

Periodogram with varying and data-driven window length

Vladimir Katkovnik^a, LJubiša Stanković^{b,*},¹

^a *Department of Statistics, University of South Africa, P.O. Box 392, Pretoria, South Africa*

^b *Signal Theory Group, Ruhr University Bochum, 44780 Bochum, Germany*

Received 24 September 1997; received in revised form 6 February 1998

Abstract

The short-time Fourier transform and the corresponding periodogram give biased estimates of the instantaneous frequency (IF) if the IF in question is a nonlinear function of time. In the case of noisy signals, the optimal choice of the window length, based on asymptotic formulae for the variance and bias, can resolve the bias–variance trade-off usual for nonparametric estimation. However, the practical value of such optimal estimator is not significant since the optimal window length depends on the unknown smoothness of the IF. The main goal of this paper is to develop an adaptive, periodogram-based IF estimator with a time-varying and data-driven window length which is able to provide the accuracy close to the one that could be achieved if the smoothness of the IF were known in advance. The developed algorithm uses only the estimates of the IF and the formula for the variance of these estimates. Simulation shows good accuracy ability of the adaptive algorithm. © 1998 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Die Kurzzeit-Fouriertransformation und das entsprechende Periodogramm liefern biasbehaftete Schätzer der Momentanfrequenz (IF), wenn diese eine nichtlineare Funktion der Zeit ist. Bei verrauschten Signalen kann eine optimale Wahl der Fensterlänge, basierend auf einer asymptotischen Formulierung für Varianz und Bias, dem üblichen Bias–Varianz Verlust bei nicht-parametrischen Schätzungen begegnen. Der praktische Wert eines solchen Optimalschätzers ist allerdings nicht signifikant, solange die optimale Fensterlänge von der unbekanntem Glätte der Momentanfrequenz (IF) abhängig ist. Ziel dieses Papiers ist es nun, einen adaptiven, auf Periodogrammen basierenden Schätzer für die Momentanfrequenz zu entwickeln, wobei eine zeitvariierende und datenabhängige Fensterlänge betrachtet wird. Mit diesem Schätzer ist man in der Lage, Aussagen über die Genauigkeit der Schätzung zu machen, und zwar in ähnlicher Weise, wie es bei Kenntnis der Glätte der IF möglich wäre. Der hier entwickelte Algorithmus benutzt nur die Schätzungen der IF und einen Ausdruck für die Varianz dieser Schätzungen. Simulationen zeigen die guten Eigenschaften des adaptiven Algorithmus in Bezug auf die Schätzgenauigkeit. © 1998 Elsevier Science B.V. All rights reserved.

Résumé

La transformation de Fourier à court terme et le périodogramme correspondant fournissent des estimées biaisées de la fréquence instantanée (IF) si l'IF en question est une fonction nonlinéaire du temps. Dans le cas de signaux bruités le

* Corresponding author. E-mail: l.stankovic@ieee.org.

¹ On leave from the University of Montenegro, Podgorica, Montenegro, Yugoslavia.

choix optimal de la longueur de fenêtre, basé sur des formulations asymptotiques de la variance et du biais, peut résoudre le compromis biais–variance usuel de l'estimation non-paramétrique. Toutefois, la valeur pratique d'un tel estimateur optimal n'est pas significative du fait que la longueur de fenêtre optimale dépend du lissage inconnu de l'IF. Le but principal de cet article est de développer un estimateur de l'IF adaptatif basé sur le périodogramme, caractérisé par une longueur de fenêtre variant dans le temps selon les données, et capable d'atteindre une précision proche de celle qui serait obtenue si le lissage de l'IF était connue à l'avance. L'algorithme développé utilise seulement les estimées de l'IF et la formule de la variance de ces estimées. Une simulation met en évidence la bonne précision atteignable par cet algorithme adaptatif. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Periodogram; Noisy signals; Instantaneous frequency; Window function; Time–frequency analysis

1. Introduction

A complex-valued harmonic with the time-varying phase is a key-model of the instantaneous frequency (IF) concept, as well as an important model in the general theory of time–frequency distributions. It has been utilized to study a wide range of signals, including speech, music, acoustic, biological, radar, sonar and geophysical ones [12].

An overview of the methods for the IF estimation, as well as the interpretation of the IF concept itself, is presented in [1]. Beside other efficient time–frequency representation-based techniques for the IF estimation, in particular the general Cohen class of quadratic distributions and the Wigner distribution [2,3,13,14], the short-time Fourier transform (STFT) represents a very efficient and commonly applied approach. Its advantage is based on the linearity of the STFT with respect to observations, a quite clear interpretation, and very efficient calculation tools using the fast Fourier transform (FFT).

Here, we will focus our attention to the standard STFT only and follow the nonparametric approach developed in [7,8,10]. The key-point of this approach is an assumption that the IF is an arbitrary nonparametric function of time, as well as that the bias of estimation is comparable with the random error. In particular, if the IF is time-invariant then its estimation, using the periodogram is unbiased. It is also unbiased for the linear in time IF, with an asymptotically small window length. However, the periodogram based estimates are biased for a general nonlinear IF. In the case of noisy signals, the periodogram based estimates are signal and noise dependent. In terms of the nonparametric approach

we can, theoretically, find the optimal window length in the periodogram and resolve the bias–variance trade-off usual for nonparametric estimation. However, the optimal window length depends on the unknown smoothness of the IF, making this approach practically useless.

The main goal of this paper is to develop an adaptive estimator with a time-varying and data-driven window length that is able to provide the accuracy close to the one what could be achieved if the smoothness of the IF was known in advance. The idea of the approach developed in this paper originated from [4], where it was proposed and justified for the local polynomial fitting of regression. For the time-varying IF estimation this approach was used in [9] and [11], where the algorithm with the time-varying and data-driven window length was presented for the local polynomial periodogram and the Wigner distribution estimators, respectively. This approach uses only the formula for the variance of the estimate, which does not require information about the smoothness of the IF to be known in advance. Simulations based on the discrete periodogram, with several noisy signal examples, show a good accuracy ability of the presented adaptive algorithm, as well as an improvement in the periodogram based time–frequency representation of signals with the nonlinear IF.

The structure of the paper is as follows. The periodogram as an IF estimator is considered in Section 2. The asymptotic bias and variance of the IF estimate, along with the optimal window length, are presented in that section, as well. The adaptive estimate of the IF with a time-varying and data-driven window length is developed in Section 3. The numerical implementation of the adaptive

algorithm, along with simulation results, is discussed in Section 4. A data-driven adjustment of the threshold parameter of the adaptive algorithm is considered in Section 5.

