

Algorithm for the Instantaneous Frequency Estimation Using Time-Frequency Distributions with Adaptive Window Width

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

Abstract—A method for minimization of the mean square error (MSE) of the instantaneous frequency estimation using time-frequency distributions, in the case of a discrete optimization parameter, is presented. It does not require knowledge of the estimation bias. The method is illustrated on adaptive window width determination in the Wigner distribution.

Index Terms— Estimation, instantaneous frequency, spectral analysis, time-frequency analysis, Wigner distribution, window optimization.

I. INTRODUCTION

INSTANTANEOUS frequency (IF) estimators based on maxima of time-frequency representations have variance and bias that are highly dependent on the lag window width. Provided that signal and noise parameters are known, by minimizing the estimation mean squared error (MSE), the optimal window width may be determined. However, those parameters are not available in advance. It is especially true for the IF derivatives that determine the estimation bias. Here, we present an adaptive algorithm, for the lag window width determination, that does not require knowledge of the estimation bias. It is assumed that the window width takes dyadic values. The discrete nature of the window width is essential for the algorithm derivation. Sliding pair-wise confidence intervals are used, instead of the intersections of all previous confidence intervals, considered in [4] and [5], from which the idea for the algorithm originated. The efficiency of the algorithm developed here is illustrated on the Wigner distribution (WD) based IF estimator, [6]. Thus, this letter may be considered as a theoretical supplement, which resulted in a modified version, of the algorithm presented in [6]. The theory and algorithm presented here are not limited to time-frequency analysis and may be quite generally used for a window (bandwidth) selection in different problems.

Manuscript received December 23, 1997. The work of Stanković was supported in part by the Alexander von Humboldt Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. D. L. Jones.

LJ. Stanković is with the Ruhr University Bochum, Signal Theory Group, D-44780 Bochum, Germany, on leave from the University of Montenegro, 81000 Podgorica, Montenegro, Yugoslavia.

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

Publisher Item Identifier S 1070-9908(98)06885-0.

II. WINDOW WIDTH OPTIMIZATION

Consider a noisy signal

$$x(nT) = s(nT) + \epsilon(nT), s(t) = A \exp(j\phi(t)) \quad (1)$$

with $s(nT)$ being a signal and $\epsilon(nT)$ being a white complex-valued Gaussian noise with mutually independent real and imaginary parts of equal variances $\sigma_\epsilon^2/2$. Consider the problem of instantaneous frequency, $\omega(t) = \phi'(t)$, estimation from discrete-time observations (1). We will assume that the IF estimation is based on maximization of a time-frequency distribution; i.e.,

$$\hat{\omega}(t) = \arg[\max_{\omega \in Q_\omega} WD(t, \omega)] \quad (2)$$

with $Q_\omega = \{\omega: 0 \leq |\omega| < \pi/(2T)\}$ being the basic interval along the frequency axis. The time-frequency distribution is denoted by $WD(t, \omega)$, since the WD is used for the algorithm demonstration. However we wish to emphasize that a wide class of time-frequency representations can be used in (2). Let $\Delta\hat{\omega}(t) = \omega(t) - \hat{\omega}(t)$ be the estimation error. The mean squared error $E\{(\Delta\hat{\omega}(t))^2\}$ is used for the accuracy characterization at the given instant t . If the estimation errors are small then provided some quite nonrestrictive assumptions the mean squared error for a wide variety of the commonly used time-frequency representations (e.g. the spectrogram, the WD and its higher order, including polynomial, versions, as well as in many nontime-frequency problems), can be represented in the following form [5]–[7]:

$$E\{(\Delta\hat{\omega}(t))^2\} = \frac{V}{h^m} + B(t)h^n \quad (3)$$

where h is a width of the symmetric lag window such that $w(t) = 0$ for $|t| > h/2$; $\sigma^2(h) = V/h^m$ and $\text{bias}(t, h) = \sqrt{B(t)h^n}$ are the variance and the bias of estimation, respectively. Parameter $B(t)$ depends on the IF derivatives. The window width h is related to the number of samples as $N = h/T$, where T is the sampling interval. In particular, for the WD with the rectangular window $m = 3, n = 4$, and $V = 6\sigma_\epsilon^2 T/A^2$ in (3) [7], [6].

