# Robust Wigner Distribution With Application to the Instantaneous Frequency Estimation

Igor Djurović and LJubiša Stanković, *Senior Member, IEEE*

*Abstract*—The Wigner distribution (WD) produces highly concentrated time–frequency (TF) representation of nonstationary signals. It may be used as an efficient signal analysis tool, including the cases of frequency modulated signals corrupted with the Gaussian noise. In some applications, a significant amount of impulse noise is present. Then, the WD fails to produce satisfactory results. The robust periodogram has been recently introduced for spectral estimation of this kind of noisy signals. It can produce good concentration for pure harmonic signals. However, it is not so efficient in the cases of signals with rapidly varying frequency. This is the motivation for introducing the robust WD. It is a reliable TF representation tool for wide class of nonstationary signals corrupted with impulse noise. This distribution produces good accuracy of the instantaneous frequency (IF) estimation. Using the Huber loss function, a generalization of the WD is presented. It includes both the standard and the robust WD as special cases. This distribution can be used for TF analysis of signals corrupted with a mixture of impulse and Gaussian noise. The presented theory is illustrated on examples, including applications on the IF estimation and time-varying filtering of signals corrupted with a mixture of the Gaussian and impulse noise. The case study analysis of the IF estimators' accuracy, based on the standard and the robust WD forms, is performed. In order to improve the IF estimation, a median filter is applied on the obtained IF estimate.

*Index Terms*—Impulse noise, instantaneous frequency estimation, spectral analysis, time–frequency analysis, time-varying filtering, Wigner distribution.

## I. INTRODUCTION

IN many practical applications, especially in communications, signals are disturbed by an impulse kind of noise. It could be caused by atmospheric or human made disturbances. This form of noise is also present in images as film-grain noise, photoelectronic noise, salt-pepper noise, etc. These noises can be modeled by heavy-tailed (long-tailed) probability density functions (pdfs). The class of generalized Cauchy noises includes almost all important heavy-tailed pdfs [1]. A fundamental approach to the estimation of signal parameters, when the signal is corrupted by an impulse noise, is given by Huber [2]. He has proposed the robust estimates as an alternative to the classical maximum likelihood (ML) estimates. The robust estimation theory has found numerous applications

in various signal processing disciplines, especially in digital image processing [3]–[5].

It is well known that time–frequency (TF) distributions are a powerful tool for analysis of nonstationary signals and the instantaneous frequency (IF) estimation [6]–[13]. Noise influence on the TF distributions is considered in [14] and [15]. It is common for these distributions to produce poor results in an impulse noise environment. This is the case for the spectrogram (with high interference reducing property), the Wigner distribution (WD) (which is highly concentrated in the TF plane), and the other distributions from the general quadratic shift-covariant class (with interference reduction and concentration properties between the previous two extreme cases). In the TF representations, impulse noise behaves quite differently from the Gaussian noise. Namely, the impulse noise is mainly concentrated around the origin and the Doppler axis in the ambiguity domain. Therefore, standard reduced interference distributions will not adequately reduce this kind of noise. In addition, some signal-dependent distributions will not be efficient here since they recognize these strong noise components as parts of the signal. Thus, this kind of noise requires a specific approach as well as reformulation of the standard TF representations into their robust forms.

The robust spectrogram (SPEC) [16] represents an extension of the standard SPEC. It is introduced by analogy with the robust $M$-periodogram [17]. It has been shown [17], [18] that the standard periodogram is an ML estimate of the periodogram of signal without noise in the case when signal is corrupted by a Gaussian white noise. The robust $M$-periodogram is an ML estimate of the periodogram of a non-noisy signal in the Laplacian noise environment. It can be successfully used for a wide class of impulse noises with other pdfs. In addition, the robust $M$-periodogram can be used for the accurate frequency estimation of harmonic signals in an impulse noise environment. However, when the signals are corrupted by the Gaussian noise, it produces slightly worse frequency estimation results than the standard periodogram. When the frequency changes in time, then the IF estimators based on both the standard and the robust $M$-periodogram are biased. The bias can be significant in the case of the FM signals with rapid nonlinear IF changes [16].

Since the robust SPEC possesses a low TF resolution, we introduce the robust WD. It can improve TF representation as well as the IF estimation for nonstationary signals in an impulse noise environment. Although the kernel of the robust WD depends both on time and frequency, the distribution is real-valued and shift-covariant. For a mixture of the Gaussian and impulse noise, a generalized robust WD has been derived by using the Huber loss function. Performance analysis, proving low sensitivity of

the proposed distribution to the impulse noise, has been done by using a case study approach. Applications of the proposed distribution on determination of the region of support in time-varying filtering, and on estimation of signal's IF, are presented.

The paper is organized as follows. Basic results concerning the robust $M$-periodogram and the robust SPEC are reviewed and presented in Section II. The robust WD and its properties are considered in Section III. An iterative procedure for calculation of the robust WD is given. In Section IV, the robust WD applications for the determination of the signal's region of support, in time-varying filtering, as well as for the IF estimation, are given. Reduction of the IF estimation error by using a median filter is demonstrated in Section IV as well.

## II. ROBUST SHORT-TIME FOURIER TRANSFORM

### A. Background and Motivation

To explain Katkovnik's motivation for introduction of the robust short-time Fourier transform (STFT), let us consider the problem of signal denoising. The general denoising task can be defined as processing of $x(n) = s(n) + \nu(n)$ with the aim to obtain $f(n)$ as close as possible to $s(n)$, where $s(n)$ is a signal, and $\nu(n)$ is an additive white noise. The values of $f(n)$ can be calculated by using several samples around the considered instant and minimizing the function $L = \sum_{k=n-N}^{n+N} F(|f(n) - x(k)|)$, where $2N + 1$ is the window length, and $F(\cdot)$ is a loss function. The loss function $F(e) = -\log p_\nu(e)$, where $p_\nu(e)$ is the pdf of noise $\nu(n)$, produces the ML estimator [2]. For $F(e) = |e|^2$, we get a linear moving average filter as the ML estimator for the signal $s(n)$ in the Gaussian noise environment. For the Laplacian noise environment, the loss function $F(e) = |e|$ results in the median filter as the ML estimator. Unfortunately, the ML estimates are quite sensitive to a deviation of the hypothetical noise distribution. Namely, even a slight deviation from the hypothetical noise distribution can result in a significant degradation of the ML estimate [2], [17], [19], [20]. Huber proposed the robust estimators as an alternative to the ML estimators. For determination of the robust estimator, it is necessary to consider a specific class of noises. The robust estimation for that class is the ML estimation for the particular noise, producing the highest outliers. The outliers can be determined based on widths of the noise pdf tail. This robust estimator then produces quite accurate estimation for all noises from this class. The robust estimation is less accurate than the ML estimation for a particular noise from the considered class, but this is the price that has to be paid for the robustness. For numerous noise classes, the Laplacian pdf comes out as "the worst" noise, resulting in $F(e) = |e|$ as the loss function for the robust estimation [2], [19]. More details about robust estimations can be found in [2] and [19].

