ЦРНОГОРСКА АКАДЕМИЈА НАУКА И УМЈЕТНОСТИ
ГЛАСНИК ОДЈЕЉЕЊА ПРИРОДНИХ НАУКА, 13, 2000.
ЧЕРНОГОРСКАЯ АКАДЕМИЯ НАУК И ИСКУССТВ
ГЛАСНИК ОТДЕЛЕНИЯ ЕСТЕСТВЕННЫХ НАУК, 13, 2000.
THE MONTENEGRIN ACADEMY OF SCIENCES AND ARTS
GLASNIK OF THE SECTION OF NATURAL SCIENCES, 13, 2000.

UDK 621.391

Ljubiša Stanković, Vladimir Katkovnik***

AN ALGORITHM FOR THE INSTANTANEOUS FREQUENCY ESTIMATION USING THE S-DISTRIBUTION

Abstract

The S-distributions are defined in order to improve the concentration of signal's time-frequency representation, along the instantaneous frequency (IF). For a finite distribution order and non-linear IF, the S-distribution gives biased IF estimates. In the case of noisy signals, the optimal choice of the window length and the distribution order, based on asymptotic formulas for the variance and bias, can resolve the bias-variance trade-off usual for nonparametric estimation. However, the practical value of such optimal estimator is not significant since the optimal window length depends on the unknown smoothness of the IF. The main goal of this paper is to develop an adaptive, the S-distribution based, IF estimator with the time-varying and data-driven window length and distribution order which are able to provide the

*Prof. dr Ljubiša Stanković, Elektrotehnicki fakultet, University of Montenegro, 81000 Podgorica, Montenegro, Yugoslavia

**Prof. dr Vladimir Katkovnik, University of South Africa, P.O.Box 392 Pretoria, Republic of South Africa

quality close to the one that could be achieved if the smoothness of the IF was known in advance. The developed algorithm uses only the formula for the variance of the estimates. Simulation shows a good accuracy ability of the adaptive algorithm.

Keywords: Time-frequency analysis, Instantaneous frequency estimation, S-distribution

1. INTRODUCTION

Since there is no distribution from the Cohen class (with signal independent kernel) which can produce the complete concentration along the instantaneous frequency (IF) when it is not a linear function of time, [4], [5], [17], various higher order distributions have been derived. For the analysis of signals with polynomial phase the Polynomial Wigner-Ville distribution are proposed by Boashash et all. in [2], [3]. The same class of signals may be efficiently treated by the local polynomial distributions, defined in [7], [8]. The L-Wigner distribution, introduced and described in [16], [18], [19], significantly reduces the influence of higher order terms in the phase function, when it is non-linear. The Polynomial Wigner-Ville distribution, as well as the L-Wigner distributions, are closely related to the time-varying higher order spectra, [3], [14]. The recently proposed S-distributions may achieve high concentration at the IF, as high as the L-Wigner distributions, while at the same time satisfying energy unbiased condition, time marginal and, for asymptotic signals (signals whose amplitude variations are much slower than its phase variations, [3]), frequency marginal.

In this paper we analyze the IF estimator, in the case of noisy signals, using the S-distribution. The estimator's variance and bias are highly dependent on the window length, as well as on the S-distribution order. Provided that the signal and noise parameters are known then, by minimizing the estimation mean squared error, an optimal window length and/or distribution order may be determined. But, those parameters are not available in advance. Especially it is true for the IF derivatives which determine the estimation bias. Here, we will present an adaptive algorithm which does not require a priori knowledge of the estimation bias at all and in particular a knowledge of the IF derivatives.

The basic idea of the method which we use for selection of the data-driven window length and order, is originated in [6] where it was proposed and justified for selection of a varying and data-driven window size in a local polynomial fitting a linear regression. The idea of this method was exploited in [9] for development of the adaptive local polynomial periodogram, giving estimates of the IF and its derivative. It was subsequently used in [10] for development of the nonparametric estimator of the IF, based on the Wigner distribution with the data-driven adaptive window size.

An analysis of the discrete nature of the optimization parameters of the $S-$distribution (window length and distribution order) as a contribution to the described algorithm itself, is done in this paper. This analysis may help in better understanding of the influence of parameters' discretization, ant it could be a basis for suitable selection of their values. A discrete nature of optimization parameters, along with a small number of their possible values, resulted also in a modification of the original algorithm, [6]. We concluded that better results, in this case, may be achieved if we use the approach of sliding pair-wise confidence intervals, instead of the intersections of all previous confidence intervals considered in [6]. The analysis, presented in this paper, leads to one more interesting conclusion that not only higher order distributions may improve the time-frequency presentation, but also "lower order" distributions may be the best choice in some cases.

The paper is organized as follows. A review of the S-distribution definition, along with a noise modeling, is done in Section 2 of the paper. In Section 3 the variance and bias of the IF estimate, using the S-distributions, are derived. The optimal window length and distribution order are also discussed in this section. A numerical implementation of the S-distribution is presented in Section 4. An adaptive IF estimator, with the data-driven window size and distribution order, is described in Section 5. Numerical examples are presented in Section 6.

2. DEFINITIONS AND NOISE MODELING

The S-distribution of a time-discrete signal $s(nT)$, at a given instant t, is defined in [17], [21]:

$$SD_L(t,\omega) = \sum_{n=-\infty}^{\infty} w_h(nT) s^{[L]}(t + n\frac{T}{L}) s^{[L]*}(t - n\frac{T}{L}) e^{-j2nT\omega}, \quad (1)$$

where $w_h(nT) = T/h \cdot w(nT/h)$, with $w(t)$ being a real-valued finite-length symmetric window, $w(t) = 0$, for $|t| > 1/2$. A modification of the signal $s(nT) = A(nT)\exp(j\phi(nT))$ denoted by $s^{[L]}(nT)$ is obtained by multiplying a phase function by L, while keeping the amplitude unchanged:

$$s^{[L]}(nT) \doteq A(nT) e^{jL\phi(nT)}. \quad (2)$$

Note that in the realization by definition (1), the signal has to be sampled with the sampling interval multiplied by factor of $1/L$ with respect to the sampling interval in the Wigner distribution[1]. Its values should be available not only at the instants defined by the Nyquist sampling rate π/ω_m, where ω_m is the maximal signals frequency, but also at the points $n\pi/(2\omega_m L)$. Note that this can be avoided, i.e., $SD_L(t,\omega)$ can be realized without oversampling by using the S-method and procedure described in [17], [20]. This realization would also produce the S-distribution which is, in the case of multicomponent signals, equal to the sum of S-distributions of each individual component, with a significant reducing of noise influence. Since in this paper the realization is not an issue, we will assume that the S-distribution is realized according to the definition (the worst case).

Consider a noisy signal:

$$x(nT) = s(nT) + \epsilon(nT), \quad s(t) = A\exp(j\phi(t)) \quad (3)$$

[1] For an integer $L > 1$, as we used in our previous papers [15,16,17,18], this means signal oversampling L times. But, in this paper, we will allow values $0 < L < 1$ (for example $L = 1/2$), which will be useful in some noisy cases and will mean signal downsampling.

with $s(nT)$ being a signal with a real-valued amplitude A and $\epsilon(nT)$ being a white complex-valued Gaussian noise with mutually independent real and imaginary parts of equal variances $\sigma^2/2$. In the analysis presented in this paper we additionally assume that the noise is small with respect to the signal, i.e. $\sigma/A \ll 1$. Note that the last assumption has also been used in particular for the Wigner distribution analysis [7], [13].