2. Background theory

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

$$y(nT) = m(nT) + \varepsilon(nT), \quad \text{with } m(t) = Ae^{j\phi(t)}, \quad (1)$$

where n is an integer, T is a sampling interval and $\varepsilon(nT)$ is a complex-valued white Gaussian noise with i.i.d real and imaginary parts. Thus, $\text{Re}(\varepsilon)$ and $\text{Im}(\varepsilon) \sim \mathcal{N}(0, \sigma_\varepsilon^2/2)$ and the total variance of the noise is equal to σ_ε^2 . The IF, by definition, is the first derivative of the phase

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

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

The STFT, in the discrete-time domain, is defined by

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

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

The IF estimator is a solution of the optimization problem:

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

where

$$I_h(\omega, t) = |F_h(\omega, t)|^2 \quad (5)$$

is the periodogram, with $Q_\omega = \{\omega : 0 \leq |\omega| < \pi/2T\}$ being the basic interval along the frequency axis. If the IF may assume only nonnegative values, then one may consider $Q_\omega = \{\omega : 0 \leq \omega < \pi/2T\}$.

Two different hypotheses about the IF are used:
 (a) $\omega(t)$ is a smooth infinitely differentiable function of time with bounded derivatives

$$|\omega^{(s)}(t)| < M < \infty, \quad s = 1, 2, \dots \quad (6)$$

(b) $\omega(t)$ is a smooth differentiable function with a uniformly bounded first derivative

$$\sup_t |\omega^{(1)}(t)| \leq M_1. \quad (7)$$

Assumption (7) is not so restrictive as (6) because it concerns only one, the first-order, derivative.

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

$$\Delta \hat{\omega}_h(t) = \omega(t) - \hat{\omega}_h(t). \quad (8)$$

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

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

$$\text{var}(\Delta \hat{\omega}_h(t)) = \frac{\sigma_\varepsilon^2}{2|A|^2} \frac{T E}{h^3 F^2}, \quad (9)$$

while the estimate's bias may be represented in the following two forms, for the hypotheses (6) and (7) respectively,

$$E(\Delta \hat{\omega}_h(t)) = \sum_{s=1}^{\infty} h^{2s} b_s \omega^{(2s)}(t), \quad (10)$$

$$|E(\Delta \hat{\omega}_h(t))| \leq h b_1^A M_1, \quad (11)$$

provided that $h^{2s} |b_s \omega^{(2s)}(t)|$ for all $s \geq 1$ and $h b_1^A M_1$ are small.

The following notation has been used in (9)–(11):

$$F = \int_{-1/2}^{1/2} w(t)t^2 dt, \quad (12)$$

$$E = \int_{-1/2}^{1/2} w^2(t)t^2 dt,$$

$$b_s = \frac{-1}{(2s+1)!F} \int_{-1/2}^{1/2} w(t)t^{2s+2} dt,$$

$$b_1^A = \frac{1}{2F} \int_{-1/2}^{1/2} |w(t)| \cdot |t|^3 dt.$$

For the rectangular window we have $F = E$. The formulae of the proposition are special cases of the more general Proposition presented in [10] for the accuracy analysis of the higher order local polynomial periodogram with the rectangular window.

Comments

1. Formulae (9)–(12) give, in a clear analytical form, a dependence of the variance and the bias on the sampling interval T , the window $w(t)$ form, the signal amplitude A and the derivatives of the IF. While formula (11) presents the upper bound of the bias, formula (10) is accurate for the estimation bias. Error-coefficients b_s determine a contribution of the derivatives $\omega^{(2s)}(t)$ to the bias.
2. The main assumption for formulae (9)–(11) is that the variance and the bias are small. If the window is rectangular and $h \rightarrow 0$, $T \rightarrow 0$, $h^3/T \rightarrow \infty$, then the following explicit formulae can be given for the variance and the error-coefficients:

$$\text{var}(\Delta\hat{\omega}_h(t)) = \frac{6\sigma_\varepsilon^2 T}{|A|^2 h^3}, \quad E = F = 1/12,$$

$$b_s = \frac{-3}{(2s + 1)!(2s + 3)2^{2s}}, \quad b_1^4 = 3/16. \quad (13)$$

The values of F , E , b_s and b_1^4 may be easily obtained for others than the rectangular window type (e.g. Table 1).

3. Let us consider the mean square error (MSE) of the estimation, provided that hypotheses (6) holds. It follows from Eqs. (10) and (13) that, for a small h , the main term of the MSE can be given in the form

$$E(\Delta\hat{\omega}_h(t))^2 = \frac{6\sigma_\varepsilon^2 T}{|A|^2 h^3} + \left(\frac{1}{40}h^2\omega^{(2)}(t)\right)^2. \quad (14)$$

Table 1
Variance and bias coefficients

Window type	F	E	b_1	b_1^4
Rectangular	0.0833	0.0833	− 0.0250	0.1875
Hanning	0.0163	0.0075	− 0.0131	0.1324
Triangular	0.0208	0.0083	− 0.0167	0.1500
Hamming	0.0217	0.0093	− 0.0168	0.1493

It is clear from Eq. (14) that decreasing the window length h results in decreasing of the bias and in increasing of the variance. Consequently, an increase of the window length h increases the bias and decreases the variance.

Optimization on h in Eq. (14), minimizing the MSE, results in

$$h_{\text{opt}}(t) = \left[\frac{7200\sigma_\varepsilon^2 T}{|A|^2(\omega^{(2)}(t))^2} \right]^{1/7}. \quad (15)$$

The optimal window length $h_{\text{opt}}(t)$ gives an optimal bias–variance trade-off, usual for nonparametric estimations. It depends on the signal-to-noise ratio A/σ_ε , the sampling interval T and the second IF derivative $\omega^{(2)}(t)$. Thus the optimal, or even reasonable choice of length h , depends on the IF second derivative $\omega^{(2)}(t)$, which is naturally unknown because the IF itself is to be estimated. At this moment, it is important to emphasize that if the second derivative $\omega^{(2)}(t)$ is quite different for different t then the optimization of the estimation accuracy requires a time-varying window length $h(t)$. The accuracy improvement from this time-varying $h(t)$ can be significant as compared with any time-invariant length h .

