ALGORITHM FOR THE INSTANTANEOUS FREQUENCY ESTIMATION USING TIME-FREQUENCY DISTRIBUTIONS WITH ADAPTIVE WINDOW WIDTH

*LJubiša Stanković**, *Vladimir Katkovnik***
*Ruhr University Bochum, Signal Theory Group, D-44780 Bochum, Germany *
on leave from the University of Montenegro. Email: l.stankovic@ieee.org
**University of South Africa, Pretoria.

ABSTRACT

A method for the minimization of mean square error 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 the adaptive window length determination in the Wigner distribution.

1 INTRODUCTION

The instantaneous frequency (IF) estimators based on maximization of time-frequency representations, have the variance and bias which are highly dependent on the lag window length. Provided that signal and noise parameters are known then, by minimizing the estimation mean squared error, the optimal window length may be determined. But, those parameters are not available in advance. Especially it is true for the IF derivatives which determine the estimation bias. Here, we present the adaptive algorithm, for the lag window length determination, which does not require knowledge of the estimation bias. It is assumed that the window length takes dyadic values. The discrete nature of the window length is essential for the algorithm derivation. The sliding pair-wise confidence intervals are used, instead of the intersections of all previous confidence intervals, considered in [4] and [5], where the idea for algorithm originated from. The efficiency of the developed algorithm 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 algorithm presented in [6]. The theory and algorithm presented here are not limited to the time-frequency analysis and may be quit generally used.

2 WINDOW LENGTH OPTIMIZATION

Consider a noisy signal:

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

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 the 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.,

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

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. But we wish to emphasize that a wide class of time-frequency representations can be used in (2). Let $\triangle\hat{\omega}(t) = \omega(t) - \hat{\omega}(t)$ be the estimation error and the mean squared error $E\{(\triangle\hat{\omega}(t))^2\}$ 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 their higher order, including polynomial, versions, as well as in many non-time-frequency problems), can be represented in the following form [5, 6, 7, 8]

$$E\left\{(\triangle\hat{\omega}(t))^2\right\} = \frac{V}{h^m} + B(t)h^n \qquad (3)$$

where h is a width of the lag-window, $\sigma^2(h) = \frac{V}{h^m}$ is the variance and $bias(t,h) = \sqrt{B(t)h^n}$ is the bias, with parameter $B(t)$ depending on the IF derivatives. Window width h ($w(t) = 0$ for $|t| > h/2$) is related with 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 $h_{opt}(t) = (\frac{mV}{nB(t)})^{1/(m+n)}$. But, this relation is not very useful in practice, mainly because, on the left 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 paper is a development of the method which produces $h_{opt}(t)$ (or due to

This work is supported by the Alexander von Humboldt foundation

discrete nature of h, a value of the window length as close as possible to $h_{opt}(t)$) without using $B(t)$. For the optimal window length, according to (3), holds

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

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

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

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

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

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

The following arguments can be given in favor of such a discrete set. First of all, the discrete scheme for window lengths is necessary for an efficient numerical realization. Realizations of the time-frequency distributions are almost absolutely based on the FFT algorithms. The most common are the radix-2 FFT algorithms which correspond to $a = 2$, when set H gives the dyadic window length scheme, $h_s = h_0 2^s$. In the realizations the smallest window length h_0 should correspond to a small number N_0 of signal samples within it. For example, for the radix-2 FFT algorithms $N_0 = 4$ or $N_0 = 8$ with $N_s = 2N_{s-1}$, $s = 1,2,...,J$.

Now we are going to derive an algorithm for the determination of the optimal window size h_{opt}, without knowing the bias, using the IF estimates (2) 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 (6). Assume that the optimal window length belongs to this set, $h_{opt} \in H$. 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 length $h = h_s$ and $\sigma(h_s)$ is its variance.

Let the window length h_{s^+} be determined as a length corresponding to the largest s ($s = 0,1,2,...,J-1$) 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 $h_{s^+} = h_{opt}$, with the corresponding probability $P(\kappa) \simeq 1$ that (5) is satisfied.