The modification $x^{[L]}(nT)$ of original observations $x(nT)$, provided the assumption about the smallness of the noise, results in the following model with the additive noise:

$$x^{[L]}(nT) = s^{[L]}(nT) + \epsilon_L(nT). \tag{4}$$

The following statement is crucial for the analysis which follows:

Let the noise $\epsilon(nT)$ in (3) be complex-valued white such that $E(\epsilon(nT)\epsilon^*(nT)) = \sigma^2$, then provided $\sigma/A \ll 1$ the noise $\epsilon_L(nT)$ in (4) is also complex-valued white with the variance

$$E(\epsilon_L(nT)\epsilon_L^*(nT)) = (L^2+1)\sigma^2/2. \tag{5}$$

Proof. The signal $s(nT) = A\exp(j\phi(nT))$ can be represented by a vector having the amplitude A and phase $\phi(nT)$. In $x(nT) = s(nT) + \epsilon(nT)$ a small vector with coordinates $(Re\{\epsilon(nT)\}, Im\{\epsilon(nT)\})$, representing the noise, is added to the signal-vector. Thus, we get a resulting phase $\phi(nT) + \Delta_\epsilon\phi(nT)$ and resulting amplitude $A + \Delta_\epsilon A$. For a moment let us assume a rotated coordinate system, with rotation angle $\phi(nT)$. The noise coordinates in this system are $\epsilon_r(nT)$ and $\epsilon_t(nT)$, which are radial and tangential components of the noise respectively. The radial component $\epsilon_r(nT)$ is collinear with $A\exp(j\phi(nT))$, while the tangential components $\epsilon_t(nT)$ is normal to vector $A\exp(j\phi(nT))$. Now, let us form $x^{[L]}(nT)$. According to the definition (2) the amplitude $A+\Delta_\epsilon A$ is kept unchanged, while the phase $\phi(nT) + \Delta_\epsilon\phi(nT)$ is multiplied by L. The new noise has the radial component whose amplitude is unchanged and equals $|\epsilon_r(nT)|$. As the noise is small, $|\Delta_\epsilon A/A| \ll 1$, the new tangential component has the amplitude $L\,|\epsilon_t(nT)|$. Therefore, one noise component is unchanged while the other is multiplied by L. Having in mind that the rotation does not change the noise statistical characteristics in this case we get (5). □

We wish to mention once more that (5) holds for a small noise. If it is not a case, then the variance is always less than given by (5). An exact expression may be derived for these cases, as well, but the analysis which follows is constrained with the small noise assumption only. If $L = 1$ the variances of the transformed and the original noise are equal to each other. For the case $0 < L < 1$ which has not been considered previously, the variance of the transformed noise is less then the variance of the original noise. For example, $L = 1/2$ reduces significantly the noise variance with respect to $L = 1$ while $L = 1/4$ or $L = 1/8$ do not lead to further significant improvement because of the summand 1.

3. INSTANTANEOUS FREQUENCY ESTIMATION

Consider the problem of the signals' instantaneous frequency (IF)

$$\omega(t) = \phi'(t) \tag{6}$$

estimation, from discrete-time observations (3). We will assume that $\omega(t)$ is a differentiable function with bounded derivatives $\left|\omega^{(r)}(t)\right| = \left|\phi^{(r+1)}(t)\right| \leq M_r(t)$, $r \geq 1$.

If the signal is not noisy then, using the Taylor's expansion of $\phi(t + n\frac{T}{L}) - \phi(t - n\frac{T}{L})$ around t, its S-distribution is of the form

$$SD_L(t,\omega) = A^2 \sum_{n=-\infty}^{\infty} w_h(nT) e^{j2\phi'(t)nT} e^{j\Delta\phi(t,n\frac{T}{L})} e^{-j2\omega nT},$$

where

$$\Delta\phi(t, n\frac{T}{L}) = 2L \sum_{s=1}^{\infty} \frac{\phi^{(2s+1)}(t)}{(2s+1)!} (n\frac{T}{L})^{(2s+1)}. \tag{7}$$

If $\phi^{(2s+1)}(t) = 0$ for all s or $L \to \infty$, then the S-distribution would have a maximum at $\omega = \phi'(t)$. Therefore, the S-distribution based IF estimator may be defined as:

$$\hat{\omega}(t) = \arg\left[\max_{\omega\in Q_\omega} SD_L(t,\omega)\right] \tag{8}$$

with $Q_\omega = \{\omega : 0 \leq |\omega| < \pi/(2T)\}$ being the basic interval along the frequency axis. As a measure of the estimate quality, at a given instant t, let us define the estimation error as:

$$\Delta\hat{\omega}(t) = \omega(t) - \hat{\omega}(t). \tag{9}$$

Proposition:

Let $\hat{\omega}(t)$ be a solution of (8), and $h \to 0$, $T \to 0$, $h/T \to \infty$. Then the variance and bias of the IF estimate are given in the form

$$var(\Delta\omega(t)) = \frac{(L^2+1)\sigma^2}{A^2}(1+\frac{(L^2+1)\sigma^2}{4A^2})\frac{E_h}{F_h^2} \tag{10}$$

$$E\left\{\Delta\omega(t)\right\} = \frac{1}{2F_h}\sum_{s=1}^{\infty}\frac{B_h(s)}{L^{2s}}\phi^{(2s+1)}(t), \tag{11}$$

where

$$\begin{aligned} F_h &= \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^2 \to h^2\int_{-1/2}^{1/2} w(t)t^2dt,\\ E_h &= \sum_{n=-\infty}^{\infty} w_h^2(nT)(nT)^2 \to Th\int_{-1/2}^{1/2} w^2(t)t^2dt,\\ B_h(s) &= \tfrac{-2}{(2s+1)!}\sum_{n=-\infty}^{\infty} w_h(nT)(nT)^{2s+2} \to \\ &\tfrac{-2h^{2s+2}}{(2s+1)!}\cdot\int_{-1/2}^{1/2} w(t)t^{2s+2}dt\,. \end{aligned} \tag{12}$$

The limits hold for $T \to 0$, and $h/T \to \infty$.

The proof is a straigthtforward extension of the one for the Wigner distribution given in [10].

Let us analyze the mean squared accuracy of estimation. Using only the first significant term in the bias, the mean squared error (MSE) can be presented in the following form:

$$E\left\{(\Delta\omega(t))^2\right\} = \frac{(L^2+1)\sigma^2}{A^2}\frac{E_h}{F_h^2} + \left[\frac{B_h(1)\phi^{(3)}(t)}{2F_hL^2}\right]^2 \tag{13}$$

provided that $\frac{(L^2+1)\sigma^2}{4A^2} \ll 1$. Note also that the recursive S-distribution realization, described in [17], [20], would significantly reduce the influence of this noise component, anyway. For the rectangular window ($E_h = Th/12$, $F_h = h^2/12$, $B_h(1) = -h^4/240$), we get:

$$E\left\{(\Delta\omega(t))^2\right\} = \frac{3\sigma^2(L^2+1)}{A^2}\frac{T}{h^3} + \left[\frac{\phi^{(3)}(t)}{40L^2}h^2\right]^2 \tag{14}$$

where h is a width of the window ($w(t) = 0$ for $|t| > h/2$) and T is the basic sampling interval.

It can be seen from (14) that the MSE has a minimum with respect to L. The optimal value of L is given by the formula

$$L_{opt}(t) = \left[\frac{A^2h^7(\phi^{(3)}(t))^2}{2400\sigma^2 T}\right]^{1/6}. \tag{15}$$

The dependence of the optimal S-distribution order L_{opt} on the parameters A, σ ,T and the derivative $\phi^{(3)}(t)$ is quite clear. In particular, the time-instants with a large $\left|\phi^{(3)}(t)\right|$ require higher distribution orders, while for small $\left|\phi^{(3)}(t)\right|$, the distribution order L should be also small. This relation will be discussed in detail in one example, as well.

