

Instantaneous Frequency Estimation Using the Wigner Distribution with Varying and Data-Driven Window Length

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

Abstract—Estimation of the instantaneous frequency (IF) of a harmonic complex-valued signal with an additive noise using the Wigner distribution is considered. If the IF is a nonlinear function of time, the bias of the estimate depends on the window length. The optimal choice of the window length, based on the asymptotic formulae for the variance and bias, can be used in order to resolve the bias-variance tradeoff. However, the practical value of this solution is not significant because the optimal window length depends on the unknown smoothness of the IF. The goal of this paper is to develop an adaptive IF estimator with a time-varying and data-driven window length, which is able to provide quality close to what could be achieved if the smoothness of the IF were known in advance. The algorithm uses the asymptotic formula for the variance of the estimator only. Its value may be easily obtained in the case of white noise and relatively high sampling rate. Simulation shows good accuracy for the proposed adaptive algorithm.

I. INTRODUCTION

A complex-valued harmonic with a time-varying phase is a key model of the instantaneous frequency (IF) concept, as well as an important model in the general theory of time–frequency distributions. It has been utilized for the study of a wide range of signals, including speech, music, biological, radar, sonar, and geophysical ones [14].

An overview of methods for the IF estimation, as well as an interpretation of the IF concept itself, is presented in [2] and [6]. Beside other efficient techniques for the IF estimation (e.g., [2], [11], [15], [16]), the time–frequency distribution approach is interesting and commonly applied [2], [6]. This approach is based on the property of time–frequency distributions to concentrate the energy of a signal, in the time–frequency plane, at and around the IF [1], [2], [4], [5], [12]. Out of the general Cohen class of time–frequency distributions with a signal-independent kernel, the Wigner distribution (WD) produces the best concentration along the linear IF [4], [5], [19], [20]. In order to improve the concentration when the IF is not a

linear function of time, various higher order time–frequency representations have been introduced [3], [8], [9], [19]. Here, we will focus our attention on the WD only. If the IF is a nonlinear function of time, then its estimate, using the WD, is biased. In the case of noisy signals, this estimate is highly signal and noise dependent [8], [17], [21]. Using the asymptotic formulae for the estimation variance and bias, we can, theoretically, find the optimal window length in the WD and resolve the bias-variance tradeoff. However, the optimal window length depends on the unknown smoothness of the IF, making this approach practically useless. The main goal of this paper is to develop an adaptive estimator with a time-varying and data-driven window length that is able to provide quality close to what could be achieved if the smoothness of the IF were known in advance. The idea of the approach developed in this paper originated from [7], where it was proposed and justified for the local polynomial fitting of regression. For the time-varying IF estimation, this approach was used in [10], where the algorithm with the time-varying and data-driven window length was presented for the local polynomial periodogram. This approach uses only the formula for the variance of the estimate, which does not require information about the IF to be known in advance. Simulations based on the discrete WD, with several noisy signal examples, show a good accuracy of the presented adaptive algorithm, as well as an improvement in the time–frequency representation of signals with a nonlinear IF.

The structure of the paper is as follows: The WD, as an IF estimator, is considered in Section II. The asymptotic bias and variance of the IF estimate, along with the optimal window size for the IF estimation, are also presented in Section II. The adaptive estimation of the IF with a time-varying and data-driven window length is developed in Section III. A numerical implementation of the adaptive algorithm, along with simulation results, is discussed in Section IV. Proofs are given in the Appendix.

II. BACKGROUND THEORY

Consider the problem of the IF estimation from the discrete-time observations

$$y(nT) = m(nT) + \varepsilon(nT)$$

with

$$m(t) = Ae^{j\phi(t)} \quad (1)$$

Manuscript received May 9, 1997; revised February 5, 1998. The work of LJ. Stanković was supported by the Alexander von Humboldt Foundation and the Montenegrin Academy of Science and Art. The associate editor coordinating the review of this paper and approving it for publication was Dr. Frans M. Coetzee.

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

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

Publisher Item Identifier S 1053-587X(98)05954-6.

where

- n integer;
- T sampling interval;
- $\varepsilon(nT)$ complex-valued white Gaussian noise with i.i.d. real and imaginary parts.

Thus, $\text{Re}(\varepsilon(nT))$ and $\text{Im}(\varepsilon(nT)) \sim \mathcal{N}(0, \sigma_\varepsilon^2/2)$ and the total variance of the noise is equal to σ_ε^2 . The IF, by definition, is the first derivative of the phase

$$\omega(t) = \phi'(t). \quad (2)$$

In the nonparametric setting of the problem, it is assumed that $\omega(t)$ is an arbitrary smooth differentiable function of time, with bounded derivatives $|\omega^{(r)}(t)| = |\phi^{(r+1)}(t)| \leq M_r(t)$.

The Pseudo-Wigner distribution (WD) in the discrete-time domain is defined by

$$W_h(\omega, t) = \sum_{n=-\infty}^{\infty} w_h(nT)y(t+nT)y^*(t-nT)e^{-j2\omega nT} \quad (3)$$

where $w_h(nT) = T/h \cdot w(nT/h)$, with $w(t)$ being a real-valued symmetric window, $w(t) = w(-t)$. The window $w_h(nT)$, whose width is denoted by $h > 0$, is used in (3) for the localization of the estimate. We also assume that $w(t)$ has a finite length, i.e., $w(t) = 0$, for $|t| > 1/2$.

As the WD (3) is a real-valued periodic function, the IF estimate is a solution of the optimization problem

$$\hat{\omega}_h(t) = \arg \left[\max_{\omega \in Q_\omega} W_h(\omega, t) \right] \quad (4)$$

where $Q_\omega = \{\omega: 0 \leq |\omega| < \pi/2T\}$ is a basic interval along the frequency axis.

The estimation error, at a time-instant t , is defined as

$$\Delta\hat{\omega}_h(t) = \omega(t) - \hat{\omega}_h(t). \quad (5)$$

The following proposition gives the asymptotic formulae for the variance and bias of the IF estimator (4).

Proposition: Let $\hat{\omega}_h(t)$ be a solution of (4), with $h \rightarrow 0$, $T \rightarrow 0$, and $h^3/T \rightarrow \infty$. Then, the variance of the IF estimate is given by

$$\text{var}(\Delta\hat{\omega}_h(t)) = \frac{\sigma_\varepsilon^2}{2|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2} \right) \frac{T}{h^3} \frac{E}{F^2} \quad (6)$$

whereas the bias of the estimate can be represented in the forms

$$E(\Delta\hat{\omega}_h(t)) = \sum_{s=1}^{\infty} h^{2s} b_s \omega^{(2s)}(t) \quad (7)$$

and

$$|E(\Delta\hat{\omega}_h(t))| \leq h^2 |b_1| M_2, \quad \text{with } M_2 = \sup_t \omega^{(2)}(t) \quad (8)$$

provided that $h^2 |b_1| M_2$ and $h^{2s} |b_s \omega^{(2s)}(t)|$ are small for all $s \geq 1$.

The notation

$$\begin{aligned} F &= \int_{-1/2}^{1/2} w(t)t^2 dt \\ E &= \int_{-1/2}^{1/2} w^2(t)t^2 dt \\ b_s &= \frac{1}{(2s+1)!F} \int_{-1/2}^{1/2} w(t)t^{2s+2} dt \end{aligned} \quad (9)$$

has been used in (6)–(8).

A proof of the proposition is given in the Appendix.

Comments:

- 1) The main assumption for formulae (6)–(8) is that the variance and the bias are small (see the Appendix). If the window is rectangular and $h \rightarrow 0$, $T \rightarrow 0$, and $h^3/T \rightarrow \infty$, then the following explicit formulae can be given for the variance and the error-coefficients:

$$\begin{aligned} \text{var}(\Delta\hat{\omega}_h(t)) &= \frac{6\sigma_\varepsilon^2}{|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2} \right) \frac{T}{h^3} \\ E &= F = 1/12 \\ b_s &= \frac{3}{(2s+1)!(2s+3)2^{2s}}. \end{aligned} \quad (10)$$

Note that the values in (10) may be obtained as special cases of the formulae for variance and bias given in [8] for the local polynomial WD.

