

The Wigner Distribution of Noisy Signals with Adaptive Time–Frequency Varying Window

LJubiša Stanković, *Senior Member, IEEE*, and Vladimir Katkovnik, *Member, IEEE*

Abstract—Time–frequency representations using the Wigner distribution (WD) may be significantly obscured by the noise in observations. The analysis performed for the WD of discrete-time noisy signals shows that this time–frequency representation can be optimized by the appropriate choice of the window length. However, the practical value of this analysis is not significant because the optimization requires knowledge of the bias, which depends on the unknown derivatives of the WD. A simple adaptive algorithm for the efficient time–frequency representation of noisy signals is developed in this paper. The algorithm uses only the noisy estimate of the WD and the analytical formula for the variance of this estimate. The quality of this adaptive algorithm is close to the one that could be achieved by the algorithm with the optimal window length, provided that the WD derivatives were known in advance. The proposed algorithm is based on the idea that has been developed in our previous work for the instantaneous frequency (IF) estimation. Here, a direct addressing to the WD itself, rather than to the instantaneous frequency, resulted in a time and frequency varying window length and showed that the assumption of small noise and bias is no longer necessary. A simplified version of the algorithm, using only two different window lengths, is presented. It is shown that the procedure developed for the adaptive window length selection can be generalized for application on multicomponent signals with any distribution from the Cohen class. Simulations show that the developed algorithms are efficient, even for a very low value of the signal-to-noise ratio.

I. INTRODUCTION

A VARIETY of tools is used for the time–frequency analysis of nonstationary signals. Many of these tools have a form of energy distributions in the time–frequency plane. The Wigner distribution (WD) is one of the most prominent members of this class. It has been defined by Wigner in quantum mechanics and then used first by Ville in signal analysis. All quadratic time–frequency distributions belonging to the general Cohen class, including the widely used spectrogram, may be written as two-dimensionally smoothed forms of the WD. The properties of these distributions have been investigated in detail during the last two decades [2], [4], [5], [7]–[9], [13], [14], [28], [34]. In particular, one of the topics of intensive study has been the influence of noise on the

time–frequency representations [1], [4], [10], [12], [18], [15], [22], [24], [25], [29], [30]. It has been concluded that the lag window length is one of the most important parameters. It has to be chosen by a tradeoff between the variance and bias of the WD estimate [25].

The aim of this paper is to present an algorithm that produces the window length close to the optimal one without knowing the bias since the bias depends on derivatives of the unknown distribution to be estimated.

The proposed algorithm is based on the idea developed in [11], [16], and [17]. The orientation of the algorithms presented in [16] and [17] on the instantaneous frequency (IF) estimation resulted in a time-varying-only window length that is the best for the IF estimation, i.e., for reconstruction of the WD at its peak on the frequency, for a given time instant. Then, this window length is used for all frequencies at this time instant. The algorithm developed in this paper directly addresses the problem of “denoising” the WD that was itself obtained from the noisy signals. This new algorithm produces the adaptive window length that is both time and frequency varying.

Two points determine the difference of the new algorithm versus the algorithms studied in [16] and [17]. First, the adaptive window in the new algorithm is assumed to be dependent both on time and frequency. Second, the variance of the WD itself, instead of the variance of the IF, is used in this algorithm. The calculation of this variance requires no assumption about the smallness of the noise. A theoretical analysis of the algorithm parameters is done in this paper as well. Furthermore, the new adaptive algorithm appears to be efficient for the analysis of multicomponent signals when it is applied in a combination with the reduced interference distributions. On the whole, the idea of the developed algorithm is quite universal and flexible and can be adapted to various time–frequency analysis problems. Finally, the algorithm is simple in implementation. In particular, a simplified version of the algorithm, based on the WD calculation with only two different (small and large) window lengths, is developed in this paper.

The paper is organized as follows. The analysis of the variance and bias of the WD of a deterministic noisy signal is reviewed in Section II. The idea of the adaptive algorithm, as well as a choice of its key parameters, is presented in Section III. The basic adaptive algorithm is described in Section IV. The simplified “two-window” version of the algorithm is presented in Section V. Sections V and VI deal with examples, including the algorithm application on the S

Manuscript received February 18, 1998; revised August 23, 1998. This work was supported in part by the Alexander von Humboldt Foundation. The associate editor coordinating the review of this paper and approving it for publication was Dr. Frans M. Coetsee.

LJ. Stanković is with Signal Theory Group, Ruhr University Bochum, Bochum, Germany, on leave from the Elektrotehnicki Fakultet, University of Montenegro, Montenegro, Yugoslavia (e-mail: l.stankovic@ieee.org).

V. Katkovnik is with the Statistics Department, University of South Africa, Pretoria, Republic of South Africa.

Publisher Item Identifier S 1053-587X(99)02149-2.

method and a generalization to the other distributions from the Cohen class.

II. THEORY

The WD of a discrete-time noisy signal $x(n) = f(n) + \nu(n)$ is defined as [3], [7], [13], [25], [35]

$$W_{xx}(n, \theta) = \sum_{k=-N/2}^{N/2-1} x(n+k)x^*(n-k)e^{-j2\theta k} \quad (1)$$

with $N \rightarrow \infty$. Note that the constant factor of 2 was used in the original definition of (1).

We assume that the signal $f(n)$ is deterministic and that the noise is complex-valued white, i.e., $\nu(n) = \nu_1(n) + j\nu_2(n)$, with

$$\begin{aligned} E\{\nu(n)\} &= 0 \\ E\{\nu(n)\nu^*(m)\} &= \sigma_\nu^2 \delta(n-m) \end{aligned} \quad (2)$$

where the asterisks denotes a complex conjugate value, and $\delta(n)$ is the Kronecker function. The Gaussian-distribution law for $\nu_1(n)$ and $\nu_2(n)$, with $E\{\nu_1(n)\nu_2(n)\} = 0$ and $E\{\nu_1^2(n)\} = E\{\nu_2^2(n)\} = \sigma_\nu^2/2$, is assumed as well.

Let

$$\Delta W_{xx}(n, \theta) = W_{xx}(n, \theta) - W_{ff}(n, \theta) \quad (3)$$

be an error of estimation of $W_{ff}(n, \theta)$ by $W_{xx}(n, \theta)$. Then, the bias of the estimate is equal to

$$E\{\Delta W_{xx}(n, \theta)\} = \sigma_\nu^2$$

because

$$E\{W_{xx}(n, \theta)\} = W_{ff}(n, \theta) + \sigma_\nu^2. \quad (4)$$

This bias is constant for all (n, θ) . It means that the WD of the noise-free signal is superimposed on a pedestal whose height is σ_ν^2 [18]. Therefore, it does not play a significant role in time-frequency analysis of the signal.

The variance of the WD is defined by

$$\begin{aligned} \sigma_{xx}^2 &= E\{W_{xx}(n, \theta)W_{xx}^*(n, \theta)\} \\ &\quad - E\{W_{xx}(n, \theta)\}E\{W_{xx}^*(n, \theta)\}. \end{aligned} \quad (5)$$

It has two components $\sigma_{xx}^2 = \sigma_{f\nu}^2 + \sigma_{\nu\nu}^2$, where the first one $\sigma_{f\nu}^2$ depends on both the signal and noise, and the second one $\sigma_{\nu\nu}^2$ depends on the noise only [25]. For the considered white complex-valued Gaussian noise, the variance of (1) is obtained as [1], [25], [30]

$$\begin{aligned} \sigma_{f\nu}^2 &= \sigma_\nu^2 \sum_{k=-N/2}^{N/2-1} |f(n+k)|^2 + |f(n-k)|^2 \\ \sigma_{\nu\nu}^2 &= N\sigma_\nu^4. \end{aligned} \quad (6)$$

For the FM signal

$$f(n) = A(n) \exp(j\phi(n)) \quad (7)$$

with a slow-varying real-valued amplitude $|f(n \pm k)|^2 \cong A^2(n)$, we get

$$\sigma_{xx}^2 = N\sigma_\nu^2(2A^2(n) + \sigma_\nu^2). \quad (8)$$

Thus, the variance of the unwindowed $W_{xx}(n, \theta)$ approaches infinity, as $N \rightarrow \infty$. In practice, truncated signals are used in the WD calculation.