Similar reasoning as in the signal denoising can be used for the estimation of the signals' transforms by using noisy samples $x(n) = s(n) + \nu(n)$. Namely, the Fourier transform (FT) $X(\omega) = \sum_{n=1}^{N} x(n)e^{-j\omega n}/N$ can be treated as an average of the samples $\{x(n)e^{-j\omega n}, n \in [1, N]\}$, i.e.,

$$X(\omega) = \text{mean}\{x(n)e^{-j\omega n}, n \in [1, N]\}. \tag{1}$$

In full analogy with the moving average, this form of the FT minimizes the functional

$$L = \sum_{n=1}^{N} |x(n)e^{-j\omega n} - X(\omega)|^2 = \sum_{n=1}^{N} |x(n) - X(\omega)e^{j\omega n}|^2. \tag{2}$$

Thus, we can conclude that the standard FT of $x(n)$ is the ML estimate of the FT of $s(n)$ for the case when signal $s(n)$ is corrupted by Gaussian noise $\nu(n)$. In a similar way, we can introduce the ML estimates and the **robust estimates of the FT**.

### B. Definition of the Robust STFT

The standard STFT, at a given point $(t, \omega)$, can be defined as a solution of the following optimization problem:

$$\text{STFT}(t, \omega) = \arg\{\min_m J(t, \omega; m)\} \tag{3}$$

$$J(t, \omega; m) = \sum_n w_h(nT)F(e(t, \omega, n)) \tag{4}$$

where $w_h(nT)$ is a window function with the length $h$, $T$ is the sampling interval, the loss function is given as $F(e) = |e|^2$, whereas the error function has the form

$$e(t, \omega, n) = x(t + nT) - m\exp(j\omega nT). \tag{5}$$

The error function can be considered to be a residuum, expressing the "similarity" between the signal and a given harmonic $\exp(j\omega nT)$. The standard STFT, for a given point $(t, \omega)$, follows from

$$\frac{\partial J(t, \omega; m)}{\partial m^*}\Big|_{m=\text{STFT}(t,\omega)} = 0. \tag{6}$$

It is equal to

$$\text{STFT}(t, \omega) = \frac{1}{\sum_n w_h(nT)} \sum_n x(t + nT)w_h(nT)e^{-j\omega nT}. \tag{7}$$

The robust STFT $\text{STFT}_R(t, \omega)$ is introduced by using the loss function $F(e) = |e|$ instead of $F(e) = |e|^2$. With this loss function, the solution of optimization problem (3) can be obtained from the following set of nonlinear equations:

$$\text{STFT}_R(t, \omega) = \sum_n d(t, \omega, n)x(t + nT)e^{-j\omega nT} \tag{8}$$

$$d(t, \omega, n) = \gamma(t, \omega, n)/\sum_n \gamma(t, \omega, n) \tag{9}$$

$$\gamma(t, \omega, n) = w_h(nT)|e(t, \omega, n)|^{-1}. \tag{10}$$

*Note:* From (9), it easily follows that $\sum_n d(t, \omega, n) = 1$. Since for a positive window $w_h(nT)$ we have $d(t, \omega, n) \geq 0$, then it follows that $0 \leq d(t, \omega, n) \leq 1$ for any form of the error $e(t, \omega, n)$.

In order to calculate $\text{STFT}_R(t, \omega)$, based on (8)–(10), it is necessary to define an appropriate iterative procedure; see [17]. The robust STFT is an estimate of the standard STFT of non-noisy signal when the signal is corrupted by an impulse kind of noise.

The robust SPEC $\text{SPEC}_R(t, \omega) = |\text{STFT}_R(t, \omega)|^2$ produces highly accurate results in the analysis of sinusoidal signals with constant IFs [17]. However, the robust SPEC, like the standard
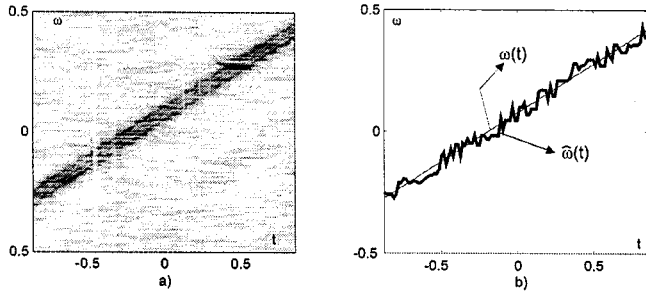
Fig. 1. Linear FM signal corrupted by heavy-tailed noise. (a) Robust SPEC. (b) IF estimation based on the robust SPEC. $\omega(t)$—true IF; $\hat{\omega}(t)$—IF estimation.



Fig. 2. TF representations of multicomponent signal. (a) Standard SM. (b) SM based on the robust STFT.

one, results in a low TF resolution in the case of nonstationary signals. This fact will be illustrated in an example. Consider signal with a linear IF

$$x(t) = \exp(j204.8\pi(t+0.5)^2 - j128\pi(t+0.5)) \quad (11)$$

corrupted by a heavy-tailed noise

$$\nu(t) = a\left(\nu_1^3(t) + j\nu_2^3(t)\right) \quad (12)$$

where $\nu_i(t)$ $i = 1, 2$ are mutually independent Gaussian noises $\mathcal{N}(0,1)$, and $a = 0.85$. The variance for this kind of noise is $\sigma_\nu^2 = 30a^2$. Thus, the signal-to-noise ratio is SNR $= -13.35$ dB. The robust SPEC, with corresponding IF estimate, is shown in Fig. 1. The IF estimate is obtained as

$$\hat{\omega}(t) = \arg\min_\omega \text{SPEC}_R(t, \omega). \quad (13)$$

For this noise, the standard SPEC failed to produce any reasonable result. It means that it was not possible to recognize the signal component at all. In addition, the IF estimation based on this SPEC was extremely inaccurate, and it was not presented here. Low TF resolution of the robust SPEC is the main reason for introducing other TF representations.

*C. Cohen Class of Distributions Based on the Robust STFT*

We have just shown that in the case of impulse noise, better estimates of the STFT of non-noisy signal could be obtained by the robust STFT than by the standard one. Once we have efficiently estimated the STFT of the non-noisy signal from the signal corrupted by an impulse nose [see (8)–(10)], we can employ this robust STFT to calculate "robust versions" of other quadratic TF distributions. Most of the commonly used quadratic TF distributions can be written in the form of the Cohen class of distributions [21]–[23]. An efficient way for the calculation of a distribution from the Cohen class, from the STFT, is based on the kernel eigenvalue decomposition [24]–[27]. A distribution from the Cohen class can be written as a sum of the SPECs [24], [26]:

$$\text{CD}(t,\omega) = \sum_k \lambda_k |\text{STFT}_{q_k}(t,\omega)|^2 \quad (14)$$

where $\lambda_k$ are eigenvalues of the rotated time-lag kernel function (for details, see [24] and [26]), and $\text{STFT}_{q_k}(n,\omega)$ are the STFTs calculated by using the eigenvectors associated with $\lambda_k$ as window functions $w_h(nT) = q_k(nT)$. Thus, by using the
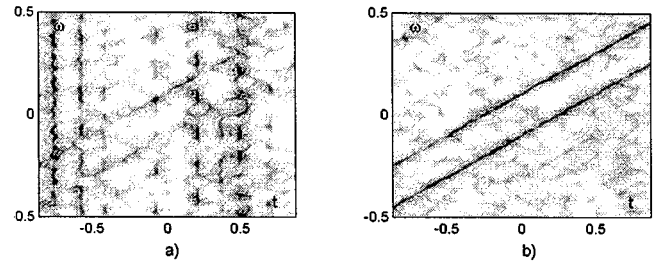
robust forms of the STFT as better estimates of the nonnoisy STFT instead of the standard ones, in the case of impulse noise, we can expect improvement in the signal representation.