It is clear that the MSE (3) has a minimum with respect to h . The corresponding optimal value of h is given by the formula $h_{\text{opt}}(t) = [mV/(nB(t))]^{1/(m+n)}$. However, this relation is not very useful in practice, mainly because on the right-hand-side it contains the bias parameter $B(t)$ depending on the derivatives of the IF which is to be estimated. The

main topic of this work is a development of the method that produces $h_{\text{opt}}(t)$ (or due to the discrete nature of h , a value of the window width as close as possible to $h_{\text{opt}}(t)$) without using $B(t)$. For the optimal window width according to (3), assuming throughout the paper without loss of generality that the bias is positive, the following holds:

$$\text{bias}(t, h_{\text{opt}}) = \sqrt{\frac{m}{n}} \sigma(h_{\text{opt}}). \quad (4)$$

Asymptotically, at least, the IF estimate $\hat{\omega}_h(t)$ is a random variable distributed around $\omega(t)$ with $\text{bias}(t, h)$ and standard deviation $\sigma(h)$. Thus we may write the relation:

$$|\omega(t) - (\hat{\omega}_h(t) - \text{bias}(t, h))| \leq \kappa \sigma(h) \quad (5)$$

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

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

$$H = \{h_s | h_s = ah_{s-1}, s = 1, 2, 3, \dots, J, a > 1\}. \quad (6)$$

The following arguments can be given in favor of such a discrete set. First of all, the discrete scheme for window widths is necessary for an efficient numerical realization. Second, implementations of the time-frequency distributions are almost always based on the FFT algorithms. The most common are the radix-2 FFT algorithms that correspond to $a = 2$, when set H gives the dyadic window width scheme, $h_s = h_0 2^s$. In the realizations the smallest window width h_0 should correspond to a small number N_0 of signal samples within it. For example, for the radix-2 fast Fourier transform (FFT) algorithms $N_0 = 4$ with $N_s = 2N_{s-1}, s = 1, 2, \dots, J$.

Now we are going to derive an algorithm for the determination of the optimal window width h_{opt} , without knowing the bias, using the IF estimates (2) and the formula for the IF estimate's variance only. It is based on the following statements.

Let H be a set of dyadic window width values, i.e., $a = 2$ in (6). Assume that the optimal window width for a given instant t belongs to this set, $h_{\text{opt}} \in H$. Define the upper and lower bounds of the confidence intervals $D_s = [L_s, U_s]$ of the IF estimates as

$$\begin{aligned} L_s &= \hat{\omega}_{h_s}(t) - (\kappa + \Delta\kappa)\sigma(h_s) \\ U_s &= \hat{\omega}_{h_s}(t) + (\kappa + \Delta\kappa)\sigma(h_s) \end{aligned} \quad (7)$$

where $\hat{\omega}_{h_s}(t)$ is an estimate of the IF, with the window width $h = h_s$ and $\sigma(h_s)$ is its variance.

Let the window width h_{s+} be determined as a width corresponding to the largest s ($s = 1, 2, \dots, J$) when two successive confidence intervals still intersect, i.e., when

$$D_s \cap D_{s+1} \neq \emptyset \quad (8)$$

is still satisfied.

Then, there exist values of κ and $\Delta\kappa$ such that $D_s \cap D_{s+1} \neq \emptyset$ and $D_{s+1} \cap D_{s+2} = \emptyset$ for $s = s^+$, when $h_{s+} = h_{\text{opt}}$, with the corresponding probability $P(\kappa) \simeq 1$ that (5) is satisfied.