Let us introduce the symmetric window $w_e(k) = w(k)w(-k)$ of a finite length N . Then, the WD of the signal truncated by the window $w(n)$ (the pseudo WD) is of the form

$$W_{xx}(n, \theta; N) = \sum_{k=-\infty}^{\infty} w(k)w(-k)x(n+k)x^*(n-k)e^{-jk2\theta}. \quad (9)$$

Calculation of the mean of $W_{xx}(n, \theta; N)$ gives [25]

$$E\{W_{xx}(n, \theta; N)\} = W_{ff}(n, \theta) *_{\theta} F_w(2\theta) + \sigma_\nu^2 w^2(0) \quad (10)$$

where $F_w(\theta) = FT[w_e(k)]$ is the Fourier transform (FT) of the window $w_e(k) = w(k)w(-k)$, and $*_{\theta}$ denotes a convolution in θ .

We will assume that $w(0)$ does not depend on N , as is true for many commonly used windows when $w(0) = 1$. Then, the term σ_ν^2 can be omitted in the following analysis since only the terms depending on N are essential for the window length optimization (formally, we may consider that $W_{ff}(n, \theta) + \sigma_\nu^2$ is the true value). The first term in (10) shows that decreasing of the lag-window length (increasing the length of its FT) causes corresponding increase of the WD bias. Using a Taylor series expansion of $W_{xx}(n, \theta; N)$ gives the approximate expression for $E\{W_{xx}(n, \theta; N)\}$ [25]

$$\begin{aligned} W_{ff}(n, \theta) *_{\theta} F_w(2\theta) &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} W_{ff}(n, \theta - \omega) F_w(2\omega) d\omega \\ &\cong W_{ff}(n, \theta) + \frac{1}{8} \frac{\partial^2 W_{ff}(n, \theta)}{\partial \theta^2} m_2. \end{aligned}$$

The amplitude moment of the window $w_e(k)$ is defined by

$$m_2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \omega^2 F_w(\omega) d\omega.$$

According to (3), the bias is obtained in the form

$$\text{bias}(n, \theta; N) = \frac{1}{8} B_f(n, \theta) m_2 \quad (11)$$

where, according to the previous discussion, the term $\sigma_\nu^2 w^2(0)$ is omitted, and

$$B_f(n, \theta) = \frac{\partial^2 W_{ff}(n, \theta; N)}{\partial \theta^2}. \quad (12)$$

For a nonrectangular window, the variance $\sigma_{xx}^2 = \sigma_{f\nu}^2 + \sigma_{\nu\nu}^2$ should be written in a more general form than (6)

$$\begin{aligned} \sigma_{f\nu}^2 &= \sigma_\nu^2 \sum_{k=-N/2}^{N/2-1} w^2(k)w^2(-k)[|f(n+k)|^2 + |f(n-k)|^2] \\ \sigma_{\nu\nu}^2 &= \sigma_\nu^4 \sum_{k=-N/2}^{N/2-1} w^2(k)w^2(-k). \end{aligned} \quad (13)$$

For the FM signal with a slowly varying amplitude

$$\sigma_{xx}^2 = E_w \sigma_\nu^2 (2A^2(n) + \sigma_\nu^2) \quad (14)$$

where

$$E_w = \sum_{k=-N/2}^{N/2-1} [w(k)w(-k)]^2$$

is an energy of the window $w_e(k) = w(k)w(-k)$.

For the Hanning window $w_e(k)$, we have

$$m_2 = 2(\pi/N)^2 \quad \text{and} \quad E_w = 3N/8. \quad (15)$$

These values may be easily calculated for any other window type [19].

The optimal window length is obtained by minimizing the mean squared error (MSE) defined by

$$e^2(n, \theta; N) = \text{bias}^2(n, \theta; N) + \sigma_{xx}^2. \quad (16)$$

Assuming that $w_e(k)$ is the Hanning window and $f(n)$ is the FM signal with a slow-varying amplitude (7), we get

$$e^2(n, \theta; N) = \frac{B_f^2(n, \theta)\pi^4}{16N^4} + \frac{3}{8}N\sigma_\nu^2(2A^2(n) + \sigma_\nu^2). \quad (17)$$

The optimal window length N_{opt} follows from the equation

$$\frac{\partial e^2(n, \theta; N)}{\partial N} = 0$$

that gives

$$N_{\text{opt}} = \sqrt[5]{\frac{2B_f^2(n, \theta)\pi^4}{3\sigma_\nu^2(2A^2(n) + \sigma_\nu^2)}}. \quad (18)$$

The optimal window length depends on $B_f(n, \theta)$, meaning that it is time–frequency varying. Formula (18) is very interesting for a theoretical analysis, but it is not useful for practical applications because $B_f(n, \theta)$ is the second derivative of $W_{ff}(n, \theta)$ on θ . It is definitely unknown in advance. Our aim is to design an adaptive algorithm that produces the value of $N(n, \theta)$ close to the optimal one N_{opt} at any given point (n, θ) without knowledge of $W_{ff}(n, \theta)$ and its derivatives.

The algorithm will be defined on the basis of the analysis that follows.

The MSE (16) can be represented in the form

$$e^2(n, \theta; N) = \sigma_{xx}^2[1 + \gamma^2(n, \theta; N)] \quad (19)$$

where

$$\gamma(n, \theta; N) = \text{bias}(n, \theta; N)/\sigma_{xx} \quad (20)$$

is a ratio of the bias to standard deviation of the WD estimate. It is easy to show that

$$\gamma(n, \theta; N_{\text{opt}}) = 1/2 \quad (21)$$

for the given formulae of the bias and variance (17).

Let us emphasize this quite interesting point. *At the optimal window length N_{opt} , the ratio of the bias to standard deviation is a constant. It does not depend on the WD or any other parameter of the problem.*

III. BASIC IDEA OF ADAPTIVE WINDOW LENGTH SELECTION

According to the presented analysis, we can conclude that the random error of the WD estimation can be represented in the form

$$\Delta W_{xx}(n, \theta; N) = W_{xx}(n, \theta; N) - \text{bias}(n, \theta; N) - W_{ff}(n, \theta)$$

where $E\{\Delta W_{xx}(n, \theta; N)\} = 0$, and $E\{\Delta W_{xx}^2(n, \theta; N)\} = \sigma_{xx}^2(N)$. Here, $W_{ff}(n, \theta)$ includes constant σ_ν^2 .

Now, we may write the inequality

$$|W_{xx}(n, \theta; N) - \text{bias}(n, \theta; N) - W_{ff}(n, \theta)| \leq \kappa\sigma_{xx}(N) \quad (22)$$

which holds with probability $P(\alpha) = 1 - \alpha$, and κ is the $(1 - \alpha/2)$ th quantile of the standard Gaussian distribution. Precisely speaking, the random error ΔW is a quadratic function of the Gaussian noise ν , i.e., it has a non-Gaussian distribution. However, it can be shown that as $N \rightarrow \infty$, the Gaussian distribution is a good approximation for the distribution of WD estimation errors. Thus, we can apply the quantile of the Gaussian distribution in (22).¹

Let us assume that for some N

$$|\text{bias}(n, \theta; N)| \leq \Delta\kappa\sigma_{xx}(N) \quad (23)$$

holds. Then, (22), for that particular N , can be strengthened to the form

$$|W_{xx}(n, \theta) - W_{ff}(n, \theta; N)| \leq (\kappa + \Delta\kappa)\sigma_{xx}(N). \quad (24)$$

We wish to make clear the idea behind assumption (23). First, it determines a proportion between the bias and the standard deviation in the window length selection. Second, $\Delta\kappa$ is a parameter of this proportion to be chosen. A few approaches could be used here.