Another way of constructing quadratic distributions from the STFT is based on the S-method (SM) [28], [29]

$$\text{SM}(t,\omega) = \int P(\theta) \text{STFT}(t, \omega + \theta) \text{STFT}^*(t, \omega - \theta) \, d\theta \quad (15)$$

where $P(\theta)$ is a window in the frequency domain. For $P(\theta) = \delta(\theta)$, the SM is reduced to the SPEC, whereas for the $P(\theta) = 1/\pi$, it is equal to the WD [28]. The SM, with a proper choice of frequency window width, can produce auto-terms close to the auto-terms of the WD while avoiding cross-terms and other interferences that exist in the WD [28], [29]. Again, by using the robust STFT, as a better estimate of the non-noisy signals STFT, we can expect that the SM will be closer to the SM of the non-noisy signal.

This concept will be illustrated by an example with the sum of two linear FM signals

$$x(t) = 2\exp(j204.8\pi t^2)\cos(102.4\pi t) \quad (16)$$

corrupted by noise (12). The standard SM is presented in Fig. 2(a), whereas the SM based on the robust STFT [with $\text{STFT}(t,\omega)$ being replaced by $\text{STFT}_R(t,\omega)$ in (15)] is shown in Fig. 2(b). The rectangular frequency window $P(\theta)$, whose width is $4\pi$, i.e., $P(\theta) = 1$ for $|\theta| \leq 2\pi$ and $P(\theta) = 0$ elsewhere, is used for the SM's calculation. Improvement in TF representation based on the robust STFT is evident.

## III. ROBUST WIGNER DISTRIBUTION

In the previous section we have used the robust STFT as a good estimate of the nonnoisy STFT in calculation of other TF distributions. Thus, the optimization (3) and (4) is done with respect to the STFT and not with respect to the desired quadratic representation. The obtained representation may not be optimal in an impulse noise environment. This is the reason for introducing the robust WD.

Consider the error function

$$e(t,\omega,n) = 2x(t+nT)x^*(t-nT) - m\exp(j2\omega nT) \quad (17)$$

and the minimization problem associated with

$$J(t,\omega; m) = \sum_n w_h(nT) F(e(t,\omega,n)) \quad (18)$$

where $F(e)$ is a loss function. By solving the equation

$$\frac{\partial J(t,\omega;m)}{\partial m^*}\Big|_{m=\mathrm{WD}(t,\omega)} = 0 \qquad (19)$$

for $F(e) = |e|^2$, the standard WD, which is normalized by a factor of $k_w = 2/\sum_n w_h(nT)$, is obtained:

$$\mathrm{WD}(t,\omega) = k_w \sum_n w_h(nT)x(t+nT)x^*(t-nT)e^{-j2\omega nT}. \qquad (20)$$

It produces an ideal concentration for the linear FM signal $x(t) = A\exp(jat^2/2 + jbt)$ [21], [22]. Note that the WD (20) produces $\mathrm{WD}(t,\omega) = A^2\delta(\omega - at - b)$ when the window $w_h(nT)$ is rectangular of an infinite duration. For any other window type, the WD value along the IF is given by $\mathrm{WD}(t, at+b) = A^2$.

The robust WD can be derived from the minimization problem (18) with the loss function $F(e) = |e|$. It assumes the form

$$\mathrm{WD}_R(t,\omega) = \frac{2}{\sum_n w_h(nT)/|e(t,\omega,n)|}$$
$$\times \sum_n \frac{w_h(nT)}{|e(t,\omega,n)|}x(t+nT)x^*(t-nT)e^{-j2\omega nT}. \qquad (21)$$

Equation (21) is an implicit definition of $\mathrm{WD}_R(t,\omega)$. Thus, it is necessary to define an iterative procedure for the robust WD calculation. Here, we will use the fixed-point iterative algorithm to solve the equation of the form $x = f(x)$, as $x_i = f(x_{i-1})$, with the stopping rule $|(x_i - x_{i-1})/x_{i-1}| < \eta$, where $\eta$ defines the solution precision. This procedure, applied to (21), can be summarized as follows.

Step 1)   $i = 0$ the standard WD calculation (20):

$$\mathrm{WD}_R^{(0)}(t,\omega) = k_w \sum_n w_h(nT)x(t+nT)x^*(t-nT)e^{-j2\omega nT}. \qquad (22)$$

Step 2)   $i = i + 1$

$$|e^{(i)}(t,\omega,n)| = |2x(t+nT)x^*(t-nT)$$
$$- \mathrm{WD}_R^{(i-1)}(t,\omega)e^{j2\omega nT}|$$
$$w_h^{(i-1)}(nT) = w_h(nT)/|e^{(i)}(t,\omega,n)| \qquad (23)$$
$$\mathrm{WD}_R^{(i)}(t,\omega) = \frac{2}{\sum_n w_h^{(i-1)}(nT)}$$
$$\times \sum_n w_h^{(i-1)}(nT)x(t+nT)$$
$$\times x^*(t-nT)e^{-j2\omega nT}. \qquad (24)$$

Step 3)   If the relative absolute difference between two iterations is smaller than $\eta$

$$\frac{|\mathrm{WD}_R^{(i)}(t,\omega) - \mathrm{WD}_R^{(i-1)}(t,\omega)|}{|\mathrm{WD}_R^{(i-1)}(t,\omega)|} \le \eta \text{ or } i = K \qquad (25)$$

then the robust WD is obtained as $\mathrm{WD}_R(t,\omega) = \mathrm{WD}_R^{(i)}(t,\omega)$. If (25) is not satisfied, then go to step i). Here, $K$ represents the maximal allowed number of iterations. Note that in all examples, we have
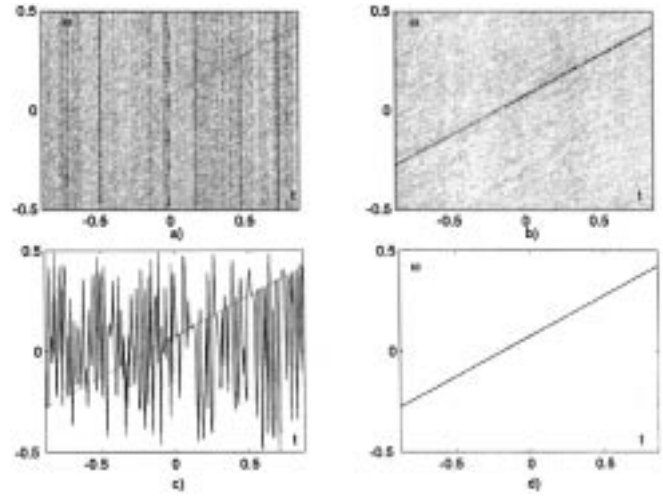


Fig. 3.   Linear FM signal corrupted by heavy-tailed noise. (a) Standard WD. (b) Robust WD. (c) IF estimation based on the standard WD. (d) IF estimation based on the robust WD.

used $K = 15$ and $\eta = 0.1$. The average number of iterations is noise dependent. It will be studied in Section IV.