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

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

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

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

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. A confidence interval of the estimate $\hat{\omega}_{h(p)}(t)$, for a given $h(p)$, is defined by $\tilde{D}(p) = [\hat{\omega}_{h(p)}(t) - \kappa\sigma(h(p)), \hat{\omega}_{h(p)}(t) + \kappa\sigma(h(p))]$. 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:

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

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., $D(p) \cap D(p+1) \neq 0$ 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 always exist 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 a point in common is $p = 0$. Such value of $\Delta\kappa$ exists because the bias and the variance are monotonic increasing and decreasing functions of h respectively. As soon as this value of $\Delta\kappa$ is found, an intersection of the confidence intervals $D(p)$ and $D(p+1)$ works as an indicator of the event $p = 0$, i.e., the event when $h_s = h(0) = h_{opt}$ 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 the relation (5) holds, consider the worst possible cases for the corresponding bounds. These, worst case conditions for $D(0)$ and $D(1)$ (assuming also, without loss of generality, that the bias is positive) are that the minimal possible value of upper bound, denoted by $\min\{U(0)\}$, is always greater or equal to the maximal possible value of lower bound denoted $\max\{L(1)\}$. The analog conditions hold for $D(1)$ and $D(2)$. These conditions may be written as:

$$\begin{array}{l} \min\{U(0)\} \geq \max\{L(1)\}; \\ \max\{U(1)\} < \min\{L(2)\}, \end{array} \tag{11}$$

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

$$\begin{array}{c} bias(h(0)) + \Delta\kappa\sigma(h(0)) \geq bias(h(1)) - \Delta\kappa\sigma(h(1)), \\ bias(h(1)) + (2\kappa + \Delta\kappa)\sigma(h(1)) < \\ bias(h(2)) - (2\kappa + \Delta\kappa)\sigma(h(2)) \end{array} \tag{12}$$

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

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

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

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

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 $\kappa + \Delta\kappa = 8$ we get that all conditions of the statement are satisfied, as well as, $P(\kappa) > 0.999$ for the Gaussian distribution of the error $\triangle\hat{\omega}(t) = \omega(t) - \hat{\omega}(t)$.

With (13),(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 of the statement. □

A search of the optimal window length over finite set H is simplified optimization, because H (6) consists of a relatively small number of elements. However, the discrete set of h inevitably leads to a suboptimal window length value due to the discretization of h effects, since, in general, the optimal window length h_{opt} does not belong to H, i.e., can not be written as $h_{opt} = 2^s T$. It is important to note that this effect, due to the discrete nature of $h \in H$, would also exist even if we knew in advance all of the parameters required for the optimal window length calculation, and decided to use radix-2 FFT algorithms in the realization. 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 efficient calculation schemes using FFT algorithms. Fortunately, this loss of the accuracy is not significant in many cases, because the MSE (3) has a stationary point for the optimal window length $h = h_{opt}$ (and the MSE varies very slowly for the window length values close to $h = h_{opt}$, see Fig.1h).

3 EXAMPLE

The discrete pseudo WD

$$WD(l,k) = \sum_{n=-N/2}^{N/2-1} w_h(nT) x(lT+nT) x^*(lT-nT) e^{-j2\frac{2\pi}{N}nk}$$

is calculated using the standard FFT routines. In the example we assumed signal of the form $x(nT) = A\exp(j\phi(nT)) + \epsilon(nT)$, with a given IF

$$\omega(nT) = 64\arctan(500(t-0.5)) + 64\pi$$

and the phase $\phi(nT) = \sum_{i=0}^{n} \omega(nT)/T$. Signal amplitude was $A = 1$ and $20\log(A/\sigma) = 10[dB]$, $(A/\sigma = 3.16)$. Considered time interval was $0 \leq nT \leq 1$.