The optimal window length, minimizing the MSE (14) provided given L, is as follows

$$h_{opt}(t) = \left[\frac{3600\sigma^2(L^2+1)T}{A^2(\phi^{(3)}(t))^2}\right]^{1/7}. \tag{16}$$

It is obvious that calculation of (15) or (16) is not possible in practice, since its requires, besides A and σ^2, the knowledge of the IF second derivative $\phi^{(3)}(t)$. It is a definitely unavailable value because the IF itself has to be estimated.

Simultaneous unrestricted minimization of the MSE with respect to L and h gives a trivial result: the MSE approaches zero as $h \to \infty$ and $h/L \to 0$. This result has no practical interest. However the minimization of the MSE with respect to a finite set of acceptable values of h and L gives optimal pairs of (h, L) which, as the simulation confirms, are able significantly improve the accuracy.

4. NUMERICAL IMPLEMENTATION

The S-distribution (1), discretized over the frequency, is implemented as:

$$SD_L(k,l) = \sum_{n=-N/2}^{N/2-1} w_h(nT)x^{[L]}(lT+n\tfrac{T}{L})x^{[L]*}(lT-n\tfrac{T}{L})e^{-j2\frac{2\pi}{N}kn} \tag{17}$$

where $N = h/T$ is a number of samples determined by the window length h and sampling interval T given as $T = \pi/(2\omega_m)$, where ω_m is the maximal signal's frequency. The IF is estimated as:

$$\hat{\omega}_h(lT) = \arg\left[\max_k SD_L(k,l)\right]\frac{\pi}{NT} \tag{18}$$

for $0 \leq k \leq N/2-1$, for signals with nonnegative-only frequencies.

Let us consider influence of a quantization error on the accuracy of the IF estimation, caused by the discretization of $SD_L(k,l)$ in (17) along the frequency axis. Note also that the quantization error may also considered as a parameter closely related to the distribution concentration (frequency resolution), which is very important for time-frequency distributions (especially in the case of multicomponent signals). For the quantization noise error, in the probability analyses, usually an uniform probability density is assumed. In (18) this probability density is uniformly distributed over the segment $(-\pi/2h, \pi/2h]$, since $\pi/(NT) = \pi/h$. Its variance is $\sigma_q^2 = \frac{1}{12}(\frac{\pi}{h})^2$, producing the resulting MSE

$$\begin{aligned} E\{(\Delta\omega(t))^2\} &= \tfrac{3\sigma^2(L^2+1)}{A^2}\tfrac{T}{h^3} + \left[\tfrac{\phi^{(3)}(t)}{40L^2}h^2\right]^2 + \tfrac{1}{12}\tfrac{\pi^2}{h^2} = \\ &\left[\tfrac{3\sigma^2}{A^2}\tfrac{(L^2+1)}{N} + \tfrac{\pi^2}{12}\right]\tfrac{1}{h^2} + \left[\tfrac{\phi^{(3)}(t)}{40L^2}h^2\right]^2. \end{aligned} \tag{19}$$

For a large signal-to-noise ratio and any reasonable number of samples N and distribution order L we have $\frac{3\sigma^2}{A^2}\frac{(L^2+1)}{N} < \frac{\pi^2}{12}$ and the estimation variance is dominated by the quantization error.

As it is well known the quantization effects of the FFT can be reduced by an appropriate zero-padding in the time domain, which means an interpolation along the frequency axis. Provided that this interpolation of the S-distribution is done up to the widest considered window length, the quantization error will be reduced and kept to a constant value.

In the next section we will first consider the case when the interpolation is not done. It is simpler for analysis and more common in the time-frequency distribution realizations. An algorithm for the optimal window length determination will be derived for this case, and then extended to the cases with high interpolation rate (when quantization error may even be neglected).

5. ALGORITHM FOR ADAPTIVE ORDER AND WINDOW DETERMINATION

A. Basic Idea of the Window Length Optimization

The basic idea follows from the accuracy analysis, given in the Proposition. Namely, at least for the asymptotic case of small noise and bias, the estimation error can be represented as a sum of the deterministic (bias) and random component, with the variance given in the Proposition. For the estimated IF $\hat{\omega}_h(t)$ we may write the following relation:

$$|\omega(t) - (\hat{\omega}_h(t) - bias(t,h))| \leq \kappa\sigma(h), \tag{20}$$

where the inequality holds with the probability $P(\kappa)$ depending on κ.

In the case of dominant quantization noise error, which will be considered now, the probability density of $\hat{\omega}_h(t)$ around $bias(t,h)$ is uniform. For this distribution value $\kappa = \sqrt{3}$ guaranties (20) with probability 1 (note that the same value of κ would guaranty only the probability of 0.93, if the distribution were Gaussian).

Using the expressions for the variance and bias

$$\begin{gathered} \sigma^2(h) = \left[\frac{3\sigma^2}{A^2}\frac{(L^2+1)}{N} + \frac{\pi^2}{12}\right]\frac{1}{h^2} \cong \frac{\pi^2}{12}\frac{1}{h^2} \\ bias(t,h) = \frac{\phi^{(3)}(t)}{40L^2}h^2 \end{gathered} \tag{21}$$

the MSE is given by

$$E\left\{(\Delta\omega(t))^2\right\} = \frac{\pi^2}{12}\frac{1}{h^2} + \left[\frac{\phi^{(3)}(t)}{40L^2}h^2\right]^2$$

Concerning the distribution order L, we may conclude that its highest value (as far as $\frac{36\sigma^2}{A^2}\frac{(L^2+1)}{N\pi^2} \ll 1$ is satisfied) should be used.

The MSE minimization with respect to h gives

$$\sigma(h_{opt}) = \sqrt{2}bias(t, h_{opt}), \tag{22}$$

where h_{opt} denotes the optimal window length,

$$h_{opt} = (200\pi^2 L^4/(3(\phi^{(3)}(t))^2))^{1/6}.$$

Let us introduce a discrete set H of window length values, $h \in H$,

$$H = \{h_s \mid h_s = ah_{s-1},\ s = 1, 2, 3, ..., J,\ a > 1\}, \tag{23}$$

The following arguments can be given in favor of such a set:

(a) Discrete scheme for window lengths is necessary for the efficient numerical realizations. The realizations of time-frequency distributions of form (17) are almost absolutely based on the FFT algorithms application (excluding only few recursive approaches, [1], [12], [11], [18]). The most common are radix-2 or radix-3 FFT algorithms which correspond to $a = 2$ or $a = 3$, when set H gives dyadic ($h_s = h_0 2^s$) or triadic ($h_s = h_0 3^s$) window length schemes. In the realizations the smallest window length h_0 should correspond to a small number N_0 of signal samples within it. For example, for radix-2 FFT algorithms $N_0 = 4$ or $N_0 = 8$ with $N_s = 2N_{s-1}$, $s = 1, 2, ..., J$.

(b) A search of the optimal window length over H is a simplified optimization, because the set (23) consist of a relatively small number of elements. However, the discrete set of h inevitably leads to suboptimal window length values due to the discretization of h effects (quantization noise error and effects due quantization of h, although with similar names are two completely different notions). It is important to note that this effect, due to the discrete nature of $h \in H$, would also exist even if we knew in advance all of the parameters required for the optimal window length calculation, and decided to use radix-2 FFT algorithms in the realization. Thus, the discretization of h effect is present in any case. It always results in a worse value of the MSE, but that is the price of efficient calculation schemes, available in this case. Fortunately this loss of the accuracy is not significant in many cases, because the MSE has a stationary point for the optimal window length $h = h_{opt}$ (and the MSE varies very slowly for the window length values close to $h = h_{opt}$).