Uniqueness and convergence of the fixed-point iterative algorithm, for solving similar problems, is studied in detail in [2, Sec 7.8] and [30]. Note that our solution (21) can be written in the form $\mathrm{WD}_{m+1} = T(\mathrm{WD}_m) = \sum_n h_n(\mathrm{WD}_m)\mathbf{x}_n/\sum_n h_n(\mathrm{WD}_m)$, where $h_n(\mathrm{WD}_m) = w_h(nT)/|e(t,\omega,n)|$, and $\mathbf{x}_n = 2x(t+nT)x^*(t-nT)e^{-j2\omega nT}$. For these kinds of solutions, it has been shown in [30] that if the initial value is within $\mathrm{WD}_0 \in [\min(\mathbf{x}_n), \max(\mathbf{x}_n)]$, then the fixed-point iterative algorithm will converge to a local minimum $\mathrm{WD}^*$ within the same range. In our case, the function $J(t,\omega;m)$ has a single minimum [30], and the proposed iterative procedure will converge to that single (global) minimum since our initial value (22) is a mean value, and it satisfies the necessary condition $\mathrm{WD}_0 \in [\min(\mathbf{x}_n), \max(\mathbf{x}_n)]$.

### A. Properties of the Robust Wigner Distribution

1) The robust WD is real-valued.
2) $\mathrm{WD}_R(t,\omega)$ is TF invariant. For a signal $y(t) = x(t - t_0)e^{j\omega_0 t}$, the robust WD is $\mathrm{WD}_R(y;t,\omega) = \mathrm{WD}_R(x;t - t_0, \omega - \omega_0)$.
3) For signal $y(t) = x(t)\exp(jat^2/2)$, we have $\mathrm{WD}_R(y;t,\omega) = \mathrm{WD}_R(x;t,\omega - at)$.

The proof of these properties is given in the Appendix.

### B. Examples

*Example 1:* For signal (11) corrupted by (12), the standard and the robust WDs are shown in Fig. 3. It is obvious that the standard WD is useless in the case of heavy-tailed noise. The IF estimate, by using the robust WD, is obtained as

$$\hat{\omega}(t) = \arg\min_\omega \mathrm{WD}_R(t,\omega). \qquad (26)$$

*Example 2:* Consider a nonlinear FM signal

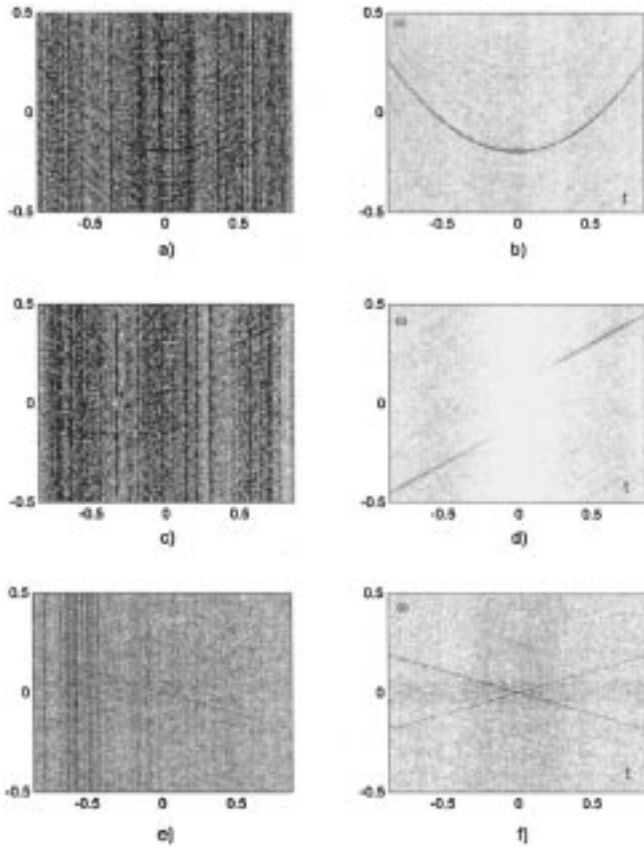$$x(t) = \exp(j204.8\pi t^3 - j204.8\pi t) \qquad (27)$$

Fig. 4.   TF representations of signals. (a) Standard WD of the nonlinear FM signal. (b) Robust WD of the nonlinear FM signal. (c) Standard WD of the sum of signals. (d) Robust WD of the sum of signals. (e) Standard WD of the real-valued linear FM signal. (f) Robust WD of the real-valued linear FM signal.

corrupted by (12). The standard WD is shown in Fig. 4(a), whereas the robust WD is shown in Fig. 4(b). The robust WD can also be used for efficient TF analysis of nonlinear FM signals corrupted by heavy-tailed noise.

*Example 3:* For two-component signal whose components are separated in time

$$x(t) = \exp(j204.8\pi t^2 + j102.4\pi t)\exp(-4(t - 0.6)^2)$$
$$+ \exp(j204.8\pi t^2 - j102.4\pi t)\exp(-4(t + 0.6)^2) \quad (28)$$

and corrupted by (12), the standard and the robust WD are shown in Fig. 4(c) and (d), respectively. Note that the lag window has eliminated the cross-terms between two components.

*Example 4:*   Consider the real-valued linear FM signal

$$x(t) = 2\cos(128\pi t^2) \quad (29)$$

corrupted by (12). The standard and the robust WD are shown in Fig. 4(e) and (f), respectively. For this specific signal, cross-terms are not emphatic. Note that the robust WD is not designed with the goal to reduce cross-terms, which exist in the original WD. Therefore, they will, in general, be present in the robust WD. The primary goal of the robust WD is to estimate the standard WD of the non-noisy signal, from a signal in high impulse environment. The multicomponent signals can also be analyzed by the approach presented in Section II-C.

## C. Wigner Distribution With Huber Loss Function

In many cases, the signal is not disturbed with a pure impulse noise, but the Gaussian noise is present, as well. Impulse noise causes relatively rarely occurring and large errors, whereas the Gaussian noise produces smaller errors but in each considered point. Knowing that the loss function $F(e) = |e|$ produces good results for impulse noise, i.e., large $|e|$ and that the loss function $F(e) = |e|^2$ produces good results for Gaussian noise, i.e., small $|e|$, we can use a combination of these two loss functions, which are known as the Huber loss function [2], [16], [31]

$$F(e) = \begin{cases} |e|^2/2, & |e| < \Delta \\ \Delta|e| - \Delta^2/2, & |e| \ge \Delta. \end{cases} \quad (30)$$

Note that the optimal value of $\Delta$ depends on the impulse noise influence. Namely, for higher impulse noise influence, the value of $\Delta$ should be close to 0, whereas for higher Gaussian noise, $\Delta$ should be close to $\infty$. In our simulation, we use $\Delta = 1$. Determination of the optimal value of $\Delta$ is a topic of our future research. The loss function (30) has been used in many applications for detection of signals in mixed Gaussian and impulse noise environments [2], [16], [31]. It can be employed to defining a more general form of the robust WD.