In a similar way, for hypotheses (7), we obtain from Eqs. (11) and (13) that

$$E(\Delta\hat{\omega}_h(t))^2 \leq \frac{6\sigma_\varepsilon^2 T}{|A|^2 h^3} + \left(\frac{3}{16}hM_1\right)^2 \quad (16)$$

and

$$h_{\text{opt}}(t) = \left[\frac{256\sigma_\varepsilon^2 T}{|A|^2 M_1^2} \right]^{1/5}, \quad (17)$$

with the optimal window length depending on the upper bound M_1 for the first derivative of the IF.

Thus, the optimization results highly depend on the hypotheses about the IF to be estimated. In this paper, we propose a data-driven choice of the time-varying window length that uses only the formula for the variance and does not need information about the unknown derivatives of the IF determining the bias. As a matter of fact, this choice of the window length is based on

quite a specific statistic that has been used in order to compare the variance of the estimate versus its bias.

4. Let the MSE be represented in two forms, corresponding to the hypotheses (6) and (7), respectively,

$$E(\Delta\hat{\omega}_h(t))^2 = \text{var}(\Delta\hat{\omega}_h(t)) + (h^2 b_s \omega^{(2)}(t))^2, \quad (18)$$

or

$$E(\Delta\hat{\omega}_h(t))^2 \leq \text{var}(\Delta\hat{\omega}_h(t)) + (hb_1^A \omega^{(1)}(t))^2, \quad (19)$$

and $h_{\text{opt}}(t)$ be the optimal window length minimizing the MSE (18) or (19). It is easy to see that the corresponding optimal values of the MSE are of the form

$$E(\Delta\hat{\omega}_{h_{\text{opt}}}(t))^2 = \text{var}(\Delta\hat{\omega}_{h_{\text{opt}}}(t)) \cdot (1 + 3/4),$$

where

$$|E(\Delta\hat{\omega}_{h_{\text{opt}}}(t))| = \sqrt{\text{var}(\Delta\hat{\omega}_{h_{\text{opt}}}(t))} \cdot \sqrt{3}/2 \quad (20)$$

or

$$E(\Delta\hat{\omega}_{h_{\text{opt}}}(t))^2 \leq \text{var}(\Delta\hat{\omega}_{h_{\text{opt}}}(t)) \cdot (1 + 3/2), \quad (21)$$

where

$$|E(\Delta\hat{\omega}_{h_{\text{opt}}}(t))| \leq \sqrt{\text{var}(\Delta\hat{\omega}_{h_{\text{opt}}}(t))} \cdot \sqrt{3}/2.$$

Thus the optimal MSE and its upper bound are proportional to the estimation variance with the coefficient equal to $(1 + 3/4)$ and $(1 + 3/2)$ according to two hypotheses for the IF.

Note also that for the optimal window length the ratio of the bias to the standard deviation of the estimate is a constant independent on t and all other parameters of the periodogram. Values of this ratio follow from Eq. (20)

$$\frac{|E(\Delta\hat{\omega}_{h_{\text{opt}}}(t))|}{\sqrt{\text{var}(\Delta\hat{\omega}_{h_{\text{opt}}}(t))}} = \sqrt{3}/2 \quad (22)$$

and from Eq. (21)

$$\frac{|E(\Delta\hat{\omega}_{h_{\text{opt}}}(t))|}{\sqrt{\text{var}(\Delta\hat{\omega}_{h_{\text{opt}}}(t))}} \leq \sqrt{3}/2, \quad (23)$$

respectively.

3. Algorithm of data-driven window length choice

3.1. Basic idea

The basic idea follows from the accuracy analysis given in the Proposition. Namely, at least for the asymptotic case, the estimation error can be represented as a sum of the deterministic (bias) and random component, with the variance given by Eq. (9). The estimation error can be written as

$$|\omega(t) - \hat{\omega}_h(t)| \leq |\text{bias}(t,h)| + \kappa\sigma(h), \quad (24)$$

with $\sigma^2(h) = \sigma_\varepsilon^2 E/(2|A|^2 h^3 F^2)$. Inequality (24) holds with probability $P(\kappa)$, where κ is the corresponding quantile of the standard Gaussian distribution $N(0,1)$. The usual choice $\kappa = 2$ gives $P(\kappa) = 0.95$.

It follows from Eqs. (10) and (11) that $|\text{bias}(t,h)| \rightarrow 0$ as $h \rightarrow 0$. Now, let $h = h_s$ be so small that

$$|\text{bias}(t,h_s)| \leq \kappa\sigma(h_s), \quad (25)$$

then

$$|\omega(t) - \hat{\omega}_{h_s}(t)| \leq 2\kappa\sigma(h_s). \quad (26)$$

It is obvious that, for a set of such small h_s , all of the segments

$$D_s = [\hat{\omega}_{h_s}(t) - 2\kappa\sigma(h_s), \hat{\omega}_{h_s}(t) + 2\kappa\sigma(h_s)], \quad (27)$$

have a point in common, namely $\omega(t)$.

Consider an increasing sequence of h_s , $h_1 < h_2 < \dots$. Let h_{s^*} be the largest of those h_s for which the segments D_{s-1} and D_s have a point in common. Let us call this window length h_{s^*} ‘optimal’ and determine the IF estimates with data-driven optimal window length as $\hat{\omega}_{h_{s^*}}(t)$.

The basic idea behind this choice is as follows. If the segments D_{s-1} and D_s do not have a point in common it means that at least one of the inequalities (26) does not hold, i.e. the bias is too large as compared with the standard deviation in Eq. (25). Thus, the statistical hypotheses to be tested for the bias is given in the form of the sequence of inequalities (26) and the largest length h_s for which these inequalities have a point in common is considered as a bias-variance compromise, when the bias and variance are of the same order (see Eqs. (22) and (23)).

3.2. Algorithm

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

$$H = \{h_s | h_1 < h_2 < h_3 < \dots < h_J\}. \quad (28)$$

In general, any reasonable choice of H is acceptable. In particular, the lengths with dyadic numbers $N_s = 2N_{s-1}$ of observations within the window length, until the largest h_J is reached, will be assumed. This scheme corresponds to the radix-2 FFT algorithms. Note that the relation between the window length and the number of observation within that length is $h_s = N_s T$.

The following steps are generated for each t .

