Instantaneous Frequency Estimation Using Higher Order L-Wigner Distributions with Data-Driven Order and Window Length

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

***Abstract*—The L-Wigner distributions are defined in order to improve the concentration of signal's time–frequency representation. For a finite distribution order and nonlinear instantaneous frequency (IF) it gives biased IF estimates. In the case of noisy signals the optimal window length and distribution order depend on the noise variance and unknown IF. In this correspondence an adaptive IF estimator with the time-varying and data-driven window length and distribution order is developed. Based on the analysis that has been done here, lower order time–frequency distributions are introduced.**

***Index Terms*—Estimation, instantaneous frequency, time–frequency analysis, Wigner distribution, window optimization.**

I. Introduction

Since there is no distribution from the Cohen class (with signal-independent kernel) which can produce the complete concentration along the instantaneous frequency (IF) when it is a nonlinear function of time, [5]–[7], [21], various higher order distributions have been derived. For the analysis of signals with polynomial phase the polynomial Wigner–Ville distributions are proposed by Boashash *et al.* in [2]–[4]. The same class of signals may be treated by the local polynomial distributions, defined in [10] and [11]. The L-Wigner distribution, introduced and described in [19], [20], [22], and [23], significantly reduces influence of higher order terms in the phase function when it is a nonlinear function of time. The polynomial Wigner–Ville distribution, and the L-Wigner distributions are closely related to the time-varying higher order spectra [3], [17].

In this correspondence we analyze the IF estimator, in the case of additive noisy signals, using the L-Wigner distribution. The estimator's variance and bias are highly dependent on the window length and distribution order. Provided that the signal and noise parameters are known, the optimal window length and/or distribution order may be determined by minimizing the estimation mean squared error. However, those parameters are not available in advance. Especially it is true for the IF derivatives that determine the estimation bias. Here, we will present an adaptive algorithm which does not require *a priori* knowledge of the estimation bias. The basic idea of the method that we use for selection of the data-driven window length and order has originated from [8] and [9], where it was proposed for the nonparametric regression. The idea of this method was exploited in [12] for development of the adaptive local polynomial periodogram, giving estimates of the IF and its derivative. It was subsequently used in [13] for development of the nonparametric estimator of the IF, based on the Wigner distribution with the data-driven adaptive window size. A discrete nature of optimization parameters of the L-Wigner distribution (window length and distribution order), along with a small number of their possible values, resulted in a modification of the original algorithm. Also, the analysis presented in this correspondence leads to an interesting conclusion that not only higher order distributions may improve time–frequency presentation, but also "lower order" distributions can be the best choice in some cases.

The correspondence is organized as follows. A review of the L-Wigner distribution definition, along with a noise modeling, is done in Section II. In Section III the variance and bias of the IF estimate, using the L-Wigner distributions, are derived. The optimal window length and distribution order are discussed in this section. A numerical implementation of the L-Wigner distribution is presented in Section IV. An adaptive IF estimator, with the data-driven window size and distribution order is described in Section V. Numerical examples are presented in Section VI.

Manuscript received March 8, 1998; revised July 1, 1999. The work of LJ. Stanković was supported by the Alexander von Humboldt Foundation.

LJ. Stanković was with with the Signal Theory Group, Ruhr University Bochum, Germany. He is now with the University of Montenegro, 81000 Podgorica, Montenegro, Yugoslavia (e-mail: l.stankovic@ieee.org).

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

Communicated by E. Soljanin, Associate Editor for Coding Techniques.
Publisher Item Identifier S 0018-9448(00)00059-6.

II. Definitions and Noise Modeling

The L-Wigner distribution of a discrete-time signal $s(nT)$, at a given instant $t = lT$, is defined by [19], [20], [22], and [23]

$$\text{LWD}_L(t, \omega) = \sum_{n=-\infty}^{\infty} w_h(nT) s^L\left(t + n\frac{T}{L}\right) s^{L*}\left(t - n\frac{T}{L}\right) e^{-j2nT\omega} \tag{1}$$

where $w_h(nT) = T/h \cdot w(nT/h)$ with $w(t)$ being a real-valued finite-length symmetric window such that $w(t) = 0$, for $|t| > 1/2$, T is the basic sampling interval, and h is the window $w_h(nT)$ width.

Note that in the realization of $\text{LWD}_L(t, \omega)$ the signal has, by definition (1), to be sampled with a sampling interval equal to $1/L$ times the sampling interval in the Wigner distribution.[1] Its values should be available not only at the instants defined by the Nyquist sampling rate π/ω_m, where ω_m is the maximal signals frequency, but also at the points $nT/L = n\pi/(2\omega_m L)$. Note that this can be avoided, i.e., $\text{LWD}_L(t, \omega)$ can be realized without oversampling by using the S-method and procedure described in [20], [22], and [23]. This realization would also produce the L-Wigner distribution which is, in the case of multicomponent signals, equal to the sum of L-Wigner distributions of each individual component, with a significant reduction of the noise influence. Since the realization is not an issue here, we will assume that the L-Wigner distribution is realized according to the definition (the worst case).

Consider a noisy signal

$$x(nT) = s(nT) + \epsilon(nT), \qquad s(nT) = A\exp(j\phi(nT)) \tag{2}$$

with $s(nT)$ being a signal with a real-valued amplitude A and $\epsilon(nT)$ being a white complex-valued Gaussian noise with mutually independent real and imaginary parts of equal variances $\sigma_\epsilon^2/2$. In the analysis presented in this correspondence we additionally assume that the noise is small with respect to the signal, i.e., $\sigma_\epsilon/A \ll 1$. Note that the last assumption has also been used for the Wigner distribution analysis [10], [16] and polynomial Wigner–Ville distribution [4].

[1]For an integer $L > 1$, as we used in our previous papers [18]–[23], this means signal oversampling L times. However, in this correspondence, we will allow values $0 < L < 1$ (for example, $L = 1/2$), which will be useful in some noisy cases and will mean signal downsampling.

The original observations $x(nT)$ can under the low-noise assumption be modified into the following model with the additive noise:

$$x^L(nT) \cong s^L(nT) + Ls^{L-1}(nT)\epsilon(nT) = s^L(nT) + \epsilon_L(nT). \quad (3)$$

The autocorrelation function of $\epsilon_L(nT) = Ls^{L-1}(nT)\epsilon(nT)$ is

$$R_{\epsilon_L \epsilon_L}(n) = A^{2L-2}\sigma_\epsilon^2 L^2 \delta(n). \quad (4)$$

If $L = 1$ the variances of the transformed and the original noise are equal to each other. For the case $0 < L < 1$, which has not been considered previously, the transformed noise-to-signal ratio in (3) is $L^2\sigma_\epsilon^2/A^2$. It can be smaller than the original σ_ϵ^2/A^2. For example, $L = 1/2$ may significantly reduce this ratio compared with $L = 1$. This will be the reason to introduce, in this correspondence, distributions of a "lower order" with respect to the Wigner distribution.

III. Instantaneous Frequency Estimation

Consider the problem of the IF $\omega(t) = \phi'(t)$ estimation, from discrete-time observations (2). We will assume that $\omega(t)$ is a differentiable function with bounded derivatives $|\omega^{(r)}(t)| = |\phi^{(r+1)}(t)| \le M_r(t)$, $r \ge 1$.