The robust WD, with the Huber loss function in (18) and (19), is equal to

$$\begin{aligned} &\mathrm{WD}_\Delta(t,\omega) \\ &= \frac{2\sum_{n,|e|\le\Delta} w_h(nT)x(t+nT)x^*(t-nT)e^{-j2\omega nT}}{\sum_{n,|e|\le\Delta} w_h(nT) + \sum_{n,|e|\triangleright\Delta} w_h(nT)|e|^{-1}} \\ &+ \frac{2\sum_{n,|e|\triangleright\Delta} w_h(nT)|e|^{-1}x(t+nT)x^*(t-nT)e^{-j2\omega nT}}{\sum_{n,|e|\le\Delta} w_h(nT) + \sum_{n,|e|\triangleright\Delta} w_h(nT)|e|^{-1}}. \end{aligned}$$
$$(31)$$

It can be calculated through a similar iterative procedure, as in the case of the robust WD (22)–(25). All properties of the robust WD (Section III-A) hold for (31) as well.

Two special cases of (31) are 1) the standard WD for $\Delta \to \infty$, and 2) the robust WD for $\Delta = 0$.

## IV. APPLICATION EXAMPLES

### A. Time-Varying Filtering

Time-varying filtering, based on the TF distributions, is an active research topic [32]–[37]. For the signal $x(t)$, a time-varying filtering relation can be defined by [32]

$$f(t) = \int_{-\infty}^{\infty} h(t+\tau/2, t-\tau/2)w(\tau)x(t+\tau)\, d\tau \quad (32)$$

where $h(t,\tau)$ is the impulse response of a time-varying system, and $w(\tau)$ is a window function. A slight modification of the filtering relation is made in (32) in order to produce unmodified output, when a non-noisy FM signal is applied at the input. Details on this modification are given in [32] and [33]. Using the Parseval's theorem, (32) becomes

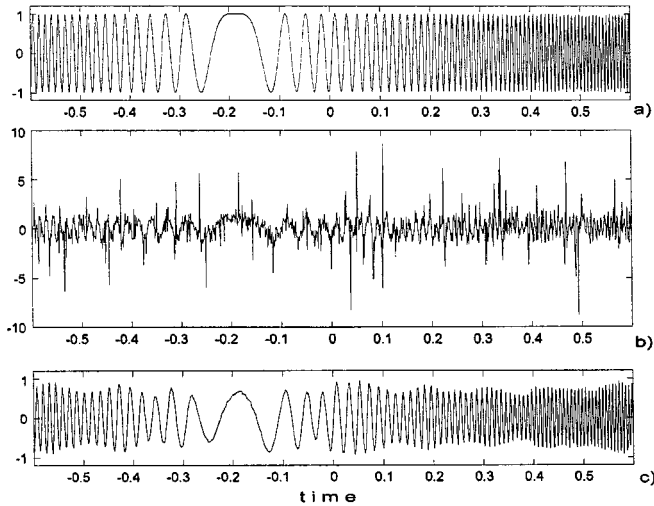$$f(t) = \int_{-\infty}^{\infty} L_H(t,\omega)\mathrm{STFT}(t,\omega)\, d\omega \quad (33)$$

Fig. 5. Time-varying filtering. (a) Original linear FM signal. (b) Signal corrupted by mixture of the Gaussian and impulse noise. (c) Filtered signal.

with the region of support $L_H(t,\omega) = \int_{-\infty}^{\infty} h(t + \tau/2, t - \tau/2)e^{-j\omega\tau}d\tau$. The value of $L_H(t,\omega)$ for the optimal time-varying filter is derived in [32]–[36] as

$$L_H(t,\omega) = \frac{E\{\mathrm{WD}_{ss}(t,\omega)\}}{E\{\mathrm{WD}_{ss}(t,\omega)\} + E\{\mathrm{WD}_{\nu\nu}(t,\omega)\}} \quad (34)$$

where $E\{\mathrm{WD}_{ss}(t,\omega)\}$ is the expected value of the WD (Wigner spectrum) of signal $s(t)$. For the cases when the signal's TF distribution is highly concentrated along the IF and the amount of noise in the TF plane at these points may be neglected, the previous relation reduces to a simple form, which has been proposed in [36], as

$$L_H(t,\omega) = \begin{cases} 1, & \text{for } E\{\mathrm{WD}_{ss}(t,\omega)\} \neq 0 \\ 0, & \text{elsewhere.} \end{cases} \quad (35)$$

We will use the robust WD for estimation of the region where $E\{\mathrm{WD}_{ss}(t,\omega)\} \neq 0$. For a monocomponent noisy signal, the filter is implemented, according to (35), in the following way: In a considered instant $t$, find the IF estimate $\hat{\omega}(t)$ as a position of the TF distribution maximum; set $L_H(t, \hat{\omega}(t)) = 1$, and $L_H(t,\omega) = 0$ for $\omega \neq \hat{\omega}(t)$; calculate $f(t)$ according to (33). Implementation of this time-varying filtering concept on multicomponent signals is discussed in [33].

This filtering approach using the robust WD in the region of support estimation has been applied on signal (11). The noise is a mixture of a Gaussian type, with $\sigma^2 = 0.1$, and an impulse type (12), with $a = 0.4$. The original signal (11), the noisy signal, and the filtered signal are presented in Fig. 5(a)–(c), respectively. The improvement in the filtered signal is obvious. Noise components that are not at the IF are suppressed.

### B. IF Estimation

Here, we will consider the IF estimation, performed by using three forms of the WD (standard, robust, and the robust WD based on the Huber loss function). The mean squared error (MSE) of the IF estimation can be written, according to [16], [38], as a sum of the variance and bias. For the SPEC, the MSE

of the IF estimate is analyzed in [16] and [18]. Following the same procedure, which was developed in the case of the robust SPEC [16], [18], we concluded that the bias for all forms of the WD is the same and equal to

$$E\{\Delta\hat{\omega}_h(t)\} = E\{\omega(t) - \hat{\omega}_h(t)\} = B_\omega h^2 \omega^{(2)}(t) + O(h^4) \quad (36)$$

where $\hat{\omega}_h(t)$ denotes the IF estimate obtained by using a WD with window width $h$; $O(h^4)$ denotes a remainder of $h^4$ order, and $B_\omega = \int_{-\infty}^{\infty} w_h(u)u^4 du / (3! \int_{-\infty}^{\infty} w_h(u)u^2 du)$. The second derivative of the IF is denoted by $\omega^{(2)}(t)$. For linear FM signals, $\omega^{(2)}(t) = 0$; the bias is equal to zero. This conclusion can be directly drawn from Property 3 of the robust WD, which states that this distribution is completely concentrated along the IF for linear FM signals. Therefore, the bias is a function of the IF higher order derivatives. The IF estimate bias is not influenced either by the form of noise or by the loss function.