- 2) Let us consider the mean squared error (MSE) of the IF estimate. It follows from (6) and (7) that, for small h , the main term of the MSE can be given in the form

$$\begin{aligned} E(\Delta\hat{\omega}_h(t))^2 &= \frac{6\sigma_\varepsilon^2}{|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2} \right) \frac{T}{h^3} \\ &+ \left(\frac{1}{40} h^2 \omega^{(2)}(t) \right)^2. \end{aligned} \quad (11)$$

From (11), it is clear that increasing the window length h increases the bias and decreases the variance. The optimization of h in (11), minimizing the MSE, results in

$$h_0(t) = \left[\frac{7200\sigma_\varepsilon^2 T \left(1 + \frac{2\sigma_\varepsilon^2}{|A|^2} \right)}{|A|^2 (\omega^{(2)}(t))^2} \right]^{1/7}. \quad (12)$$

This optimal window length $h_0(t)$ gives the optimal bias-variance tradeoff, which is usual for nonparametric estimation, depending on the signal-to-noise ratio $|A|/\sigma_\varepsilon$, the sampling interval T , and the second IF derivative $\omega^{(2)}(t)$. Thus, the optimal choice of length h depends on the IF second derivative $\omega^{(2)}(t)$. This derivative is unknown because the IF itself is to be estimated. Note that if the second derivative $\omega^{(2)}(t)$ is significantly different for different t , then the optimization of the estimation accuracy requires a time-varying window length $h(t)$. The improvement in accuracy from this time-varying $h(t)$ can be significant when compared with any time-invariant length h .

In this paper, we propose a data-driven choice of the time-varying window length that uses only the formula for the variance and does not need information about the unknown derivatives of the IF. Moreover, this choice of the window length is based on specific statistics that are used in order to compare the variance of the estimate versus its bias.

III. ALGORITHM FOR THE DATA-DRIVEN WINDOW LENGTH CHOICE

A. Basic Idea

The basic idea follows from the accuracy analysis, which was given in the proposition. For the asymptotic case, when the estimation error is small, it can be represented as a sum of the deterministic (bias) and random component, with the variance given by (6). The estimation error can be written as

$$\begin{aligned} |\omega(t) - \hat{\omega}_h(t)| &\leq |\text{bias}(t, h)| + \kappa\sigma(h) \\ \sigma^2(h) &= \text{var}(\Delta\hat{\omega}_h(t)) \\ &= \frac{6\sigma_\varepsilon^2}{|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2}\right) \frac{T}{h^3} \end{aligned} \quad (13)$$

where the inequality holds with probability $P(\kappa)$, where κ is the corresponding quantile of the standard Gaussian distribution. Thus, $P(\kappa) \rightarrow 1$ as κ increases.

It follows from (7) and (8) that $|\text{bias}(t, h)| \rightarrow 0$ as $h \rightarrow 0$.

Now, let $h = h_s$ be so small that

$$|\text{bias}(t, h_s)| \leq \kappa\sigma(h_s). \quad (14)$$

It follows from (13) that with probability $P(\kappa)$

$$|\omega(t) - \hat{\omega}_{h_s}(t)| \leq 2\kappa\sigma(h_s) \quad (15)$$

i.e.,

$$\hat{\omega}_{h_s}(t) - 2\kappa\sigma(h_s) \leq \omega(t) \leq \hat{\omega}_{h_s}(t) + 2\kappa\sigma(h_s). \quad (16)$$

Let us introduce the confidence intervals

$$D_s = [\hat{\omega}_{h_s}(t) - 2\kappa\sigma(h_s), \hat{\omega}_{h_s}(t) + 2\kappa\sigma(h_s)]. \quad (17)$$

Then, we can say that for a set of h_s , which are so small that (14) and (15) are true, all of the segments D_s have a point in common, namely, $\omega(t)$.

Consider an increasing sequence of h_s , $h_1 < h_2 < \dots < h_J$. Let h_{s^+} be the largest of those h_s for which the segments D_{s-1} and D_s have a point in common. Let us call this window length h_{s^+} “optimal.” We propose to use this optimal h_{s^+} as a reasonable choice of the window length h . Then, the IF estimate with the data-driven adaptive window length is determined as $\hat{\omega}_{h_{s^+}}(t)$.

The idea behind this choice is clear. If the segments D_{s-1} and D_s do not have a point in common, then at least one of the inequalities (15) does not hold, i.e., the bias is too large as compared with the standard deviation in (14). Thus, the statistical hypothesis to be tested for the bias is given in the form of the sequence of inequalities (15). The largest length h_s for which these inequalities have a point in common is considered as a bias-variance compromise, where the bias and standard deviation are of the same order. For more details, see [22].

B. Algorithm

Let us initially assume that the amplitude of signal A and the standard deviation of noise σ_ε are known. Let H be an increasing sequence of the window length values

$$H = \{h_s | h_1 < h_2 < h_3 < \dots < h_J\}. \quad (18)$$

In general, any reasonable choice of H is acceptable. The relation $h_s = 2N_s T$ gives a link between the window length h_s and the number of observations N_s . Here, the lengths corresponding to the dyadic numbers $N_s = 2N_{s-1}$ will be used.

For every t , the following steps are generated.

- 1) The WD is calculated for all of $h_s \in H$. Thus, we obtain a set of distributions for a fixed time instant t

$$\{W_{h_s}(\omega, t)\}, \quad h_s \in H$$

and the IF estimates are found as

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

- 2) The upper and lower bounds of the confidence intervals D_s in (17) are built as

$$\begin{aligned} U_s(t) &= \hat{\omega}_{h_s}(t) + 2\kappa\sigma(h_s) \\ L_s(t) &= \hat{\omega}_{h_s}(t) - 2\kappa\sigma(h_s) \end{aligned} \quad (19)$$

where, according to (13)

$$\sigma(h_s) = \sqrt{\frac{6\sigma_\varepsilon^2}{|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2}\right) \frac{T}{h_s^3}}. \quad (20)$$

For the sake of simplicity, we will use $\kappa = 2$ (or the values just slightly less than this value) in (19), which, for the Gaussian estimation error, corresponds to the standard two-sigma rule with the probability $P(2) = 0.95$ of inequality (13).

- 3) The optimal window length h_{s^+} is determined as the largest s ($s = 1, 2, \dots, J$) when

$$[L_{s-1}(t), U_{s-1}(t)] \cap [L_s(t), U_s(t)] \neq \emptyset \quad (21)$$

is still satisfied, i.e.,

$$|\hat{\omega}_{h_s}(t) - \hat{\omega}_{h_{s-1}}(t)| \leq 2\kappa[\sigma(h_s) + \sigma(h_{s-1})]$$

still holds. Here, $L_s(t)$ and $U_s(t)$ are determined by (19). 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 optimal window length is chosen as

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

and $\hat{\omega}_{\hat{h}(t)}(t)$ is the adaptive IF estimator with the data-driven window for a given instant t .