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

$$\text{LWD}_L(t, \omega) = A^{2L} \sum_{n=-\infty}^{\infty} w_h(nT)\, e^{j2\phi'(t)nT}\, e^{j\Delta\phi(t, n(T/L))} \cdot e^{-j2\omega nT},$$

where

$$\Delta\phi\left(t, n\frac{T}{L}\right) = 2L \sum_{s=1}^{\infty} \frac{\phi^{(2s+1)}(t)}{(2s+1)!} \left(n\frac{T}{L}\right)^{2s+1}. \quad (5)$$

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

$$\hat{\omega}(t) = \arg\left[\max_{\omega \in Q_\omega} \text{LWD}_L(t, \omega)\right] \quad (6)$$

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

$$\Delta\hat{\omega}(t) = \hat{\omega}(t) - \omega(t). \quad (7)$$

Proposition: Let $\hat{\omega}(t)$ be a solution of (6), and $h \to 0$, $T \to 0$, and $h/T \to \infty$. Then the variance and bias of the IF estimation error $\Delta\hat{\omega}(t)$ are given by

$$\text{var}\,(\Delta\hat{\omega}(t)) = \frac{L^2\sigma^2}{2A^2}\, \frac{E_h}{F_h^2} \quad (8)$$

$$E\{\Delta\hat{\omega}(t)\} = \frac{1}{2F_h} \sum_{s=1}^{\infty} \frac{B_h(s)}{L^{2s}}\, \phi^{(2s+1)}(t), \quad (9)$$

where

$$F_h = \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^2 \to h^2 \int_{-1/2}^{1/2} w(t)t^2\, dt$$

$$E_h = \sum_{n=-\infty}^{\infty} w_h^2(nT)(nT)^2 \to Th \int_{-1/2}^{1/2} w^2(t)t^2\, dt$$

$$B_h(s) = \frac{-2}{(2s+1)!} \sum_{n=-\infty}^{\infty} w_h(nT)(nT)^{2s+2} \to \frac{-2h^{2s+2}}{(2s+1)!} \cdot \int_{-1/2}^{1/2} w(t)t^{2s+2}\, dt. \quad (10)$$

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

Proof: The stationary point of $\text{LWD}_L(t, \omega)$ is defined by the zero value of the derivative $\partial\text{LWD}_L(t, \omega)/\partial\omega$, given by

$$\frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega} = \sum_{n=-\infty}^{\infty} w_h(nT) x^L\left(t + n\frac{T}{L}\right) x^{L^*}\left(t - n\frac{T}{L}\right) \cdot (-j2nT)\, e^{-j2nT\omega}. \quad (11)$$

The values of $\partial\text{LWD}_L(t, \omega)/\partial\omega$ around the stationary point may be approximated by a linear model for small values of estimation error $\Delta\hat{\omega}$, phase residual $\Delta\phi(t, n\frac{T}{L})$, and noise $\epsilon_L(nT)$. The linearization gives

$$\frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega} \cong \frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega}_{|0} + \frac{\partial^2\text{LWD}_L(t, \omega)}{\partial\omega^2}_{|0} \Delta\hat{\omega} + \frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega}_{|0} \delta_{\Delta\phi} + \frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega}_{|0} \delta_\epsilon \quad (12)$$

where the index $|0$ means that the expressions are calculated at the point where $\omega = \phi'(t)$, $\Delta\phi(t, n(T/L)) = 0$, and $\epsilon_L(nT) \equiv 0$. The terms $\partial\text{LWD}_L(t, \omega)/\partial\omega\delta_{\Delta\phi}$ and $\partial\text{LWD}_L(t, \omega)/\partial\omega\delta_\epsilon$ represent variations of the derivative $\partial\text{LWD}_L(t, \omega)/\partial\omega$ caused by small $\Delta\phi(t, nT)$ and noise $\epsilon_L(nT)$, respectively. For example, in the calculation, of $\partial\text{LWD}_L(t, \omega)/\partial\omega\delta_{\Delta\phi}$ we assume that the other two disturbances caused by the noise $\epsilon_L(nT)$ and frequency variation $\Delta\hat{\omega}$ can be neglected and that the phase variation $\Delta\phi(t, nT)$ is small. Then

$$\frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega}_{|0} \delta_{\Delta\phi} = \sum_{n=-\infty}^{\infty} w_h(nT) A^{2L}\, e^{j2nT\phi'(t) + j\Delta\phi(t, nT)} \cdot (-j2nT)\, e^{-j2nT\omega}.$$

For $\omega = \phi'(t)$ and small $\Delta\phi(t, nT)$, when

$$\exp(j\Delta\phi(t, nT)) \cong 1 + j\Delta\phi(t, nT)$$

we easily get the third line in the next equation, as a variation of $\partial\text{LWD}_L(t, \omega)/\partial\omega$ due to small $\Delta\phi(t, nT)$. In the same way we get the other terms as

$$\frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega}_{|0} = A^{2L} \sum_{n=-\infty}^{\infty} w_h(nT)(-j2nT) = 0$$

$$\frac{\partial^2\text{LWD}_L(t, \omega)}{\partial\omega^2}_{|0} = -A^{2L} \sum_{n=-\infty}^{\infty} w_h(nT)(4nT)^2 = -4A^{2L}F_h$$

$$\frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega}_{|0} \delta_{\Delta\phi} = A^{2L} \sum_{n=-\infty}^{\infty} w_h(nT)\Delta\phi\left(t, n\frac{T}{L}\right)(2nT) = 2A^{2L}L_h(t)$$

$$2\Xi = \frac{\partial\text{LWD}_L(t, \omega)}{\partial\omega}_{|0} \delta_\epsilon = \sum_{n=-\infty}^{\infty} w_h(nT)\left[s^L\left(t + n\frac{T}{L}\right) s^{L^*} \cdot \left(t - n\frac{T}{L}\right) - x^L\left(t + n\frac{T}{L}\right) \cdot x^{L*}\left(t - n\frac{T}{L}\right)\right](-j2nT)e_{|0}^{-jnT\omega}. \quad (13)$$

From $\partial\text{LWD}_L(t, \omega)/\partial\omega = 0$ and the previous equations follows that

$$-4A^{2L}F_h\Delta\hat{\omega} + 2A^{2L}L_h(t) + 2\Xi = 0$$

or

$$\Delta\hat{\omega}(t) = \omega(t) - \hat{\omega}(t) = \frac{1}{2F_h}\left(L_h(t) + \frac{\Xi}{A^{2L}}\right). \tag{14}$$

The error $\Delta\hat{\omega}$ has two components: deterministic $L_h(t)/2F_h$ and random $\Xi/(2A^{2L}F_h)$. Straightforward calculations, similar to the ones used in [18] for the Wigner distribution, give

$$\begin{aligned}\mathrm{var}(\Delta\hat{\omega}(t)) = \frac{1}{4A^4F_w^2}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} & w_h(mT)w_h(nT) \\ \cdot & \left[s^L\left(t+n\frac{T}{L}\right)s^{L^*}\left(t+m\frac{T}{L}\right)R_{\epsilon_L\epsilon_L}\right. \\ & \cdot\left((n-m)\frac{T}{L}\right) + s^L\left(t-m\frac{T}{L}\right) \\ & \cdot s^{L^*}\left(t-n\frac{T}{L}\right)R_{\epsilon_L\epsilon_L}\left((n-m)\frac{T}{L}\right) \\ & \left. + R^2_{\epsilon_L\epsilon_L}\left((n-m)\frac{T}{L}\right)\right](nmT^2) \\ & \cdot e^{-j2(m-n)T\omega}_{|0}.\end{aligned}$$

For the white noise $\epsilon_L(nT)$ it results in

$$\mathrm{var}\,(\Delta\hat{\omega}(t)) = \frac{R_{\epsilon_L\epsilon_L}(0)}{2A^{2L}}\frac{E_h}{F_h^2} \tag{15}$$

which proves (8).

