Adaptive Instantaneous Frequency Estimation Using TFDs

LJubiša Stanković

Instantaneous frequency (IF) estimators based on maxima of time-frequency representations have variance and bias which are highly dependent on the lag window width. The optimal window width may be determined by minimizing the estimation mean squared error (MSE), provided that some signal and noise parameters are explicitly known. However, these parameters are not available in advance. This is especially true for the IF derivatives which determine the estimation bias. In this article, an adaptive algorithm for the lag window width determination, based on the confidence intervals intersection, will be presented [3]-[7]. This algorithm does not require knowledge of the estimation bias value. The theory and algorithm presented here are not limited to the IF estimation and time-frequency analysis. They may be applied to a parameter value selection in various problems.

A. Optimal Window Width

Consider a noisy signal:

$$x(n\Delta t) = s(n\Delta t) + \epsilon(n\Delta t),$$
$$s(t) = a\exp(j\phi(t)), \qquad (1)$$

with $s(n\Delta t)$ being a signal and $\epsilon(n\Delta t)$ being a white complex-valued Gaussian noise with mutually independent real and imaginary parts of equal variances $\sigma_\epsilon^2/2$. Sampling interval is denoted by Δt. Consider the problem of the IF, $f_i(t) = \phi'(t)/2\pi$, estimation from the discrete-time observations $x(n\Delta t)$, based on maxima of a time-frequency distribution $\rho_x(t, f)$,

$$\hat{f}(t) = \arg\{\max_f \rho_x(t, f)\}. \qquad (2)$$

Let $\triangle\hat{f}(t) = f_i(t) - \hat{f}(t)$ be the estimation error. The MSE, $E\{(\triangle\hat{f}(t))^2\}$, is used

for the accuracy characterization at a given time instant t. Asymptotically, the MSE for commonly used time-frequency representations (e.g. the spectrogram, the pseudo Wigner distribution (WD), and its higher order versions) can be expressed in the following form [1], [4]-[7] [Articles 10.3, 10.4]

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

where h is a lag window $w_h(t)$ width, such that $w_h(t) = 0$ for $|t| > h/2$. It is related to the number of samples N by $h = N\Delta t$. The variance and the bias of estimate, for a given h, are

$$\sigma^2(h) = V/h^m, \quad \text{bias}(t, h) = \sqrt{B(t)h^n}. \quad (4)$$

The expression for $B(t)$ is a function of the IF derivatives.

For example, for the WD with a rectangular lag window we have [5]

$$E\left\{(\triangle\hat{f}(t))^2\right\} = \frac{6\sigma_\epsilon^2\Delta t}{(2\pi a)^2}\frac{1}{h^3} + \left(\frac{\phi^{(3)}(t)}{80\pi}\right)^2 h^4, \quad (5)$$

corresponding to $m = 3$ and $n = 4$ in (3). Values of m and n for some other distributions are indicated in Table 1, according to the results from [4]-[7].

The MSE in (3) has a minimum with respect to h. This minimum occurs for the optimal value of h given by

$$h_{opt}(t) = [mV/(nB(t))]^{1/(m+n)}. \qquad (6)$$

Note that this relation is not useful in practice, because its right hand-side contains $B(t)$ which depends on derivatives of the unknown IF.

B. Adaptive Algorithm

Here, we present an adaptive method which can produce an estimate of $h_{opt}(t)$ without having to know the value of $B(t)$. For the optimal window width, according to (3), holds

$$\frac{\partial E\left\{(\triangle \hat{f}(t))^2\right\}}{\partial h} =$$

$$-m\frac{V}{h^{m+1}} + nB(t)h^{n-1} = 0_{|h=h_{opt}}. \quad (7)$$

Multiplying (7) by h, we get the relationship between the bias and standard deviation, (4), for $h = h_{opt}$,

$$\text{bias}(t, h_{opt}) = \sqrt{\frac{m}{n}}\sigma(h_{opt}). \quad (8)$$

It will be assumed, without loss of generality, that the bias is positive. The IF estimate $\hat{f}_h(t)$ (obtained from (2) by using the lag window of width h) is a random variable distributed around the true IF $f_i(t)$ with the bias bias (t, h) and the standard deviation $\sigma(h)$. Thus, we may write the relation:

$$\left| f_i(t) - \left(\hat{f}_h(t) - \text{bias}(t, h)\right) \right| \leq \kappa\sigma(h), \quad (9)$$

where the inequality holds with probability $P(\kappa)$ depending on parameter κ.[1] We will assume that κ is such that $P(\kappa) \to 1$.