Now we are going to derive an algorithm for the determination of the optimal window size h_{opt}, without knowing the bias, using the IF estimates (18) and the formula for the IF estimate's variance only.

It is based on the following statement:

Let H be a set of dyadic window length values, i.e., $a = 2$ in (23), and $\kappa = \sqrt{3}$. Define the upper and lower bounds of confidence intervals $D_s = [L_s, U_s]$ as

$$\begin{aligned} L_s &= \hat{\omega}_{h_s}(t) - (\kappa + \Delta\kappa)\ \sigma(h_s), \\ U_s &= \hat{\omega}_{h_s}(t) + (\kappa + \Delta\kappa)\ \sigma(h_s), \end{aligned} \tag{24}$$

where $\hat{\omega}_{h_s}(t)$ is an estimate of the IF, with $h = h_s$ and $\sigma(h_s)$ given by (21).

Let the window length h_{s+} be determined as a length corresponding to the largest s $(s = 2, ..., J)$ when

$$L_s(t) \leq U_{s-1}(t) \quad \text{or} \quad U_s(t) \geq L_{s-1}(t) \tag{25}$$

is still satisfied, i.e., when $D_s \cap D_{s-1} \neq \emptyset$.

Let $\Delta\kappa = 1/2$, and $\sqrt{2}h_{opt} \in H$, then with probability 1

$$h_{s+} = \sqrt{2}h_{opt}, \tag{26}$$

where h_{s+} is determined by the algorithm (24)-(25). For such a determined h_{s+} holds $2\sigma(h_{s+}) = bias(t, h_{s+})$.

Proof: Let us denote by b the unknown bias

$$bias(t, h_{opt}) = b,$$

when the window length has its optimal value $h = h_{opt}$. Without a loss of generality we will assume that $b > 0$. The window lengths belonging to H, having in mind that we assumed $\sqrt{2}h_{opt} \in H$, can be represented as follows

$$h(p) = h_{opt}a^{p+1/2}, \quad p = ..., -2, -1, 0, 1, 2, ...$$

where $p = 0$ corresponds to the window length $\sqrt{2}h_{opt}$, we are looking for ($a = 2$). Note that the optimal window length h_{opt} is a geometric mean of lengths $h(p)$ for $p = -1$ and $p = 0$, $h_{opt} = \sqrt{h(0)h(-1)} = h(0)/\sqrt{2}$. The reason why we decided to look for $h(0) = \sqrt{2}h_{opt}$ and not for h_{opt} will be clarified later. Note also that we use two indexes for the window lengths, one s (i.e., h_s) which denotes the indexing which starts from the narrowest window length, and the other p (used is form of an argument i.e., $h(p)$) where the indexing starts from the $\sqrt{2}h_{opt}$ window length (when $p = 0$), with narrower windows having negative p and wider window lengths having positive p.

The bias and variance values for any $h(p)$, according to (21), may be rewritten as:

$$bias(t, h(p)) = a^{2p+1}b, \qquad \sigma(h(p)) = \sqrt{2}a^{-p-1/2}b. \tag{27}$$

From (27) we can conclude that for $p \ll 0$ and $a = 2$ the bias is much smaller as compared to the variance since $a^{2p} \ll \sqrt{2}a^{-p}$, thus the estimate $\hat{\omega}_h(t)$ is spread around the exact value $\omega(t)$ with a small bias ($bias(t, h(p)) \to 0$ as $h(p) \to 0$) and large variance ($\sigma(h(p)) \to \infty$ as $h(p) \to \infty$). A confidence interval of the estimate $\hat{\omega}_{h(p)}(t)$, for a given $h(p)$, is defined by

$$D_p = [\hat{\omega}_{h(p)}(t) - \kappa\sigma(h(p)), \hat{\omega}_{h(p)}(t) + \kappa\sigma(h(p))].$$

For $\kappa = \sqrt{3}$ we have that $\omega(t) \in D_p$ with probability 1, when $bias(t, h(p)) = 0$.

Now consider a confidence interval, modified in order to take into account the biased estimate $\hat{\omega}_{h(p)}(t)$ in the following way:

$$\tilde{D}_p = [\hat{\omega}_{h(p)}(t) - (\kappa + \Delta\kappa)\sigma(h(p)), \hat{\omega}_{h(p)}(t) + (\kappa + \Delta\kappa)\sigma(h(p))], \tag{28}$$

where $\Delta\kappa > 0$ is to be found.

It is obvious that $\omega(t) \in \tilde{D}_p$ for $p \ll 0$ because in this case the bias is small and the segment $\tilde{D}_p$ is wider than D_p as $\Delta\kappa > 0$. Note also that all of the confidence intervals $\tilde{D}_p$, with p such that the bias is very small, have the true IF value $\omega(t)$ in common, or to be precise have at least region $[\omega(t) - \Delta\kappa\,\sigma(h(p)), \omega(t) + \Delta\kappa\,\sigma(h(p))]$ in common, i.e., $[\omega(t) - \Delta\kappa\,\sigma(h(p)), \omega(t) + \Delta\kappa\,\sigma(h(p))] \subseteq \tilde{D}_p \cap \tilde{D}_{p-1}$ for any $p \ll 0$.

For $p \gg 0$ the variance is small but the bias is large as $a^{2p+1} \gg \sqrt{2}a^{-p-1/2}$. It is clear that always exist such a large p that $\tilde{D}_p \cap \tilde{D}_{p-1} = \emptyset$ for any given $\Delta\kappa$.

The idea behind of the algorithm (24)-(25) is that $\Delta\kappa$ in $\tilde{D}_p$ can be found in such a way that the largest p for which a sequence of the pairs of the confidence intervals $\tilde{D}_{p-1}$ and $\tilde{D}_p$ has a point in common is $p = 0$. Such value of $\Delta\kappa$ exists because the bias and the variance are monotonic increasing and decreasing functions of h respectively. As soon as this value of $\Delta\kappa$ is found an intersection of the confidence intervals $\tilde{D}_{p-1}$ and $\tilde{D}_p$ works as an indicator of the event $p = 0$, i.e., the event when $h_s = \sqrt{2}h_{opt}$ is found. The algorithm given in the form (24)-(25) tests the intersection of the confidence intervals, where (25) is a condition that two sequential intervals $\tilde{D}_{s-1}$ and $\tilde{D}_s$ is the last pair of the confidence intervals having a point in common (note again that indexes s and p only indicate if we assume the first confidence interval or the confidence interval when $h = \sqrt{2}h_{opt}$ as the one having index 0).