In a similar way, the bias from (14) is of the form

$$L_h(t) = -2\sum_{s=1}^{\infty}\frac{\phi^{(2s+1)}(t)}{L^{2s}(2s+1)!}\sum_{n=-\infty}^{\infty}w_h(nT)(nT)^{2s+2}$$

which proves (9) of the proposition. □

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

$$E\{(\Delta\hat{\omega}(t))^2\} = \frac{2L^2\sigma_\epsilon^2}{A^2}\frac{E_h}{F_h^2} + \left[\frac{B_h(1)\phi^{(3)}(t)}{2F_hL^2}\right]^2. \tag{16}$$

For the rectangular window $(E_h = Th/12,\, F_h = h^2/12,\, B_h(1) = -h^4/240)$, we get

$$\begin{aligned}\mathrm{var}\,(\Delta\hat{\omega}(t)) &= \frac{6\sigma_\epsilon^2L^2}{A^2}\frac{T}{h^3} \\ E(\Delta\hat{\omega}(t)) &= \frac{\phi^{(3)}(t)}{40L^2}h^2 \end{aligned}\tag{17}$$

$$E\{(\Delta\hat{\omega}(t))^2\} = \frac{6\sigma_\epsilon^2L^2}{A^2}\frac{T}{h^3} + \left[\frac{\phi^{(3)}(t)}{40L^2}h^2\right]^2. \tag{18}$$

Note that the values of E_h, F_h, and $B_h(1)$ can be easily calculated for any other window type.

It can be seen from (18) that the MSE has a minimum with respect to L. The optimal value of L is given by

$$L_{\mathrm{opt}}(t) = \left[\frac{A^2h^7(\phi^{(3)}(t))^2}{4800\sigma_\epsilon^2T}\right]^{1/6}. \tag{19}$$

The dependence of the optimal L-Wigner distribution order L_{opt} on the parameters A, σ_ϵ,T, and the derivative $\phi^{(3)}(t)$ is quite clear. At times when $|\phi^{(3)}(t)|$ is large, higher distribution orders are required, while for small $|\phi^{(3)}(t)|$, the distribution order L should also be small. This relation will be discussed in detail in examples.

The optimal window length, minimizing the MSE (18) for a given L is

$$h_{\mathrm{opt}}(t) = \left[\frac{7200\sigma_\epsilon^2L^2T}{A^2(\phi^{(3)}(t))^2}\right]^{1/7}. \tag{20}$$

It is obvious that the calculation of (19) or (20) is not possible in practice, since its requires, besides A and σ_ϵ^2, the knowledge of the IF second derivative $\phi^{(3)}(t)$. It is a definitely unavailable value because the IF itself has to be estimated.

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

IV. Numerical Implementation

The L-Wigner distribution (1), discretized over the frequency, is implemented as

$$\begin{aligned}\mathrm{LWD}_L(k,\, l) = \sum_{n=-N/2}^{N/2-1} & w_h(nT)x^L\left(lT+n\frac{T}{L}\right)x^{L^*} \\ & \cdot\left(lT-n\frac{T}{L}\right)e^{-j2(2\pi/N)kN}\end{aligned} \tag{21}$$

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

$$\hat{\omega}_h(lT) = \arg\left[\max_k\, \mathrm{LWD}_L(k,\, l)\right]\frac{\pi}{NT} \tag{22}$$

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

Let us consider the influence of the quantization error on the accuracy of the IF estimation, caused by the discretization of $\mathrm{LWD}_L(k,\, l)$ in (21) along the frequency axis. Note that the quantization error may also be considered as a parameter closely related to the distribution concentration (frequency resolution), which is very important for time–frequency distributions (especially in the case of multicomponent signals). For the quantization noise error a uniform probability density is usually assumed. In (22) this probability density is uniform over the segment $(-\pi/2h,\, \pi/2h]$, since $\pi/(NT) = \pi/h$. Its variance is $\sigma_q^2 = 1/12\,(\pi/h)^2$, producing the resulting MSE in (18) of the form

$$\begin{aligned}E\{(\Delta\hat{\omega}(t))^2\} &= \frac{6\sigma_\epsilon^2L^2}{A^2}\frac{T}{h^3} + \left[\frac{\phi^{(3)}(t)}{40L^2}h^2\right]^2 + \frac{1}{12}\frac{\pi^2}{h^2} \\ &= \left[\frac{6\sigma^2}{A^2}\frac{L^2}{N} + \frac{\pi^2}{12}\right]\frac{1}{h^2} + \left[\frac{\phi^{(3)}(t)}{40L^2}h^2\right]^2.\end{aligned} \tag{23}$$

For a large signal-to-noise ratio and any reasonable number of samples N and distribution order L we have $(6\sigma_\epsilon^2/A^2)(L^2/N) < \pi^2/12$ and the estimation variance is dominated by the quantization error.

As is well known, the quantization effects of the fast Fourier transform (FFT) can be reduced by appropriate zero-padding in the time domain, i.e., by interpolation along the frequency axis. Provided that this interpolation of the L-Wigner distribution is done up to the widest considered window length, the quantization error will be reduced and kept to a constant value.

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

V. Algorithm for Adaptive Order and Window Length Determination

A. Basic Idea of the Window Length Optimization

The basic idea follows from the accuracy analysis, given in the Proposition. Namely, at least for the asymptotic case of small noise and bias, the estimation error can be represented as a sum of the deterministic (bias) and random component, with the variance given in the Proposition and (17). For the estimated IF $\hat{\omega}_h(t)$ as a random variable distributed around the exact IF $\omega(t)$ with the bias $bias(t, h)$ and the variance $\sigma^2(h)$, we may write the following relation:

$$|\omega(t) - (\hat{\omega}_h(t) - bias(t, h))| \le \kappa\sigma(h) \tag{24}$$

where $\sigma^2(h) \triangleq \mathrm{var}(\Delta\hat{\omega}(t))$, and the inequality holds with the probability $P(\kappa)$ depending on κ. The values of κ will be discussed later. Here we will only mention that it should be such that $P(\kappa)$ is close to 1. In the case of dominant quantization noise error, which will be considered now, the probability density of $\hat{\omega}_h(t)$ around $bias(t, h)$ is uniform. For this distribution, the value $\kappa = \sqrt{3}$ guarantees (24) with probability 1. The same value of κ would guarantee only the probability of 0.93, if the distribution were Gaussian.

Using the expressions for the variance and bias

$$\sigma^2(h) = \left[\frac{6\sigma_\epsilon^2}{A^2}\,\frac{L^2}{N} + \frac{\pi^2}{12}\right]\frac{1}{h^2} \cong \frac{\pi^2}{12}\,\frac{1}{h^2}$$
$$bias(t, h) = \frac{\phi^{(3)}(t)}{40L^2}\,h^2 \tag{25}$$

the MSE is given by

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

Concerning the distribution order L, *we may conclude that its highest value should be used,* as long as $\frac{72\sigma_\epsilon^2}{A^2}\,\frac{L^2}{N\pi^2} \ll 1$.

The MSE minimization with respect to h gives

$$\sigma(h_{\mathrm{opt}}) = \sqrt{2}\,bias(t, h_{\mathrm{opt}}), \tag{26}$$

where h_{opt} denotes the optimal window length,

$$h_{\mathrm{opt}} = (50\pi^2L^4/(3(\phi^{(3)}(t))^2))^{1/6}.$$

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

$$H = \{h_s | h_s = ah_{s-1}, \quad s = 1, 2, 3, \cdots, J, a > 1\}. \tag{27}$$

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

a) A discrete scheme for window lengths is necessary for efficient numerical realizations. Realizations of time–frequency distributions of the form (21) are almost exclusively based on the FFT algorithms (excluding only a few recursive approaches, [1], [14], [15]). The most common are radix-2 or radix-3 FFT algorithms, which correspond to $a = 2$ or $a = 3$, respectively, in which case the set H gives dyadic $(h_s = h_0 2^s)$ or triadic $(h_s = h_0 3^s)$ window length schemes. In the realizations, the smallest window length h_0 should correspond to a small number N_0 of contained signal samples. For example, for radix-2 FFT algorithms, $N_0 = 4$ or $N_0 = 8$ with $N_s = 2N_{s-1}$, $s = 1, 2, \cdots, J$.

b) A search for the optimal window length over H is a simplified optimization, because the set (27) consist of a relatively small number of elements. However, the discrete set of h inevitably leads to suboptimal window length values due to the discretization of h (quantization noise error and effects due to quantization of h, which, despite similar names, are two completely different notions). It is important to note that this effect, due to the discrete nature of $h \in H$, would exist even if we knew in advance all of the parameters required for the optimal window length calculation, and decided to use radix-2 FFT algorithms in the realization. Fortunately, the loss of accuracy is not significant in many cases, because the MSE has a stationary point for the optimal window length $h = h_{\mathrm{opt}}$ (and the MSE varies very slowly for the window length values close to $h = h_{\mathrm{opt}}$). The MSE as a function of the window length is presented in Fig. 1(a), where both axeshave logarithmic scales. Fig. 1(b) illustrates the regions where the IF may take its values, according to bias and variance relations (25). This dependence will be discussed in the sequel. Here we will only mention that for small window lengths the IF bias is very small, but variance is large (25). Therefore, the IF lies within the wide region $[-\sqrt{3}\,\sigma(h), \sqrt{3}\,\sigma(h)]$, around the exact IF value. As the window length increases, these IF limits are closer to each other (since $\sigma(h)$ decreases), but at the same time the bias is increasing, see Fig. 1.