- a) The simplest choice $\Delta\kappa = 1$ assumes the equality of the random error and the bias, as we did in the IF estimation in [17].
- b) $\Delta\kappa = \gamma(n, \theta, N_{\text{opt}}) = 1/2$ assumes a proportion between the bias and the standard deviation corresponding to that which holds for the optimal window length minimizing the MSE (21). This value will be a good intuitive choice for $\Delta\kappa$ in the simplified “two-window” algorithm that is going to be presented.
- c) Finding $\Delta\kappa$ through an optimal estimation of the unknown N_{opt} , the last approach is considered in the sequel of this section.

In order to derive the algorithm, we have to define a finite set of the window length values for which the WD will be calculated. The widest length N_1 is determined by the computational time we may allow. For the last length N_J , we may assume any reasonably small value, for example, $N_J = 8$ or 16. In between these two values, a finite number of window lengths has to be assumed. Let us introduce a dyadic set \mathbf{N} of descending window lengths $N_s \in \mathbf{N}$

$$\mathbf{N} = \{N_s \mid N_s = N_{s-1}/2, \quad s = 1, 2, 3, \dots, J\}. \quad (25)$$

¹Note that the Gaussian assumption for ΔW is not crucial for the analysis that follows.

A simplicity and efficiency of the WD calculation for $N_s \in \mathbf{N}$, using the standard radix-2 FFT algorithms, is one of the best reasons in favor of selecting \mathbf{N} as a dyadic set.

Now, we are going to derive the algorithm for determination of the adaptive window length close to the optimal $N_{\text{opt}}(n, \theta)$, using $W_{xx}(n, \theta; N_s)$ as the estimate of $W_{ff}(n, \theta)$ and the variance of this estimate only. We assume from the beginning that the noise variance σ_v^2 and the amplitude $A(n)$ of (7) are known.

Let us consider confidence intervals for (24). They play a key role in the adaptive algorithm. Denote the upper and lower bounds of these interval's confidence $D_s = [L_s, U_s]$ as U_s and L_s

$$\begin{aligned} L_s &= W_{xx}(n, \theta; N_s) - (\kappa + \Delta\kappa)\sigma_{xx}(N_s) \\ U_s &= W_{xx}(n, \theta; N_s) + (\kappa + \Delta\kappa)\sigma_{xx}(N_s) \end{aligned} \quad (26)$$

where $W_{xx}(n, \theta; N_s)$ is an estimate of $W_{ff}(n, \theta)$, using the window length $N = N_s$, and $\sigma_{xx}(N_s)$ is its variance. Consider these confidence intervals for the successive values of s , i.e., the successive window lengths.

The algorithm is based on the following statement.

Let us assume that (22) holds with the probability $P(\alpha) = 1 - \alpha \simeq 1$ and that the optimal window length belongs to the set \mathbf{N} , $N_{\text{opt}} \in \mathbf{N}$.

Determine N_{s^+} as a window length corresponding to the largest $s = s^+$ ($s = 1, 2, \dots, J$) when two successive confidence intervals still intersect, i.e.,

$$D_s \cap D_{s+1} \neq \emptyset, \quad \forall s = 1, 2, \dots, s^+ \quad (27)$$

is still satisfied.

Then, there exists the value of $\Delta\kappa$ in (26) such that $D_{s^+} \cap D_{s^++1} \neq \emptyset$ and $D_{s^++1} \cap D_{s^++2} = \emptyset$ when $N_{s^+} = N_{\text{opt}}$.

Proof: It is clear that because of the assumption that $N_{\text{opt}} \in \mathbf{N}$, any N_s can be represented as

$$\begin{aligned} N(p) &= 2^{-p}N_{\text{opt}} \\ p &= \dots, -2, -1, 0, 1, 2, \dots \end{aligned} \quad (28)$$

where $p = 0$ corresponds to the optimal window length N_{opt} . From here, we start to use two different indices for the window lengths:

- s in the form N_s , which denotes the indexing starting from the largest toward the narrowest window length;
- p in the form $N(p)$, where the indexing starts from N_{opt} window length (when $p = 0$).

Confidence intervals (26), as functions of p , are defined by

$$\begin{aligned} D(p) &= [W_{xx}(n, \theta; N(p)) - (\kappa + \Delta\kappa)\sigma_{xx}(N(p)) \\ &W_{xx}(n, \theta; N(p)) + (\kappa + \Delta\kappa)\sigma_{xx}(N(p))] \end{aligned} \quad (29)$$

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

The bias and variance for any $N(p)$, according to (17), (21), and (28), may be rewritten as

$$\begin{aligned} \sigma_{xx}(N(p)) &= 2^{-p/2}\sigma_{xx}(N_{\text{opt}}) \\ \text{bias}(n, \theta; N(p)) &= 2^{(2p-1)}\sigma_{xx}(N_{\text{opt}}). \end{aligned} \quad (30)$$

In the last expression, we assumed, without loss of generality, that the bias is positive. Note again that N_{opt} is an (n, θ) dependent parameter.

From the monotonicity property of functions in (30), we can conclude that for $p \ll 0$, the bias is much smaller compared with the variance, whereas for $p \gg 0$, the bias is much larger than the variance. Since the bias is small and the variance is large for all $p \ll 0$, we have $W_{ff}(n, \theta; N(p)) \in D(p)$, and therefore, $D(p) \cap D(p+1) \neq \emptyset$. Contrary to this case, for all $p \gg 0$, we have an inverse picture. The variance is small, and the bias is large; therefore, $D(p) \cap D(p+1) = \emptyset$ for any given $\Delta\kappa$.

From the definition of N_{s^+} and from the monotonicity of variance and bias in (30), it follows that $N_{s^+} \equiv N(0) = N_{\text{opt}}$, if and only if the intervals $D(0)$ and $D(1)$ have at least a point in common, and the intervals $D(1)$ and $D(2)$ do not have a point in common. Thus, the intersection of three intervals $D(0)$, $D(1)$, and $D(2)$ only will be considered. It can be seen that for the positive bias, the intersection of $D(0)$ and $D(1)$ and the nonintersection of $D(1)$ and $D(2)$ will occur, provided that

$$\begin{aligned} \min\{U(0)\} &\geq \max\{L(1)\} \\ \max\{U(1)\} &< \min\{L(2)\}. \end{aligned} \quad (31)$$

Minimization and maximization in (31) are produced with respect to the random values of the estimate $W_{xx}(n, \theta; N_s)$, which determine a middle point of the confidence intervals $D(p)$.