- 4) The WD with the optimal window length is

$$W^+(\omega, t) = W_{\hat{h}(t)}(\omega, t). \quad (22)$$

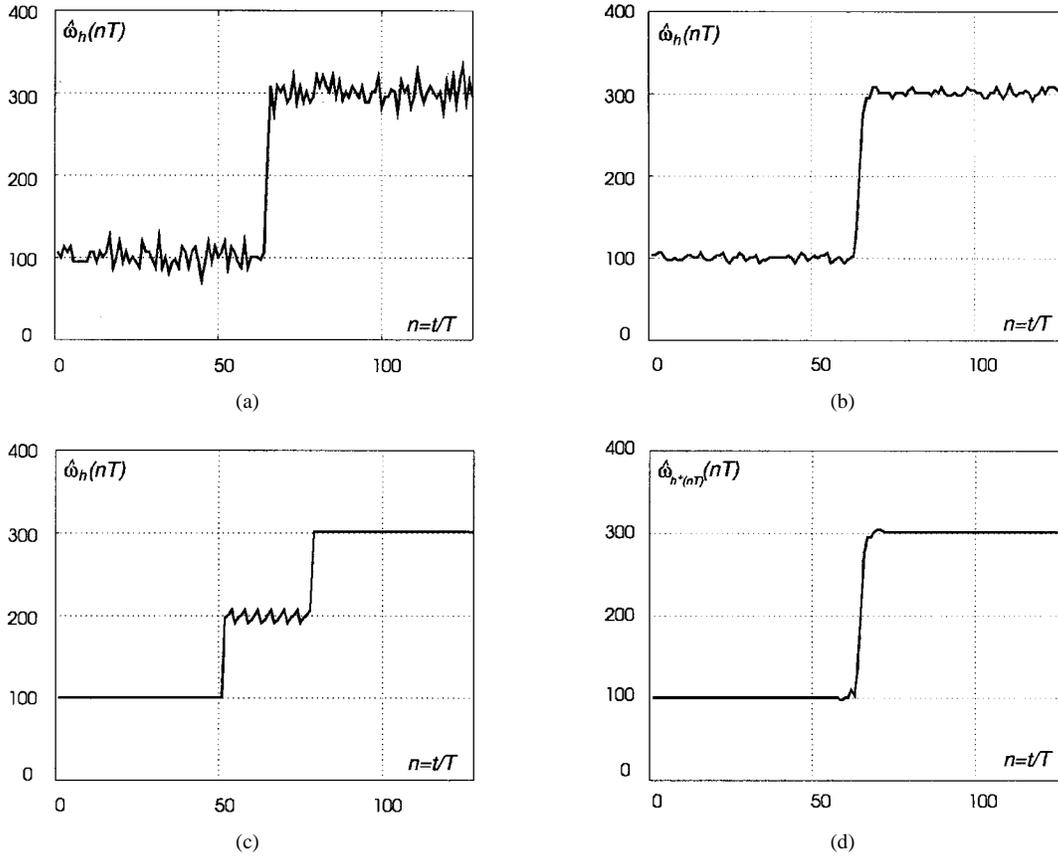


Fig. 1. Estimation of the step-wise IF. (a) Estimation with the window length $2N_s = 8$. (b) Estimation with the window length $2N_s = 16$. (c) Estimation with the window length $2N_s = 256$. (d) IF estimation using the adaptive window length.

Steps 1–4 are repeated for every time instant t .

Comments on the Algorithm:

- a) Provided that T is small, the following estimates of A and σ_ε^2 can be used:

$$\hat{A}^2 + \hat{\sigma}_\varepsilon^2 = \frac{1}{N} \sum_{n=1}^N |y(nT)|^2 \quad (23)$$

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

$$\begin{aligned} \hat{\sigma}_{\varepsilon r} &= \frac{\{\text{median}(|y_r(nT) - y_r((n-1)T)| : n = 2, \dots, N)\}}{0.6745} \\ \hat{\sigma}_{\varepsilon i} &= \frac{\{\text{median}(|y_i(nT) - y_i((n-1)T)| : n = 2, \dots, N)\}}{0.6745} \\ \hat{\sigma}_\varepsilon^2 &= (\hat{\sigma}_{\varepsilon r}^2 + \hat{\sigma}_{\varepsilon i}^2)/2 \end{aligned} \quad (24)$$

where $y_r(nT)$ and $y_i(nT)$ are the real and imaginary parts, respectively, of $y(nT)$, and T is sufficiently small. The average $(1/2N) \sum_{n=1}^N |y(nT) - y((n-1)T)|^2$ could also be used as an estimate of σ_ε^2 . However, we prefer the median in (24) as a robust estimate.

- b) We wish to emphasize that the intersection of the segments D_s is a specific method for the bias examination. Originally, the intersection of the confidence intervals D_s for the window length selection has been proposed in [7], where it was used for the local polynomial fitting of regression.

In contrast to the algorithm described above, the algorithm given in [7] is based on the intersection of all of the segments D_s starting from $s = 1$ up to the largest s^+ , when all of those segments have a point in common. The strong convergence results given in [7] prove that the estimator with the proposed time-varying and data-driven window length is able to provide the quality close to the one that could be achieved if the smoothness of the function to be estimated was known in advance.

In our algorithm, we use only the sliding pair-wise intersections (21) of the pairs of segments D_s and D_{s-1} for $s = 2, 3, \dots, J$. The simulation shows that the results for the IF estimation are accurate and promising. Simulation produced for the algorithm [7], using the intersection of all confidence intervals as in the regression estimation, showed worse accuracy, compared with our algorithm.

IV. NUMERICAL IMPLEMENTATION AND EXAMPLES

The discrete-time WD given by (3) is discretized over the frequency and calculated as

$$W_h(k, l) = \sum_{n=-N_s}^{N_s-1} w_h(nT) y(lT + nT) y^*(lT - nT) \cdot e^{-j2(\pi/2N_s)kn} \quad (25)$$

where $2N_s$ is the number of samples determined by the sampling interval T [i.e., signal's maximal expected frequency

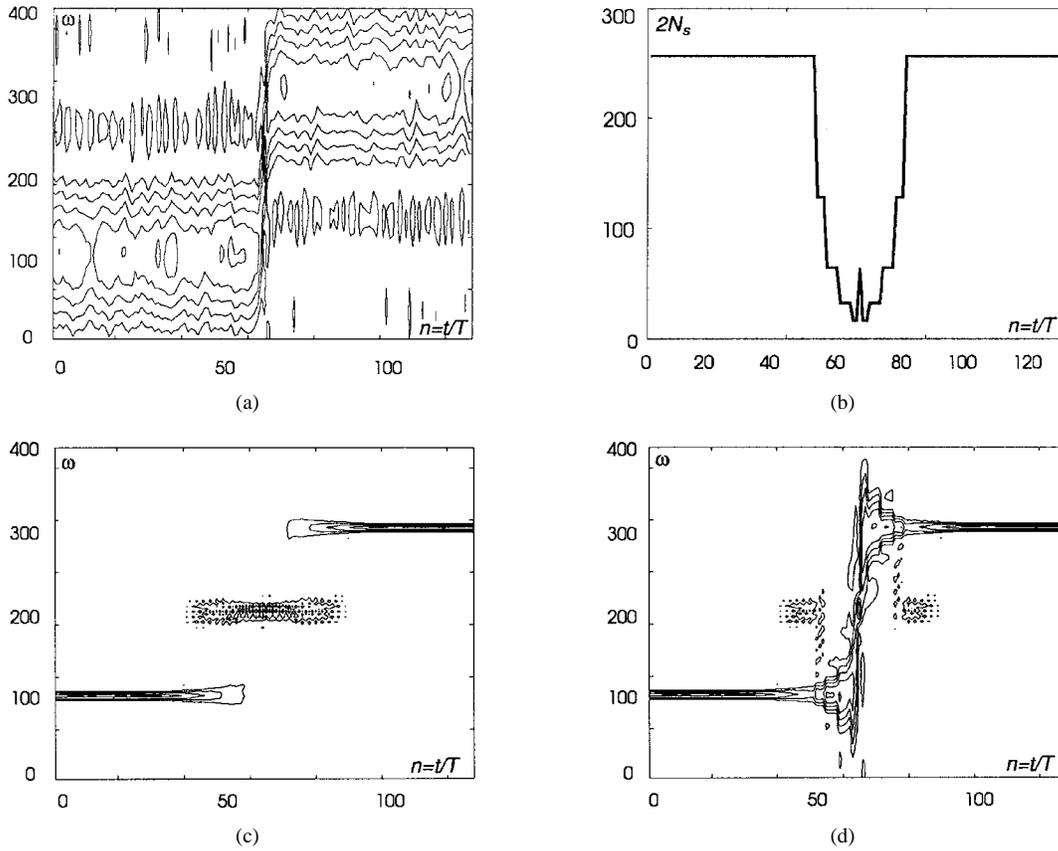


Fig. 2. Wigner distribution of the signal with a step-wise IF. (a) WD with the window length $2N_s = 16$. (b) Adaptive window length as a function of time $n = t/T$. (c) WD with the window length $2N_s = 256$. (d) the WD with the adaptive window length.