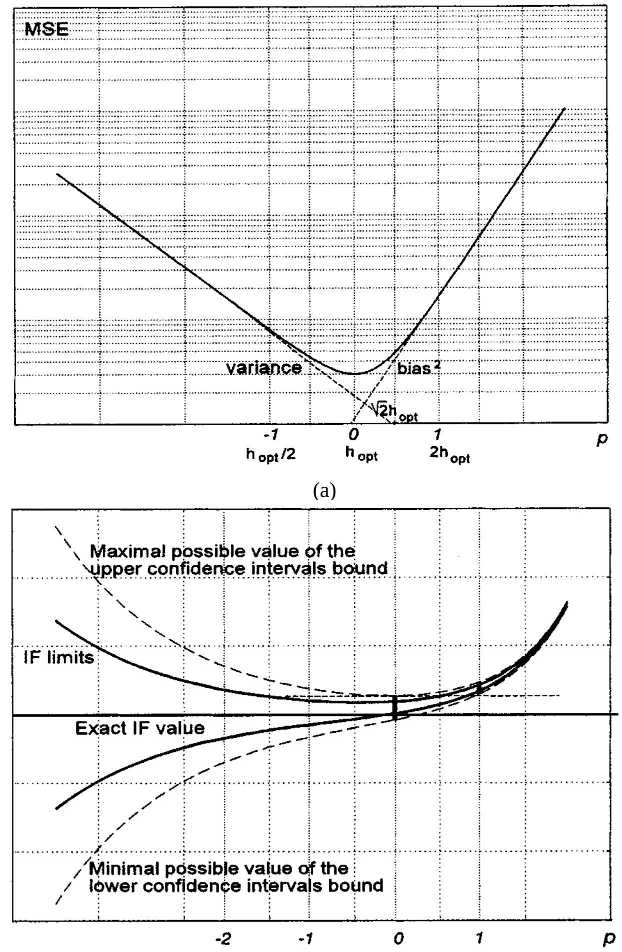

Fig. 1. (a) Illustration of the MSE. (b) Exact IF value, bias, and confidence intervals limits as a function of the window width illustration.

Now we will derive an algorithm for the determination of the optimal window size h_{opt}, without knowing the bias, using the IF estimates (22) and the formula for the IF estimate's variance only.

It is based on the following statement.

Let H be a set of dyadic window length values, i.e., $a = 2$ in (27), and $\kappa = \sqrt{3}$. Assume that $h_{\rm opt} \in H$. Define the upper and lower bounds of confidence intervals $D_s = [L_s, U_s]$ as

$$\begin{aligned} L_s &= \hat{\omega}_{h_s}(t) - (\kappa + \Delta\kappa)\sigma(h_s) \\ U_s &= \hat{\omega}_{h_s}(t) + (\kappa + \Delta\kappa)\sigma(h_s) \end{aligned} \tag{28}$$

where $\hat{\omega}_{h_s}(t)$ is an estimate of the IF, with $h = h_s$ and $\sigma(h_s)$ given by (25).

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

$$\begin{aligned} D_s \cap D_{s+1} &\neq \emptyset \\ D_{s+1} \cap D_{s+2} &= \emptyset \end{aligned} \tag{29}$$

is satisfied, i.e., when the last two successive confidence interval still intersect.

There exists value of $\Delta\kappa$ such that, with probability close to 1

$$h_{s^+} = h_{\rm opt}. \tag{30}$$

Proof: Let us denote by b the unknown bias

$$bias\,(t, h_{\rm opt}) = b \tag{31}$$

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

$$h(p) = h_{\rm opt} a^p, \qquad a = 2,\ p = \cdots, -2, -1, 0, 1, 2, \cdots \tag{32}$$

where $p = 0$ corresponds to the desired window length $h_{\rm opt}$. Note also that we use two indices for the window lengths s (i.e., h_s) for indexing which starts from the narrowest window length, and p (used in form of an argument, i.e., $h(p)$) starting from the $h_{\rm opt}$ window length (when $p = 0$), with narrower windows having negative p and wider window lengths having positive p.

The bias and variance values given by (25) can be written for any $h(p)$ of the form (32) as

$$bias\,(t, h(p)) = a^{2p} b, \qquad \sigma(h(p)) = \sqrt{2}\, a^{-p} b \tag{33}$$

according to (26) and (31).

From (33) we can conclude that for $p \ll 0$ and $a = 2$ the bias is much smaller than the variance since $a^{2p} \ll \sqrt{2}\, a^{-p}$. Thus the estimate $\hat{\omega}_h(t)$ is spread around the exact value $\omega(t)$ with a small bias $(bias(t, h(p)) \to 0$ as $h(p) \to 0)$ and a large variance $(\sigma(h(p)) \to \infty$ as $h(p) \to \infty)$. A confidence interval of the estimate $\hat{\omega}_{h(p)}(t)$, for a given $h(p)$, is defined by

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

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

A confidence interval, that takes into account the biased estimate $\hat{\omega}_{h(p)}(t)$ is

$$\tilde{D}(p) = [\hat{\omega}_{h(p)}(t) - (\kappa + \Delta\kappa)\sigma(h(p)),\ \hat{\omega}_{h(p)}(t) + (\kappa + \Delta\kappa)\sigma(h(p))] \tag{34}$$

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

$$[\omega(t) - \Delta\kappa\sigma(h(p)),\ \omega(t) + \Delta\kappa\sigma(h(p))] \subseteq \tilde{D}(p) \cap \tilde{D}(p+1)$$

for any $p \ll 0$.

For $p \gg 0$ the variance is small but the bias is large, since $a^{2p} \gg \sqrt{2}\, a^{-p}$. It is clear that for a large enogh p, $\tilde{D}(p) \cap \tilde{D}(p+1) = \emptyset$ for any given $\Delta\kappa$.

The idea behind the algorithm (28) and (29) is that $\Delta\kappa$ in $\tilde{D}(p)$ can be found in such a way that the largest p for which two consecutive confidence intervals $\tilde{D}(p)$ and $\tilde{D}(p+1)$ have a point in common is $p = 0$. Such a value of $\Delta\kappa$ exists because the bias and the variance are monotonically increasing and decreasing functions of h, respectively. As soon as this value of $\Delta\kappa$ is found, an intersection of the confidence intervals $\tilde{D}(p)$ and $\tilde{D}(p+1)$ works as an indicator of the event $p = 0$, i.e., the event when $h_s = h_{\rm opt}$ is found. The algorithm given in the form (28) and (29) tests the intersection of the confidence intervals, where (29) is a condition that two sequential intervals $\tilde{D}_{s+1}$ and $\tilde{D}_s$ are the last pair of the confidence intervals having a point in common (note again that indices s and p only indicate whether we assume the first confidence interval, or the confidence interval for which $h = h_{\rm opt}$ has index 0).