The IF estimator variance, based on the standard WD, has two components: The first one represents noise influence, and the second one results from the noise-to-signal interaction [16], [38]. It has been derived in [16] and [38] for high SNR values $\sigma_\nu^2 \ll |A|^2$.

*Case Study Analysis:* The impulse noise usually has the variance that is significantly greater than the signal amplitude. It prevents one from getting a reliable expression for the IF estimator variance. That is the reason we will perform a numerical analysis of the IF estimator accuracy. In particular, we will use the case study approach [39]. For this purpose, we will consider signal (11), which is embedded with a mixed Gaussian and impulse noise

$$\nu(t) = \sigma[(1-\beta)(\nu_1(t) + j\nu_2(t)) + \beta(\nu_3^3(t) + j\nu_4^3(t))] \quad (37)$$

where $\nu_i(t), i = 1, \ldots, 4$ are mutually independent Gaussian noises $\mathcal{N}(0,1)$. The variance of considered noise is equal to $\sigma_\nu^2 = 2\sigma^2[(1-\beta)^2 + 15\beta^2]$. Since signal (11) has a unity amplitude, the SNR is

$$\begin{aligned} \mathrm{SNR} &= 10\log_{10}[A^2/2\sigma^2[(1-\beta)^2 + 15\beta^2]] \\ &= -3 \text{ dB} - 20\log_{10}\sigma - 10\log_{10}(1 - 2\beta + 16\beta^2). \end{aligned} \quad (38)$$

The parameter $\beta$ determines the amount of impulse noise in the noise mixture. For $\beta = 0$, we get the Gaussian white noise, and for $\beta = 1$, we get a pure impulse noise. The parameter $\sigma$ (noise amount) varies within $\sigma \in [0,2]$. The resolution step is 0.05 in both directions. For each pair $(\beta, \sigma)$, we performed 4480 tests and found the percentage of hits at the correct IF. For various WDs ($\Delta \to \infty, \Delta = 0$ and $\Delta = 1$), this percentage is shown in Fig. 6. One can conclude from this figure that for a pure Gaussian noise, $\beta = 0$, the standard WD, which is shown in Fig. 6(a), gives the best accuracy since its percentage of hits remains close to 1, within the widest region of $\sigma$, starting from $\sigma = 0$. The accuracy of the robust WDs is slightly lower for $\beta = 0$. However, when the impulse noise amount increases, $\beta > 0$, and then, the robust WDs [see Fig. 6(b) and (c)] remain almost insensitive to this component of noise, whereas the standard WD performance is significantly worsened; see Fig. 6(a). Areas of
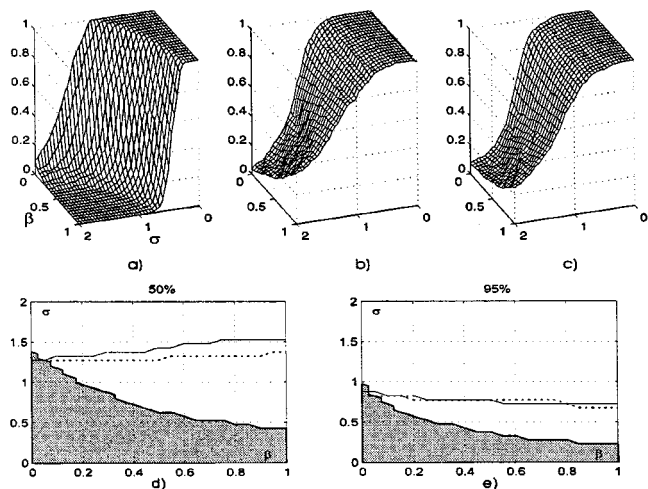
Fig. 6. Accuracy of the IF estimates based on the different forms of the WD. (a) Standard WD. (b) Robust WD. (c) WD with Huber loss function. (d) Contour plot for 50% accuracy. (e) Contour plot for 95% accuracy. Shaded area—the standard WD, solid line—the robust WD for $\Delta = 0$, and dotted line—the robust WD for $\Delta = 1$.

50% and 95% of accuracy of the WD-based IF estimators are shown in Fig. 6(d) and (e). They are obtained as contour plots of Fig. 6(a)–(c) for the given percentage of hits. The shaded area represents the range where the standard WD has a percentage of hits greater than the given value. For $\beta = 0$ and $\sigma = 1$, the standard WD has accuracy greater than 95%, whereas for $\beta = 1$ and $\sigma = 1$, the accuracy of this distribution is less than 50%. The area of the IF estimator accuracy with the robust WD, and the WD with the Huber loss function, are shown by solid and dotted lines, respectively. For $\beta \approx 0$, the standard WD shows slightly better accuracy than the robust forms of the WD. By an increase of the impulse noise component, accuracy of the standard WD becomes poor. The robust WD's accuracy remains good. For example, for $\beta = 1$, the robust WD has accuracy greater than 70% for $\sigma = 1.5$ and greater than 95% for $\sigma = 0.75$.

Note that in the impulse noise environment, by smoothing the WD (using narrower window or kernel function), we get a significant increase of the MSE and worse accuracy than in the case of the nonsmoothed WD. This effect is analyzed in detail for the robust SPEC in [16].

*Median Filter in the IF Estimation:* The large noise values could cause the position of the TF distribution maximum to be outside the distribution auto-term [40]. This effect results in the IF estimation error that has an impulse nature. By implementing a median filter directly on the estimated IF values, we can decrease the MSE. Details on the high noise influence on the IF estimation, in the case of the additive Gaussian noise, can be found in [40]. The IF estimation by using the robust WD, for $\beta = 0.85$ and $\sigma = 1$, is shown in Fig. 7(a). From this figure, we can see that errors in the IF estimation have an impulse nature. The IF estimate from Fig. 7(a), after applying median filter with length 3

$$\hat{\omega}_{\mathrm{med}}(t) = \mathrm{median}\{\hat{\omega}(t-T), \hat{\omega}(t), \hat{\omega}(t+T)\} \qquad (39)$$

is shown in Fig. 7(b). By $\hat{\omega}_{\mathrm{med}}(t)$, we denoted the median of the set whose elements are the estimated IFs at the instants $t -$
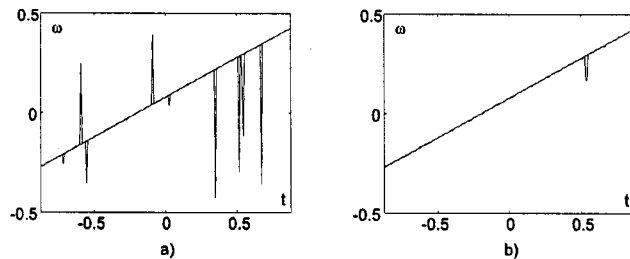


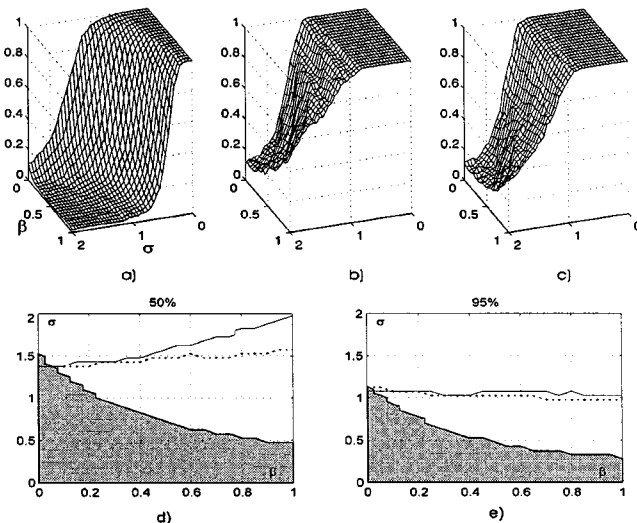Fig. 7. IF estimate. (a) Without median filter. (b) With median filter.