Let us, for example, consider $U(1)$. The value of $W_{xx}(n, \theta; N(p))$, according to (22), is within the interval $W_{xx}(n, \theta; N(1)) \in [W_{ff}(n, \theta) + \text{bias}(n, \theta, N(1)) - \kappa\sigma_{xx}(N(1)), W_{ff}(n, \theta) + \text{bias}(n, \theta, N(1)) + \kappa\sigma_{xx}(N(1))]$. Consequently, the upper bound $U(1)$ of the confidence interval $D(1)$, according to (29), takes a value from the interval $[W_{ff}(n, \theta) + \text{bias}(n, \theta, N(1)) + \Delta\kappa\sigma_{xx}(N(1)), W_{ff}(n, \theta) + \text{bias}(n, \theta, N(1)) + (2\kappa + \Delta\kappa)\sigma_{xx}(N(1))]$. The maximal possible value of $U(1)$ is, therefore, given by $\max\{U(1)\} = W_{ff}(n, \theta) + \text{bias}(n, \theta, N(1)) + (2\kappa + \Delta\kappa)\sigma_{xx}(N(1))$. In the same way, we get other bound limits required by (31), thus giving

$$\begin{aligned} &\text{bias}(n, \theta, N(0)) + \Delta\kappa\sigma_{xx}(N(0)) \\ &\geq \text{bias}(n, \theta, N(1)) - \Delta\kappa\sigma_{xx}(N(1)) \\ &\text{bias}(n, \theta, N(1)) + (2\kappa + \Delta\kappa)\sigma_{xx}(N(1)) \\ &< \text{bias}(n, \theta, N(2)) - (2\kappa + \Delta\kappa)\sigma_{xx}(N(2)). \end{aligned} \quad (32)$$

According to (30), it follows from (32) that

$$\begin{aligned} \frac{1}{2} + \Delta\kappa &\geq 2 - \Delta\kappa 2^{-1/2} \\ 2 + (2\kappa + \Delta\kappa)2^{-1/2} &< 8 - (2\kappa + \Delta\kappa)/2. \end{aligned} \quad (33)$$

The above relation gives the interval for $\Delta\kappa$

$$0.8787 \leq \Delta\kappa < 4.9706 - 2\kappa. \quad (34)$$

For example, for $\kappa = 2$, i.e., the probability $P(\kappa) \simeq 0.95$, we obtain

$$0.8787 \leq \Delta\kappa < 0.9706. \quad (35)$$

Inequality (34) shows that $\Delta\kappa > 0$ exists if $\kappa \leq 2.046$. It completes the proof of the statement and gives the inequality for the possible values of $\Delta\kappa$. \square

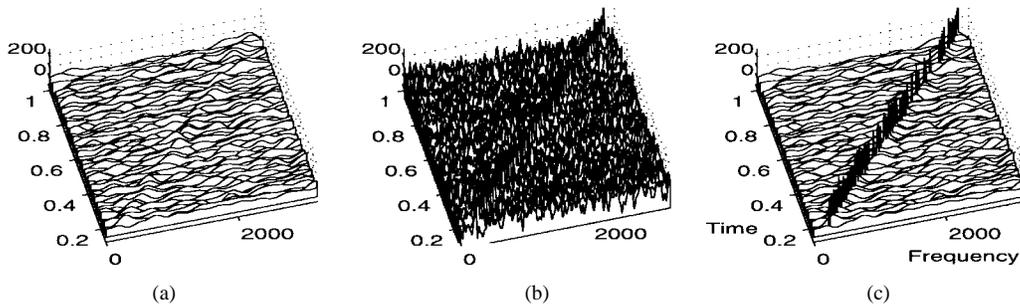


Fig. 1. Wigner distribution of the noisy signal. (a) With the constant window length $N_1 = 64$. (b) With the constant window length $N_2 = 512$. (c) With the adaptive time–frequency varying window length.

A search of the optimal window length over the finite set \mathbf{N} simplifies the optimization. However, it inevitably leads to a suboptimal solution due to the discretization effects since, in general, the optimal window length N_{opt} does not belong to the dyadic set \mathbf{N} . It always results in worse values of the MSE, but that is the price of efficient calculation schemes and the simplicity of the presented algorithm. Fortunately, this loss of accuracy is not significant in many cases because the MSE (19) varies slowly around the stationary point, i.e., in the neighborhood of $N = N_{\text{opt}}$. Note that the other window length schemes may also be assumed. We may go in two directions. One is to increase the number of discrete window lengths between N_1 and N_J . If we may computationally allow a large number of window lengths, i.e., small difference between N_s and N_{s+1} , theoretically $N_{s+1} = N_s - 1$, then a similar approach and algorithm can be used. The only difference in the case when two successive window lengths are very close to each other is that the intersection of the confidence interval D_{s+1} with all previous intervals D_s, D_{s-1}, \dots, D_1 should be used as an indicator when the optimal window length is achieved [11]. However, we will go in the other direction. Our aim is to reduce the number of considered window lengths between N_1 and N_J while keeping the high quality of the presented adaptive approach. This is possible due to a specific form of time–frequency representation. This topic is discussed in Section V.

IV. ALGORITHM

Let us initially assume that the amplitude A and the standard deviation σ_ν^2 of the noise are known. Define a dyadic set \mathbf{N} , for example, $\mathbf{N} = \{512, 256, 128, 64, 32, 16\}$. The following steps are then generated for each value of the time-instant n and the frequency θ .

- 1) The WD is calculated for all of $N_s \in \mathbf{N}$. Thus, we obtain a set of distributions

$$W_{xx}(n, \theta, N_s), \quad N_s \in \mathbf{N}.$$

- 2) The optimal window length $N_{s^+}(n, \theta)$ for every (n, θ) is determined by the largest s , from $s = 1, 2, \dots$, when (27) is satisfied, i.e.,

$$\begin{aligned} & |W_{xx}(n, \theta; N_s) - W_{xx}(n, \theta; N_{s+1})| \\ & \leq (\kappa + \Delta\kappa)(\sigma_{xx}(N_s) + \sigma_{xx}(N_{s+1})) \end{aligned} \quad (36)$$

still holds.

This s^+ is the smallest of those s for which the segments D_s and D_{s+1} , have a point in common.

- 3) The WD with the adaptive window length, for a point (n, θ) , is

$$W_{xx}^+(n, \theta) = W_{xx}(n, \theta; N_{s^+}(n, \theta)). \quad (37)$$

Comments on the Algorithm.

- a) The standard deviation

$$\sigma_{xx}(N_s) = \sqrt{E_w(N_s)\sigma_\nu^2(2A^2 + \sigma_\nu^2)} \quad (38)$$

depends on the amplitude of the signal A , the standard deviation of the noise σ_ν , and the energy of the window, E_w .

- 1) Provided that *the sampling period is small*, the following estimates of A and σ_ν^2 can be used:

$$\hat{A}^2 + \hat{\sigma}_\nu^2 = \frac{1}{N} \sum_{n=1}^N |x(n)|^2 \quad (39)$$

where the sum is calculated over all N observations, and N is assumed to be large. The variance can be estimated by

$$\begin{aligned} \hat{\sigma}_{vr} &= \frac{\{\text{median}(|x_r(n) - x_r(n-1)| : n = 2, \dots, N)\}}{0.6745} \\ \hat{\sigma}_{vi} &= \frac{\{\text{median}(|x_i(n) - x_i(n-1)| : n = 2, \dots, N)\}}{0.6745} \\ \hat{\sigma}_\nu^2 &= (\hat{\sigma}_{vr}^2 + \hat{\sigma}_{vi}^2)/2 \end{aligned} \quad (40)$$

where $x_r(n)$ and $x_i(n)$ are the real and imaginary part of $x(n)$, respectively.

The average $\sum_{n=1}^N |x(n) - x(n-1)|^2/N$ could also be used as an estimate of σ_ν^2 .

- 2) In the case of a *very high noise*, i.e., of a low signal-to-noise ratio, as we will consider in the simulations in this paper, we may use $\sigma_\nu^2(2A^2 + \sigma_\nu^2) \cong (A^2 + \sigma_\nu^2)^2 = [\sum_{n=1}^N |x(n)|^2/N]^2$.

The window energy $E_w(N_s)$ is a signal independent constant. For example, for the Hanning window, $E_w(N_s) = 3N_s/8$. Thus, we may easily get the estimate of (38).