Proof: Provided that $h_{\text{opt}} \in H$, the window widths belonging to H , can be represented as follows

$$h(p) = h_{\text{opt}} 2^p, \quad p = \dots, -2, -1, 0, 1, 2, \dots$$

where $p = 0$ corresponds to the window width h_{opt} , we are looking for. Note also that we use two indices for the window widths, one s (in the form h_s) which denotes the indexing which starts from the narrowest window width, and the other p (used in the form of an argument; i.e., $h(p)$ or $D(p)$) where the indexing starts from h_{opt} window width (when $p = 0$). The bias and variance for any $h(p)$, according to (3), (4), may be rewritten as

$$\begin{aligned} \sigma(h(p)) &= 2^{-pm/2} \sigma(h_{\text{opt}}), \\ \text{bias}(t, h(p)) &= 2^{pn/2} \sqrt{\frac{m}{n}} \sigma(h_{\text{opt}}). \end{aligned} \quad (9)$$

From (9) we can conclude that for $p \ll 0$ the bias is much smaller as compared to the variance, thus the estimate $\hat{\omega}_h(t)$ is spread around the exact value $\omega(t)$ with a small bias and large variance. The standard definition of a confidence interval of the estimate $\hat{\omega}_{h(p)}(t)$, for a given $h(p)$, is $\tilde{D}(p) = [\hat{\omega}_{h(p)}(t) - \kappa\sigma(h(p)), \hat{\omega}_{h(p)}(t) + \kappa\sigma(h(p))]$. In order to take into account the biasedness of the estimate $\hat{\omega}_{h(p)}(t)$ the confidence interval $\tilde{D}(p)$ is modified in the following way:

$$\begin{aligned} D(p) &= [\hat{\omega}_{h(p)}(t) - (\kappa + \Delta\kappa)\sigma(h(p)) \\ &\quad \cdot \hat{\omega}_{h(p)}(t) + (\kappa + \Delta\kappa)\sigma(h(p))] \end{aligned} \quad (10)$$

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

It is obvious that $\omega(t) \in D(p)$ for $p \ll 0$ because in this case the bias is small and the segment $D(p)$ is wider than $\tilde{D}(p)$ as $\Delta\kappa > 0$; i.e., $D(p) \cap D(p+1) \neq \emptyset$ for all $p \ll 0$ (with probability $P(\kappa)$). For $p \gg 0$ the variance is small but the bias is large. It is clear that there always exists such a large p that $D(p) \cap D(p+1) = \emptyset$ for any given $\Delta\kappa$.

The idea behind of the algorithm is that $\Delta\kappa$ in $D(p)$ can be found in such a way that the largest p for which the sequence of the pairs of the confidence intervals $D(p)$ and $D(p+1)$ has at least 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 $D(p)$ and $D(p+1)$ works as an indicator of the event $p = 0$, i.e., the event when $h_s = h_{\text{opt}}$ is found. The algorithm given in the form (7), (8) tests the intersection of the confidence intervals, where (8) is a condition that two sequential intervals D_s and D_{s+1} is the last pair of the confidence intervals having at least a point in common.

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 $D(0), D(1)$, and $D(2)$ should be considered. The confidence intervals $D(0)$ and $D(1)$ should have and the intervals $D(1)$ and $D(2)$ should not have at least a point in common. Assuming that relation (5) holds, consider the worst possible cases for the corresponding bounds. These worst-case conditions for $D(0)$ and $D(1)$ are that the minimal possible value of upper bound, denoted by $\min\{U(0)\}$, is always greater than or equal to the maximal possible value of