Fig. 8. Accuracy of the IF estimates based on the different forms of the WD with output median filter. (a) Standard WD. (b) Robust WD. (c) WD with Huber loss function $\Delta = 1$. (d) Contour plot for 50% accuracy. (e) Contour plot for 95% accuracy. Shaded area—the standard WD, solid line—the robust WD for $\Delta = 0$, and dotted line—the robust WD for $\Delta = 1$.

$T, t, t + T$. The percentage of hits at the correct IF, by using the WDs with $\Delta \to \infty, \Delta = 0$ and $\Delta = 1$, after applying the median filter, is shown in Fig. 8(a)–(c). Here, we assume that the correct hits are those within a range of one sample around the exact IF. Areas where the IF estimator accuracy is greater than 50 and 95%, after applying the median filter, are shown in Fig. 8(d) and (e). The area of applicability of the robust WD is wider than for the standard WD case. Comparing Fig. 6 and 8, one can see the improvement of accuracy, which is achieved by using output median filter. Note that the median filter could not be efficiently used for FM signal filtering since the high-frequency components would be completely degraded.

*Number of Iterations:* A statistical analysis of the average number of iterations in the robust WD calculation is performed as well. The number of iterations depends on the noise parameters. For pure Gaussian noise ($\beta = 0$), the number of iterations is shown in Fig. 9(a), as a function of $\sigma$. Fig. 9(b) presents the number of iterations for $\beta = 1$, whereas Fig. 9(c) shows the number of iterations as a function of $\beta$ for fixed $\sigma = 1$. In all examples, it is approximately twice as high for $\Delta = 0$ than for $\Delta = 1$. We can conclude that the convergence with the Huber loss function is faster. At the instants where the impulse noise assumes relatively small values, the error is small, and thus, the number of points where $|e| \geq \Delta$ is also small. Therefore, the
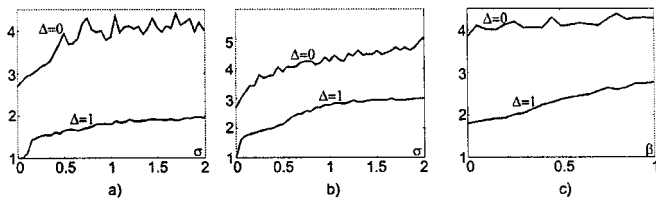
Fig. 9. Average number of iterations for the robust WD calculation. (a) $\beta = 0$ Gaussian white noise. (b) $\beta = 1$ heavy-tailed noise. (c) Fixed $\sigma = 1$ as function of $\beta$.

second sum in (31) does not require numerous iterations, as it was the case for the robust WD.

## V. CONCLUSION

The robust WD is introduced for processing of signals corrupted by additive impulse noise. It produces significantly better results than the standard WD in the impulse noise environment, whereas the results are slightly worse in a pure Gaussian environment. This distribution is calculated from the standard one through an iterative procedure. The improvement in TF representation is paid by an increase of computational complexity. Two different forms of the robust WD are considered. The robust WD, based on a combination of the squared absolute error and the absolute error, improves iteration convergence with respect to the case when only the absolute error is used as a loss function. Both of these forms behave similarly, and both are better than the standard WD. The IF estimation based on the robust WD is highly accurate for considered signals. It has been shown that an output median filter can directly be used for the IF estimated values. This may further improve the results.

## APPENDIX

*1) Realness:* The complex conjugate of expression (21) $\mathrm{WD}_R^*(t, \omega)$, for $w_h(nT) = w_h^*(-nT)$, is equal to $\mathrm{WD}_R(t, \omega)$ since equations $x = f(x)$ and $x^* = f(x^*)$ have the same solution $x_o = x_o^*$.

The property of realness holds for each iteration in the procedure (23)–(24). This can easily be proven by using mathematical induction and the fact that the initial distribution $\mathrm{WD}_R^{(0)}(t, \omega)$ is the standard WD, which is real-valued.

*2) TF Shift Property:* This property can be proven by considering (21) and making appropriate substitutions $t - t_0 \to t$, and $\omega - \omega_0 \to \omega$; therefore, we get (40), shown at the bottom of the page. Expression (40) is the same as (21). It can be concluded that $\mathrm{WD}_R(y; t + t_0, \omega + \omega_0) = \mathrm{WD}_R(x; t, \omega)$, and $\mathrm{WD}_R(y; t, \omega) = \mathrm{WD}_R(x; t - t_0, \omega - \omega_0)$.

## REFERENCES

[1] S. A. Kassam, *Signal Detection in Non-Gaussian Noise.* New York: Springer-Verlang, 1988.
[2] P. J. Huber, *Robust Statistics.* New York: Wiley, 1981.
[3] P. M. Narendra, "A separable median filter for image noise smoothing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI–3, pp. 20–29, Jan. 1981.
[4] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications.* Boston, MA: Kluwer, 1990.
[5] J. B. Bednar and T. L. Watt, "Alpha-trimmed means and their relationship to the median filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 145–153, Feb. 1984.
[6] C. Wang and M. G. Amin, "Zero-tracking time-frequency distributions," *Proc. IEEE ICASSP*, 1997.
[7] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—Parts 1, and 2," *Proc. IEEE*, vol. 80, pp. 521, 539–538, 560, Apr. 1992.
[8] D. L. Jones and R. G. Baraniuk, "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2361–2371, Oct. 1995.
[9] ——, "A simple scheme for adapting time-frequency representations," *IEEE Trans. Signal Processing*, vol. 42, pp. 3530–3535, Dec. 1994.
[10] H. K. Kwok and D. L. Jones, "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform," *IEEE Trans. Signal Processing*, vol. 48, pp. 2864–2972, Oct. 2000.
[11] J. C. O'Neill and P. Flandrin, "Virtues and vices of quadratic time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 48, pp. 2641–2650, Sept. 2000.
[12] S. Barbarossa and O. Lemoine, "Analysis of nonlinear FM signals by pattern recognition of their time-frequency representations," *IEEE Signal Processing Lett.*, vol. 3, pp. 112–115, Apr. 1996.
[13] A. Francos and M. Porat, "Analysis and synthesis of multicomponent signals using positive time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 47, pp. 493–504, Feb. 1999.
[14] M. G. Amin, "Minimum variance time-frequency distributions kernels for signals in additive noise," *IEEE Trans. Signal Processing*, vol. 44, pp. 2352–2356, Sept. 1996.
[15] S. B. Hearon and M. G. Amin, "Minimum variance time-frequency distribution kernel," *IEEE Trans. Signal Processing*, vol. 43, pp. 1258–1262, May 1995.
[16] V. Katkovnik, I. Djurović, and LJ. Stanković, "Instantaneous frequency estimation using robust spectrogram with varying window length," *AEUE*, vol. 54, no. 4, pp. 193–202, 2000.
[17] V. Katkovnik, "Robust $M$-periodogram," *IEEE Trans. Signal Processing*, vol. 46, pp. 3104–3109, 1998.
[18] ——, "Robust $M$-estimates of the frequency and amplitude of a complex-valued harmonic," *Signal Process.*, vol. 77, no. 1, pp. 71–84, Aug. 1999.
[19] B. T. Poljak and J. Z. Tsypkin, "Robust identification," *Automatika*, vol. 16, pp. 53–63, 1980.
[20] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, *Robust Estimation of Location: Survey and Advances.* Princeton, NJ: Prinston Univ. Press, 1972.
[21] L. Cohen, "Time-frequency distributions:—A review," *Proc. IEEE*, vol. 77, pp. 941–981, July 1989.