1. The periodogram is calculated for all of $h_s \in H$. Thus, we obtain a set of periodograms for a fixed time instant t

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

The IF estimates are found as

$$\hat{\omega}_{h_s}(t) = \arg \left[\max_{\omega \in Q_\omega} I_{h_s}(\omega, t) \right]. \quad (29)$$

2. The upper and lower bounds of the confidence intervals D_s in Eq. (27) are built as follows:

$$U_s(t) = \hat{\omega}_{h_s}(t) + 2\kappa\sigma(h_s), \quad (30)$$

$$L_s(t) = \hat{\omega}_{h_s}(t) - 2\kappa\sigma(h_s).$$

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

$$L_s(t) \leq U_{s-1}(t) \quad \text{and} \quad U_s(t) \geq L_{s-1}(t), \quad (31)$$

i.e.,

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

is still satisfied. Here $L_s(t)$ and $U_s(t)$ are determined by Eq. (30).

Then this s^+ is the largest of those s for which the segments D_{s-1} and D_s , $s \leq J$, have a point in common. The optimal window length is

chosen as

$$\hat{h}(t) = h_{s^+}(t) \quad (32)$$

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

4. The periodogram with the optimal window length is

$$I^+(\omega, t) = I_{\hat{h}(t)}(\omega, t). \quad (33)$$

Steps 1–4 are repeated for each considered time-instant t .

Comments on the algorithm

- (a) Provided that T is small the following estimates of $|A|$ and σ_e^2 can be used:

$$|\hat{A}|^2 + \hat{\sigma}_e^2 = \frac{1}{N} \sum_{n=1}^N |y(nT)|^2, \quad (34)$$

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

$$\hat{\sigma}_{er} = \frac{1}{0.6745\sqrt{2}} \{ \text{median}(|y_r(nT) - y_r((n-1)T)|; n=2; \dots, N) \}, \quad (35)$$

$$\hat{\sigma}_{ei} = \frac{1}{0.6745\sqrt{2}} \{ \text{median}(|y_i(nT) - y_i((n-1)T)|; n=2; \dots, N) \}, \quad (36)$$

$$\hat{\sigma}_e^2 = \hat{\sigma}_{er}^2 + \hat{\sigma}_{ei}^2, \quad (37)$$

where $y_r(nT)$ and $y_i(nT)$ are the real and imaginary part of $y(nT)$, respectively, as well as T is sufficiently small. The average $\frac{1}{N} \sum_{n=1}^N |y(nT) - y((n-1)T)|^2$ could also be used as an estimate of σ_e^2 . However, we prefer a median in Eq. (37) as a robust estimate.

(b) We wish to emphasize that the intersection of the segments D_s is quite a specific statistic for examining the bias. The point is that the segment's shift, defined by the IF estimate $\hat{\omega}_{\hat{h}(t)}(t)$, which determine the intersection (and as a result the s^+ value) is quite sensitive with respect to the estimation bias. It increases when the window size is increased. Thus, s^+ gives a desirable bias–variance compromise.

Originally the intersection of the confidence intervals D_s as a statistic for window length selection

has been proposed in [4] (see also [5]), where it was done for the local polynomial fitting of regression. The algorithm given in [4] is different from the one described above because it is based on the intersection of all of the segments D_s starting from $s = 1$ up to the largest s^+ when all of those segments have a point in common. The strong convergence of results, given in [4], proves that the estimator with the proposed time-varying and data-driven window length is able to provide a nearly optimal quality, close to the one what could be achieved if the smoothness of the IF was known in advance.

In contrast to [4], in our algorithm we apply the sliding pair-wise intersections (31) of the pairs of segments D_s and D_{s-1} for $s = 2, 3, \dots, J$. The simulation shows that the results for the IF estimation in this way are quite accurate and prospective.

(c) The confidence interval, in form (27), corresponds to the bias–variance compromise determined by inequality (25). This sort of compromise results in the product by a factor of 2 for $\kappa\sigma(h_s)$ in Eqs. (26) and (27). It is clear that this compromise can be quite different from those which appears in the MSE optimization (20)–(21).

Note also that as a matter of fact the product 2κ , where the quantile κ depends on the probability P , is a special threshold parameter of the algorithm. Let us denote this parameter $\Gamma = 2\kappa$. A selection of the value of Γ can significantly influence the accuracy of estimation. The constant value of Γ will be assumed in Section 4, while the data-drive selection of this parameter is discussed in Section 5.

4. Numerical implementation and examples

The discrete-time STFT, given by Eq. (3), is discretized over frequency and calculated as

$$F_h(k, l) = \sum_{n=-N/2}^{N/2-1} w_h(nT)y(lT + nT)e^{-j2\pi kn/N}, \quad (38)$$

where N is a number of samples determined by the sampling interval T (i.e. signal’s maximal expected frequency $\omega_m = \pi/T$) and the window length $h = NT$.

For the dyadic window length $h_s = N_s T$, $h_s \in H$, we have the corresponding number of samples $N_s \in N_H = \{N_s | N_s = 2N_{s-1}\}$. Starting with a narrow

window, for example $N_1 = 4$, we calculate $F_h(k, l)$ for all $N_s \in \{4, 8, 16, 32, 64, 128\}$ up to the widest window (in this case $N_J = 128$). For a given instant lT , the IF is estimated as

$$\hat{\omega}(lT, N_s) = \frac{\pi}{N_s T} \cdot \arg \left[\max_{k \in Q_k} I_{h_s}(k, l) \right], \quad (39)$$

where $Q_k = \{k : 0 \leq k < N/2\}$ for nonnegative-only IF values. In order to reduce the FFT quantization error, in the cases of small N_s , as well as to have distributions of the same length for different N_s , the sequence $y(lT + nT)$ is zero-padded up to the length N_J , i.e. $F_h(k, l)$ is interpolated in the frequency domain up to N_J samples within the period. Then, the discrete STFT, given by Eq. (38), is calculated using the standard FFT routines. Having in mind the interpolation, the IF estimation is given by

$$\hat{\omega}(lT, N_s) = \frac{\pi}{N_J T} \cdot \arg \left[\max_{k \in Q_k} I_{h_s}(k, l) \right]. \quad (40)$$

The quantization error has the variance $\frac{1}{12}(\pi/(N_J T))^2 = \pi^2/(12h_J^2)$. This quantization error variance can be large as compared with the variance (9) of the IF estimate. In order to reduce the quantization error, we may additionally interpolate distribution $F_h(k, l)$. In the numerical realizations we have done the interpolation with factor of 2.

After h_{s^+} , i.e. N_{s^+} , is found, according to the IF estimate (40) and algorithms (30)–(31), the calculation for the time-instant $t = lT$ can be stopped in order to save computational time. Note also that for each next (twice wider) window length we just replace $N_{s-1} = N_s/2$ zero values (added by zero padding) by the values of $y(lT + nT)$. Therefore, the computations with various window lengths are not completely independent. This may be a source of an additional computational time saving.