Now let us find this crucial value of $\Delta\kappa$. According to the above analysis, only three values of $p = 0$, 1, and 2 along with the corresponding intervals $\tilde{D}(0)$, $\tilde{D}(1)$, and $\tilde{D}(2)$ should be considered, in this case. The intervals $\tilde{D}(0)$ and $\tilde{D}(1)$ *should have* and the intervals $\tilde{D}(1)$ and $\tilde{D}(2)$ *should not have* at least one point in common. Since $\hat{\omega}_{h(p)}(t)$ is a random (uniformly distributed) variable, then the confidence interval bounds are also random and uniformly distributed. Thus we must consider the worst possible cases for the corresponding bounds. These worst case conditions, for $b > 0$, are given by

$$\begin{aligned} \min\{U(0)\} &\geq \max\{L(1)\} \\ \max\{U(1)\} &< \min\{L(2)\}. \end{aligned} \tag{35}$$

Let us, for example, consider $U(1)$. The estimated IF $\hat{\omega}_{h(1)}(t)$maay, according to (34), assume values within the interval

$$\begin{aligned} \hat{\omega}_{h(p)}(t) \in &[\omega(t) + bias\,(h(1)) - \kappa\sigma(h(1)), \\ &\omega(t) + bias\,(h(1)) + \kappa\sigma(h(1))]. \end{aligned}$$

Consequently, the upper confidence interval bound $U(1)$, according to (34), may take values from the interval

$$\begin{aligned} U(1) \in &[\omega(t) + bias\,(h(1)) + \Delta\kappa\sigma(h(1)), \\ &\omega(t) + bias\,(h(1)) + (2\kappa + \Delta\kappa)\sigma(h(1))]. \end{aligned}$$

The maximal possible value of $U(1)$ is

$$\max\{U(1)\} = \omega(t) + bias\,(h(1)) + (2\kappa + \Delta\kappa)\sigma(h(1)).$$

In the same way we get the other bounds required by (35)

$$\begin{aligned} bias\,(h(0)) + \Delta\kappa\sigma(h(0)) &\geq bias\,(h(1)) - \Delta\kappa\sigma(h(1)) \\ (h(1)) + (2\kappa + \Delta\kappa)\sigma(h(1)) &< bias\,(h(2)) - (2\kappa + \Delta\kappa)\sigma(h(2)) \end{aligned} \tag{36}$$

or

$$\begin{aligned} 1 + \Delta\kappa\,\sqrt{2} &\geq a^2 - \Delta\kappa\,\sqrt{2}\, a^{-1} \\ 1 + (2\kappa + \Delta\kappa)\,\sqrt{2}\, a^{-1} &< a^4 - (2\kappa + \Delta\kappa)\,\sqrt{2}\, a^{-2}. \end{aligned} \tag{37}$$

It can be verified that $\Delta\kappa = (a^2 - 1)/\left[\sqrt{2}\,(1 + a^{-1})\right]$ is the smallest $\Delta\kappa > 0$ that satisfies the first inequality in (37). For $a = 2$ we get $\Delta\kappa = 1.414$. This value of $\Delta\kappa$ with $\kappa = \sqrt{3}$ satisfies the second inequality in (37).

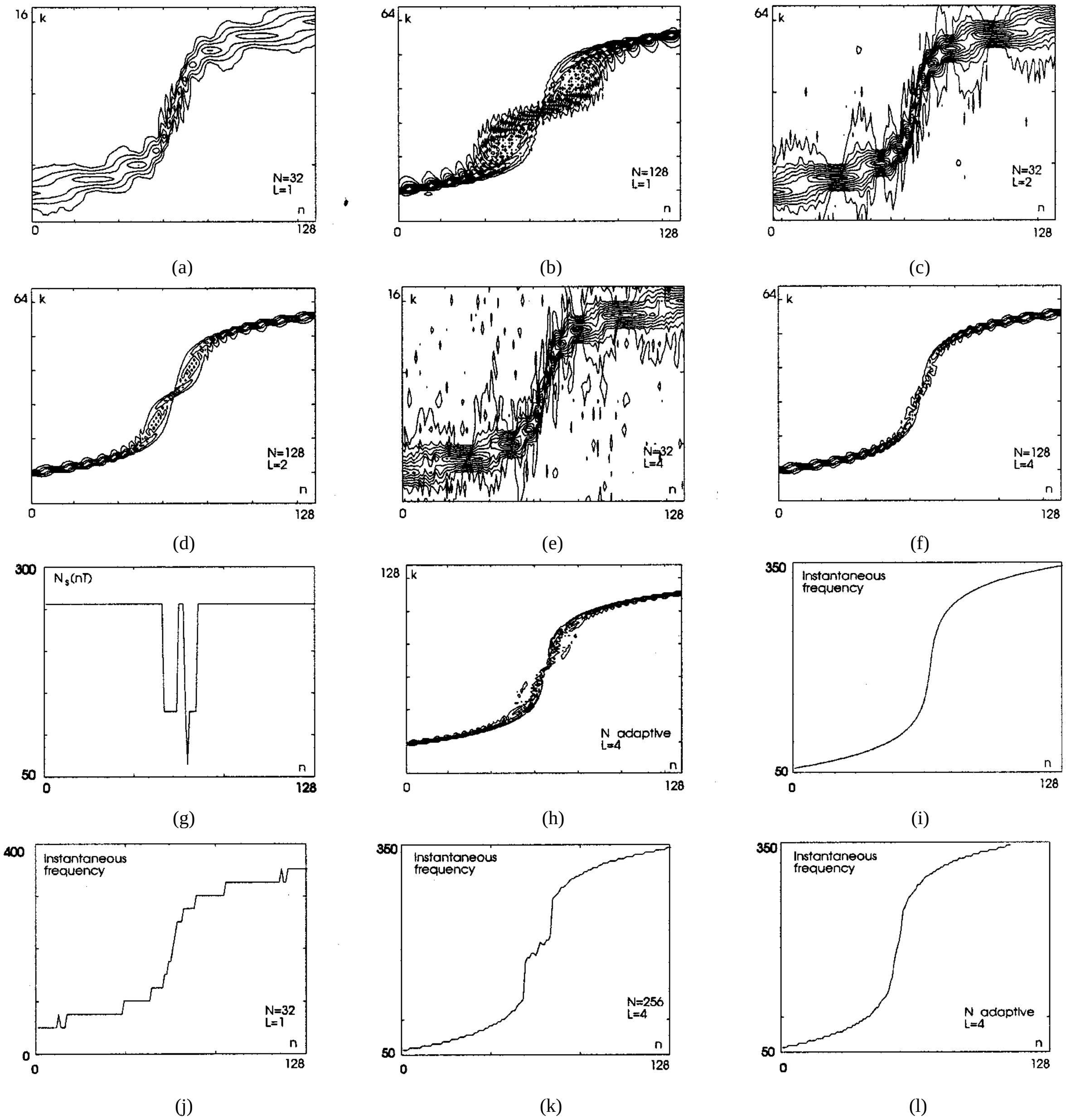

Fig. 2. Time–frequency representation and the IF estimation of a signal with the nonlinear IF. (a) Wigner distribution with $N = 32$. (b) Wigner distribution with $N = 128$. (c) L-WD with $L = 2$ and $N = 32$. (d) L-WD with $L = 2$ and $N = 128$. (e) L-WD with $L = 4$ and $N = 32$. (f) L-WD with $L = 4$ and $N = 128$. (g) Adaptive window length. (h) L-WD with $L = 4$ and adaptive window length. (i) The IF of signal. (j) The IF estimation with the Wigner distribution and $N = 32$. (k) The IF estimation with $L = 4$ and $N = 256$. (l) The IF estimation with $L = 4$ and adaptive window length.