$\omega_m = \pi/(2T)]$ and the window length h_s , $h_s = 2N_s T$. For the dyadic window lengths $h_s \in H$, we have corresponding $N_s \in N_H = \{N_s | N_s = 2N_{s-1}\}$. Starting with a narrow window, for example $N_1 = 4$, we calculate $W_h(k, l)$ for all $N_s \in \{4, 8, 16, 32, 64, 128\}$ up to the widest window (in this case $N_J = 128$). For a given instant lT , the IF is estimated as

$$\hat{\omega}_h(lT) = \frac{\pi}{2N_s T} \cdot \arg \left[\max_{k \in Q_k} W_h(k, l) \right] \quad (26)$$

where $Q_k = \{k: 0 \leq k < N_s\}$ for non-negative-only IF values. In order to reduce the FFT quantization error in the cases of small N_s , as well as to have distributions of the same length for different N_s , the product $y(lT+nT)y^*(lT-nT)$ is zero padded up to $2N_J$ length. That is, $W_h(k, l)$ is interpolated in the frequency domain up to $2N_J$ samples within the period. Then, the discrete WD given by (25) is calculated using the standard FFT routines, and

$$\hat{\omega}_h(lT) = \frac{\pi}{2N_J T} \cdot \arg \left[\max_{k \in Q_k} W_h(k, l) \right].$$

The quantization error has the variance $\pi^2/12(\pi/2N_J T)^2 = \pi^2/(12h_J^2)$. This variance can be even larger than the variance in (6). In order to reduce it, we may additionally interpolate the distribution $W_h(k, l)$. In the numerical realizations, we did this interpolation with a factor of 2.

After h_{s+} is found, according to the IF estimate (26) and the algorithm (19)–(22), the calculation for the time-instant $t = lT$ can be stopped in order to save computational time. Note also

TABLE I
NORMALIZED MEAN ABSOLUTE ERROR FOR THE STEP-WISE CONSTANT IF

$N_s = 4$	$N_s = 8$	$N_s = 16$	$N_s = 32$	$N_s = 64$	$N_s = 128$	Adaptive N_s
1.4883	0.5078	0.3750	0.8398	1.7266	3.2344	0.1250

that for each next (twice wider) window length, we simply replace $2N_{s-1} = N_s$ zero values (added by zero padding) by the real values of $y(lT+nT)y^*(lT-nT)$. Therefore, the computations with various window lengths are not completely independent, which may be a source of an additional savings in computation time.

The algorithm is tested on several examples. For each of these, we plotted $2N_s = \hat{h}(nT)/T$ for the adaptive time-varying window as a function of time instant n and presented the IF estimates for the various time-invariant and adaptive time-varying window lengths. Furthermore, the mean absolute errors of the estimations are given for all of the considered cases. In all examples, we assumed the signal of the form

$$y(nT) = A \exp(j\phi(nT)) + \varepsilon(nT)$$

with the IF $\omega(nT)$ and the phase $\phi(nT) = T \cdot \sum_{i=0}^n \omega(iT)$. We also assumed that $A = 1$ and $20 \log(A/\sigma_\varepsilon) = 15$ (in decibels) ($A/\sigma_\varepsilon = 5.62$). The estimation time interval was $0 \leq nT \leq 1$. A value 1.75 of the parameter κ , which is slightly less than $\kappa = 2$, turned out to be a good choice for the practical realizations.

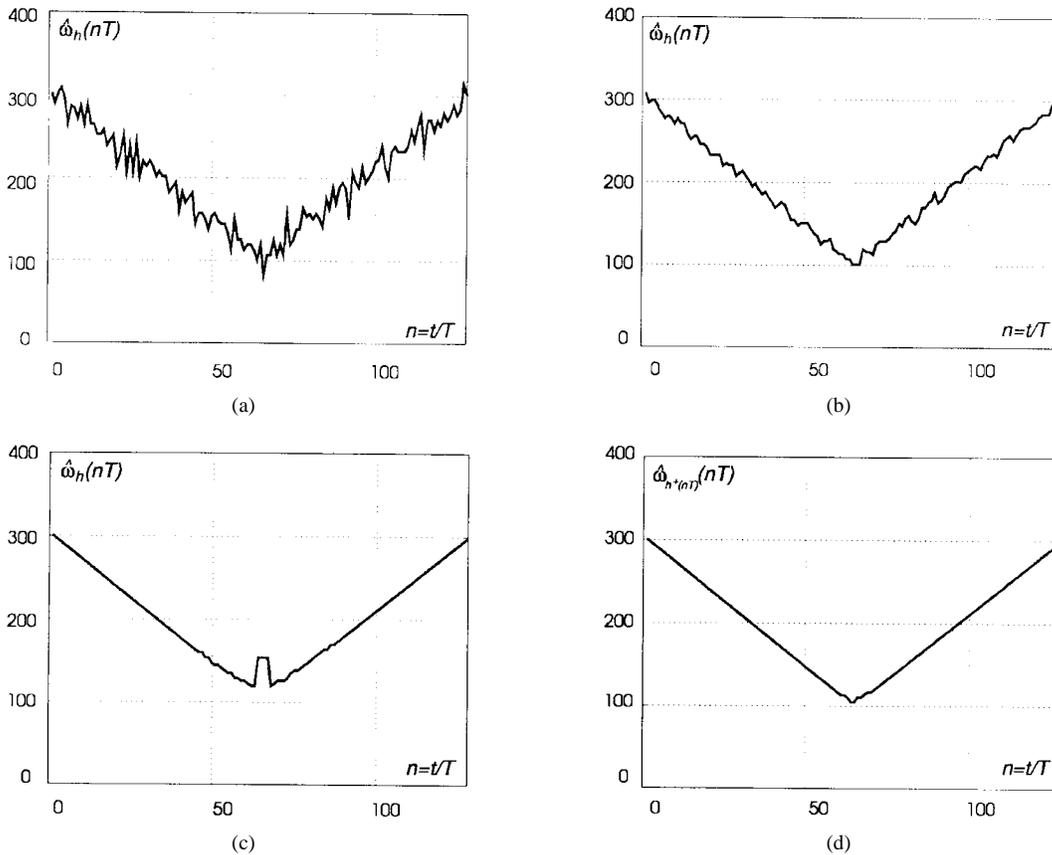


Fig. 3. Estimation of the piecewise linear IF. (a) Estimation with the window length $2N_s = 8$. (b) Estimation with the window length $2N_s = 16$. (c) Estimation with the window length $2N_s = 256$. (d) IF estimation using the adaptive window length.