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

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

with the given IF $\omega(nT)$ and the phase $\phi(nT) = T \cdot \sum_{i=0}^n \omega(iT)$. It has been assumed that $A = 1$ and $20 \log(A/\sigma_\varepsilon) = 15$ [dB], ($A/\sigma_\varepsilon = 5.62$) and the estimation time interval $0 \leq nT \leq 1$.

The value 1.75 of parameter κ , slightly less than $\kappa = 2$, has turned out to be a good choice for the practical realizations. This value of κ corresponds to probability 0.93 of the inequality (24) for the Gaussian estimation error.

Example 1. The step-wise IF

$$\omega(nT) = 64\pi + 32\pi \operatorname{sgn}(nT - 0.5), \quad (41)$$

with a step in the IF at the instant $nT = 0.5$ is considered first.

The estimates $\hat{\omega}_h(nT)$ of the IF with $N_s = 4$, $N_s = 8$ and the widest window length $N_s = 128$ are presented in Fig. 1(a)–Fig. 1(c), respectively. The IF estimate $\hat{\omega}_{h^+}(t)$ with the adaptive, time-varying, window length is given in Fig. 1(d). The adaptive window length $N^+(nT) = h_s^+(nT)/T$, as a function of time instant n , is presented in Fig. 2(b). We may conclude that at the instants which are comparatively far from the jump-point $t = 0.5$, the adaptive window definitely shows a tendency to be as wide

as possible. However, at and around the jump-point $t = 0.5$ the adaptive window is narrowed up to the smallest length value. Thus, it demonstrates that the data-driven window length is really quite sensitive with respect to the fast IF variations.

This is in a complete agreement with Eq. (15). When the IF is fast varying then the bias is dominant and the narrowest window is selected by the algorithm.

The mean absolute error for the estimate $\hat{\omega}_{h^+}(t)$ and for the estimates with the time-invariant window lengths $N_s = [4, 8, 16, 32, 64, 128]$, normalized with the frequency step $\pi/(N_s T)$, are given in Table 2. Fig. 1(d) shows that the IF estimate $\hat{\omega}_{h^+}(t)$ gives a quite accurate IF estimation for all time instants $n = t/T$. Fig. 1(c), corresponding to $N_s = 128$, demonstrates the estimate curve that looks to be close to the one presented in Fig. 1(d). But, the figures in Table 2 show that the accuracy of the adaptive estimator is much higher.

The time–frequency signal representation using the periodogram with the constant window lengths $N_s = 8$, $N_s = 128$ and the adaptive window length, is shown in Fig. 2(a), 2(c) and 2(d), respectively. We may see that, in the case of the window with adaptive

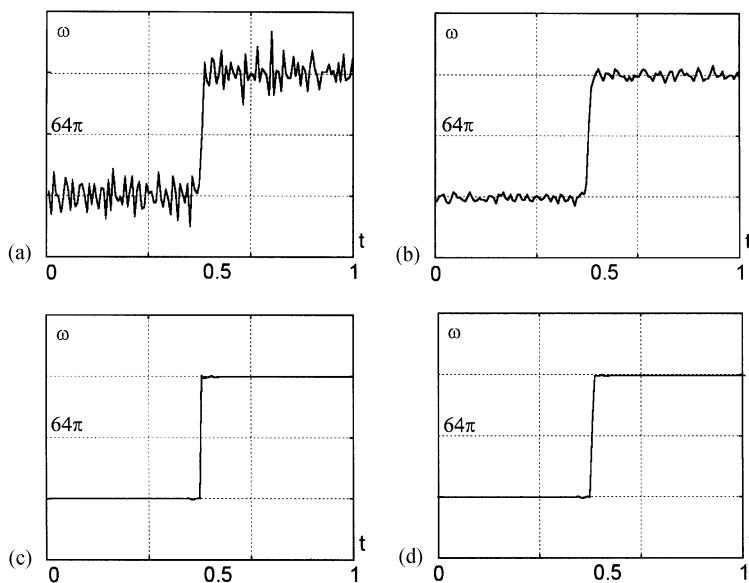


Fig. 1. Instantaneous frequency estimation using the periodogram with: (a) $N_s = 4$, (b) $N_s = 8$, (c) $N_s = 128$, (d) adaptive window length.

Table 2
Normalized mean absolute error for the stepwise constant IF

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

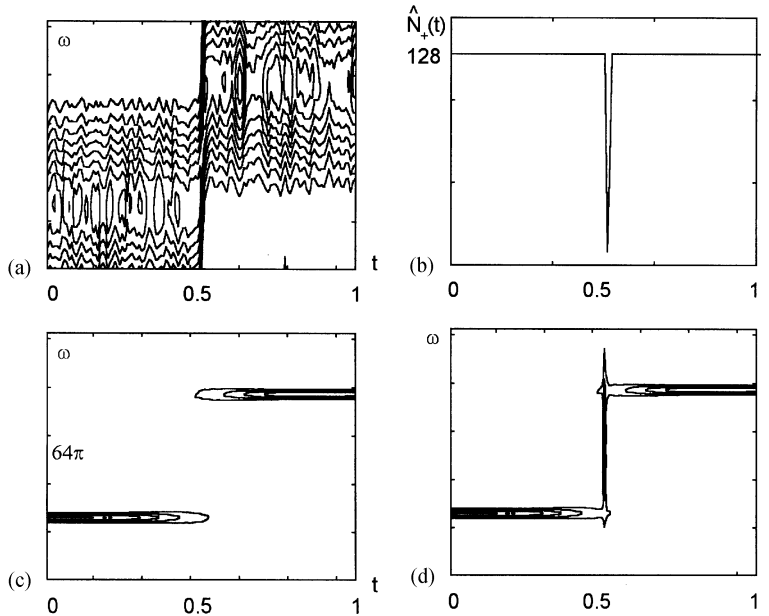


Fig. 2. (a) Periodogram with $N_s = 4$, (b) adaptive window length, (c) periodogram with $N_s = 128$, (d) periodogram with adaptive window length.

length, the periodogram is highly concentrated along the instantaneous frequency within the entire time interval considered.

Example 2. The piecewise linear frequency modulated signal with a step in the IF's first derivative,

$$\omega(nT) = 128\pi|nT - 0.5| + 32\pi, \quad (42)$$

is considered in this example. The results of the IF estimation are shown in Fig. 3 and Table 3.

Here, the explanations are similar to those in Example 1.