A set H of window lengths h_s corresponding to the following number of signal samples $N = \{8, 16, 32, 64, 128, 256, 512\}$ is considered. The WD is calculated from the smallest toward the wider window lengths. The IF is estimated using (2). According to the estimated IF $\hat{\omega}_{h_s}(t)$ and $\sigma(h_s) = \sqrt{\frac{6\sigma_\varepsilon^2 T}{A^2 h_s^3}}$, the segments (10) are defined with, for example, $\kappa + \Delta\kappa = 8$, when $P(\kappa) > 0.999$. The estimation of signal and noise parameters A and σ_ε^2 can be done using

$$\left|\hat{A}\right|^2 + \hat{\sigma}_\varepsilon^2 = \frac{1}{N} \sum_{n=1}^{N} |x(nT)|^2,$$

where N is assumed to be large, as well as T is small. The variance is estimated by

$$\begin{array}{c} \hat{\sigma}_{\varepsilon r} = \frac{\{\text{median}(|x_r(nT) - x_r((n-1)T)| : n=2,..,N)\}}{0.6745\sqrt{2}}. \\ \hat{\sigma}_{\varepsilon i} = \frac{\{\text{median}(|x_i(nT) - x_i((n-1)T)| : n=2,..,N)\}}{0.6745\sqrt{2}} \\ \hat{\sigma}_\varepsilon^2 = \hat{\sigma}_{\varepsilon r}^2 + \hat{\sigma}_{\varepsilon i}^2 \end{array}$$

where $x_r(nT)$ and $x_i(nT)$ are real and imaginary part of signal $x(nT)$, respectively. The adaptive window length h_{s^+} is determined as the length corresponding to the largest s $(s = 1, 2, ..., J)$ when (8) is still satisfied , i.e., when still

$$\left|\hat{\omega}_{h(p)}(t) - \hat{\omega}_{h(p+1)}(t)\right| \leq (\kappa + \Delta\kappa)(\sigma(h(p) + \sigma(h(p+1)).$$

The WDs with constant window lengths $(N = 32, 256)$ are presented in Fig.1a),c). The adaptive window lengths, determined by the algorithm, are shown in Fig.1d. 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 WD with adaptive window length is presented in Fig.1b. The IF estimates using constant window widths $N = 16$, and $N = 256$, and estimate with adaptive window length, are given in Fig.1e),g),f), respectively. Absolute mean error, normalized to the minimal discretization step, for each considered window length, is shown in Fig.1h.

4 CONCLUSION

We can conclude that the instantaneous frequency estimation using the adaptive window length, according to the algorithm derived in this letter, has lower MSE than the best constant window case, which also is not known in advance. The algorithm may be used in a wide range of time-frequency representations.

Figure 1: a) Time-frequency representation using constant window length N=32, b) Using adaptive window, c) Using constant window length N=512, d) Adaptive window length as a function of time, e) IF estimated using window length N=16, f) IF estimated using adaptive window length, g) IF estimated using N=256, h) Mean absolute error for various fixed window lengths (denoted by *) and for the adaptive window length (line).

References

[1] M.G.Amin, "Minimum variance time-frequency distribution kernels for signals in additive noise", *IEEE Trans.SP*, vol.44, Sept.1996, pp.2352-2356.

[2] B.Boashash: "Estimating and interpreting the instantaneous frequency of a signal-Part 1: Fundamentals", *Proc.IEEE*, vol.80, April 1992, pp.519-538.

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

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

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

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

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

[8] D.C.Reid, A.M.Zoubir, B.Boashash:"Aircraft flight parameter estimation based on passive acoustic techniques using the polynomial Wigner-Ville distribution", *J.Acoust.Soc.Am.,* vol.102, July 1997, pp.207-223.

[9] LJ.Stanković, S.Stanković: "On the Wigner distribution of discrete-time noisy signals with application to the study of quantization effects", *IEEE Trans.SP*, vol-42, July 1994, pp.1863-1867.