Example 1: Piecewise constant IF with a step at the instant $nT = 0.5$

$$\omega(nT) = 64\pi + 32\pi \text{sign}(nT - 0.5) \quad (27)$$

is considered in this example. The estimates $\hat{\omega}_h(nT)$ of the IF, with $N_s = 4$, $N_s = 8$, and the widest window length $N_s = 128$, are presented in Fig. 1(a)–(c), respectively. The IF estimate $\hat{\omega}_{h^*(t)}(t)$ with the adaptive, time-varying window length is given in Fig. 1(d). The adaptive window length $2N_s = h_{s^*}(nT)/T$, as a function of time instant n , is presented in Fig. 2(b). We may conclude that at times far from the point of the jump ($t = 0.5$), the adaptive window definitely tends to a maximum size. However, at and around the jump-point $t = 0.5$, the window is narrowed up to the smallest length value. Thus, it demonstrates that the data-driven window length is really sensitive with respect to a fast variation of the IF. These results are in complete agreement with (12). When the IF is fast varying, then the bias is dominant, and the narrowest window is selected by the algorithm. Fig. 1(d) shows that the IF estimate $\hat{\omega}_{h^*(t)}(t)$ gives an accurate IF estimation for all time instants $n = t/T$. The mean absolute error for the adaptive estimate $\hat{\omega}_{h^*(t)}(t)$ and for the estimates with the time-invariant window lengths $N_s = [4, 8, 16, 32, 64, 128]$ is given in Table I. Note that the mean absolute error in Table I is normalized by the frequency step $\pi/(2N_s T)$.

The WD itself, for the constant window lengths with $N_s = 8$, $N_s = 128$, and for the adaptive window length, is shown

TABLE II
NORMALIZED MEAN ABSOLUTE ERROR FOR THE PIECEWISE LINEAR IF

$N_s = 4$	$N_s = 8$	$N_s = 16$	$N_s = 32$	$N_s = 64$	$N_s = 128$	Adaptive N_s
1.6484	0.5234	0.1719	0.0430	0.1953	0.5859	0.0352

in Fig. 2(a), (c), and (d), respectively. We see that in the case of the adaptive window length, the WD is highly concentrated along the instantaneous frequency within the entire considered time interval.

Example 2: The results of the IF estimation for the piecewise linear frequency modulated signal with a step in the first derivative of the IF at the instant $nT = 0.5$

$$\omega(nT) = 128\pi|nT - 0.5| + 32\pi \quad (28)$$

are shown in Fig. 3 and Table II.

Here, the explanations are similar to those given for Example 1. Note that in these two examples, we assumed the IF values that mainly belong to the discrete time–frequency grid (after the described interpolations). Therefore, the quantization error was negligible. The WD's with constant and adaptive window lengths are presented in Fig. 4.

Example 3: For the nonlinear frequency modulated signal

$$\omega(nT) = 10\pi a \sinh(100(nT - 0.5)) + 64\pi \quad (29)$$

the IF estimate is presented in Fig. 5 and Table III. The effects, which are similar to those in Figs. 1 and 3, appear in this

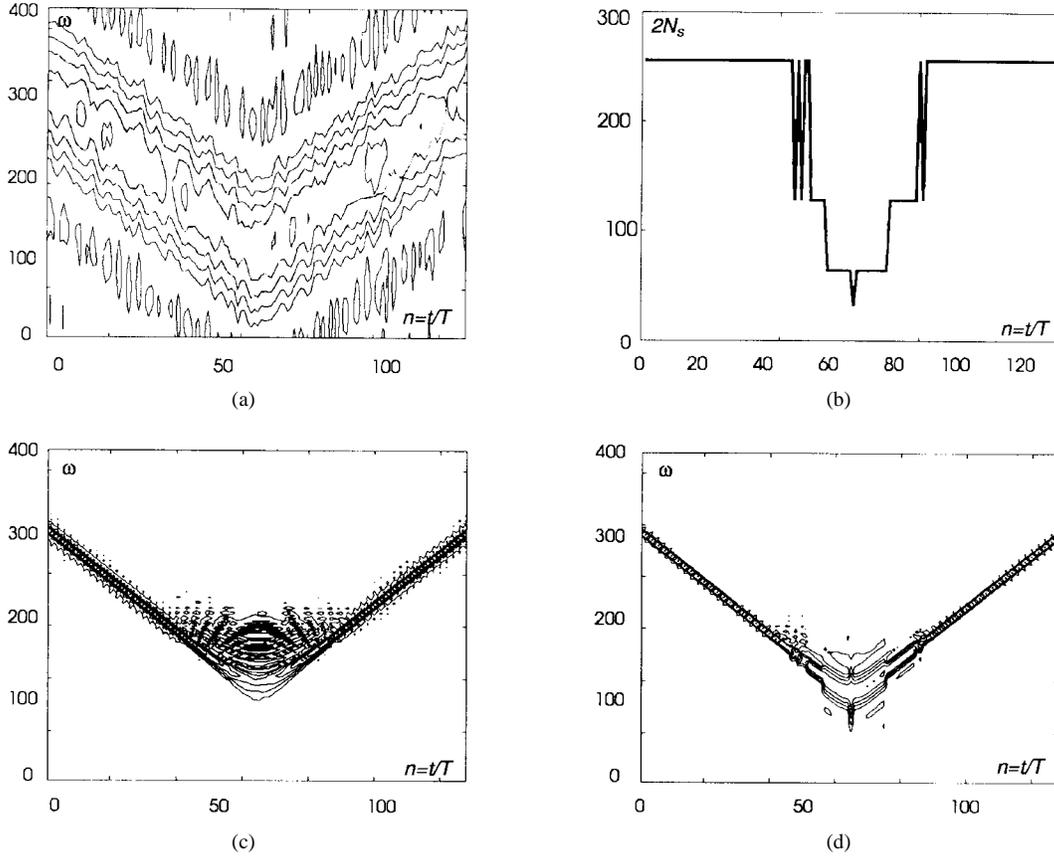


Fig. 4. Wigner distribution of the signal with a piecewise linear IF. (a) WD with the window length $2N_s = 16$. (b) Adaptive window length as a function of time $n = t/T$. (c) WD with the window length $2N_s = 256$. (d) WD with the adaptive window length.

example, although not in such a clear form. Note that the normalized minimal mean absolute error is greater than in the previous two examples. This result is due to the discretization error, which appears here in full scale. This was not the case in Examples 1 and 2, where the special IF values were used. The quantization error may be further decreased by the additional interpolation. The WD's and adaptive window length are given in Fig. 6.

V. GENERALIZATION

The presented approach and algorithm may be directly extended to the form of the discrete-time WD [8] in

$$W_h(\omega, t) = \sum_{n=-\infty}^{\infty} w_h(nT)y(t + \chi nT)y^*(t - \chi nT)e^{-j2\omega\chi nT} \quad (30)$$

where $\chi \geq 1$ is a constant.

The parameter χ can be used for an additional improvement of estimation. For the rectangular window and conditions similar to those considered in (11), the MSE of the IF estimate is of the form [8]

$$E(\Delta\hat{\omega}_h(t))^2 = \frac{6\sigma_\varepsilon^2}{|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2}\right) \frac{T}{\chi^2 h^3} + \left(\frac{\omega^{(2)}(t)}{40}(\chi h)^2\right)^2. \quad (31)$$

This expression may be derived in a straightforward manner, following the derivations presented in the Appendix. The simulation shows that a significant accuracy improvement can be achieved by the time-varying and data-driven choice of h and χ (i.e., when both of them are used in the optimization).

The following modification of the technique given in Section III was developed for this problem. Let us introduce two sets of values of the window length h and the parameter χ as in

$$\begin{aligned} H &= \{h_s \mid h_1 < h_2 < h_3 < \dots < h_J\} \\ \Gamma &= \{\chi_r \mid \chi_1 < \chi_2 < \dots < \chi_K\}. \end{aligned} \quad (32)$$

Note that in order to avoid the aliasing effects in (30), the sampling interval should be chosen such that $\omega_m = \pi/(2\chi_K T)$, where ω_m is the maximal frequency in the signal's spectrum. Note that consequently, the case with $\chi = \chi_K$ will correspond to the WD (Section IV).