Which one of the previous methods will be used for the variance $\sigma_{xx}^2(N_s)$ estimation depends on the specific situation. While the first one is restricted by a small sampling rate for an accurate σ_ν^2 estimation, the other is valid for low signal to noise ratio cases only.

b) *A note on the realization*: The WD with $N_{s+1} = N_s/2$ samples can be easily obtained from the distribution with N_s samples as its smoothed version. Let us, for example, suppose

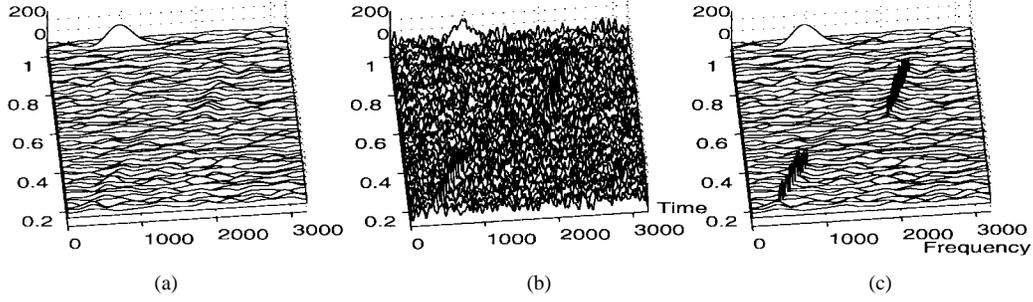


Fig. 2. Wigner distribution of the noisy signal: (a) With the constant window length $N_1 = 64$, (b) With the constant window length $N_2 = 512$, (c) With the adaptive time-frequency varying window length.

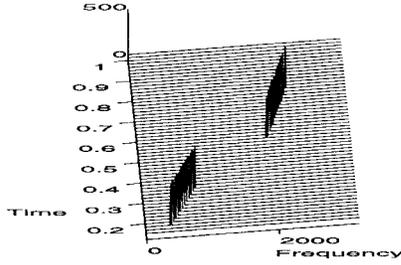


Fig. 3. Time-frequency varying window length $N_{s+}(n, \theta)$ in Fig. 2(c) as a function of (n, θ) obtained using “two-window” algorithm $\mathbf{N} = \{64 \ 512\}$. The algorithm has used the small variance distribution from Fig. 2(a), calculated with $N_{s+}(n, \theta) = N_1 = 64$, for the entire time-frequency plane, except the domains where the first two components exist. Since the distribution for these two components has significant variations along the frequency axis, producing large bias coefficient, the algorithm has only here used the lowbiased distribution values from Fig. 2(b) with $N_{s+}(n, \theta) = N_2 = 512$.

that the distribution $W_{xx,P}(n, \theta; N_s)$ is calculated using the rectangular window of the length N_s . The distribution with the Hanning window of the same length can be obtained as

$$W_{xx,H}(n, \theta; N_s) = \sum_{k=-1}^1 a_{|k|} W_{xx,P}\left(n, \theta - k \frac{4\pi}{N_s}; N_s\right)$$

where $\{a_0, a_1\} = \{0.5, 0.25\}$. Now, using twice narrower Hanning window width means further smoothing of the values of $W_{xx,P}(n, \theta; N_s)$. Assuming that the samples of the Hanning window Fourier transform may be neglected outside the main lobe, we get

$$W_{xx,H}(n, \theta; N_{s+1}) = \sum_{k=-3}^3 a_{|k|} W_{xx,P}\left(n, \theta - k \frac{4\pi}{N_s}; N_s\right)$$

where $a_{|k|}$ are samples of the Fourier transform of the Hanning window within its main lobe $\{a_0, a_1, a_2, a_3\} = \{0.5, 0.42, 0.25, 0.08\}$. This procedure may significantly save the computation time. The previous procedure may be easily generalized to any set of Wigner distributions where $W_{xx,H}(n, \theta; N_{s+1})$ may be treated as a smoothed form of $W_{xx,H}(n, \theta; N_s)$.

V. TWO-WINDOW ALGORITHM

Let us consider only two window lengths $\mathbf{N} = \{N_1, N_2\}$ such that N_1 is small enough so that the variance of distribution is small, and N_2 is large so that bias is small (17), i.e., $N_1 \ll N_2$. If the bias at the points $(n, \theta; N_s)$, $s = 1, 2$ is

very small, i.e., the factor $B(n, \theta; N_s)$ is close to zero, then the corresponding two confidence intervals will intersect. Thus, for these points (n, θ) , we will use the distribution $W_{xx}(n, \theta; N_1)$ calculated with the smaller window length N_1 as a better choice with respect to the variance. Otherwise, for a large bias, the confidence intervals do not intersect. This fact indicates that the algorithm has to use the distribution $W_{xx}(n, \theta; N_2)$, which is the better choice with respect to the bias, since the bias is now dominant. Therefore, we may define a very simple “two-window” algorithm. This algorithm has the following form:

$$W_{xx}^+(n, \theta) = \begin{cases} W_{xx}(n, \theta; N_1), & \text{if } \Phi = \text{true} \\ W_{xx}(n, \theta; N_2), & \text{otherwise} \end{cases} \quad (41)$$

where $\Phi = \text{true}$ means that $|W_{xx}(n, \theta; N_1) - W_{xx}(n, \theta; N_2)| \leq (\kappa + \Delta\kappa)\sigma_{xx}(N_2)(1 + \sqrt{N_1/N_2})$ holds. Using only two distributions, we can expect a significant improvement of the time-frequency representation. We may expect that this “black-and-white” approach is very suitable for this kind of problem since a time-frequency representation is usually either zero or very highly concentrated, i.e., very fast varying 2-D function. This very simple algorithm will be demonstrated on several examples.

Note that the variance $\sigma_{xx}(N_1)$ can be calculated from the better estimated $\sigma_{xx}(N_2)$ as $\sigma_{xx}^2(N_1) = \sigma_{xx}^2(N_2)N_1/N_2$ since if we scale any window along the time axis, say a times, its energy E_w is also scaled a times. Variance $\sigma_{xx}^2(N_2)$ is estimated according to the comments in Section IV. In practical realizations, the distribution $W_{xx}(n, \theta; N_1)$, with a narrower window, can be treated just as a very frequency domain-smoothed form of $W_{xx}(n, \theta; N_2)$. The “two-window” algorithm may be very efficiently implemented in this way.

VI. EXAMPLES

In the examples, we will present results obtained by the two-window method for signals with a very low signal-to-noise ratio (SNR).

1) First, consider a linear FM signal

$$x(n) = Ae^{j1600(nT-0.05)^2} + \nu(n) \quad (42)$$

within the time interval $0 \leq nT \leq 1$. The time-frequency analysis is done using the WD with the Hanning window $w_e(k)$ and the discretization period $T = 1/2048$. The variance of the noise $\nu(n)$ was such that $20 \log(A/\sigma_\nu) = -5$ [dB].

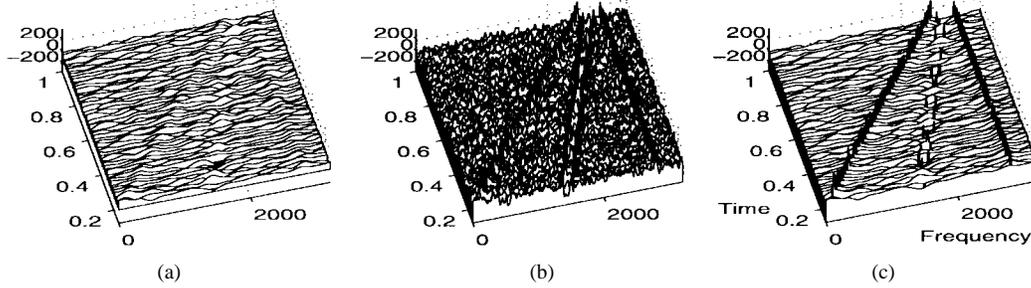


Fig. 4. Wigner distribution of the noisy multicomponent signal. (a) With the constant window length $N_1 = 64$. (b) With the constant window length $N_2 = 512$. (c) With the adaptive time–frequency varying window length.