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

$$H = \{h_s \mid h_s = 2h_{s-1}, \ s = 1, 2, ..., J\}. \quad (10)$$

Define the confidence intervals $D_s = [L_s, U_s]$ of the IF estimates, with the following upper and lower bounds

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

$$U_s = \hat{f}_{h_s}(t) + (\kappa + \Delta\kappa)\sigma(h_s), \quad (11)$$

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

[1]If we assume, for example, that the random variable $\hat{f}_h(t)$ is Gaussian, with the mean value $M = f_i(t) +$ bias (t, h) and the standard deviation $\sigma(h)$, then the probability that $\hat{f}_h(t)$ takes a value within the interval $[M - \kappa\sigma(h), M + \kappa\sigma(h)]$ is $P(\kappa) = 0.95$ for $\kappa = 2$, and $P(\kappa) = 0.997$ for $\kappa = 3$.

standard deviation. Assume that a window width denoted by $h_{s^+} \in H$ is of h_{opt} order, $h_{s^+} \sim h_{opt}$. Since h_{opt} does not correspond to any h_s from the set H, for the analysis that follows we can write $h_{s^+} = 2^p h_{opt}$, where p is a constant close to 0. According to (10) all other windows can be written as a function of h_{s^+} as

$$h_s = h_{s^+}2^{(s-s^+)} = h_{opt}2^{s-s^++p},$$

$$(s - s^+) = ..., -2, -1, 0, 1, 2, ... \quad (12)$$

With this notation, having in mind (8), the standard deviation and the bias from (4) can be expressed by

$$\sigma(h_s) = \sqrt{V/h_s^m} = \sigma(h_{opt})2^{-(s-s^++p)m/2},$$

$$\text{bias}(t, h_s) = \sqrt{B(t)h_s^n} =$$

$$\sqrt{m/n}\,\sigma(h_{opt})2^{(s-s^++p)n/2}. \quad (13)$$

For small window widths h_s, when $s \ll s^+$, the bias of $\hat{f}_{h_s}(t)$ is negligible, thus $f_i(t) \in D_s$ (with probability $P(\kappa + \Delta\kappa) \to 1$). Then, obviously, $D_{s-1} \cap D_s \neq \emptyset$, since at least the true IF, $f_i(t)$, belongs to both confidence intervals. For $s \gg s^+$ the variance is small, but the bias is large. It is clear that for bias $(t, h_s) \neq 0$ there exists such a large s that $D_s \cap D_{s+1} = \emptyset$ for a finite $\kappa + \Delta\kappa$.

The **idea behind the algorithm** is that $\Delta\kappa$ in D_s can be found in such a way that the largest s, for which the sequence of the pairs of the confidence intervals D_{s-1} and D_s has at least a point in common, is $s = s^+$. Such a value of $\Delta\kappa$ exists because the bias and the variance are monotonically increasing and decreasing functions of h, respectively, (13). As soon as this value of $\Delta\kappa$ is found, an intersection of the confidence intervals D_{s-1} and D_s,

$$\left|\hat{f}_{h_{s-1}}(t) - \hat{f}_{h_s}(t)\right| \leq (\kappa + \Delta\kappa)[\sigma(h_{s-1}) + \sigma(h_s)], \quad (14)$$

works as an indicator of the event $s = s^+$, i.e., the event $h_s = h_{s^+} \sim h_{opt}$. The value of h_{s^+} is the last h_s when (14) is still satisfied.

B.1 Parameters in the Adaptive Algorithm

There are three possible approaches to choosing algorithm parameters κ, $\Delta\kappa$, and p. Their performance do not differ significantly.

1) When our knowledge about the variance and bias behavior, given by (3), is not quite reliable, an approximative approach for κ, $\Delta\kappa$, and p determination can be used. Then, we can assume a value of $\kappa \cong 2.5$, such that $P(\kappa) \cong 0.99$ for Gaussian distribution of estimation error. The value of $\Delta\kappa$ should take into account the bias for the expected optimal window width (8). It is common to assume that, for the optimal value of h, the bias and variance are of the same order, resulting in $\Delta\kappa \cong 1$. Then we can expect that the obtained value h_{s^+} is close to h_{opt}, thus $p \cong 0$, and all parameters for **the key algorithm equation (14)** are defined. This simple heuristic form has been successfully used in [4],[5], and *it is highly recommended for most of practical applications.* Estimation of the standard deviation $\sigma(h_s)$ will be discussed within the Numerical example.