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

$$\Phi = \{g_q = 1/(\chi_r^2 h_s^3) \mid g_1 \geq g_2 \geq \dots \geq g_{JK}\}. \quad (33)$$

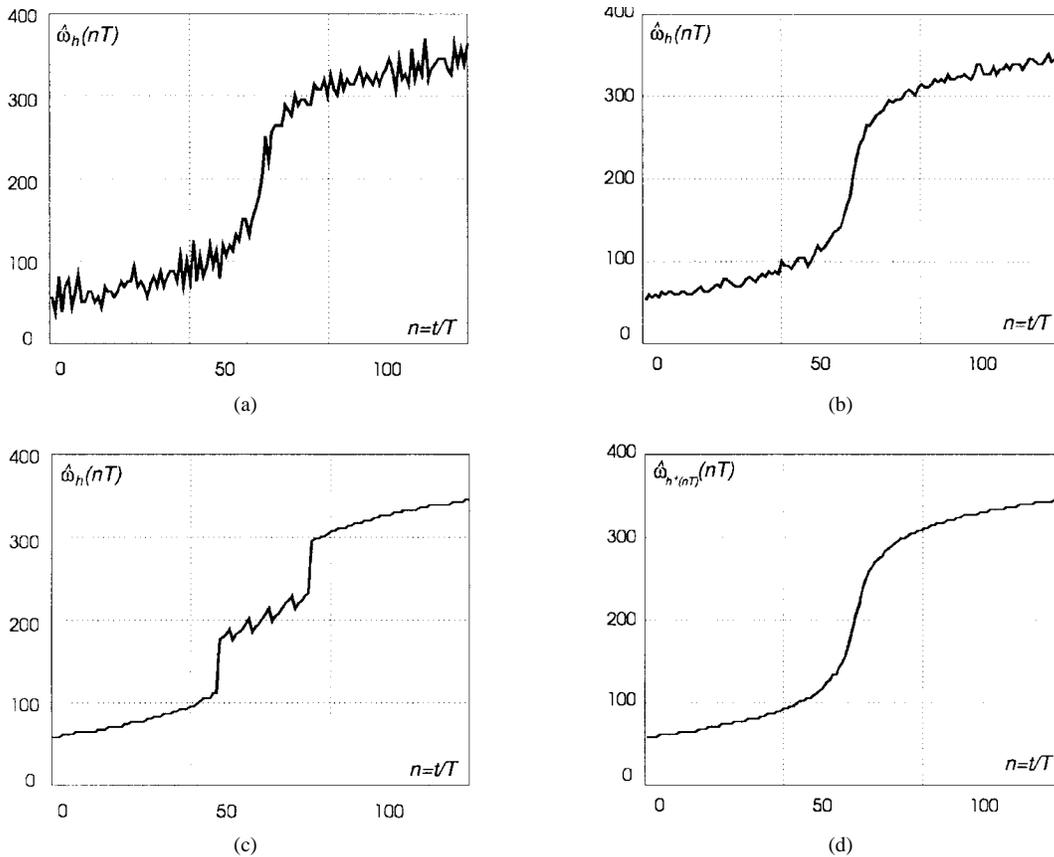


Fig. 5. Estimation of the nonlinear IF. (a) Estimation with the window length $2N_s = 8$. (b) Estimation with the window length $2N_s = 16$. (c) Estimation with the window length $2N_s = 256$. (d) IF estimation using the adaptive window length.

TABLE III
NORMALIZED MEAN ABSOLUTE ERROR FOR THE NONLINEAR IF

$N_s = 4$	$N_s = 8$	$N_s = 16$	$N_s = 32$	$N_s = 64$	$N_s = 128$	Adaptive N_s
1.5311	0.5444	0.2451	0.3220	0.7532	2.2509	0.2009

In this case, the standard deviations

$$\sigma(g_q) = \sqrt{\frac{6\sigma_\varepsilon^2}{|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2}\right) g_q T} \quad (34)$$

also form a decreasing sequence.

From the basic idea presented in Section III, it is clear that the sequence g_q can be used instead of the window length sequence h_s without any other changes in the algorithm. Thus, the algorithm described in Section III can be used in a straightforward manner. The confidence intervals $D_q = [L_q(t), U_q(t)]$, as in (19), are found for the sequence g_q , and the largest q for which $D_{q-1} \cap D_q \neq \emptyset$, $q = 2, 3, \dots$, JK still holds gives the optimal adaptive boundary value q^+ , which, according to (33), determines a corresponding pair (h_{s^+}, χ_{r^+}) .

Let us demonstrate the accuracy improvement when both h and χ are used in the optimization. For the IF considered in Example 1, the adaptive values for the pairs $2N_s = h_{s^+}(nT)/T$ and $\chi_{r^+}(nT)$ are presented in Fig. 7(a) and (b). The values of the mean absolute error for different N_s and

parameter $\chi_r \in \{1, 2, 4\}$ showing an additional improvement of the mean absolute error are presented in Fig. 7(c).

In the discrete realization, in addition to the already described interpolation with respect to the window length, an interpolation by factor χ_K is done in order to reduce the quantization error and to have the same number of samples within the basic frequency period.

VI. CONCLUSION

The WD with a data-driven and time-varying window length is developed as an adaptive estimator of the IF. The choice of the window length is based on the intersection of the confidence intervals of the IF estimates with the increasing window lengths. The developed algorithm uses only the formula for the variance of the estimate obtained for the relatively high sampling rate and white noise. Simulations show a good accuracy ability of the adaptive algorithm.

APPENDIX

Proof of the Proposition: Let us assume that the estimate $\hat{\omega}_h(t)$ in (4) is an interior point of Q_ω . The stationary point of $W_h(t, \omega)$ is determined by the zero value of the derivative of $W_h(t, \omega)$, which is given as

$$\frac{\partial W_h(t, \omega)}{\partial \omega} = \sum_{n=-\infty}^{\infty} w_h(nT) y(t+nT) y^*(t-nT) (-j2nT) \cdot e^{-j2nT\omega} \quad (A.1)$$

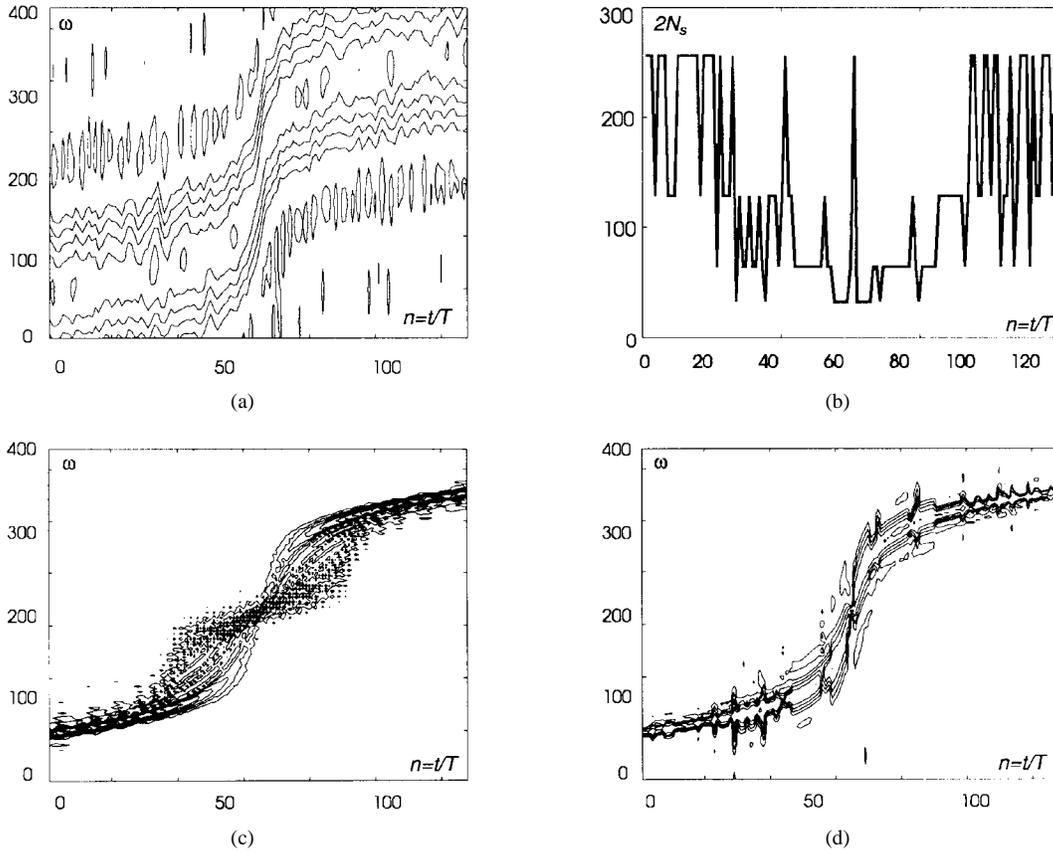


Fig. 6. Wigner distribution of the signal with a nonlinear IF. (a) WD with the window length $2N_s = 16$. (b) Adaptive window length as a function of time $n = t/T$. (c) WD with the window length $2N_s = 256$. (d) WD with the adaptive window length.

