a (end of the vertical lines, starting from the optimal window width line) for $a = 2$, $m = 3$, $n = 4$, $V = 1$. The variance-to-bias ratio $V/B(t)$ is logaritmicaly changed, in 1000 points. The adaptive width $h_a = h_{s+}/2^{p_1}$ is obtined by correcting h_{s+} by 2^{-p_1}, Table I. False results are indicated by "x"

estimated by using

$$\hat{\sigma}_\epsilon = \frac{\text{median}\,|x(n) - x(n-1),\, n = 1, ..., N|}{0.6745\sqrt{2}}. \tag{18}$$

In our case it resulted in $\hat{\sigma}_\epsilon = 0.101$, what is very close to the original σ_ϵ value. Values of $a = 2$ and $\kappa = 2$ are used, while $\Delta\kappa = 0.86$ is taken from Table I. Mean absolute errors (MAE) are: $MAE = 0.041$ for constant h with 129 smoothing points, Fig.4d, $MAE = 0.047$ for constant h with 3 smoothing points, Fig.4f, and $MAE = 0.009$ for the adaptive parameter h, Fig.4e. Note that in Fig.4 the total number of smoothing points was $2h(t) + 1$.

VII. Conclusion

The algorithm for parameter optimization, in a quite general formulation of the estimation problem, is considered. Presented study enables fine adjustment of the algorithm parameters. Reliability study for a general form of estimation error is done. It has been shown that even in the cases of some heavy tailed types of the estimation noise, like the Laplacian one, the algorithm can produce accurate and reliable results.

Acknowledgment: This work is supported by the Volkswagen Stiftung, Federal Republic of Germany.

References

[1] V. Katkovnik, LJ. Stanković: "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length", *IEEE Trans. on Signal Processing,* vol.46, no.9, Sept. 1998, pp.2315-2325.

[2] V. Katkovnik, LJ. Stanković: "Periodogram with varying and data-driven window length", *Signal Processing,* vol. 67, no.3, 1998, pp. 345-358.

[3] LJ. Stanković and V. Katkovnik: "Algorithm for the instantaneous frequency estimation using time-frequency distributions with variable window width", *IEEE Sig. Proc. Let.*, Vol.5, No.9, Sept. 1998. pp.224-228.

[4] LJ. Stanković, V. Katkovnik: "Instantaneous frequency estimation using higher order distributions with adaptive order and window length", *IEEE Trans. Information Theory*, Vol.46, No.1, Jan.2000, pp.302-311.

[5] LJ. Stanković: "Adaptive Instantaneous frequency estimation using TFDs", in *Time-Frequency Signal Analysis and Processing*, ed. B. Boashash, Elsevier, 2003.

[6] LJ. Stanković, V. Katkovnik: "The Wigner distribution of noisy signals with adaptive time-frequency varying window", *IEEE Trans. on Signal Processing*, vol.47, no.4, April 1999, pp.1099-1108.

[7] B. Krstajić, LJ. Stanković, Z. Uskoković: "An approach to variable step-size LMS algorithm", *IEE Electronics Letters*, Vol. 38, No. 16, Aug. 2002. pp.927-928.

[8] A.B. Gershman, LJ. Stanković, V. Katkovnik: "Sensor array signal tracking using a data-driven window approach", *Signal Processing*, Vol.80, No.12, Dec.2000, pp.2507-2515.

[9] LJ. Stanković, S. Stanković, I. Djurović: "Space/spatial-frequency analysis based filtering", *IEEE Trans. on Signal Processing*, Vol.48, No.8, Aug.2000, pp.2343-2352.