2) When the knowledge about the variance and bias behavior is reliable, i.e., when (3) accurately describes estimation error, then we can calculate all algorithm parameters. According to the algorithm basic idea, only three confidence intervals, D_{s^+-1}, D_{s^+}, and D_{s^++1}, should be considered. The confidence intervals D_{s^+-1} and D_{s^+} **should have,** while D_{s^+} and D_{s^++1} **should not have,** at least one point in common. Assuming that relation (9) holds, and that the bias is positive, this condition means that the minimal possible value of upper D_{s^+-1} bound, (11), denoted by $\min\{U_{s^+-1}\}$, is always greater than or equal to the maximal possible value of the lower D_{s^+} bound, denoted by $\max\{L_{s^+}\}$, i.e., $\min\{U_{s^+-1}\} \geq \max\{L_{s^+}\}$. The condition that D_{s^+} and D_{s^++1} do not intersect is given by $\max\{U_{s^+}\} < \min\{L_{s^++1}\}$. According to (9) and (11) the above analysis results in

$$\text{bias}\,(h_{s^+-1}) + \Delta\kappa\sigma(h_{s^+-1}) \geq$$
$$\text{bias}\,(h_{s^+}) - \Delta\kappa\sigma(h_{s^+}),$$
$$\text{bias}\,(h_{s^+}) + (2\kappa + \Delta\kappa)\sigma(h_{s^+}) <$$
$$\text{bias}\,(h_{s^++1}) - (2\kappa + \Delta\kappa)\sigma(h_{s^++1}). \qquad (15)$$

Since the inequalities are written for the worst case, we can calculate the algorithm parameters by using the corresponding equalities. With (13) we get

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

$$2^p = \left[\frac{\Delta\kappa\sqrt{n/m}\left(2^{m/2} + 1\right)}{1 - 2^{-n/2}} \right]^{2/(m+n)}. \qquad (16)$$

Values of the parameters $\Delta\kappa$ and p for various distributions, i.e., for various values of m and n, are given in Table 1.

For further, and very **fine tuning of the algorithm parameters**, one may want that the adaptive window is unbiased in logarithmic, instead of in linear scale (due to definition (10)). The estimation bias and variance are exponential functions with respect to m and n, (13). Thus the confidence interval limits vary as $2^{(s-s^+)(m+n)/2}$. The mean value for this exponential function, for two successive confidence intervals, for example $(s - s^+) = 0$ and $(s - s^+) = 1$, is $(1 + 2^{(m+n)/2})/2$. It is shifted with respect to the geometrical mean $\sqrt{2^{(m+n)/2}}$ of these two intervals, by approximately $\Delta p \cong \left[\log_2\left((1 + 2^{(m+n)/2})/2\right)\right]\frac{2}{m+n} - \frac{1}{2}$, resulting in the total logarithmic shift $p_1 = p + \Delta p$, presented in Table 1. Therefore the adaptive window width (as an estimate of the optimal window width) should be $\hat{h}_{opt} = h_{s^+}/2^{p_1}$.

Note that the set H of window widths h is a priori assumed. Therefore, as long as we can calculate p_1, we can use it in the following ways: a) To calculate distribution with the new window width $h_a = h_{s^+}/2^{p_1}$ as the best estimate of h_{opt}, b) To remain within the assumed set of $h_s \in H$, and to decide only whether to correct the obtained h_{s^+} or not. For example, if $|p_1| \leq 1/2$ the correction is smaller than the window discretization step. Thus, we can use $h_a = h_{s^+}$. For $1/2 < p_1 \leq 3/2$ it is better to use $h_a = h_{s^+}/2 = h_{s^+-1}$, as the adaptive window width value. Fortunately, the loss of accuracy for the adaptive widths h_a, as far as they are of h_{opt} order, is not significant since the MSE varies slowly around its stationary point. Thus, in numerical implementations we can use only the lag windows from the given set H.