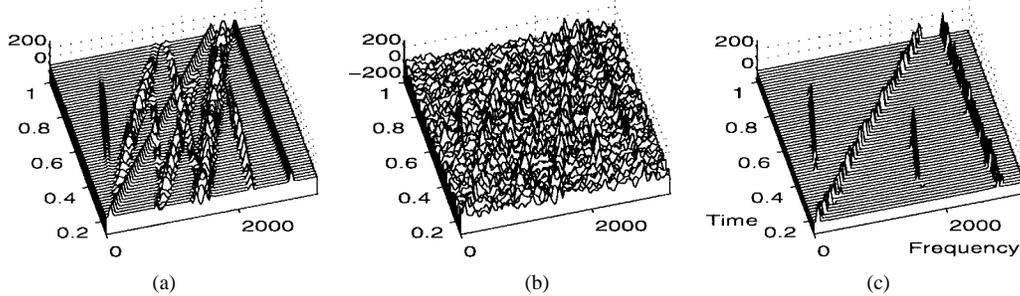


Fig. 5. Time–frequency representation of the multicomponentnoisy signal. (a) Wigner distribution of the signal without noise. (b) Wigner distribution of noisy signal with the constant window length $N_2 = 256$. (c) S -method of noisy signal with the adaptive time–frequency varying window length.

We use two window lengths $N_1 = 64$ and $N_2 = 512$. Applying algorithm (41), we get the adaptive distribution as it is presented in Fig. 1(c), whereas Fig. 1(a) and (b) present the WD with the constant windows N_1 and N_2 , respectively. The improvement is evident. The adaptive algorithm uses the low-variance distribution everywhere except in the area where the wider window significantly improves the distribution concentration.

2) Consider a sum of three noisy chirp signals:

$$\begin{aligned} x(n) = & A e^{j1200(nT)^2} e^{-25(nT-0.25)^2} \\ & + A e^{-20(nT-0.65)^2} e^{j750(nT+0.75)^2} \\ & + 3.5 A e^{-22.500(nT-0.96875)^2} e^{j1000nT} + \nu(n). \end{aligned} \quad (43)$$

The parameters of the algorithm are the same as in the first example. The WD's with the constant and adaptive window lengths, are presented in Fig. 2(a)–(c), respectively. The adaptive algorithm uses the distribution with $N_1 = 64$, i.e., with the low variance for all regions where the bias is small, including the third signal component. The distribution with $N_2 = 512$ is chosen by the adaptive algorithm only for the first two signal components. A 3-D plot of the adaptive time–frequency varying window length is given in Fig. 3.

Note that the signals do not significantly overlap in time; therefore, the lag-window was sufficient to reduce the cross-terms in the WD.

3) The following are the zero bias and zero noise cases.

a) If the bias is zero, the adaptive algorithm will always choose the smaller window length N_1 since, according to (11) and (12), $\text{bias}(n, \theta; N) = 0$ for any n , and the confidence intervals intersect (41). For example, let $x(n) = A\delta(n - n_0) + \nu(n)$; then, $W_{xx}(n, \theta) = A^2\delta(n -$

$n_0) + \text{Noise}$, with the variance $\sigma_{xx}^2 = \sigma_\nu^2(A^2 + N\sigma_\nu^2)$, and the bias is not dependent on N . Approximately, this situation has appeared for the third signal's component in Example 2.

b) If there is no noise and the difference between $W_{xx}(n, \theta; N_1)$ and $W_{xx}(n, \theta; N_2)$ is not equal to zero, then according to (41), the algorithm chooses the distribution with the larger window length. This choice results in a better concentration of the distribution.

4) Finally, consider a two-component signal that is a combination of the linear frequency modulated signal and the signal with a constant frequency

$$x(n) = A e^{j1360(nT-0.1)^2} + A e^{j2800nT} + \nu(n).$$

The WD's with $N_1 = 64$ and $N_2 = 512$ window lengths, as well as the distribution with the adaptive variable window length, are presented in Fig. 4(a)–(c), respectively. The same parameters as in the previous examples are used.

The adaptive algorithm gives a very concentrated distribution with a reduced noise influence. The cross-term effects are here, of course, present.

VII. GENERALIZATION TO THE COHEN'S CLASS OF DISTRIBUTIONS

Consider an arbitrary quadratic time–frequency distribution from the Cohen class [7], [13], [14], [34]

$$\begin{aligned} C_{xx}(t, \omega; \Pi) \\ = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_{xx}(t-u, \omega-v) \Pi(u, v) du dv \end{aligned} \quad (44)$$

where $\Pi(t, \omega)$ is a kernel in the analog time–frequency domain. It is a two-dimensional (2-D) Fourier transform of the usual ambiguity domain kernel $c(\Theta, \tau)$.

The bias of estimate (44) is derived using a Taylor expansion of $W_{xx}(t - u, \omega - v)$ and the following facts.

- 1) For all product kernels, $c(\Theta, \tau) = c(\tau\Theta)$ holds $\frac{\partial^p c(\tau\Theta)}{\partial \tau^p_{(0,0)}} = \frac{\partial^p c(\tau\Theta)}{\partial \Theta^p_{(0,0)}} = 0$.
- 2) Kernel $c(\tau\Theta)$ is an even function with respect to the coordinate axis.
- 3) It satisfies the condition $c(0, 0) = c(0) = 1$.

Then, we get

$$E\{C_{xx}(t, \omega; \Pi)\} = W_{ff}(t, \omega) + \frac{\partial^4 W_{ff}(t, \omega)}{\partial t^2 \partial \omega^2} M_4^{(t^2 \omega^2)} + \dots \quad (45)$$

where

$$\begin{aligned} M_4^{(t^2 \omega^2)} &= \frac{1}{8\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t^2 \omega^2 \Pi(t, \omega) dt d\omega \\ &= \frac{1}{4} \frac{\partial^4 c(\Theta, \tau)}{\partial \Theta^2 \partial \tau^2} \Big|_{(\tau, \Theta)=(0,0)}. \end{aligned}$$

The constant term σ_v^2 is neglected as in (10).

For example, for the Choi–Williams distribution [5], [7], when $c(\Theta, \tau) = \exp(-\tau^2 \Theta^2 / \alpha^2)$, we get

$$\text{bias}(t, \omega; \alpha) \cong \frac{\partial^4 W_{ff}(t, \omega)}{\partial t^2 \partial \omega^2} M_4^{(t^2 \omega^2)} \sim \frac{1}{\alpha^2}. \quad (46)$$

The mean of the estimation variance in the Cohen class of distribution is proportional to the kernel energy, [1], [12], [29]

$$\sigma_{xx}^2 \sim \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |c(\Theta, \tau)|^2 d\tau d\Theta. \quad (47)$$

For the Choi–Williams distribution, we have $\sigma_{xx}^2 \sim \alpha$. Therefore, the total MSE has the same form with respect to the parameter α as (19) with respect to N

$$e^2(t, \omega; \alpha) = \frac{B(t, \omega)}{\alpha^4} + V\alpha. \quad (48)$$

Now, taking only two values for $\alpha \in \{\alpha_1, \alpha_2\}$ (one very small such that the variance may be neglected, and the other much larger $\alpha_1 \ll \alpha_2$ that the bias is small but still within a region where the crossterms are sufficiently suppressed [28]), we may directly apply the algorithm described by (41) in Section V.