Note that in these two examples we have assumed that the IF values are mainly placed on the discrete

time–frequency grid (after the described interpolation up to the largest window length, and then additionally with a factor of 2). It significantly decreases the quantization errors and sometimes gives zero-error estimates. The periodograms (with parameters as those in Fig. 2), along with the adaptive window lengths, are presented in Fig. 4.

Example 3. Consider now a nonlinear frequency modulated signal:

$$\omega(nT) = 10\pi \operatorname{asinh}(100(nT - 0.5)) + 64\pi. \quad (43)$$

The IF estimates and accuracy results are presented in Fig. 5 and Table 4. The effects, similar to those in Figs. 1–3, appear in this example, although not

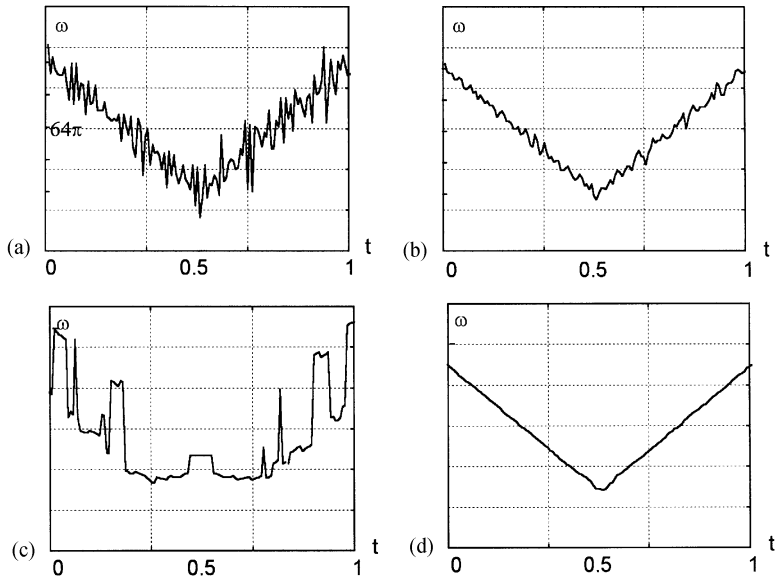


Fig. 3. Instantaneous frequency estimation using the periodogram with (a) $N_s = 4$, (b) $N_s = 8$, (c) $N_s = 128$, (d) adaptive window length.

Table 3

Normalized mean absolute error for the piecewise linear IF

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

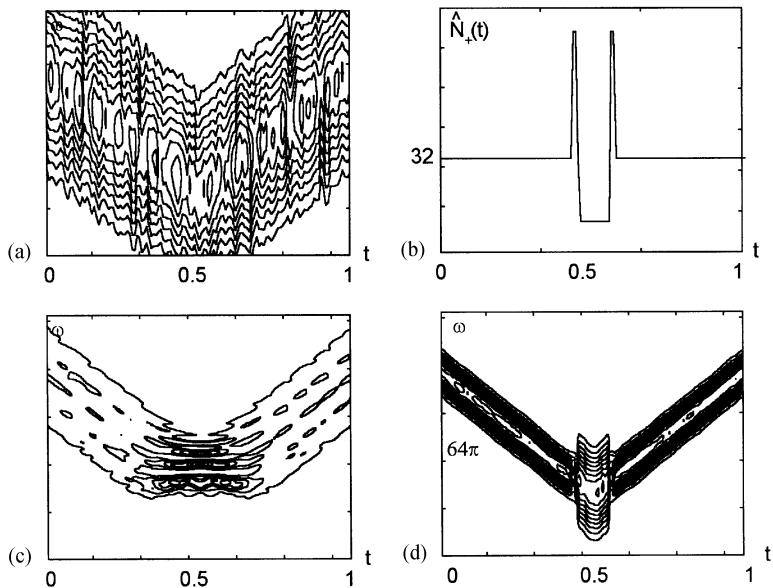


Fig. 4. (a) Periodogram with $N_s = 4$, (b) adaptive window length, (c) periodogram with $N_s = 128$, (d) periodogram with adaptive window length.

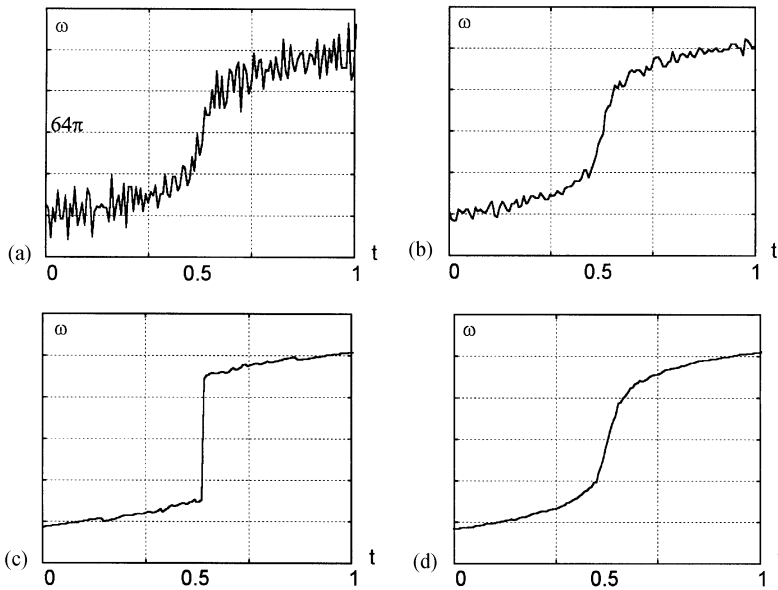


Fig. 5. Instantaneous frequency estimation using the periodogram with (a) $N_s = 4$, (b) $N_s = 8$, (c) $N_s = 128$, (d) Adaptive window length.

Table 4
Normalized mean absolute error for the nonlinear IF

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

in such a pure form as there. Note that the normalized minimal mean absolute error is here greater than in the previous two examples. The reason is that the quantization errors appeared in a full scale, what was not a case in Examples 1 and 2, due to the discrete values of the IF assumed in these examples. The quantization error may be further decreased by the additional interpolation.

The periodograms for $N_s = 4$, $N_s = 128$ and adaptive window length are shown in Figs. 6(a), 6(c) and 6(d), along with the adaptive window length, Fig. 6(b).