TABLE I

PARAMETERS IN THE ADAPTIVE ALGORITHM FOR VARIOUS m, n, κ: $m = 3, n = 4$ FOR THE SPECTROGRAM, WIGNER AND L-WIGNER DISTRIBUTION BASED IF ESTIMATORS; $m = 3, n = 8$ FOR THE FOURTH ORDER POLYNOMIAL WIGNER-VILLE DISTRIBUTION, AND LOCAL POLYNOMIAL DISTRIBUTION BASED IF ESTIMATORS; $m = 1, n = 4$ FOR THE WIGNER DISTRIBUTION AS A SPECTRUM ESTIMATOR.

m	1	1	3	3	3	3	3	3
n	4	4	4	4	4	8	8	8
κ	2	3	2	3	5	2	3	5
$\Delta\kappa$	0.86	1.29	0.39	0.58	0.97	0.09	0.14	0.23
p	0.99	1.22	0.34	0.51	0.72	-0.13	-0.03	0.11
p_1	1.18	1.41	0.59	0.76	0.97	0.19	0.30	0.43

C. Illustration

We have simulated the IF estimates as a random variable

$$\hat{f}_h(t) = \mathbf{a}\sqrt{V/h^m} + \sqrt{B(t)h^n} + f_i(t), \quad (17)$$

having the MSE given by (3), where $\mathbf{a} = \mathcal{N}(0,1)$ is a Gaussian (zero-mean, unity-variance) random variable, $m = 3$, $n = 4$, and $V = 1$. For the true IF value $f_i(t)$, at a given t, any constant can be assumed. The bias parameter $B(t)$ in $\hat{f}_h(t)$ logarithmically varies within $\frac{1}{7}\log_2(mV/nB(t)) \in [-4, 3]$, with step 0.05.

-For each value of parameter $B(t)$ we have calculated optimal window width according to (6), and plotted $\log_2(h_{opt})$ as a thick line in Fig. 1.

-The value of $\hat{f}_h(t)$ was simulated for each $B(t)$ and $h_s \in H$. The assumed set of possible window widths was $H = \{1/16, 1/8, 1/4, 1/2, 1, 2, 4, 8, 16, 32\}$, and $\kappa = 2$. The key algorithm relation (14) was tested each time, with the known standard deviation $\sigma(h_s) = \sqrt{V/h_s^m}$. The largest value of h_s when the key equation (14) was still satisfied was denoted by h_{s+}. Value $\Delta\kappa = 0.39$, corresponding to $m = 3$, $n = 4$, $\kappa = 2$, was used (Table 1). The adaptive values $h_a = h_{s+}/2^{p_1}$, $p_1 = 0.59$ (Table 1), produced in this way, are connected with the optimal window line, by thin vertical lines in Fig. 1.

-The same simulation is repeated with $\kappa = 3$ and $\kappa = 5$.

-We can conclude that the presented algorithm almost always chooses the width h_s from H which is the nearest to the optimal one.

However, for relatively small $\kappa = 2$ there are few complete misses of the optimal window width, since (9) is satisfied only with probability $P(2) = 0.95$. For $\kappa = 2$, two successive confidence intervals do not intersect when the bias is small, producing false result, with probability of $2(0.05)^2 \sim 10^{-2}$ order.

3) The third approach for the parameter $(\kappa + \Delta\kappa)$ estimation is based on the statistical nature of confidence intervals, and a posteriori check of the fitting quality [4]. This approach is beyond the scope of this article.

D. Numerical Example

In the example we assumed a signal of (1) form, with the given IF,

$$f_i(n\Delta t) = 128 \arctan(250(n\Delta t - 0.5))/\pi + 128,$$

and the phase $\phi(n\Delta t) = 2\pi\Delta t \sum_{m=0}^{n} f_i(m\Delta t)$. The signal amplitude was $a = 1$, and $20\log(a/\sigma_\epsilon) = 10[dB]$ ($a/\sigma_\epsilon = 3.16$). Considered time interval was $0 \leq n\Delta t \leq 1$, with $\Delta t = 1/1024$. The IF is estimated by using the discrete WD with a rectangular lag-window, $W_x^h(t, f) = DFT_{n \to f}[w_h(n\Delta t)x(t + n\Delta t) x^*(t - n\Delta t)]$, calculated with the standard FFT routines.