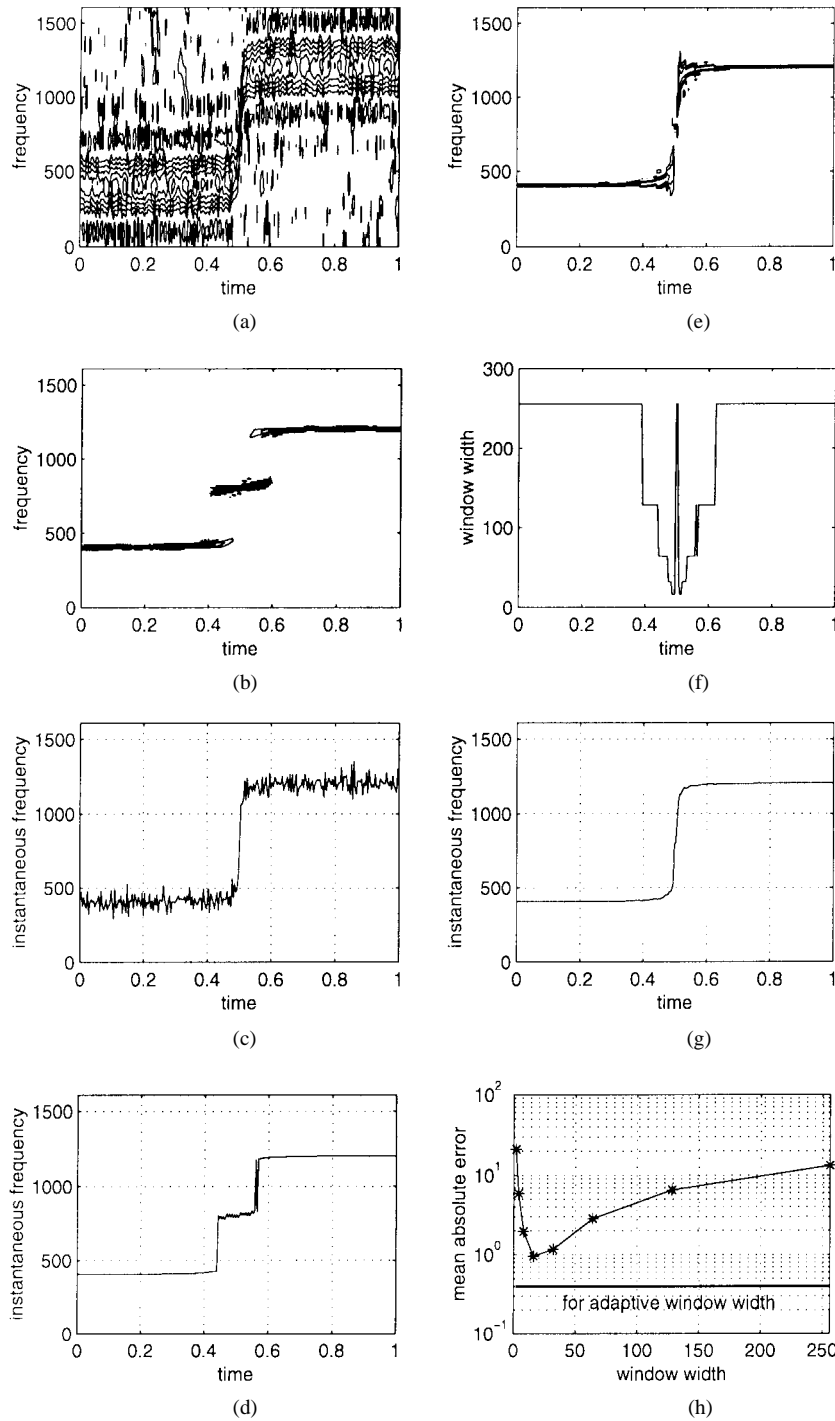


Fig. 1. Time-frequency analysis of a noisy signal. (a) Wigner distribution with $N = 16$. (b) Wigner distribution with $N = 256$. (c) Estimated instantaneous frequency using the Wigner distribution with $N = 8$. (d) Estimated instantaneous frequency using the Wigner distribution with $N = 256$. (e) Wigner distribution with adaptive window width. (f) Adaptive window width as function of time. (g) Estimated instantaneous frequency using the Wigner distribution with the adaptive window width. (h) Absolute mean error as a function of the window width; line represents the mean absolute error value for the adaptive window width.