Comments on the Estimation Variance: The above analysis has been presented for the mean (over frequency θ) value of the estimation variance. Here, however, we have a more complex situation than in the Wigner distribution case. The estimation variance may be time and frequency dependent. Since the value of estimation variance at a specific point (t, ω) is a relevant parameter in the algorithm, instead of the mean variance value, we will consider this parameter in more detail. In the case of the reduced interference distributions, the estimation variance has two parts: one that depends on the noise only and the other, which is signal and noise dependent [1], [29]. The noise only dependent part of the estimation variance is given by [1], [29]

$$\sigma_{\nu\nu}^2(t, \omega) = \sigma_\nu^4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\varphi(t, \tau)|^2 dt d\tau$$

where $\varphi(t, \tau)$ is the Fourier transform of $c(\Theta, \tau)$ with respect to Θ . In the discrete time-lag form, this variance is given by [1], [29]

$$\sigma_{\nu\nu}^2(n, \theta) = \sigma_\nu^4 \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |\varphi(m, k)|^2.$$

It is important to point out that this part of the estimation variance is time and frequency invariant. Since the kernel energy is a constant for a given distribution, in order to estimate $\sigma_{\nu\nu}^2(n, \theta)$, we will need to estimate the noise variance σ_ν^2 only. This procedure has already been described in Section IV.

The second part of the estimation variance $\sigma_{f\nu}^2(n, \theta)$ has the form [1], [29]

$$\begin{aligned} \sigma_{f\nu}^2(n, \theta) &= \sigma_\nu^2 \sum_{m_1=-\infty}^{\infty} \sum_{k_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \varphi(m_1, k_1) \varphi^*(m_2, k_2) \\ &\quad \times [f(n + m_1 + k_1) f^*(n + m_2 + k_2) \\ &\quad \times \delta(m_1 - m_2 - k_1 + k_2) \\ &\quad + f^*(n + m_1 - k_1) f(n + m_2 - k_2) \\ &\quad \times \delta(m_1 - m_2 + k_1 - k_2)] e^{-2j\theta(k_1 - k_2)}. \end{aligned}$$

Note that $\sigma_{f\nu}^2(n, \theta)$ contains the time- and frequency-invariant part obtained by summation for $m_1 = m_2$ and for any k_1, k_2 , as well as $k_1 = k_2$ and any m_1, m_2 . This part of variance, for frequency-modulated signals, is equal to $2\sigma_\nu^2 A^2 \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |\varphi(m, k)|^2$. It is very interesting to note that the mean of $\sigma_{f\nu}^2(n, \theta)$, over frequency θ , is also equal to the same value [1]

$$\overline{\sigma_{f\nu}^2(n, \theta)} = 2\sigma_\nu^2 A^2 \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |\varphi(m, k)|^2.$$

These were the reasons why, in the adaptive algorithm, we have used the mean value of the variance over the frequency [1], as an optimization-relevant parameter

$$\sigma_{xx}^2(n, \theta) = \sigma_{\nu\nu}^2(n, \theta) + \overline{\sigma_{f\nu}^2(n, \theta)} = E_\varphi \sigma_\nu^2 (2A^2 + \sigma_\nu^2) \quad (49)$$

where $E_\varphi = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |\varphi(m, k)|^2$ is the energy of distribution's kernel. Thus, we came up to a formally the same expression as (14), and we may directly apply variance estimation techniques described in Section IV: “Comments on the algorithm”. Here, however, we want to emphasize, once more, that the variance $\sigma_{f\nu}^2(n, \theta)$ beside the time–frequency-invariant components contains one component that has zero mean value and is time–frequency dependent. Although the analysis of this part is very complicated, it has been shown [30] that this time–frequency-variant part of variance increases the value of $\sigma_{f\nu}^2(n, \theta)$ within the region where the time–frequency distribution of a nonnoisy signal has significant values and, therefore, decreases value of $\sigma_{f\nu}^2(n, \theta)$, where the distribution of nonnoisy signal assumes values close to zero (since it has zero mean value). Thus, using the mean value of σ_{xx}^2 , given by (49), we will use lower values for the estimation variance than the true ones in the regions where the nonnoisy time–frequency distribution has significant values. It will increase the probability that the algorithm takes the parameter value α_2 , producing

a low bias since it will conclude that the variance is already small. For the cases with the small distribution values, the mean of estimation variance will be higher than the true one so that the algorithm will chose the lower variance parameter value α_1 . Fortunately, this is a safe side error, and it will not degrade the results in the *two-parameter algorithm*.

An interesting distribution, from the point of view of the algorithm presented in this paper, is the Butterworth distribution with kernel [33]

$$c(\Theta, \tau) = \frac{1}{1 + (\tau\Theta/\alpha)^{2M}}. \quad (50)$$

For this distribution, all factors (partial derivatives) in series (45) expansion are equal to zero up to the term

$$\frac{\partial^{4M} c(\Theta, \tau)}{\partial \Theta^{2M} \partial \tau^{2M}} \sim \frac{1}{\alpha^{4M}}. \quad (51)$$

Thus, for this distribution $e^2(n, \theta; \alpha) = B(n, \theta)/\alpha^{4M} + V\alpha$. For $M = 1$, we have the same expression as (19), but for $M > 1$, we have a much faster decreasing bias that allows us to use closer values of α_1 and α_2 for better final results (sense of the derivations in Section III).

The multicomponent signal case will be illustrated on the adaptive version of the S method [26], [27], [30], [31]. The S -method belongs to the Cohen class of distributions [28]. It is defined as

$$S_{xx}(n, k) = \sum_{i=-L_d}^{L_d} P(i) STFT(n, k+i) STFT^*(n, k-i). \quad (52)$$

For the multicomponent signals

$$x(n) = \sum_{m=1}^M f_m(n) + \nu(n)$$

the S method can produce a sum of the WD's of each of these signal components separately

$$S_{xx}(n, \theta) = \sum_{m=1}^M W_{f_m f_m}(n, \theta) + \text{Noise} \quad (53)$$

provided that the components $f_m(n)$ of the signal $x(n)$ do not overlap in the time–frequency plane. The short-time Fourier transform, at the frequency $\theta = \pi k/N$, is denoted by $STFT(n, k)$ in (52). It is defined by $STFT(n, k) = DFT_k\{w(k)x(n+k)\}$. An appropriate rectangular window $P(i)$ eliminates (reduces) cross-terms and reduces the noise influence, as explained in detail in [26], [27], [30], and [31]. The S -method with simplified algorithm (41) is applied in order to produce a time–frequency representation of the signal

$$x(n) = Ae^{j1360(nT-0.1)^2} + Ae^{-16(nT-13/32)^2 + j375(nT+2)^2} + Ae^{j2800nT} + Ae^{-16(nT-5/8)^2 + j375(nT)^2} + \nu(n). \quad (54)$$

Using the same parameters as in the first example and the window lengths $\mathbf{N} = \{32\ 256\}$ with the rectangular window $P(i)$, which includes only two samples around the central frequency $\theta = \pi k/N$ on each side ($L_d = 2$), we get

the distribution presented in Fig. 5(c). The WD of original *nonnoisy signal* with the constant window length $N_2 = 256$ is given in Fig. 5(a), whereas Fig. 5(b) shows the WD of noisy signal with the same constant window length. The variance of (41) is estimated by techniques described in Section IV (Comments on the Algorithm) and (49). From Fig. 5(c), we may conclude that we get the distribution in accordance with (53) with a very suppressed noise influence, as expected.

VIII. CONCLUSION

The algorithm with the varying and adaptive window length is developed in order to obtain the improved time–frequency distributions of noisy signals. The algorithm is simple and uses only the estimate of the distribution and its variance. The simplified version of the algorithm, called the two-window algorithm, using only two different values of the window lengths, is proposed. The numerical and qualitative efficiency of the proposed simplified algorithm is demonstrated on the time–frequency representations of monocomponent and multicomponent signals corrupted by the noise with the very low SNR.