The algorithm is implemented as follows:

1) A set H of window widths h_s, corresponding to the following number of signal samples $\mathbf{N} = \{4, 8, 16, 32, 64, 128, 256, 512\}$, is assumed. In order to have the same number of frequency samples, as well as to reduce the quantization error, all windows are zero-padded up to the maximal window width.
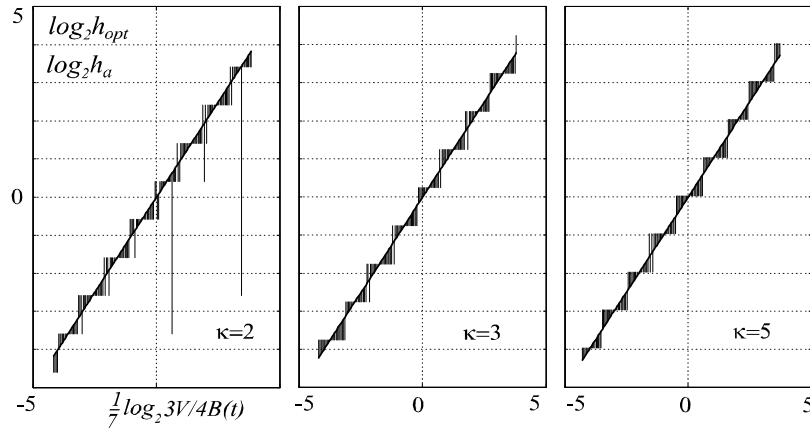
Fig. 1. Optimal window width (straight thick line), and adaptive window widths (end of the vertical lines, starting from the optimal window width line) for $m = 3$, $n = 4$, $V = 1$. The variance to bias ratio $V/B(t)$ is logarithmically varied. The adaptive width $h_a = h_{s+}/2^{p_1}$ is obtained from h_{s+}, according to (14), after correction for the corresponding values of p_1, given in Table 1.

2) For a given time instant $t = n\Delta t$, the WDs are calculated starting from the smallest toward the wider window widths h_s.

3) The IF is estimated using equation (2) and $W_x^{h_s}(t, f)$.

4) The confidence intervals intersection, (14), is checked for the estimated IF, $\hat{f}_{h_s}(t)$, and $\sigma(h_s) = \sqrt{3\sigma_\epsilon^2 \Delta t/(2\pi^2 a^2 h_s^3)}$ with, for example, $\kappa + \Delta\kappa = 6$, when $p_1 \cong 1$, and $P(\kappa) \to 1$ (see Table 1, and the Comment that follows).

5) The adaptive window width $h_a = h_{s+}/2$ is obtained from the last $h_s = h_{s+}$ when (14) is still satisfied. Back to 2).

Comment: Estimation of the signal and noise parameters a and σ_ϵ^2 can be done by using $|\hat{a}|^2 + \hat{\sigma}_\epsilon^2 = \frac{1}{N}\sum_{n=1}^{N} |x(n\Delta t)|^2$. The variance is estimated by $\hat{\sigma}_\epsilon^2 = \hat{\sigma}_{\epsilon r}^2 + \hat{\sigma}_{\epsilon i}^2$, where

$$\hat{\sigma}_{\epsilon r, i} = \frac{\underset{n=2,..,N}{\text{median}}(|x_{r,i}(n\Delta t) - x_{r,i}((n-1)\Delta t)|)}{0.6745\sqrt{2}}, \quad (18)$$

with $x_r(n\Delta t)$ and $x_i(n\Delta t)$ being the real and imaginary part of $x(n\Delta t)$. It is assumed that N is large, and Δt is small [4]-[7]. For this estimation we oversampled the signal by factor of four.

The WDs with constant window widths $N_s = 16$ and $N_s = 256$ are presented in Fig.2(a), and Fig.2(b), respectively. The IF estimates using the WDs with constant window widths $N_s = 8$, and $N_s = 256$ are given in Fig.2(c) and Fig.2(d). Fig.2(e) shows the

WD with adaptive window width. Values of the adaptive window width, determined by the algorithm, are presented in Fig.2(f). We can see that when the IF variations are small the algorithm uses the widest window in order to reduce the variance. Around the point $n\Delta t = 0.5$, where the IF variations are fast, the windows with smaller widths are used. The IF estimate with adaptive window width is presented in Fig.2(g). Mean absolute error, normalized to the discretization step, is shown in Fig.2(h) for each considered window width. The line represents value of the mean absolute error for the adaptive window width.