With (37) being satisfied we have, for $p \leq 0$, that $\tilde{D}(p) \cap \tilde{D}(p+1) \neq \emptyset$ with probability close to 1, and for $p \geq 1$, $\tilde{D}(p) \cap \tilde{D}(p+1) = \emptyset$ with probability close to 1. This completes the proof of the statement. □

We emphasize that the statement is derived provided that $h_{\rm opt}$ assumes one of the dyadic values from H and the bias and variance are given by the asymptotic formulas (25). In applications, due to the discrete nature of h, we will never have $h_{\rm opt} \in H$, which results in the already described discretization of h and slightly suboptimal MSE values. This means that the values κ and $\Delta\kappa$, given in the statement, should be interpreted as a reasonable approximation which can be used in the algorithm (28) and (29), at least as long as (25) holds for the bias and variance. This was the reason for the expression "probability close to 1."

B. Algorithm

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

1) Assume that a set H is given by (27).
2) For a given instant t, perform the L-Wigner distribution calculation for increasing window lengths $h_s \in H$ beginning with the smallest.
3) Estimate IF using

$$\hat{\omega}_{h_s}(t) = \arg\left[\max_{\omega \in Q_\omega} \mathrm{LWD}_L(t, \omega)\right]. \tag{38}$$

4) With
$\sigma(h_s) = \sqrt{[(6\sigma_\epsilon^2/A^2)(L^2/N) + \pi^2/3]\,1/h_s^2} \cong \pi/h_s\,\sqrt{3}$
and $\hat{\omega}_{h_s}(t)$, define the segments

$$\tilde{D}_s(t) = [L_s(t),\, U_s(t)] \tag{39}$$

where

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

with, for example, $\kappa + \Delta\kappa \approx 3.2$.

5) The adaptive window length h_{s+} is determined as the length corresponding to the largest $s (s = 1, 2, \cdots, J)$ when $D_s \cap D_{s+1} \neq \emptyset$, i.e.,

$$|\hat{\omega}_{h_s}(t) - \hat{\omega}_{h_{s+1}}(t)| \le (\kappa + \Delta\kappa)[\sigma(h_s) + \sigma(h_{s+1})] \tag{40}$$

is still satisfied.

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

$$\hat{h}(t) = h_{s+}(t) \tag{41}$$

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

6) Take next t.

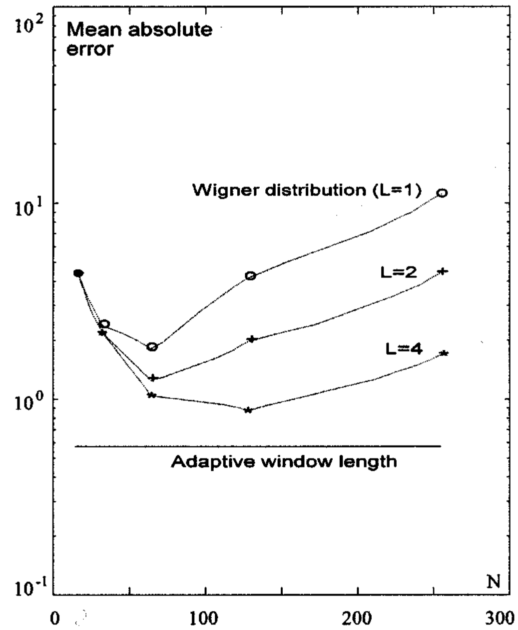

Fig. 3. Mean absolute error for various window lengths and L-Wigner distribution orders in the case of dominant quantization error.

C. Comments on the Algorithm

If the noise-dependent part in the standard deviation $\sigma(h_s)$ cannooot be neglected (as will be the case in further analysis) then the estimation of signal and noise parameters $|A|$ and σ_ϵ^2 can be done using

$$|\hat{A}|^2 + \hat{\sigma}_\epsilon^2 = \frac{1}{N}\sum_{n=1}^{N} |x(nT)|^2 \tag{42}$$

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

$$\hat{\sigma}_{\epsilon r} = \frac{\{\text{median}(|x_r(nT) - x_r((n-1)T)|\}}{0.6745}$$
$$\hat{\sigma}_{\epsilon i} = \frac{\{\text{median}(|x_i(nT) - x_i((n-1)T)|\}}{0.6745}$$
$$\hat{\sigma}_\epsilon^2 = (\hat{\sigma}_{\epsilon r}^2 + \hat{\sigma}_{\epsilon i}^2)/2 \tag{43}$$

where $x_r(nT)$ and $x_i(nT)$ are the real and imaginary parts of $x(nT)$, respectively, and T is sufficiently small.

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

D. Estimation with Interpolation

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

a) Let the order L of an L-Wigner distribution be fixed and the bias and variance of estimation be determined by (18). It can be seen that for the optimal window size h_{opt}, the relation

$$\sigma(h_{\text{opt}}) = \sqrt{4/3}\, bias(t, h_{\text{opt}}) \tag{44}$$

between the bias and the variance holds instead of (26). It is obtained by equating the first derivative of (18) with respect to h with zero. We will follow the same reasoning as in Section V-A, with the assumption that for a certain κ we may assume that

$$|\omega(t) - (\hat{\omega}_h(t) - bias(t, h))| \le \kappa\sigma(h) \tag{45}$$

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

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

Then the conditions that $\tilde{D}(0) \cap \tilde{D}(1) \neq \emptyset$ and $\tilde{D}(1) \cap \tilde{D}(2) = \emptyset$, with $b > 0$, are of the form (36) and produce the inequalities similar to (37)

$$1 + \frac{2}{\sqrt{3}}\,\Delta\kappa \ge a^2 - \Delta\kappa\,\frac{2}{\sqrt{3}}\,a^{-3/2}$$
$$1 + (2\kappa + \Delta\kappa)\,\frac{2}{\sqrt{3}}\,a^{-3/2} < a^4 - (2\kappa + \Delta\kappa)\,\frac{2}{\sqrt{3}}\,a^{-3}. \tag{46}$$

For $a = 2$ these inequalities give the smallest positive value $\Delta\kappa = 1.919$ and the largest $(2\kappa + \Delta\kappa) = 27.15$. Note that for the value of $\kappa \approx 3$, i.e., $\kappa + \Delta\kappa \approx 5$, the inequality (45) holds with the probability $P(\kappa) = 0.997$.

With these hints and parameter values we can now use the algorithm (38)–(41) as described in Section V-B for the adaptive window length determination.

b) Consider the simultaneous optimization with respect to both parameters in question, L and h. Let us define the two sets

$$H = \{h_s\,|h_1 < h_2 < h_3 < \cdots < h_J\}$$
$$\Lambda = \{L_r | L_1 < L_2 < \cdots < L_K\} \tag{47}$$

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

Consider a direct product of H and Λ as a set

$$H \times \Lambda = \{(h_s,\, L_r)| s = 1, 2, \cdots, J,\ r = 1, 2, \cdots, K\}$$