Note that

$$\begin{aligned} y(t+nT)y^*(t-nT) &= m(t+nT)m^*(t-nT) + m(t+nT)\varepsilon^*(t-nT) \\ &\quad + \varepsilon(t+nT)m^*(t-nT) + \varepsilon(t+nT)\varepsilon^*(t-nT) \end{aligned}$$

where the main deterministic term is of the form

$$m(t+nT)m^*(t-nT) = |A|^2 e^{j(\phi'(t) \cdot 2nT + \Delta\phi(t, nT))}$$

because $\phi(t+nT) - \phi(t-nT) = \phi'(t) \cdot 2nT + \Delta\phi(t, nT)$. The Taylor series of $\phi(t+nT)$ and $\phi(t-nT)$ around t can be used to get two different representations for $\Delta\phi(t, nT)$

$$\Delta\phi(t, nT) = 2 \sum_{s=1}^{\infty} \omega^{(2s)}(t) \cdot \frac{(nT)^{2s+1}}{(2s+1)!},$$

and

$$\Delta\phi(t, nT) = \frac{(nT)^3}{3!} [\omega^{(2)}(t+\tau_1) + \omega^{(2)}(t-\tau_2)] \quad (\text{A.2})$$

where $0 \leq \tau_1, \tau_2 \leq nT$.

The linearization of $\partial W_h(t, \omega)/\partial\omega = 0$ with respect to the small

- 1) estimation error $\Delta\hat{\omega}_h(t)$;
- 2) residual of the phase deviation $\Delta\phi$;
- 3) noise ε ;

gives

$$\begin{aligned} \frac{\partial W_h(t, \omega)}{\partial\omega} \Big|_0 + \frac{\partial^2 W_h(t, \omega)}{\partial\omega^2} \Big|_0 \Delta\omega \\ + \frac{\partial W_h(t, \omega)}{\partial\omega} \Big|_0 \delta_{\Delta\phi} + \frac{\partial W_h(t, \omega)}{\partial\omega} \Big|_0 \delta_\varepsilon = 0 \quad (\text{A.3}) \end{aligned}$$

where $|_0$ means that the corresponding derivatives are calculated at the point $\omega = \phi'(t)$, $\Delta\phi(t, nT) = 0$, and $\varepsilon(nT) = 0$. The terms $\partial W_h(t, \omega)/\partial\omega|_0 \delta_{\Delta\phi}$ and $\partial W_h(t, \omega)/\partial\omega|_0 \delta_\varepsilon$ in (A.3) determine the variations of $\partial W_h(t, \omega)/\partial\omega$ caused by $\Delta\phi$ and ε , respectively.

The elements of (A.3) are

$$\begin{aligned} \frac{\partial W_h(t, \omega)}{\partial\omega} \Big|_0 &= |A|^2 \sum_{n=-\infty}^{\infty} w_h(nT) (-j2nT) \\ \frac{\partial^2 W_h(t, \omega)}{\partial\omega^2} \Big|_0 &= -|A|^2 \sum_{n=-\infty}^{\infty} w_h(nT) (2nT)^2 \\ \frac{\partial W_h(t, \omega)}{\partial\omega} \Big|_0 \delta_{\Delta\phi} &= |A|^2 \sum_{n=-\infty}^{\infty} w_h(nT) \Delta\phi(t, nT) (2nT) \\ \frac{\partial W_h(t, \omega)}{\partial\omega} \Big|_0 \delta_\varepsilon &= \sum_{n=-\infty}^{\infty} w_h(nT) [\varepsilon(t+nT)\varepsilon^*(t-nT) \\ &\quad + \varepsilon(t+nT)m^*(t-nT) \\ &\quad + \varepsilon^*(t-nT)m(t+nT)] (-j2nT) \\ &\quad \cdot e^{-j2nT\omega} \end{aligned} \quad (\text{A.4})$$

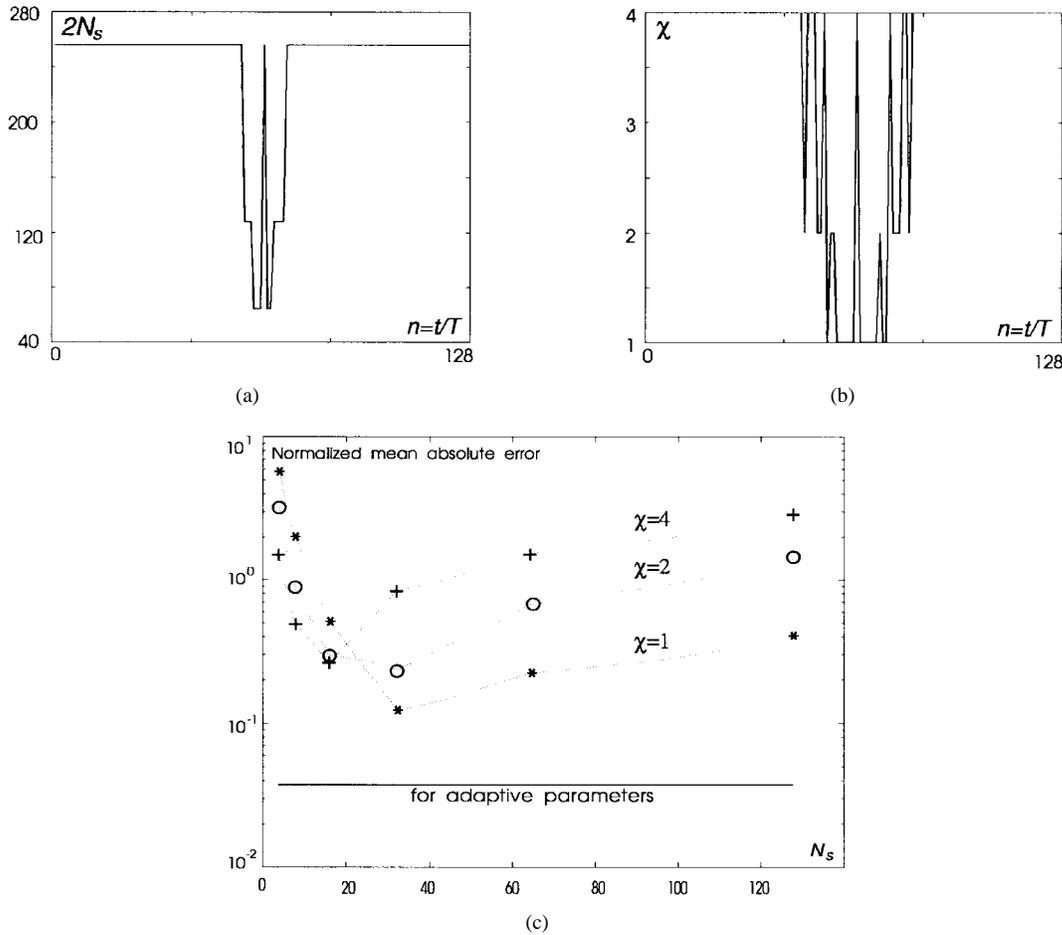


Fig. 7. Parameters and accuracy in the case of the adaptive generalized Wigner distribution. (a) Adaptive window length $2N_s$. (b) Parameter χ as a function of time $n = t/T$. (c) Normalized mean absolute errors of the IF estimation for different window lengths and parameter χ versus time $n = t/T$: “*” for $\chi = 1$, “o” for $\chi = 2$, “+” for $\chi = 4$. The line represents normalized mean absolute error for the adaptive pairs.