E. Conclusion

The algorithm that can produce accurate estimate of the optimal window width, without using the bias value, is presented. The IF estimates obtained by using this algorithm and the WD have lower error than by using the best constant-window width, which also is not known in advance. Additional examples, including distributions with adaptive order, the WD as a spectrum estimator, algorithm application to the sensor array signal tracking, as well as other realization details can be found in [2], [4]-[7]. The presented algorithm can be used in various other similar problems.
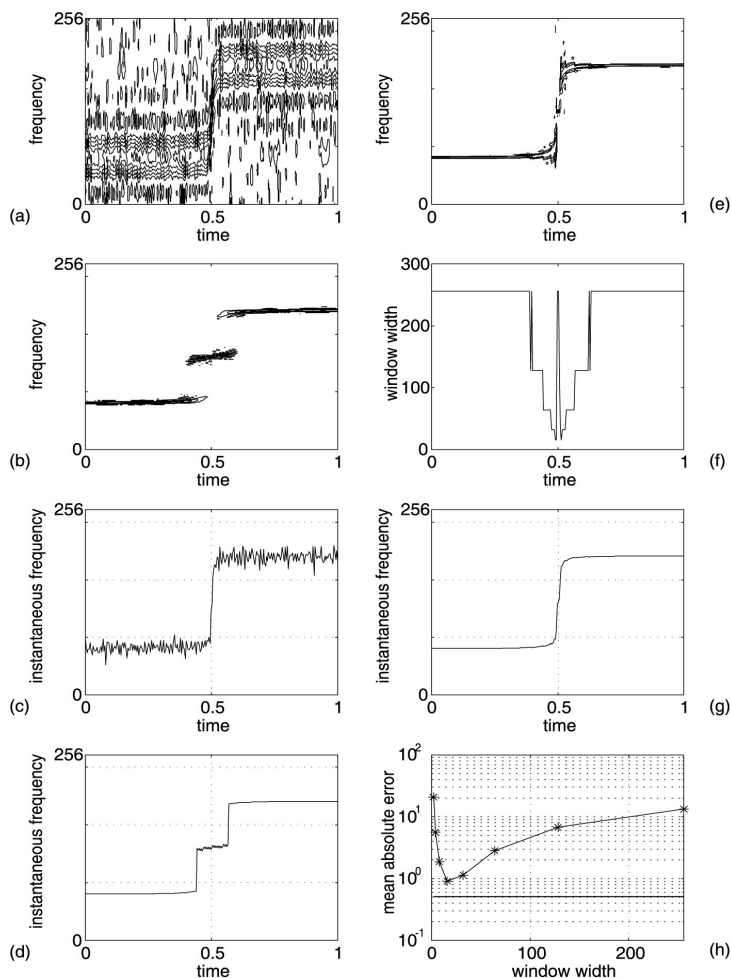
Fig. 2. Time-frequency analysis of a noisy signal: a) Wigner distribution with $N = 16$, b) Wigner distribution with $N = 256$, c) Estimated instantaneous frequency using the Wigner distribution with $N = 8$, d) Estimated instantaneous frequency using the Wigner distribution with $N = 256$, e) Wigner distribution with adaptive window width, f) Adaptive window width as a 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; The line represents the mean absolute error value for the adaptive window width.

REFERENCES

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

[2] A.B. Gershman, LJ. Stanković and V.Katkovnik, "Sensor array signal tracking using a data-driven window approach," *Sig. Proc.*, vol.80, pp.2507-2515, Dec. 2000.

[3] A. Goldenshluger and A. Nemirovski, "Adaptive denoising of signals satisfying differential inequalities," *IEEE Trans. IT*, vol.43, pp.873-889, May 1997.

[4] V. Katkovnik and LJ. Stanković, "Periodogram with varying and data-driven window length," *Sig. Proc.*, vol. 67, pp. 345-358, no.3, 1998.

[5] V. Katkovnik and LJ. Stanković, "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length," *IEEE Trans. SP,* vol.46, pp.2315-2325, Sept. 1998.

[6] LJ. Stanković and V. Katkovnik, "Instantaneous frequency estimation using the higher order L-Wigner distributions with the data driven order and window length," *IEEE Trans. IT*, vol.46, pp.302-311, Jan. 2000.

[7] LJ. Stanković and V. Katkovnik, "The Wigner distribution of noisy signals with adaptive time-frequency varying window," *IEEE Trans. SP*, vol.47, pp.1099-1108, April 1999.