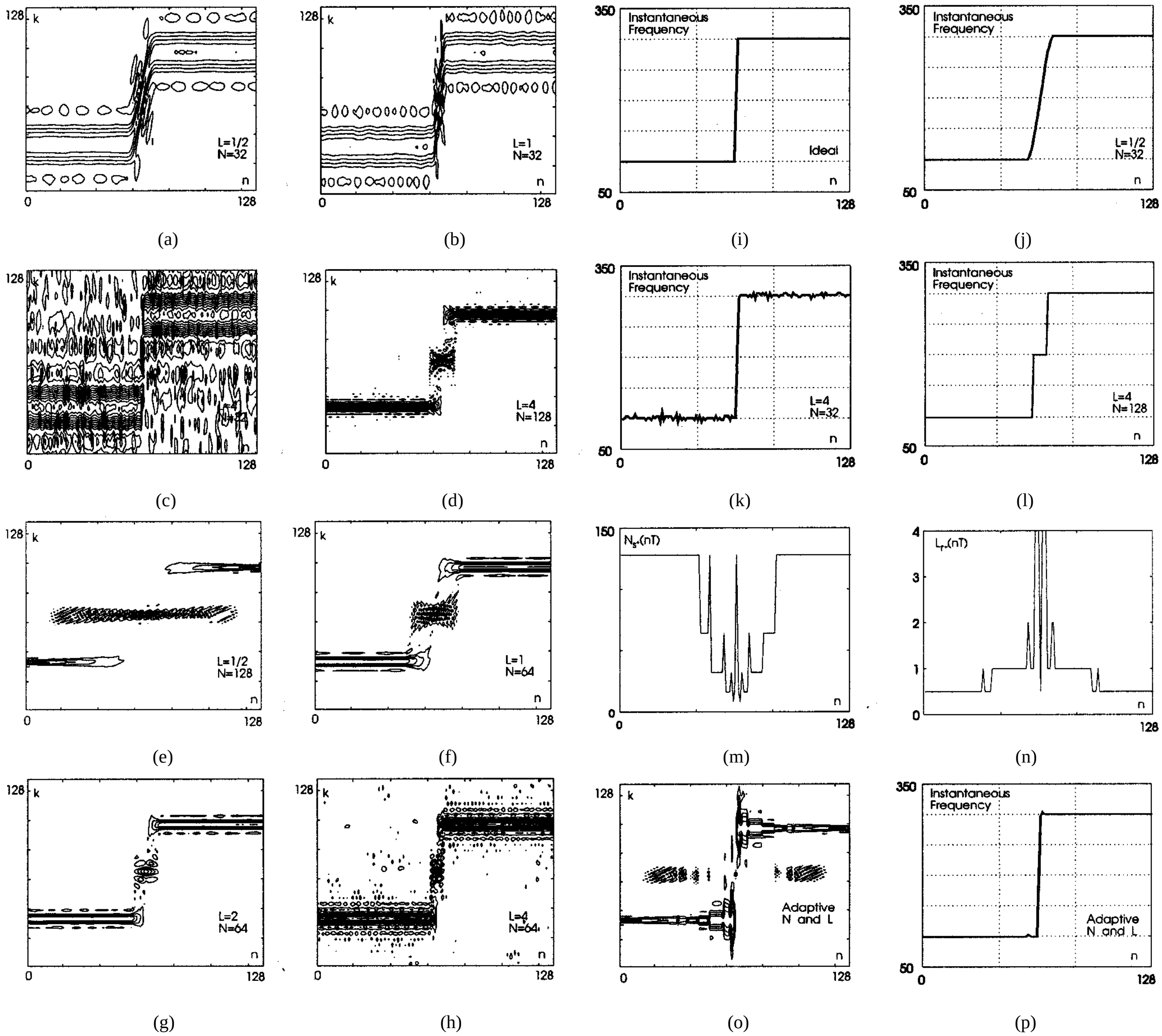

Fig. 4. Time–frequency representation and IF estimation of a signal with the step-wise IF. (a) L-WD with $L = 1/2$ and $N = 32$. (b) Wigner distribution with $N = 32$. (c) L-WD with $L = 4$ and $N = 32$. (d) L-WD with $L = 4$ and $N = 128$. (e) L-WD with $L = 1/2$ and $N = 128$. (f) Wigner distribution with $N = 64$. (g) L-WD with $L = 2$ and $N = 64$. (h) L-WD with $L = 4$ and $N = 64$. (i) Signal's IF. (j) Estimated IF using the L-WD with $L = 1/2$ and $N = 32$. (k) Estimated IF using the L-WD with $L = 4$ and $N = 32$. (l) Estimated IF using the L-WD with $L = 4$ and $N = 128$. (m) Adaptive window length. (n) Adaptive distribution order. (o) L-WD with adaptive order and window length. (p) Estimated IF using the L-WD with adaptive order and window length.

of all possible pairs (h_s, L_r). Now let us reorder the elements of $H \times \Lambda$ in such a way that we get a new set Φ whose elements $g_q = L_r^2/h_s^3$, $q = 1, 2, 3, \cdots, JK$ form an decreasing sequence

$$\Phi = \{g_q = L_r^2/h_s^3 | g_1 \geq g_2 \geq \cdots \geq g_{JK}\}. \tag{48}$$

The elements g_q form a decreasing sequence of the estimation variance (17)

$$\sigma(g_q) = \sqrt{\frac{6\sigma_\epsilon^2}{A^2}\, g_q T}. \tag{49}$$

The confidence intervals corresponding to the sequence g_q are

$$D_q = [\hat{\omega}(t) - (\kappa + \Delta k)\sigma(g_q),\, \hat{\omega}(t) + (\kappa + \Delta k)\sigma(g_q)] \tag{50}$$

and the algorithm (38)–(41) can be applied in a straightforward manner. The only difference is that the set H is replaced by the set Φ and instead of the window size selection, we find q^+ which immediately determines a pair of the corresponding (h_{s+}, L_{r+}).

The set Λ can be determined by any reasonable method. In simulation we use a dyadic set with $L_r = 2^{r-2}$, $r = 1, 2, 3, 4$. Note that the distribution with $L_r = 1$ $(r = 2)$ is the Wigner distribution, the distributions with $L_r = 2, 4$ are higher order distributions, and $L = 1/2$ would be a "lower order" distribution. (The notions "higher order" and "lower order" are used with respect to the Wigner distributions.)

VI. Examples

The discrete L-Wigner distribution is calculated using the standard FFT routines. Note that if we use $L < 1$ (for example, $L = 1/2$ as we did in this correspondence) then we have to take care about the phase continuity of $x^{1/2}(nT)$ over the π borders.

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

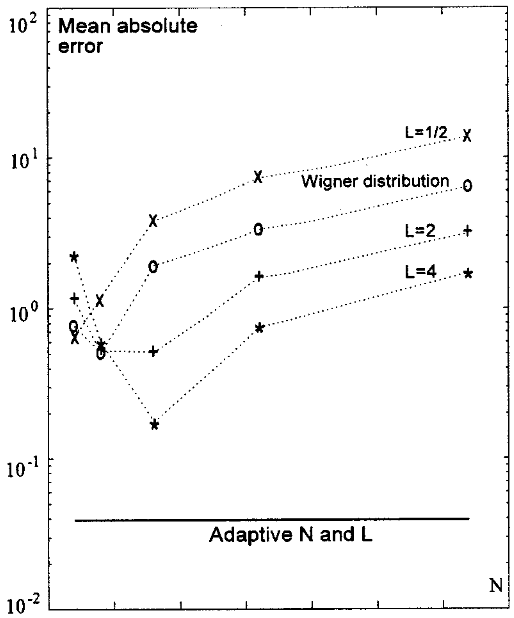

Fig. 5. Mean absolute error for various window lengths and L-Wigner distribution orders.

The signal amplitude was $A = 1$ and $20 \log(A/\sigma_\epsilon) = 15$ [dB] $(A/\sigma_\epsilon = 5.62)$. The considered time interval was $0 \le nT \le 1$.

Example 1: Signal with a nonlinear IF defined by

$$\omega(nT) = 10\pi \text{ a sinh}(100(nT - 0.5)) + 64\pi.$$

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

Since in this example we have not done any additional interpolation to find an adaptive distribution, then according to the results in Section V-A, we considered distributions with a maximal order of $L = 4$ and various window lengths h_s corresponding to $N_s = 16, 32, 64, 128, 256$ signal samples within h_s. The adaptive window lengths, determined by the algorithm (38)–(41) with $\kappa + \Delta\kappa = 3.5$, are shown in Fig. 2(g). We can see that when the IF variations are small then the algorithm uses the widest window length in order to reduce the variance. Around the point $nT = 0.5$, where the bias is large, the windows with smaller lengths are used. The L-Wigner distribution with adaptive window length is presented in Fig. 2(h). The IF, as well as its estimates with $(L = 1, N = 32)$, $(L = 4, N = 256)$, and an adaptive window length using $L = 4$, are given in Fig. 2(i)–(l), respectively. The mean absolute IF estimation error $E\{|\Delta\hat{\omega}(nT)|\}$, calculated by averaging over $0 \le nT \le 1$ and normalized to the minimal discretization step is shown in Fig. 3 for each considered distribution order and window length. This figure confirms that for each window length, the L-Wigner distribution with $L = 4$ produces the smallest error, as well as that the closest distribution to the distribution with adaptive window length (given by the solid line) is the one with $L = 4$, $N = 128$ presented in Fig. 2(f).