$$\mathrm{WD}_R(y; t + t_0, \omega + \omega_0)$$
$$= \frac{2}{\sum_n \frac{w_h(nT)}{|2x(t+nT)x^*(t-nT) - \mathrm{WD}_R(y; t+t_0, \omega+\omega_0)e^{2j\omega nT}|}}$$
$$\times \sum_{n=-\infty}^{\infty} \frac{w_h(nT)x(t+nT)x^*(t-nT)e^{-j2\omega nT}}{|2x(t+nT)x^*(t-nT) - \mathrm{WD}_R(y; t+t_0, \omega+\omega_0)e^{2j\omega nT}|}. \tag{40}$$

[22] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Mag.*, pp. 21–67, Apr. 1992.

[23] J. Jeong and W. Williams, "Kernel design for reduced interference distributions," *IEEE Trans. Signal Processing*, vol. 40, pp. 402–412, Feb. 1992.

[24] G. S. Cunningham and W. J. Williams, "Kernel decomposition of time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 42, pp. 1425–1442, June 1994.

[25] J. R. O'Hair and B. W. Suter, "The Zak transform and decimated time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 44, pp. 1099–1110, May 1996.

[26] G. S. Cunningham and W. J. Williams, "Fast implementation of generalized time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 42, pp. 1496–1508, June 1994.

[27] M. G. Amin, "Spectral decomposition of time-frequency distribution kernels," *IEEE Trans. Signal Processing*, vol. 42, pp. 1156–1165, May 1994.

[28] LJ. Stanković, "A method for time-frequency analysis," *IEEE Trans. Signal Processing*, vol. 42, pp. 225–229, Jan. 1994.

[29] L. L. Scharf and B. Friedlander, "Toeplitz and Hankel kernels for estimating time-varying spectra of discrete-time random processes," *IEEE Trans. Signal Processing*, vol. 49, pp. 179–189, Jan. 2001.

[30] S. Kalluri and G. R. Arce, "Fast algorithms for weighted myriad algorithm computation by fixed point search," *IEEE Trans. Signal Processing*, vol. 48, pp. 159–171, Jan. 2000.

[31] X. Wang and H. V. Poor, "Robust multiuser detection in non-Gaussian channels," *IEEE Trans. Signal Processing*, vol. 47, pp. 289–305, Feb. 1999.

[32] LJ. Stanković, "On the time-frequency analysis based filtering," *Ann. Telec.*, vol. 55, no. 5–6, pp. 216–225, May/June 2000.

[33] LJ. Stanković, S. Stanković, and I. Djurović, "Space/spatial-frequency analysis based filtering," *IEEE Trans. Signal Processing*, vol. 48, pp. 2343–2352, Aug. 2000.

[34] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectral analysis and the Weyl spectrum of nonstationary random processes," *IEEE Trans. Signal Processing*, vol. 45, pp. 1520–1534, June 1997.

[35] H. Kirchauer, F. Hlawatsch, and W. Kozek, "Time-frequency formulation and design of nonstationary Wiener filters," in *Proc. ICASSP*, 1995, pp. 1529–1552.

[36] W. Kozek, "Time-frequency processing based on the Wigner-Weyl framework," *Signal Process.*, vol. 29, no. 1, pp. 77–92, Oct. 1992.

[37] G. F. Boudreaux-Bartels and T. W. Parks, "Time-varying filtering and signal estimation using the Wigner distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 442–451, June 1986.

[38] P. Rao and F. J. Taylor, "Estimation of the instantaneous frequency using the discrete Wigner distribution," *Elect. Lett.*, vol. 26, pp. 246–248, 1990.

[39] P. Petrus, "Robust Huber adaptive filter," *IEEE Trans. Signal Processing*, vol. 47, pp. 1129–1133, Apr. 1999.

[40] I. Djurović and LJ. Stanković, "Influence of high noise on the instantaneous frequency estimation using quadratic time-frequency distributions," *IEEE Signal Processing Let.*, vol. 7, pp. 317–319, Nov. 2000.

**Igor Djurović** was born in Montenegro in 1971. He received the B.S., M.S., and Ph.D. degrees, all in electrical engineering, from the University of Montenegro, Podgorica, in 1994, 1996, and 2000, respectively.

During 2002, he will be on leave at the Kyoto Institute of Technology, Kyoto, Japan, supported by the JSPS foundation. His current research interests include application of virtual instruments, time-frequency analysis-based methods for signal estimation and filtering, fractional Fourier transform applications, image processing, robust estimation, and digital watermarking. He is the co-author of three books.



**LJubisa Stanković** (M'91–SM'96) was born in Montenegro on June 1, 1960. He received the B.S. degree from the University of Montenegro, Podgorica, in 1982, with the honor of "the best student at the University," the M.S. degree in 1984 from the University of Belgrade, Belgrade, Yugoslavia, and the Ph.D. degree in 1988 from the University of Montenegro, all in electrical engineering.

As a Fulbright grantee, he spent 1984 to 1985 at the Worcester Polytechnic Institute, Worcester, MA. He was also active in politics, as a Vice President of the Republic of Montenegro from 1989 to 1991 and then as leader of democratic (anti-war) opposition in Montenegro from 1991 to 93. Since 1982, he has been on the faculty at the University of Montenegro, where he presently holds the position of Full Professor. From 1997 to 1999, he was on leave at the Ruhr University Bochum, Bochum, Germany, with Signal Theory Group, supported by the Alexander von Humboldt foundation. In 2001, he spent a month with the Technische Universiteit Eindhoven, Eindhoven, The Netherlands, as a Visiting Professor. His current interests are in signal processing and electromagnetic field theory. He has published about 200 technical papers, 50 of them in leading international journals. He has published several textbooks such as *Signal Processing* (in Serbo-Croat) and the monograph *Time-Frequency Signal Analysis* (in English).

Dr. Stanković was awarded the highest state award of the Republic of Montenegro for scientific achievements in 1997. He is a member of the National Academy of Science and Art of Montenegro (CANU).