REFERENCES

- [1] M. G. Amin, "Minimum variance time-frequency distribution kernels for signals in additive noise," *IEEE Trans. Signal Processing*, vol. 44, pp. 2352–2356, Sept. 1996.
- [2] L. E. Atlas, Y. Zhao, and R. J. Marks, II, "The use of cone shape kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1084–1091, 1990.
- [3] B. Boashash and P. J. Black, "An efficient real-time implementation of the Wigner-Vile distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1611–1618, Nov. 1987.
- [4] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—Part 1: Fundamentals," *Proc. IEEE*, vol. 80, pp. 519–538, Apr. 1992.
- [5] H. Choi and W. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 43, pp. 862–871, June 1989.
- [6] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [7] L. Cohen, "Time-frequency distributions—A review," *Proc. IEEE*, vol. 77, pp. 941–981, July 1989.
- [8] L. Cohen and C. Lee, "Instantaneous bandwidth," in *Time-Frequency Signal Analysis*, B. Boashash, Ed. London, U.K.: Longman Cheshire, 1992.
- [9] L. Cohen, "Distributions concentrated along the instantaneous frequency" *SPIE, Adv. Signal Process. Alg., Arch., Implementations*, vol. 1348, pp. 149–157, 1990.
- [10] C. Griffin, P. Rao, and F. Taylor, "Roundoff error analysis of the discrete Wigner distribution using fixed-point arithmetic," *IEEE Trans. Signal Processing*, vol. 39, pp. 2096–2098, Sept. 1991.
- [11] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," Res. Rep., 5/94, Technion—Israel Inst. Technol., Haifa, Nov. 1994.
- [12] S. B. Hearon and M. G. Amin, "Minimum-variance time-frequency distributions kernels," *IEEE Trans. Signal Processing*, vol. 43, pp. 1258–1262, May 1995.
- [13] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representation," *IEEE Signal Processing Mag.*, pp. 21–67, Apr. 1992.
- [14] J. Jeong and W. J. Williams, "Alias-free generalized discrete-time time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 40, pp. 2757–2765, Nov. 1992.
- [15] V. Katkovnik, "Nonparametric estimation of instantaneous frequency," *IEEE Trans. Inform. Theory*, vol. 43, pp. 183–189, Jan. 1997.
- [16] V. Katkovnik, "Adaptive local polynomial periodogram for time-varying frequency estimation," in *Proc. IEEE-SP Time-Freq. Time-Scale Anal.*, Paris, France, June 1996, pp. 183–189.

- [17] V. Katkovnik and LJ. Stanković, "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length," *IEEE Trans. Signal Processing*, vol. 46, pp. 2315–2325, Sept. 1998.
- [18] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 1461–1470, Dec. 1985.
- [19] A. Papoulis, *Signal Analysis*. New York: McGraw-Hill, 1977.
- [20] D. Petranović, S. Stanković, and LJ. Stanković, "Special purpose hardware for time-frequency analysis," *Electron. Lett.*, vol. 33, pp. 464–466, Mar. 1997.
- [21] K. J. R. Liu, "Novel parallel architectures for short-time Fourier transform," *IEEE Trans. Circuits Syst.*, vol. 40, pp. 464–466, Dec. 1993.
- [22] P. Rao and F. J. Taylor, "Estimation of the instantaneous frequency using the discrete Wigner distribution," *Electron. Lett.*, vol. 26, pp. 246–248, 1990.
- [23] B. Ristic and B. Boashash, "Relationship between the polynomial and the higher order Wigner-Ville distribution," *IEEE Signal Processing Lett.*, vol. 2, pp. 227–229, Dec. 1995.
- [24] LJ. Stanković and S. Stanković, "Wigner distribution of noisy signals," *IEEE Trans. Signal Processing*, vol. 41, pp. 956–960, Feb. 1993.
- [25] ———, "On the Wigner distribution of discrete-time noisy signals with application to the study of quantization effects," *IEEE Trans. Signal Processing*, vol. 42, pp. 1863–1867, July 1994.
- [26] LJ. Stanković, "A method for time-frequency analysis," *IEEE Trans. Signal Processing*, vol. 42, pp. 225–229, Jan. 1994.
- [27] ———, "A method for improved distribution concentration in the time-frequency analysis of multicomponent signals the L -Wigner distribution" *IEEE Trans. Signal Processing*, vol. 43, pp. 1262–1268, May 1995.
- [28] ———, "Auto-term representation by the reduced interference distributions; A procedure for kernel design," *IEEE Trans. Signal Processing*, vol. 44, pp. 1557–1564, June 1996.
- [29] LJ. Stanković and V. Ivanović, "Further results on the minimum variance time-frequency distributions kernels," *IEEE Trans. Signal Processing*, vol. pp. 1650–1655, June 1997.
- [30] LJ. Stanković, V. Ivanović, and Z. Petrović, "Unified approach to the noise analysis in the Wigner distribution and spectrogram using the S -method," *Ann. des Telecomm.*, nos. 11/12, pp. 585–594, Nov./Dec. 1996.
- [31] S. Stanković and LJ. Stanković, "Architecture for the realization of a system for time-frequency analysis," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 600–604, July 1997.
- [32] LJ. Stanković and V. Katkovnik, "Algorithm for the instantaneous frequency estimation using time-frequency representations with adaptive window length," *IEEE Signal Processing Lett.*, vol. 5, pp. 224–227, Sept. 1998.
- [33] D. Wu and J. M. Morris, "Time-frequency representation using a radial Butterworth kernel," in *Proc. IEEE-SP Time-Freq. Time-Scale Anal.*, Philadelphia, PA, Oct. 1994, pp. 60–63.
- [34] D. Wu and J. M. Morris, "Discrete Cohen's class of distributions," in *Proc. IEEE-SP Time-Freq. Time-Scale Anal.*, Philadelphia, PA, Oct. 1994, pp. 532–535.
- [35] Y. M. Zhu, F. Peyrin, and R. Goutte, "Transformation de Wigner-Ville: Description d'un nouvel outil de traitement du signal et des images," *Ann. Telecomm.*, vol. 42, nos. 3–4, pp. 105–117, 1987.



Ljubiša Stanković (M'91–SM'96) was born in Montenegro, Yugoslavia, on June 1, 1960. He received the B.S. degree in electrical engineering from the University of Montenegro in 1982 with the honor "the best student at the University," the M.S. degree in electrical engineering in 1984 from the University of Belgrade, Belgrade, Yugoslavia, and the Ph.D. degree in electrical engineering in 1988 from the University of Montenegro.

As a Fulbright grantee, he spent the 1984–1985 academic year at the Worcester Polytechnic Institute, Worcester, MA. He was also active in politics as a Vice President of the Republic of Montenegro from 1989 to 1991 and then as the leader of democratic (anti-war) opposition in Montenegro from 1991 to 1993. Since 1982, he has been on the Faculty at the University of Montenegro, where he presently holds the position of a Full Professor. From 1997 to 1998, he was on leave at the Signal Theory Group, Ruhr University Bochum, Bochum, Germany, supported by the Alexander von Humboldt Foundation. His current interests are in the signal processing and electromagnetic field theory. He has published more than 100 technical papers, about 40 of them in leading international journals, mainly the IEEE editions. He has also published several textbooks in signal processing (in Serbo-Croat) and the monograph *Time-Frequency Signal Analysis* (in English).

Prof. Stanković was awarded the highest state award of the Republic of Montenegro for scientific achievements in 1997. He is a member of the National Academy of Science and Art of Montenegro (CANU).



Vladimir Katkovnik (M'96) received the M.Sc., Ph.D., and Doctor of Sc. from Leningrad Polytechnic Institute, Leningrad, Russia, in 1960, 1964, and 1974, all in technical cybernetics.

From 1961 to 1986, he held positions of Assistant, Associate, and Professor with the Department of Mechanics and Control Processes, Leningrad Polytechnic Institute. From 1981 until 1991, he was with the Department of Automatic Machines of the same institute. Since 1991, he has been a Professor of Statistics with the Department of the University of South Africa, Pretoria. His research interests include linear and nonlinear filtering, nonparametric and robust estimation, nonstationary systems, time-frequency analysis, stochastic optimization, and adaptive stochastic control. He has published five books and more than 150 papers.