the lower bound denoted $\max\{L(1)\}$. Analogous conditions hold for $D(1)$ and $D(2)$. These conditions may be written as

$$\begin{aligned} \min\{U(0)\} &\geq \max\{L(1)\} \\ \max\{U(1)\} &< \min\{L(2)\}. \end{aligned} \quad (11)$$

According to (5) and (10) this results in

$$\text{bias}(h(0)) + \Delta\kappa\sigma(h(0)) \geq \text{bias}(h(1)) - \Delta\kappa\sigma(h(1));$$

$$\begin{aligned} \text{bias}(h(1)) + (2\kappa + \Delta\kappa)\sigma(h(1)) \\ < \text{bias}(h(2)) - (2\kappa + \Delta\kappa)\sigma(h(2)). \end{aligned} \quad (12)$$

Having in mind (9), it can be verified that

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

is the smallest $\Delta\kappa > 0$ satisfying the first inequality in (12). With $\Delta\kappa$ from (13) the second inequality in (12) is satisfied for

$$\kappa < \sqrt{\frac{m}{n}} 2^{(m/2-1)} \frac{2^{n/2} - 1}{2^{m/2} + 1} (2^{(m+n)/2} - 1). \quad (14)$$

For the WD, which is considered as an example, we have $m = 3, n = 4$. It gives $\Delta\kappa = 1.9194$ and $\kappa < 9.8983$. The lower bound for κ is determined by the condition that $P(\kappa) \simeq 1$. Thus, we see that the conditions (11), along with the condition that $P(\kappa) \simeq 1$, can be easily satisfied. Taking, for example, a value of κ such that $6 < \kappa + \Delta\kappa < 11.8$ we get that all conditions of the statement are satisfied, as well as, $P(\kappa) > 0.9999$ for the Gaussian distribution of the error $\Delta\hat{\omega}(t) = \omega(t) - \hat{\omega}(t)$.

With (13) and (14) being satisfied we have $D(p) \cap D(p+1) \neq \emptyset$, for $p \leq 0$ and $D(p) \cap D(p+1) = \emptyset$, for $p \geq 1$, with probability $P(\kappa) \simeq 1$. This completes the proof. ■

A search of the optimal window width over a finite set H is a simplified optimization, because H consists of a relatively small number of elements. However, the discrete set of h inevitably leads to a suboptimal window width value due to the discretization of h since, in general, the optimal window width h_{opt} does not belong to H ; i.e., it cannot be written as $h_{\text{opt}} = 2^s T$. It is important to note that this discretization of h effect would also exist even if we knew in advance all of the parameters required for the optimal window width calculation, and decided to use radix-2 FFT algorithms in the realization. Then we should find h_{opt} and then use the nearest one of the form $2^s T$. Thus, the discretization of h effect is present in any case. It always results in worse values of the MSE, but that is the price of the algorithm efficiency. Fortunately, this loss of the accuracy is not significant in many cases, because the MSE (3) has a stationary point for the optimal window width $h = h_{\text{opt}}$ (and the MSE varies very slowly for the window width values close to $h = h_{\text{opt}}$).

III. EXAMPLE

The discrete pseudo-WD with the rectangular lag-window is calculated using the standard FFT routines, as $WD(k, l) = FT\{w_h(nT)x(lT + nT)x^*(lT - nT)\}$. In the example, we assumed a signal of the form $x(nT) = A \exp(j\phi(nT)) + \epsilon(nT)$, with a given IF, $\omega(nT) = 256 \arctan(250(nT - 0.5)) + 256\pi$, and the phase $\phi(nT) = T \sum_{i=0}^n \omega(iT)$. The signal amplitude was $A = 1$ and $20 \log(A/\sigma) = 10$ [dB], ($A/\sigma = 3.16$). The time interval considered was $0 \leq nT \leq 1$ with $T = 1/1024$. A set H of window widths h_s corresponding to the following number of signal samples $N = \{4, 8, 16, 32, 64, 128, 256, 512\}$ is considered. The WD is calculated from the smallest toward the wider window widths. All distributions are interpolated up to the largest number of samples in order to have the same number of frequency samples and to reduce the quantization error whose variance is $\sigma_q^2 = \pi^2/(12(NT)^2)$ and may also be included as a part of total estimation variance.