Now let us find this crucial value of $\Delta\kappa$. According to the above analysis, only three values of $p = -1$, 0, and 1 along with the corresponding intervals $\tilde{D}_{-1}$, $\tilde{D}_0$, and $\tilde{D}_1$ should be considered, in this case. The intervals $\tilde{D}_{-1}$ and $\tilde{D}_0$ **should have** and the intervals $\tilde{D}_0$ and $\tilde{D}_1$ **should not have** at least a point in common. Since $\hat{\omega}_{h(p)}(t)$ is a random (uniformly distributed) variable, then the confidence interval bounds are also random and uniformly distributed. Thus, we must consider the worst possible cases for the corresponding bounds. These worst case conditions, for $b > 0$, are given by:

$$\begin{aligned} \min\{U_{-1}\} &\geq \max\{L_0\}, \\ \max\{U_0\} &< \min\{L_1\}, \end{aligned} \tag{29}$$

Let us, for example, consider U_{-1}. The estimated IF $\hat{\omega}_{h(-1)}(t)$ may assume values within the interval $\hat{\omega}_{h(p)}(t) \in [\omega(t) + bias(h(-1)) - \kappa\sigma(h(-1)), \omega(t) + bias(h(-1)) + \kappa\sigma(h(-1))]$. Consequently, the upper confidence interval bound U_{-1}, according to (28), may take values from the interval $U_{-1} \in [\omega(t) + bias(h(-1)) + \Delta\kappa\sigma(h(-1)), \omega(t) + bias(h(-1)) + (2\kappa + \Delta\kappa)\sigma(h(-1))]$. The minimal possible value of U_{-1} is $\min\{U_{-1}\} = \omega(t) + bias(h(-1)) + \Delta\kappa\sigma(h(-1))$. In the same way we may get other bound limits, required by (29), what results in

$$\begin{aligned} bias(h(-1)) + \Delta\kappa\sigma(h(-1)) &\geq bias(h(0)) - \Delta\kappa\sigma(h(0)), \\ bias(h(0)) + (2\kappa+\Delta\kappa)\sigma(h(0)) &< bias(h(1)) - (2\kappa+\Delta\kappa)\sigma(h(1)) \end{aligned} \tag{30}$$

or

$$\begin{aligned} b(a^{-1} + \Delta\kappa\sqrt{2}a^{1/2}) &\geq b(a - \Delta\kappa\sqrt{2}a^{-1/2}), \\ b(a + (2\kappa + \Delta\kappa)\sqrt{2}a^{-1/2}) &< b(a^3 - (2\kappa + \Delta\kappa)\sqrt{2}a^{-3/2}). \end{aligned} \tag{31}$$

It can be verified that $\Delta\kappa = 1/2$ is smallest $\Delta\kappa > 0$ satisfying first inequality in (31), for $a = 2$. This value of $\Delta\kappa$, with $\kappa = \sqrt{3}$, closely satisfies second inequality in (31), which requires $\kappa + \Delta\kappa < 1.75$. Therefore $\kappa + \Delta\kappa = \sqrt{3} + 1/2 \simeq 2.25$, with $a = 2$, satisfies both inequalities in (31).

With (31) being satisfied we have that $\tilde{D}_p \cap \tilde{D}_{p-1} \neq \emptyset$, for $p \leq 0$, with probability 1, and $\tilde{D}_p \cap \tilde{D}_{p-1} = \emptyset$, for $p \geq 1$, with probability 1. This completes the proof of the statement. □

Note that the inequalities in (31) are true for arbitrary a,κ and $\Delta\kappa$ and they can be used in particular for a choice of $\Delta\kappa$ for $a \neq 2$. It is easy to check that if, in the analysis, we used more natural discrete scheme $h_p = h_{opt}a^p$, then we could not satisfy both inequalities in (31) with $a = 2$ and $\kappa = \sqrt{3}$. We should then use larger a (for example, radix-3 FFT algorithms with $a = 3$), what would significantly increase the MSE due to the significant discretization of h effects.

We wish to emphasize that results of the statement are derived provided that $\sqrt{2}h_{opt}$ assumes one of the dyadic values from H and the bias and variance are given by the asymptotic formulas (21). In applications, due to discrete nature of h we will never have that $\sqrt{2}h_{opt} \in H$, what will result in the already described discretization of

h effects, producing slightly suboptimal MSE values. This means that the values κ and $\Delta\kappa$, given in the statement, should be interpreted as a reasonable approximate values of these parameters which can be used in the algorithm (24)-(25), at least as far as the formulas (21) for the bias and variance holds.

B. *Algorithm*

According to the statement and the analysis in the previous subsection, we may define the following algorithm for the adaptive IF estimation:

1. Assume a set H is given by (23).

2. For a given instant t, start the S-distribution calculation from the smallest toward the wider window lengths $h_s \in H$.

3. Estimate IF using

$$\hat{\omega}_{h_s}(t) = \arg\left[\max_{\omega \in Q_\omega} SD_L(t,\omega)\right] \tag{32}$$

4. With $\sigma(h_s) = \sqrt{\left[\frac{3\sigma^2}{A^2}\frac{(L^2+1)}{N} + \frac{\pi^2}{3}\right]\frac{1}{h_s^2}} \cong \frac{\pi}{h_s\sqrt{3}}$ and $\hat{\omega}_{h_s}(t)$ define the segments

$$\tilde{D}_s(t) = [\hat{\omega}_{h_s}(t) - (\kappa+\Delta\kappa)\sigma(h_s),\ \hat{\omega}_{h_s}(t) + (\kappa+\Delta\kappa)\sigma(h_s)], \tag{33}$$

with their upper and lower bounds built as follows

$$U_s(t) = \hat{\omega}_{h_s}(t) + (\kappa+\Delta\kappa)\sigma(h_s)$$
$$L_s(t) = \hat{\omega}_{h_s}(t) - (\kappa+\Delta\kappa)\sigma(h_s)$$

with $\kappa + \Delta\kappa \approx 2.25$.

5. The adaptive window length h_{s^+} is determined as the length corresponding to the largest s $(s = 1, 2, ..., J)$ when

$$L_s(t) \leq U_{s-1}(t) \quad \text{or} \quad U_s(t) \geq L_{s-1}(t) \tag{34}$$

is still satisfied..

Then, this s^+ is the largest of those s for which the segments D_{s-1} and D_s, $s \leq J$, have a point in common. The adaptive window length is chosen as

$$\hat{h}(t) = h_{s^+}(t) \tag{35}$$

and $\hat{\omega}_{\hat{h}(t)}(t)$ is the adaptive IF estimator with data driven window for a given instant t. We have not used the optimal window length correction $\hat{h}_{opt}(t) = \hat{h}(t)/\sqrt{2}$, which follows from (26), since this correction falls within the discretization of h error and, more importantly, would require one non radix-2 FFT calculation.

6. Take next t.

In contrast to [6], in our algorithm we apply the sliding pair-wise intersections (34) of the pairs of segments D_s and D_{s-1} for $s=2,3,...,J$. The simulation shows that the results for the IF estimation on the base of this sliding pair-wise intersection are more accurate as compared with the algorithm using the intersection of all of the segments D_s starting from $s = 1$ as it is done in [6]. The convergence analysis given in [6] using intersections of all previous intervals is based on a large number of confidence intervals D_s, while in this paper we have a relatively small number of possible window lengths, and therefore small number of considered confidence intervals. That may be one of the reasons why our two segments approach, produced better results here.

Further we will consider how a compromise, corresponding to the MSE minimization, can be achieved for the IF estimation with the S-distribution implemented with an appropriate interpolation mentioned above.

C. Estimation With Interpolation

In the case when the quantization error may be neglected, i.e. an appropriate interpolation is done, we have the variance and the bias strongly depending on both: the window length h and the distribution order L. Consider the cases of optimization with respect to h provided a fixed order L, as well as the simultaneous optimization with respect to both h and L.

a) Let the order L of a $S-$distribution be fixed and the bias and variance of estimation be determined by (14). It can be seen that for the optimal window size h_{opt} we obtain, instead of (22), the following relation between the bias and the variance:

$$\sigma(h_{opt}) = \sqrt{4/3}bias(t, h_{opt}). \tag{36}$$

The significant difference from the previous case is that the inequality $|\omega(t) - (\hat{\omega}_h(t) - bias(t,h))| \leq \kappa\sigma(h)$ holds with the probability $P(\kappa)$ depending on κ and there is no κ such that $P(\kappa) = 1$, as it was the case earlier.

The accurate analysis in this case is quite complex because the intersections of the intervals $\tilde{D}_{p-1}$ and $\tilde{D}_p$ are random events having a place with probabilities depending on $h(p)$.