Example 2: A signal with step-wise IF

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

In this example we did additional interpolation for each window length, up to $N = 128$, and choose IF values such that the quantization error can be neglected. The L-Wigner distributions with some constant orders and window lengths are presented in Fig. 4(a)–(h), along with the estimated IF obtained using some of them, Fig. 4(i)–(l).

The adaptation is done with respect to both window length and distribution order, according to the algorithm (38)–(41) and hints in Section V-D. The adaptive window length $N_{s+}(nT)$ and distribution order $L_{r+}(nT)$, as well the L-Wigner distribution with the same parameters using $\kappa + \Delta\kappa = 5$, are presented in Fig. 4(m)–(o), along with the estimated IF, Fig. 4(p). As expected, the algorithm produced the smallest possible variance (with $L_{r+}(nT) = 1/2$ and $N_{s+}(nT) = 128$) in the regions where the instantaneous frequency estimator is not biased (i.e., the IF is constant). The application resulted in small window lengths and high distribution orders in the region where the bias is large, around the point $nT = 0.5$. The absolute mean error, normalized to the minimal discretization step, is shown in Fig. 5. It further illustrates our considerations about the influence of window lengths and distribution orders on the accuracy of the IF estimation. Here we will also discuss the the dependence of the optimal distribution order dependence on the window length. From Fig. 5 we see that for the narrowest window length the smallest mean absolute error is obtained with the distribution having order $L = 1/2$. The best distribution order increases with window length. For a reasonably large window length (which is important for the distribution's frequency resolution), the best results are obtained for the highest distribution order $L = 4$. This is in complete agreement with (19).

In all considered examples a normalized signal with unity amplitude is assumed. If that were not the case the IF estimation would not be influenced at all, but we would not be able to form a resulting distribution in order to combine the distributions with different orders into a resulting one. Then a normalized version of the L-Wigner distribution, introduced as the S-distribution [21], with the same properties as the L-Wigner distribution concerning the if estimation, could be used.

VII. Conclusion

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

Acknowledgment

The authors wish to thank the reviewers for very thorough reading of the correspondence and the valuable comments.

References

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

[2] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—Part 1: Fundamentals," *Proc. IEEE*, vol. 80, no. 4, 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] D. Reid, A. Zoubir, and B. Boashash, "Aircraft flight parameter estimation based on passive acoustic techniques using the polynomial Wigner–Ville distribution," *J. Acoust. Soc. Amer.*, vol. 102, pp. 207–223, July 1997.

[5] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[6] L. Cohen and C. Lee, "Instantaneous bandwidth," in *Time-Frequency Signal Analysis*, B. Boashash, Ed. London, U.K.: Longman Cheshire, 1992.
[7] L. Cohen, "Distributions concentrated along the instantaneous frequency," *SPIE, Adv. Signal Proc. Alg., Arch., Implement.*, vol. 1348, pp. 149–157, 1990.
[8] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," Dept. Elec. Eng., Technion, Haifa, Israel, Res. Rep., 5/94, Nov. 1995.
[9] ——, "Adaptive de-noising of signals satisfying differential inequalities," *IEEE Trans. Inform. Theory*, vol. 43, pp. 873–889, May 1997.
[10] V. Katkovnik, "Nonparametric estimation of instantaneous frequency," *IEEE Trans. Inform. Theory*, vol. 43, pp. 183–189, Jan. 1997.
[11] ——, "Local polynomial approximation of the instantaneous frequency: Asymptotic accuracy," *Signal Processing*, vol. 52, no. 3, pp. 343–356, 1996.
[12] ——, "Adaptive local polynomial periodogram for time-varying frequency estimation," in *Proc. IEEE-SP IS-TFTSA*, Paris, France, June 1996, pp. 329–332.
[13] 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.
[14] D. Petranović, S. Stanković, and LJ. Stanković, "Special purpose hardware for time-frequency analysis," *Electron. Lett.*, vol. 33, no. 6, pp. 464–466, Mar. 1997.
[15] K. J. Ray Liu, "Novel parallel architectures for short-time Fourier transform," *IEEE Trans. Circuits Syst.*, vol. 40, pp. 786–789, Dec. 1993.
[16] P. Rao and F. J. Taylor, "Estimation of the instantaneous frequency using the discrete Wigner distribution," *Electron. Lett.*, vol. 26, pp. 246–248, 1990.
[17] 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.
[18] LJ. 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] ——, "An analysis of the instantaneous frequency presentation using time-frequency distributions—Generalized Wigner distribution," *IEEE Trans. Signal Processing*, vol. 43, pp. 549–552, Feb. 1995.
[20] LJ. 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. 1262–1268, May 1995.
[21] ——, "Highly concentrated time-frequency distribution: Pseudo quantum signal representation," *IEEE Trans. Signal Processing*, vol. 45, pp. 543–552, Mar. 1997.
[22] ——, "A multitime definition of the Wigner higher order distribution: L–Wigner distribution," *IEEE Signal Processing Lett.*, vol. 1, pp. 106–109, July 1994.
[23] ——, "L-class of time-frequency distributions," *IEEE Signal Processing Lett.*, vol. 3, pp. 22–25, Jan. 1996.

Polynomial Cancellation Coding and Finite Differences

Katherine A. Seaton and Jean Armstrong, *Member, IEEE*

***Abstract*—We give here a mathematical context for polynomial cancellation coding, proposed recently to reduce intercarrier interference in orthogonal frequency division multiplexing (OFDM). In particular, we analyze polynomial cancellation coding (PCC) in terms of finite differences.**

***Index Terms*—Finite differences, orthogonal frequency division multiplexing, polynomial cancellation coding.**

I. Introduction

A technique has been proposed recently [1] which provides several benefits for orthogonal frequency division multiplexing (OFDM). In particular, this technique, termed *polynomial cancellation coding* (PCC), reduces the intercarrier interference (ICI) due to frequency shift between transmitter and receiver [2]. Further, PCC has been discussed in the context of reduction to out-of-band power and intersymbol interference in OFDM systems [3]. Used in its simplest form, PCC achieves these advantages at the cost of bandwidth efficiency. However, the advantages can be retained while maintaining, or even increasing, bandwidth efficiency if the symbol periods of the PCC coded data are overlapped [3] and an equalizer is used at the receiver to recover the transmitted data from the overlapped symbols.

The main idea of PCC is to map each complex number which is to be transmitted onto a group of k subcarriers, with appropriate weightings, rather than to a single subcarrier. In a previous article [2], the weightings have been given as the coefficients of the polynomial $(1-x)^{k-1}$. It has been claimed that if the same weightings are applied in decoding the received signal, polynomial variation of order $(2k-3)$ in the ICI is canceled.

In this correspondence, we explain how ICI cancellation is achieved by relating PCC to finite-difference techniques, well known in numerical analysis. In Section II, we give mathematical expressions for polynomial cancellation coding as described in [2]. In Section III we list some standard results in finite difference theory, and recast our equations in this language. Finally, in Section IV we discuss the reduced ICI obtained by use of PCC.

II. Polynomial Cancellation Coding

In the ith symbol period (length T) of an OFDM communications system, the complex numbers $a_{0,i}, \cdots, a_{N-1,i}$ modulate the N subcarriers. If we assume an ideal channel, and the local oscillator at the

Manuscript received July 17, 1998; revised June 1, 1999.
K. A. Seaton is with the School of Mathematics, La Trobe University, Bundoora VIC 3083, Australia.
J. Armstrong is with the School of Electronic Engineering, La Trobe University, Bundoora VIC 3083, Australia.
Communicated by M. L. Honig, Associate Editor for Communications.
Publisher Item Identifier S 0018-9448(00)00078-X.