[10] LJ. Stanković: "On the time-frequency analy-

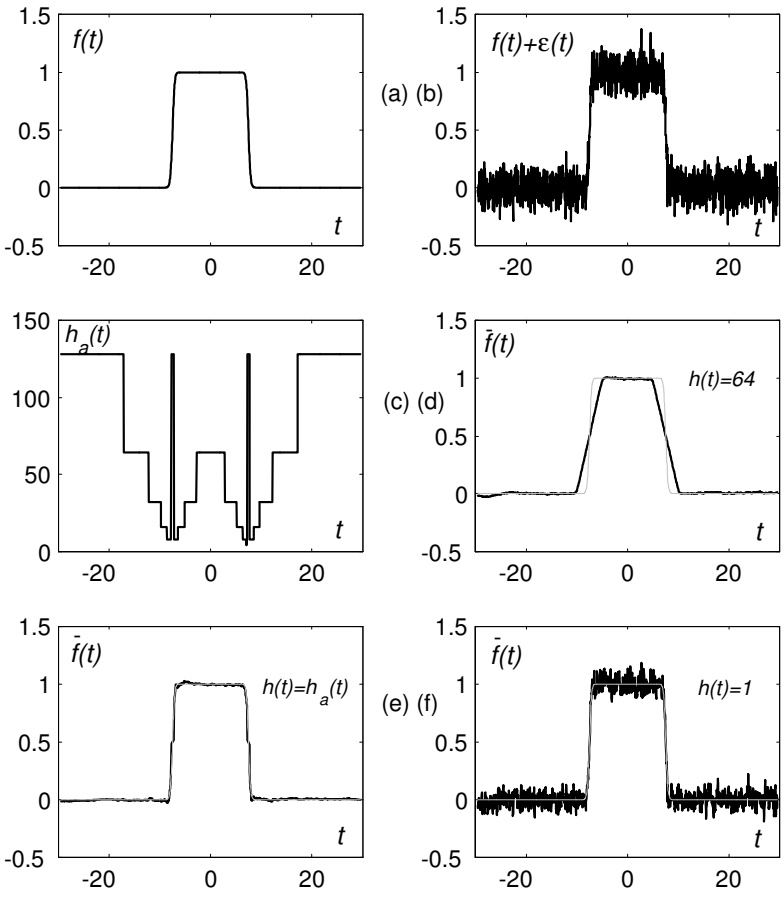

Fig. 5. Illustration of noisy signal smoothing by using the considered algorithm: a) Original signal, b) Noisy signal, c) Adaptive parameter $h_a(t)$, d) Signal smoothed in a wide range with constant $2h(t)+1=129$, e) Adaptive smoothed signal by using the widths $h_a(t)$ from c), f) Signal smoothed in a narrow range with constant $2h(t)+1=3$.

sis based filtering", *Annales des Telecommunications*, Vol.55, No.5-6, May 2000, pp.216-225.

[11] V. Katkovnik: "A new method for varying adaptive bandwidth selection", *IEEE Trans. on Signal Processing*, Vol.47, No.9, Sept.1999, pp.2567 -2571.

[12] J. Astola, K. Egiazarian, V. Katkovnik: "Adaptive window size image denoising based on ICI rule", in *Proc. Int. Conf. ASSP,* Vol. 3, 2001, pp.1869 -1872.

[13] I. Djurović, LJ. Stanković: "Adaptive windowed Fourier transform", *Signal Processing*, Vol.83, 2003, No.1, Jan.2003, pp.91-100.

[14] A. Goldenshluger and A. Nemirovski: "Adaptive de-noising of signals satisfying differential inequalities," *IEEE Trans. IT*, Vol. 43, May 1997, pp. 873-889.

[15] V. Katkovnik, "Adaptive local polynomial periodogram for time-varying frequency estimation", in *Proc. IEEE-SP on TFTS Analysis*, Paris, pp.329-332, June 1996

[16] B. Barkat: "Instantaneous frequency estimation of nonlinear frequency-modulated signals in the presence of multiplicative and additive noise", *IEEE Trans. on Signal Processing*, Vol.49, No.10, Oct 2001, pp 2214 -2222.

[17] Z.M. Hussain, B. Boashash: "Adaptive instantaneous frequency estimation of multicomponent FM signals using quadratic time-frequency distributions", *IEEE Trans. on Signal Processing*, Vol.50, No.8, Aug.2002, pp.1866-1676.

[18] V. Katkovnik, K. Egiazarian, J. Astola: "Application of the ICI principle to window size adaptive median filtering", *Signal Processing*, Vol.83, 2003, No.1, Feb.2003, pp.251-257.