Nevertheless, we will present the analysis which may help to determine the parameters a, κ and $\Delta\kappa$ of the defined algorithm in this case.

We will follow the same reasoning as in the subsection 5. A, with the assumption that for a certain κ we may assume that

$$|\omega(t) - (\hat{\omega}_h(t) - bias(t,h))| \leq \kappa\sigma(h) \tag{37}$$

holds with a probability close to 1. Again assume that the window length is dyadic $h_s \in H$ and $h(p) = \sqrt{2}h_{opt}a^p \in H$ with $a = 2$. The bias and variance as functions of unknown parameter $b = bias(t, h_{opt})$ are

$$bias(t, h(p)) = ba^{2p+1}, \qquad \sigma(h(p)) = \frac{2}{\sqrt{3}}ba^{-3(p+1/2)/2}$$

then the conditions that $\tilde{D}_{-1} \cap \tilde{D}_0 \neq \emptyset$ and $\tilde{D}_0 \cap \tilde{D}_1 = \emptyset$, with $b > 0$, are of the form (30) and produce the inequalities similar to (31):

$$\begin{gathered} a^{-1} + \tfrac{2}{\sqrt{3}}\Delta\kappa a^{3/4} \geq a - \Delta\kappa\tfrac{2}{\sqrt{3}}a^{-3/4} \\ a + (2\kappa + \Delta\kappa)\tfrac{2}{\sqrt{3}}a^{-3/4} < a^3 - (2\kappa + \Delta\kappa)\tfrac{2}{\sqrt{3}}a^{-9/4}. \end{gathered} \tag{38}$$

For $a = 2$ these inequalities give as a smallest positive value $\Delta\kappa = 0.57$ and largest $\kappa = 2.942$. Thus, for the confidence interval we obtain finally

$$\kappa + \Delta\kappa \approx 3.5. \tag{39}$$

Note that this values would result in $\kappa \approx 3$ i.e., the inequality (37) holds with the probability $P(\kappa) = 0.997$. Therefore our assumption that we are working with probability $P(\kappa)$ close to 1 is almost absolutely true.

We wish to mention once more that this analysis has more a qualitative meaning than a quantitative one. Nevertheless, in the numerical realizations (experimenting with various $\kappa + \Delta\kappa$), we found that the value $\kappa + \Delta\kappa \approx 3.5$, given by (39), produces the best results in all examples we considered.

With these hints and parameter values we can now use the algorithm (32)-(35) as described in subsection 5.B for the adaptive window length determination.

In this case the standard deviation $\sigma(h_s)$ could not be neglected, thus the estimation of signal and noise parameters $|A|$ and σ^2 can be done using $\left|\hat{A}\right|^2 + \hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}|y(nT)|^2$, where the sum is calculated over all N observations and N is assumed to be large, as well as T is small. The variance is estimated by $\hat{\sigma} = \{\text{median}(|y(nT) - y((n-1)T)| : n = 2, .., N)\}/1.349$. where T is sufficiently small.

b) Consider the simultaneous optimization with respect to both parameters in question L and h. Let us start with an introduction of two sets

$$\begin{aligned} H &= \{h_s \mid h_1 < h_2 < h_3 < ... < h_J\}, \\ \Lambda &= \{L_r \mid L_1 < L_2 < ... < L_K\}, \end{aligned} \tag{40}$$

where $h \in H$ is a set of values of the window lengths h, and Λ is a set of distribution orders, denoted by L_r.

Consider a direct product of H and Λ as a set $H \times \Lambda = \{(h_s, L_r) \mid s = 1, 2, ..., J, r = 1, 2, .., K\}$ of all possible pairs (h_s, L_r). Now let us reorder the elements of $H \times \Lambda$ in such a way that we get a new set Φ whose elements $g_q = (L_r^2 + 1)/h_s^3$, $q = 1, 2, 3, ..., JK$ form an decreasing sequence:

$$\Phi = \{g_q = (L_r^2 + 1)/h_s^3 \mid g_1 \geq g_2 \geq ... \geq g_{JK}\}. \tag{41}$$

The elements g_q forms a decreasing sequence of the estimation variance

$$\sigma(g_q) = \sqrt{\frac{3\sigma^2}{A^2} g_q T}. \tag{42}$$

The confidence intervals corresponding to the sequence g_q are as follows

$$D_q = [\hat{\omega}(t) - (\kappa + \Delta k)\sigma(g_q),\ \hat{\omega}(t) + (\kappa + \Delta k)\sigma(g_q)] \qquad (43)$$

and the algorithm (32)-(35) can be applied in a straightforward manner. The only difference is that the set H is replaced by the set Φ and instead of the window size selection, we find q^+ which immediately determines a pair of the corresponding (h_{s^+}, L_{r^+}).

The set Λ can be determined by any reasonable way. In simulation we use a dyadic set with $L_r = 2^{r-2}$, $r = 1, 2, 3, 4$, as well as $\kappa + \Delta\kappa$ given by (39). Note that distribution with $L_r = 1$ ($r = 2$) is the Wigner distribution, distributions with $L_r = 2, 4$ are higher order distributions, while $L = 1/2$ would be a "lower order" distribution (notion "higher order" and "lower order" are used with respect to the Wigner distributions)

6. EXAMPLES

The discrete S-distribution is calculated using the standard FFT routines, after the signal is modified according to (2) and product $w_h(nT)x^{[L]}(lT + n\frac{T}{L})x^{[L]*}(lT - n\frac{T}{L})$ is calculated. Note that if we use $L < 1$ (for example $L = 1/2$ as we did in this paper) then we have to take care about the phase continuity of $x^{[1/2]}(nT)$ over the π borders. We have found such a function only in some older versions of the MATLAB, where the m-file "phase.m" does exactly what we needed.

The algorithm is tested on two examples. In all of them we assumed signal of the form $x(nT) = A\exp(j\phi(nT)) + \epsilon(nT)$, with a given IF $\omega(nT)$ and the phase $\phi(nT) = \sum_{i=0}^{n} \omega(nT)/T$.

Signal amplitude was $A = 1$ and $20\log(A/\sigma) = 15[dB]$, $(A/\sigma = 5.62)$. Considered time interval was $0 \leq nT \leq 1$.

Example 1: Signal with IF defined by

$$\omega(nT) = 10\pi \operatorname{asinh}(100(t - 0.5)) + 64\pi$$

Several S-distributions with constant orders ($L = 1, 2, 4$) and window lengths ($N = 32, 128$) are presented in Fig.1a)-f).