Note that $\partial W_h(t, \omega)/\partial \omega|_0 = 0$ because of the symmetry of window $w(t)$. Equation (A.3) then gives

$$\Delta \hat{\omega}_h(t) = \frac{1}{2F_h} \left(L_h(t) + \frac{\Xi_h}{|A|^2} \right) \quad (A.5)$$

where the notation

$$\begin{aligned} F_h &= \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^2 \\ L_h(t) &= \sum_{n=-\infty}^{\infty} w_h(nT) \Delta \phi(t, nT) nT \\ \Xi_h &= \frac{1}{2} \left. \frac{\partial W_h(t, \omega)}{\partial \omega} \right|_0 \delta_\varepsilon \end{aligned} \quad (A.6)$$

is used. It can be verified that the expectation of Ξ_h is equal to zero $E(\Xi_h) = 0$. According to (A.5), for the estimation bias, we get

$$E(\Delta \hat{\omega}_h(t)) = \frac{L_h(t)}{2F_h}. \quad (A.7)$$

Using the same calculations as in [17] and [18], it can be shown that the estimation variance, following from (A.5), is of the form

$$\begin{aligned} \text{var}(\Delta \hat{\omega}_h(t)) &= \frac{1}{4F_h^2 |A|^4} E(\Xi_h^2) \\ &= \frac{\sigma_\varepsilon^2}{2|A|^2} \left(1 + \frac{\sigma_\varepsilon^2}{2|A|^2} \right) \frac{E_h}{F_h^2}. \end{aligned} \quad (A.8)$$

As $h \rightarrow 0$, $T \rightarrow 0$, and $h/T \rightarrow \infty$

$$\begin{aligned} F_h &= \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^2 \rightarrow h^2 F \\ E_h &= \sum_{n=-\infty}^{\infty} w_h^2(nT)(nT)^2 \rightarrow hTE \end{aligned} \quad (A.9)$$

where the constants F and E , depending on the window type only, are defined in (9).

Substituting (A.9) into (A.8) proves (6) of the proposition. Now, let us consider the estimation bias (A.7), where, according to (A.6) and (A.2), $L_h(t)$ may be represented in

two different forms:

$$\begin{aligned} L_h(t) &= 2 \sum_{s=1}^{\infty} \frac{\phi^{(2s+1)}(t)}{(2s+1)!} \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^{2s+2} \\ &= \sum_{s=1}^{\infty} B_h(s) \phi^{(2s+1)}(t) \\ B_h(s) &= \frac{2}{(2s+1)!} \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^{2s+2} \end{aligned} \quad (\text{A.10})$$

or

$$\begin{aligned} L_h(t) &\leq \frac{M_2}{3} \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^4 \\ &= M_2 |B_h(1)| \\ \sup_t |\omega^{(2)}(t)| &\leq M_2. \end{aligned} \quad (\text{A.11})$$

It is easy to verify that as $h \rightarrow 0$, $T \rightarrow 0$, and $h/T \rightarrow \infty$

$$B_h(s) \rightarrow 2F b_s h^{2s+2} \quad (\text{A.12})$$

where b_s is given by (9). Substitution of (A.10)–(A.12) into (A.7) gives (7) and (8). These conclusions complete the proof of the proposition.

ACKNOWLEDGMENT

The authors are very thankful to the Associate Editor and reviewers for the comments and the additional references that helped to improve the paper.

REFERENCES

- [1] M. Biddash and K. B. Yu, "Analysis and filtering using the optimally smoothed Wigner distribution functions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 41, Apr. 1993.
- [2] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—Part 1: Fundamentals," *Proc. IEEE*, vol. 80, pp. 519–538, Apr. 1992.
- [3] B. Boashash and P. O'Shea, "Polynomial Wigner–Ville distributions and their relationship to time-varying higher order spectra," *IEEE Trans. Signal Processing*, vol. 42, pp. 216–220, Jan. 1994.
- [4] L. Cohen and C. Lee, "Instantaneous bandwidth," in *Time-Frequency Signal Analysis*, B. Boashash, Ed. Cheshire, U.K.: Longman, 1992.
- [5] L. Cohen, "Distributions concentrated along the instantaneous frequency," *SPIE, Advanced Signal Process. Alg., Architect., Implement.*, vol. 1348, pp. 149–157, 1990.
- [6] ———, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [7] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Res. Rep.*, 5/94, Technion, Haifa, Israel, Nov. 1994.
- [8] V. Katkovnik, "Nonparametric estimation of instantaneous frequency," *IEEE Trans. Inform. Theory*, vol. 43, pp. 183–189, Jan. 1997.
- [9] ———, "Local polynomial approximation of the instantaneous frequency: Asymptotic accuracy," *Signal Process.*, vol. 52, no. 3, 1996.
- [10] ———, "Adaptive local polynomial periodogram for time-varying frequency estimation," in *Proc. IEEE-SP Time-Freq. Time-Scale Anal.*, Paris, France, June 1996, pp. 329–332.
- [11] S. M. Kay, "Statistically/computationally efficient frequency estimation," in *Proc. IEEE ICASSP*, New York, 1988, pp. 2292–2295.
- [12] B. C. Lovell, R. C. Williamson, and B. Boashash, "The relationship between instantaneous frequency and time-frequency representations," *IEEE Trans. Signal Processing*, vol. 41, pp. 1458–1461, Mar. 1993.
- [13] ———, "The statistical performance of some instantaneous frequency estimators," *IEEE Trans. Signal Processing*, vol. 40, pp. 1708–1723, July 1992.
- [14] *Proc. IEEE*, Special Issue on Time–Frequency Analysis, vol. 84, Sept. 1996.
- [15] B. G. Quinn, "Estimating frequency by interpolation using Fourier coefficients," *IEEE Trans. Signal Processing*, vol. 42, pp. 1264–1268, May 1994.
- [16] B. G. Quinn and J. M. Fernandes, "A fast efficient technique for the estimation of frequency," *Biometrika*, vol. 78, pp. 489–497, Sept. 1991.
- [17] P. Rao and F. J. Taylor, "Estimation of the instantaneous frequency using the discrete Wigner distribution," *Electron. Lett.*, vol. 26, pp. 246–248, 1990.
- [18] L.J. Stanković and S. Stanković, "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.
- [19] L.J. Stanković, "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. 1262–1269, May 1995.
- [20] S. Stanković and L.J. Stanković, "An architecture for realization of a method for time-frequency analysis," *IEEE Trans. Circuits Syst.*, vol. 44, pp. 600–604, July 1997.
- [21] K. M. Wong, "Estimation of the time-varying frequency of a signal: The Cramér–Rao bound and the application of Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 519–535, May 1990.
- [22] L.J. Stanković and V. Katkovnik, "Algorithm for the instantaneous frequency estimation using time-frequency distributions with adaptive window width," *IEEE Signal Processing Lett.*, vol. 5, Sept. 1998.



Vladimir Katkovnik (M'96) received the M.Sc., Ph.D., and Doctor of Sc. degrees from the 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 at 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, 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.



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 1984 and 1985 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. He is now on leave with the Signal Theory Group, Ruhr University Bochum, Bochum, Germany, where he is 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, 35 of them in leading international journals, mainly IEEE publications. He has published several textbooks in signal processing (in Serbo-Croat) and the monograph *Time-Frequency Signal Analysis* (in English). Dr. 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).