The IF is estimated using (2). According to the estimated IF $\hat{\omega}_{h_s}(t)$ and $\sigma(h_s) = \sqrt{6\sigma_\epsilon^2 T / (A^2 h_s^3)}$, the segments (10) are defined with, for example, $\kappa + \Delta\kappa = 11$, when $P(\kappa) \simeq 1$.

The estimation of signal and noise parameters A and σ_ϵ^2 can be done using $|\hat{A}|^2 + \hat{\sigma}_\epsilon^2 = (1/N) \sum_{n=1}^N |x(nT)|^2$. The variance is estimated by $\hat{\sigma}_\epsilon^2 = (\hat{\sigma}_{\epsilon_r}^2 + \hat{\sigma}_{\epsilon_i}^2)/2$ where $\hat{\sigma}_{\epsilon_r, i} = \{\text{median}(|x_{r, i}(nT) - x_{r, i}((n-1)T)| : n = 2, \dots, N)\} / 0.6745$, with $x_r(nT)$ and $x_i(nT)$ being the real and imaginary part of $x(nT)$. It is assumed that N is large, as well as T is small. For this estimation, we used signal oversampled by factor of four. The adaptive window width h_{s+} is determined as the width corresponding to the largest s ($s = 1, 2, \dots, J$) when (8) is still satisfied, i.e., when still

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

The WD's with constant window widths $N_s = 16$ and $N_s = 256$ are presented in Fig. 1(a) and (b), respectively. The IF estimates using the WD's with constant window widths $N_s = 8$ and $N_s = 256$ are given in Fig. 1(c) and (d). Fig. 1(e) shows the WD with adaptive window width whose values determined by the algorithm are given in Fig. 1(f). We can see that when the IF variations are small, the algorithm uses the widest window width in order to reduce the variance. Around the point $nT = 0.5$, where the IF variations are large, the windows with smaller widths are used. The IF estimate with adaptive window width is presented in Fig. 1(g). Mean absolute error, normalized to the discretization step, is shown in Fig. 1(h) for each considered window width. Line represents its value for the adaptive window width. Additional examples and realization details may be found in [6].

We can conclude that the adaptive window width estimation, using the algorithm derived in this letter, has lower error than the best constant-window case, which, by the way, is also not known in advance.

ACKNOWLEDGMENT

The authors are very thankful to the Associate Editor and reviewers for remarks that helped in preparing the final version of this letter.

REFERENCES

- [1] M. G. Amin, "Minimum variance time-frequency distribution kernels for signals in additive noise," *IEEE Trans. Signal Processing*, vol. 44, pp. 2352–2356, Sept. 1996.
- [2] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal-Part 1: Fundamentals," *Proc. IEEE*, vol. 80, pp. 519–538, Apr. 1992.
- [3] L. Cohen and C. Lee, "Instantaneous bandwidth," in *Time-Frequency Signal Analysis*, B. Boashash, Ed. Australia: Longman Cheshire, 1992.
- [4] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," Res. Rep., 5/94, Technion—Israel Inst. Technol., Haifa, Nov. 1995.
- [5] V. Katkovnik, "Adaptive local polynomial periodogram for time-varying frequency estimation," in *Proc. IEEE-SP IS-TFTSA*, Paris, France, June 1996, pp. 329–332.
- [6] V. Katkovnik and L.J. 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.
- [7] P. Rao and F. J. Taylor, "Estimation of the instantaneous frequency using the discrete Wigner distribution," *Electron. Lett.*, vol. 26, pp. 246–248, 1990.
- [8] L.J. Stanković, 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.