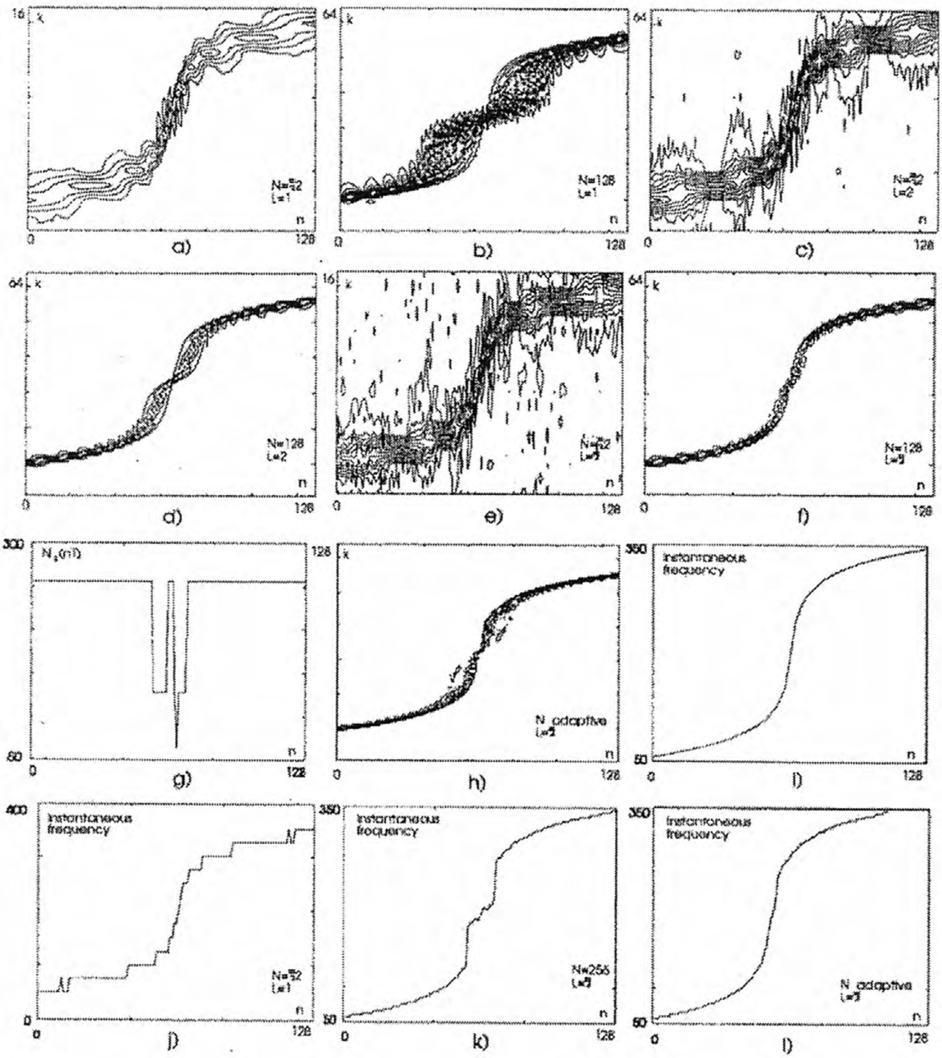

Fig. 1: Time-freqeuncy representation and IF estimation of a signal with non-linear IF. S-distribution with: a) L=1, N=32, b) L=1, N=128, c) L=2, N=32, d) L=2, N=128, e) L=4, N=32, f) L=4, N=128, g) Adaptive window length, h) S-distribution with adaptive window length, i) Signal's IF, j) IF estimation with L=1 and N=32, k) with L=4 and N=128, l) IF estimation with L=4 and adaprive window length.

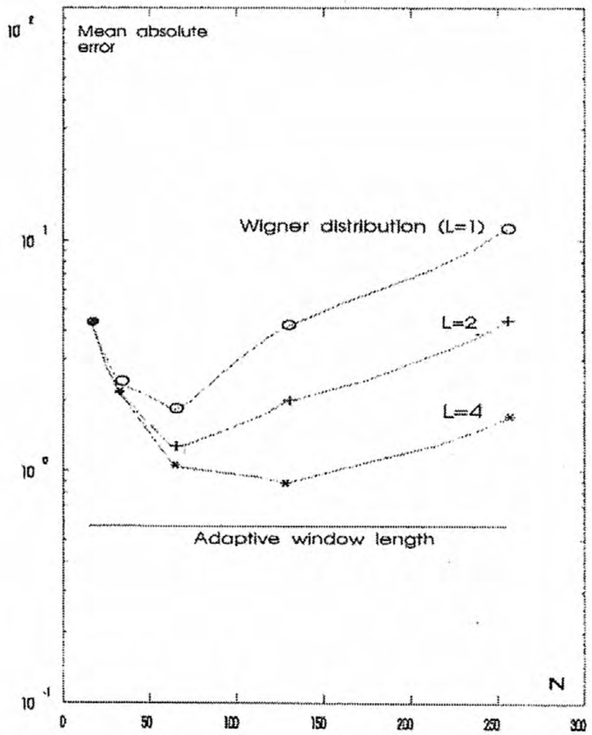

Fig. 2: Mean absolute error for various window lengths and S-distribution orders in the case of dominant quantization error.

Since, in this example, we have not done any additional interpolation, in order to find an adaptive distribution, then according to results in 5.A, we considered distributions with maximal order $L = 4$ and various window lengths h_s corresponding to the following number of signal samples within it: $N_s = 16, 32, 64, 128, 256$. The adaptive window lengths, determined by the algorithm (32)-(35) with $\kappa + \Delta\kappa = 2.25$, are shown in Fig.1g. We can see that when the IF variations are small then algorithm uses the widest window length in order to reduce the variance. Around the point $nT = 0.5$, where the bias is large, the

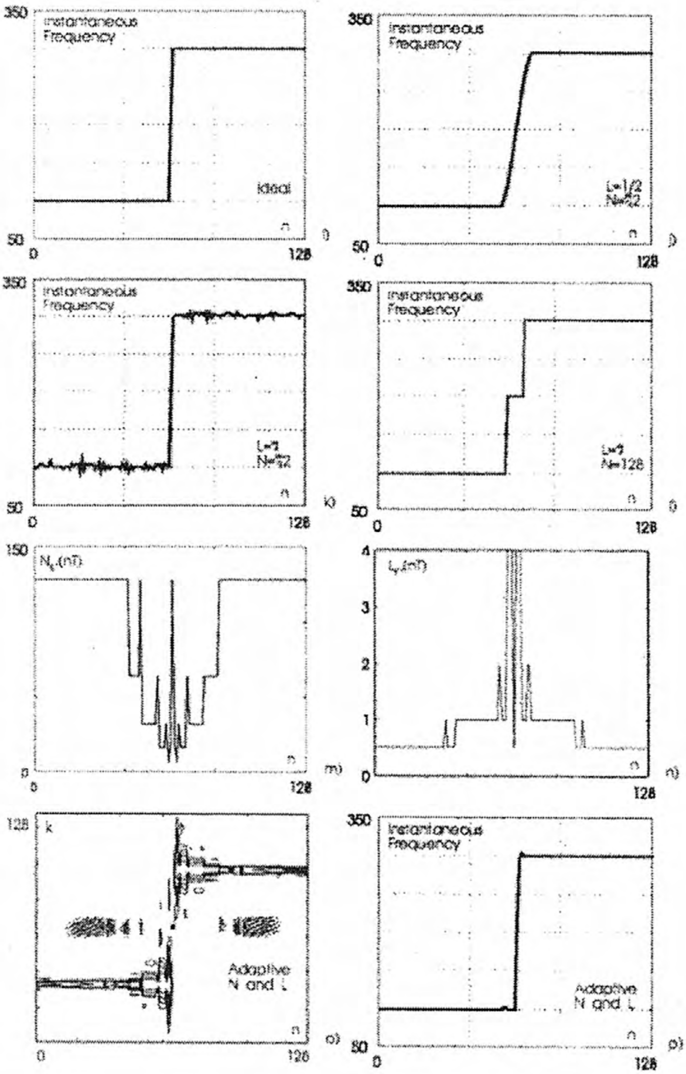

Fig. 3: Time-freqeuncy representation and IF estimation of a signal with step-wise IF: a) Signal's IF, b) IF estimation with L=1/2 and N=32, c) with L=4 and N=32, d) with L=4 and N=128, e) Adaptive window length, f) Adaptive distribution order, g) The S-distribution with adaptive window length and distribution order, h) Estimated IF using the S-distribution with adaptive order and window length.