Table 2 shows that the adaptive window improves the estimation accuracy in $0.3750/0.1250 = 3.0$ and in $3.2344/0.1250 = 25.875$ times as compared with the best and the worst time-invariant window length for the stepwise constant IF. The

similar results for the linear and nonlinear IF in Tables 3 and 4 show that the accuracy improvement produced by the adaptive window length is given by the figures $0.0430/0.0352 = 1.2216$ and $0.2451/0.2009 = 1.22$, and by the figures $1.6484/0.0352 = 46.83$ and $1.5311/0.2009 = 7.62$, respectively, for the best and the worst choices of the time-invariant windows. Keeping in mind that the best window length is unknown in advance we can conclude that the adaptive window can really improve the accuracy in a valuable way.

5. Algorithm with the adaptive threshold Γ

The simulation results discussed above show that the standard value of $\kappa \approx 2$ and the corresponding

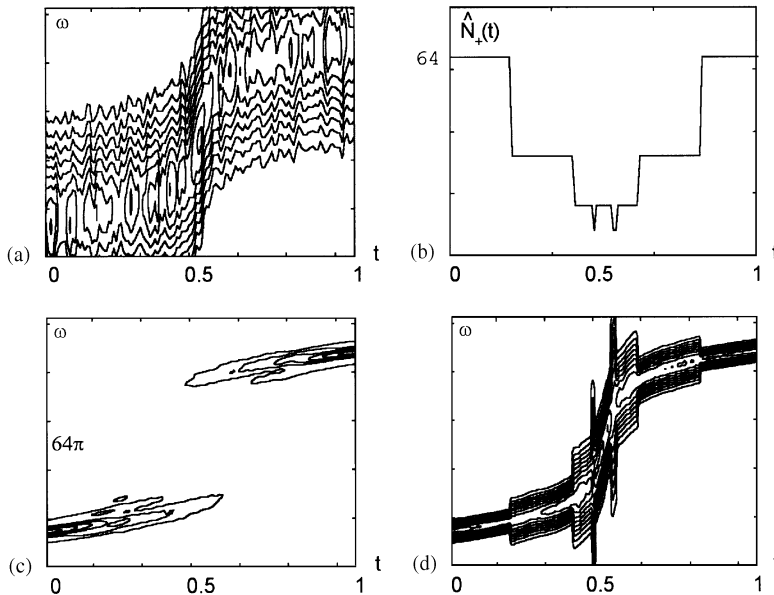


Fig. 6. (a) Periodogram with $N_s = 4$, (b) adaptive window length, (c) periodogram with $N_s = 128$, (d) periodogram with adaptive window length.

value of the threshold $\Gamma \simeq 2 \cdot 2 = 4$ results in quite a good accuracy of IF estimation. However, further accuracy improvement can be achieved by an appropriate adjustment of the value of the threshold.

The point is that a proper correction of the value of Γ in confidence intervals (27) gives:

1. A desirable proportion between the variance and bias for the minimal MSE value. It is different from the one determined by a factor of 2, as we have used in the formula $\Gamma = 2\kappa$.
2. Probability $P(\kappa)$ used in the inequalities (24).

A theoretical analysis of the link between the threshold Γ and the corresponding MSE value is a difficult problem, in particular because the optimal window length h_{s+} in the algorithm is random, and it is determined by the sequential statistical rule given by Eq. (31). Nevertheless, a quite efficient algorithm has been proposed and tested here for a data-driven choice of the adaptive Γ -value. This algorithm uses the cross-validation approach based on the quality-of-fit statistics, to be optimized with respect to Γ .

Let $\hat{\omega}_{h^+(lT)}^{[l]}$ be the adaptive estimate of the IF at a time-instant $t = lT$ given by algorithms (29)–(32), provided that the observation $y(lT)$ is

omitted from the set used for the estimation. For the sake of simplicity we will use, for this IF estimate, a notation

$$\hat{\omega}^{[l]} = \hat{\omega}_{h^+(lT)}^{[l]}(lT). \tag{44}$$

The corresponding complex-valued amplitude, given by the DFT, is determined by the following formula:

$$\hat{A}^{[l]} = \sum_{n=-\infty, n \neq l}^{\infty} w_h(nT)y(t+nT)e^{-j\omega nT}|_{\omega=\hat{\omega}^{[l]}}. \tag{45}$$

The quality-of-fit cross-validation statistic is of the form

$$I_{CV} = \frac{1}{N_A} \sum_{l \in A} |y(lT) - \hat{A}^{[l]}|^2, \tag{46}$$

where A is a set of observations used in this statistic. In particular, it can include all of the observations or exclude edge points. N_A is a size of the set A .

The optimal choice of Γ is determined as a solution of the problem

$$\hat{\Gamma} = \arg \min_{\Gamma \in \Theta} I_{CV}, \tag{47}$$

where $\Theta = \{\Gamma_r\}$ is a set of considered Γ -values. It is assumed that I_{CV} is calculated for all $\Gamma_r \in \Theta$. Afterwards, the final estimate of the IF is determined by the algorithm (29)–(32), using $\Gamma = \hat{\Gamma}$.

Now let us explain a nature of statistic (46). The STFT, given by Eq. (3), may be understood as a fitting of the signal y by a complex exponent, determined by the loss function

$$J = \sum_{n=-\infty}^{\infty} w_h(nT) |y(t + nT) - Ce^{-j\omega nT}|^2. \quad (48)$$

A minimization of Eq. (48) with respect to C gives $\hat{C} = F_h(\omega, t)$, where $F_h(\omega, t)$ is given by Eq. (3). In order to make clear what is estimated by C consider the case without noise and rewrite J as follows:

$$\begin{aligned} J &= \sum_{n=-\infty}^{\infty} w_h(nT) |A \exp(j\phi(t)) \cdot \exp(\phi(t + nT)) \\ &\quad - \phi(t)) - Ce^{-j\omega nT}|^2 \\ &= \sum_{n=-\infty}^{\infty} w_h(nT) |A \exp(j\phi(t)) \cdot \exp(\phi'(t)nT) \\ &\quad + \delta\phi(t, n)) - Ce^{-j\omega nT}|^2. \end{aligned}$$

Let us also assume that $\delta\phi(t, n)$, which is an error of approximation of $\phi(t + nT) - \phi(t)$ by $\phi'(t)nT$, is negligible within the considered window. Then, it is clear that ω is an estimate of the IF (given by $\phi'(t)$) and that C is an estimate of the product $A \exp(j\phi(t))$. It can also be seen that $Ce^{j\omega nT}$ is an estimate of $m(t + nT)$ with the corresponding n . The last conclusion means that C without the exponent $e^{j\omega nT}$ is a fitting $m(t)$.