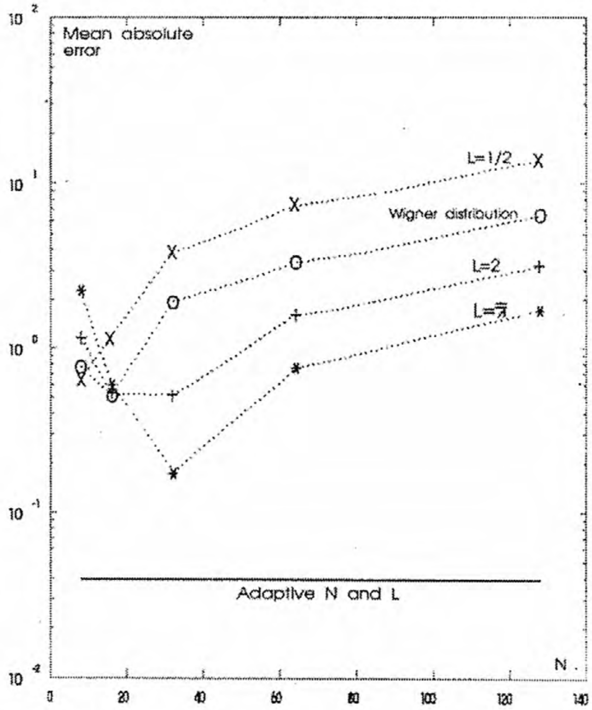

Fig. 4: Mean absolute error for various window lengths and S-distribution orders.

windows with smaller lengths are used. The S-distribution with adaptive window length is presented in Fig.1h. The IF, as well as its estimates with ($L = 1, N = 32$), ($L = 4, N = 256$), and estimate with adaptive window length using $L = 4$, are given in Fig.1i)-l), respectively. Absolute mean error, normalized to the minimal discretization step, for each considered distribution order and window length, is shown in Fig.2. This figure confirms that for each window length, S-distribution with $L = 4$ produces smallest error, as well as that the closest one to the distribution with adaptive window length (given by solid line) is distribution with $L = 4, N = 128$ presented in Fig.1f. Example 2: Signal with step-wise IF

$$\omega(nT) = 32\pi \mathrm{sign}(t - 0.5) + 64\pi$$

In this example we did additional interpolation for each window length, up to the widest $N = 128$, as well as choose such IF values, that the quantization error can be even neglected. The estimated IF using the S-distributions with some constant orders and window lengths are presented in Fig.3a)-d).

The adaptation is done with respect to both window length and distribution order, according to the algorithm (32)-(35) and hints in 5.D. The adaptive window length $N_{s^+}(nT)$ and distribution order $L_{r^+}(nT)$, as well the S-distribution with that parameters, using $\kappa + \Delta\kappa = 3.5$, are presented in Fig.3e)-g), along with the estimated IF, Fig.3h). As expected, the algorithm produced the smallest possible variance (with $L_{r^+}(nT) = 1/2$ and $N_{s^+}(nT) = 128$) in the regions where the instantaneous frequency estimator is not biased (i.e., IF is constant). Algorithm application resulted in small window lengths and high distribution orders in the region where the bias is large, around the point $nT = 0.5$. Absolute mean error, normalized to the minimal discretization step, is shown in Fig.4. It further illustrates our considerations about window lengths and distribution orders influence to the accuracy of the IF estimation. Here we will also discuss the optimal distribution order dependence on the window length. From Fig.4 we see that for the narrowest window length the smallest mean absolute error is obtained with the distribution having $L = 1/2$ order. Increasing the window length the best distribution order also increases. For a reasonably large window lengths (what is important from the distributions' resolution point of view) the best results are obtained for the highest distribution order $L = 4$. This is in complete agreement with (15).

7. CONCLUSION

The S-distribution with the data-driven and time-varying window length and order is presented, as an adaptive estimator of the IF. The choice of the window length and the distribution order is based on the intersection of the confidence intervals of the IF estimates. The developed algorithm uses only the formula for the asymptotic variance of the IF estimates. Simulation shows a quite valuable accuracy improvement of the adaptive algorithm.

References

[1] M. G. Amin, "A new approach to recursive Fourier transform", *Proc. IEEE*, vol.75, Nov. 1987, p.1537.

[2] B. Boashash: "Estimating and interpreting the instantaneous frequency of a signal-Part 1: Fundamentals", *Proc. of the IEEE*, vol-80, no.4, April 1992, pp.519-538.

[3] B.Boashash and P. O'Shea: "Polynomial Wigner-Ville distributions and their relationship to time-varying higher order spectra", *IEEE Trans. on SP*, vol-42, no.1, Jan. 1994, pp.216-220.

[4] L. Cohen and C. Lee: "Instantaneous bandwidth", in *Time-frequency signal analysis*, B. Boashash ed., Longman Cheshire, 1992.

[5] L. Cohen: "Distributions concentrated along the instantaneous frequency" *SPIE*, vol-1348, Adv. Signal Proc. Alg., Arch., and Implementations, 1990, pp.149-157.

[6] A. Goldenshluger, A.Nemirovski:"On spatial adaptive estimation of nonparametric regression", Res. report, 5/94, Technion, Israel, Nov.1995.

[7] V. Katkovnik: "Nonparametric estimation of instantaneous frequency", *IEEE Trans. on IT*, vol-43, no.1, Jan.1997, pp.183-189.

[8] V. Katkovnik: "Local polynomial approximation of the instantaneous frequency: Asymptotic accuracy", *Signal Processing,* vol-52, no.3, 1996, pp.343-356.

[9] V. Katkovnik: "Adaptive local polynomial periodogram for time-varying frequency estimation", in *Proc. IEEE-SP IS-TFTSA*, Paris June 1996, pp.329-332.

[10] V. Katkovnik, LJ. Stanković: "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length", submitted to the *IEEE Trans. on SP.*

[11] D. Petranović, S. Stanković, LJ. Stanković: "Special purpose hardware for time-frequency analysis", *Electronics Letters*, vol-33, no.6, March 1997, pp.464-466.

[12] K.J. Ray Liu: "Novel parallel architectures for Short-time Fourier transform", *IEEE Trans. on CAS*, vol-40, no.12, Dec. 1993, pp.786-789.

[13] P. Rao, F.J.Taylor:"Estimation of the instantaneous frequency using the discrete Wigner distribution", *Electronics Letters,* vol-26, 1990, pp.246-248.

[14] B. Ristic, B. Boashash:"Relationship between the polynomial and the higher order Wigner-Ville distribution", *IEEE SP Letters*, vol-2, no.12, Dec.1995, pp.227-229.

[15] LJ. Stanković, S.Stanković: "On the Wigner distribution of discrete-time noisy signals with application to the study of quantization effects", *IEEE Trans. on SP*, vol-42, no.7, July 1994, pp.1863-1867.

[16] LJ. Stanković: "A method for improved distribution concentration in the time-frequency analysis of multicomponent signals the L-Wigner distribution" *IEEE Trans. on SP*, vol-43, no.5, May1995.

[17] LJ. Stanković: "Highly concentrated time-frequency distribution: Pseudo quantum signal representation", *IEEE Trans. on SP*, vol-45, no.3, March 1997.

[18] LJ. Stanković: "A method for improved distribution concentration in the time-frequency signal analysis using the L-Wigner distribution" *IEEE Trans. on SP,* vol-43, no.5, May 1995.

[19] LJ. Stanković: "A multitime definition of the Wigner higher order distribution: L-Wigner distribution" *IEEE SP Letters*, vol.1, no.7, July 1994, pp. 106-109.

[20] LJ. Stanković: "S-class of time-frequency distributions", *IEE Proceedings: Vision, Image and Signal Processing,* vol.144, no.2, April 1997, pp.57-64.

[21] LJ. Stanković: "A time-frequency distribution concentrated along instantaneous frequency", *IEEE SP Letters*, vol-3, no.2, Feb. 1996.