It is important to emphasize that in the local approximation the parameter C gives an estimate of the signal $m(t)$ rather than its amplitude A only. This fact explains that the differences between the observations $y(lT)$ and $\hat{A}^{[l]}$, used in Eq. (46), are residuals of the fitting $y(lT)$ by its complex-valued approximation $\hat{A}^{[l]}$. Formulae (44)–(47) present the idea and procedure of the cross-validation.

Some more notes about the cross-validation method

The basic motivation is in a simple splitting the data into two parts. One part is used for estimation and then the second for the goodness judging, by matching the estimates with the rest of the data. Hence the name ‘cross-validation’. The method described above is usually referred to as a ‘simple leave-one-out cross-validation’.

The residuals in Eq. (46) can be interpreted as predictive errors between the signal $m(lT)$ and its estimate using all data but $y(lT)$. The leave-one-out idea is of importance in the cross-validation because if we use all data in the estimate of the IF, the statistic (46) will give the minimum value from Θ for $\hat{\Gamma}$. More details about the cross-validation and related methods can be found in [6,15,16].

Let us show how efficient can be the adaptive choice of κ . Let the set Θ consists of six values of Γ , $\Theta = \{1.0, 3.0, 4.0, 5.0, 6.0, 7.0\}$. Table 5 demonstrates how the cross-validation statistic I_{CV} , Eq. (46), works. The results are obtained from 100 independent runs of the algorithm with the adaptive Γ for the estimation of the IF from Example 1. The second row of the table presents the mean absolute error of the IF estimation, averaged over the 100 simulation runs, provided that Γ has a corresponding value shown in the first row.

A distribution of the random $\hat{\Gamma}$, chosen according to rule (47), is presented in the third row of Table 5. The figures in this row give a number of events from 100 independent trails (simulation runs) when $\hat{\Gamma}$ takes the corresponding value. The high distribution concentration in the area of the minimum values of the mean absolute error is obvious. It demonstrate clearly that the cross-validation statistic I_{CV} (46) works reasonably well and determines the effective rule in choosing the optimal, or close to the optimal, values of the threshold Γ .

Table 5
Normalized mean absolute error of IF estimation and distribution of the random $\hat{\Gamma}$ given by the algorithm (47)

Γ	1.0	2.0	3.0	4.0	5.0	6.0	7.0
Mean abs error	2.774	0.4452	0.1066	0.0947	0.0955	0.0997	1.0207
Frequency (%)	0	1	26	24	49	0	0

6. Conclusions

The periodogram with a data-driven and time-varying window length is developed as an adaptive estimator of the IF. The choice of the window length is based on the intersection of the confidence intervals of the IF estimates, with increasing window lengths. The developed algorithm uses only the formula for the asymptotic variance of the IF estimates. The formula for the estimate variance is derived under the condition of relatively high signal-to-noise ratio and white Gaussian noise. Note that the assumption that the noise is white is not crucial. As far as the correlation of the noise is known, the formula for the standard deviation of the estimate can be derived and used in the presented algorithm. Simulations show good accuracy ability of the adaptive algorithm.

The theory and examples, in this paper, are presented for the monocomponent signals. In the cases of multicomponent signals we may distinguish two cases. One case is when the number of signal components is known in advance, and their corresponding IFs are well separated. In this case we can directly apply the presented algorithm, by finding corresponding number of the periodogram's maxima for each time instant. The correct amplitude of each signal component has to be used in the formula for the estimate variance. It may be obtained by an inversion of the short time Fourier transform, using its values around the detected maximum for that component, only. Note that here a frequency-varying window length, corresponding to each component, will be obtained. Other case is when we do not know the number of components, as well as cannot assume that they are well separated in the time–frequency plane. In this case, in our opinion, it is better to apply a version of the algorithm presented here to the periodogram as a whole, rather than to the periodogram as an IF estimator. We are conducting the research in this direction using the Wigner distribution, as a time–frequency representation tool.

Acknowledgements

The authors thank the anonymous referees for helpful comments. The work of L.J. Stanković is

partially supported by the Alexander von Humboldt grant and Montenegrin Academy of Science and Art.

References

- [1] B. Boashash, Estimating and interpreting the instantaneous frequency of a signal – Part 1: Fundamentals, *Proc. IEEE* 80 (4) (April 1992) 519–538.
- [2] L. Cohen, C. Lee, Instantaneous bandwidth, in: B. Boashash (Ed.), *Time–frequency Signal Analysis*, Longman Cheshire, 1992.
- [3] L. Cohen, Distributions concentrated along the instantaneous frequency, *SPIE, Advanced Signal Proc. Alg., Architect. and Implementations* 1348 (1990) 149–157.
- [4] A. Goldenshluger, A. Nemirovski, On spatial adaptive estimation of nonparametric regression, Research report, 5/94, Technion, Israel, November 1994.
- [5] A. Goldenshluger, A. Nemirovski, Adaptive de-noising of signals satisfying differential inequalities, *IEEE Trans. Inf. Theory* 43 (3) (1997) 873–889.
- [6] P. Hall, I. Johnstone, Empirical functionals and smoothing parameter selection, *Journal of Royal Statistical Society, Series B* 54 (2) (1992) 475–530.
- [7] V. Katkovnik, Local polynomial periodogram for time-varying frequency estimation, *South African Statistical Journal* 29 (1995) 169–198.
- [8] V. Katkovnik, Local polynomial approximation of the instantaneous frequency: Asymptotic accuracy, *Signal Processing* 52 (3) (1996) 343–356.
- [9] V. Katkovnik, Adaptive local polynomial periodogram for time-varying frequency estimation, in: *Proc. IEEE-SP on TFTS Analysis*, Paris, June 1996, pp. 329–332.
- [10] V. Katkovnik, Nonparametric estimation of instantaneous frequency, *IEEE Trans. Inform. Theory* 43 (1) (January 1997) 183–189.
- [11] V. Katkovnik, L.J. Stanković, Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length, *IEEE Trans. on Signal Processing* (1997) in print.
- [12] *Proceedings of the IEEE, Special Issue on Time–Frequency Analysis*, Vol. 84, No. 9, September 1996.
- [13] L.J. Stanković, A method for improved distribution concentration in the time–frequency analysis of multicomponent signals the L-Wigner distribution, *IEEE Trans. Signal Processing* 43 (5) (May 1995) 1262–1268.
- [14] 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* 42 (7) (July 1994) 1863–1867.
- [15] G. Wahba, *Spline models for observational data*, Chapter 2, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
- [16] C.Z. Wei, On predictive least squares principles, *The Annals of Statistics* 20 (1) (1992) 1–42.