Sparsity and Concentration Measures for Optimum Quadratic Time-frequency Distributions of Doppler Signals

Branka Jokanovic and Moeness Amin
Center for Advanced Communications, Villanova University
Villanova, PA 19085, USA
Email:{branka.jokanovic, moeness.amin}@villanova.edu

***Abstract*—Reduced interference distributions are well-known methods used for quadratic time-frequency signal analysis of nonstationary data including Doppler signals. These distributions use kernels that determine the quality of a time-frequency signal representation. Optimal reduced interference distribution kernels are determined based on the employed design criteria, including sparsity and concentration measures that are applied in the time-frequency domain. This paper demonstrates that sparsity, compared to common concentration measures, satisfies desirable performance attributes and, as such, is considered effective in revealing the signal local frequency structure and its time-frequency signature, including the signal instantaneous frequency.**

I. INTRODUCTION

Fourier transform is a standard method for analyzing stationary signals. However, the majority of signals in practice are nonstationary, i.e., their spectral contents change with time. Time-frequency representations (TFRs) are typically used to properly examine the signal local frequency characteristics and reveal the signal time-varying spectral behaviour [1], [2]. Various TFRs have been defined in the literature, but none provides ideal time-frequency signature for all signal types [1]. The simplest TFR is the Short-time Fourier transform (STFT) which is defined as the Fourier transform of the data over sliding windows in time. The role of these windows is to partition the signals into stationary segments over which Fourier transform becomes meaningful. Even though STFT is appealing due to its simplicity, this TFR and its corresponding energetic distribution, spectrogram, suffer from the resolution dependency on the window size. Namely, short windows provide good temporal resolution, but poor frequency resolution and vice versa.

Improvement of the time-frequency resolution can be achieved using the Wigner-Ville distribution (WVD)[2]. This distribution is obtained as the Fourier transform of the signal instantaneous autocorrelation function (IAF). WVD provides an ideal time frequency signature for linearly frequency modulated signals, i.e., chirps. However, for the case of multiple components, WVD suffers from cross-terms which are induced by the bilinear products of the IAF. In order to suppress the cross terms, reduced interference distributions (RIDs) were introduced [1], [2]. These distributions seek to reduce the cross-terms in the WVD using low-pass kernels which multiply the signal ambiguity function. For the majority of signals, the auto-terms reside around the origin while cross-terms appear distant from both the origin and the two axes in the ambiguity domain. After filtering the ambiguity function, the two dimensional (2D) Fourier transform is applied in order to obtain the signal TFR.

Even though RIDs assume two important and desirable qualities, namely, high concentration and cross-term suppression, the tradeoffs between these two qualities can significantly vary depending on the kernel employed. With parameterized kernels, the parameters would need to be tuned so accurate time-frequency signatures for multi-component signals can be obtained. Various criteria can be used for adjusting the kernel parameters. The signal power concentration in the time-frequency plane is often used to qualitatively describe the signal time-frequency structure. There are several well-established concentration measures which exist for evaluating the suitability of the time-frequency signal representation [3]-[6]. A rival property which is linked to the efficiency of the TFR is sparsity. Namely, desirable RIDs can be considered sparse since only a portion of the time-frequency plane is occupied, even in the presence of multi-component signals [7]-[9]. It is noted that in radar and sonar applications, these complex nonstationary signals represent target Doppler signals under translation, rotation, vibration, and oscillation motions.

In this paper, we compare the concentration and the sparsity measures in view of their roles in producing accurate TFR close to the ground truth or the analytical formulation. We demonstrate that the sparsity measure provides more efficient TFR, where efficiency is ultimately defined as cross-terms free distribution with resolution resembling that of the WVD.

The paper is organized as follows. In Section II, RIDs are briefly described. Section III describes the effect of a kernel on RID sparsity and concentration. Also, in this section, different sparsity and concentration measures are compared. Simulation results are given in Section IV, whereas conclusion is given in Section V.

II. REDUCED INTERFERENCE DISTRIBUTIONS

The RID of a signal $s(t)$ is defined as

$$RID(t,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\theta,\tau)C(\theta,\tau)e^{-\jmath\theta t}e^{-\jmath\tau\omega}d\theta d\tau, \quad (1)$$

where t and ω denote time and frequency, respectively, $C(\theta,\tau)$ is the kernel, whereas $A(\theta,\tau)$ is the ambiguity function of signal $s(t)$. The ambiguity function can be formulated as

$$A(\theta,\tau)=\int_{-\infty}^{\infty} s(t+\tau/2)s^*(t-\tau/2)e^{-\jmath\theta t}dt. \tag{2}$$

The variables θ and τ represent Doppler frequency and time lag, respectively. Various kernels have been designed in order to suppress cross-terms while preserving the signal auto-terms. Kernels are commonly designed as low-pass filters which are intended to capture the signal auto-terms and leave out most of the signal cross-terms.

III. Determining the optimal distribution

A. Sparsity and concentration of RIDs

In this section, we analyze the effect of the kernel passband on the RID sparsity. Based on (1), we observe that by changing the kernel parameters, we change the auto and cross-terms shape in the RID. This shape determines various properties of the RID, including sparseness and concentration.

Without loss of generality, we consider a signal composed of two sinusoids, i.e., $s_1(t)=e^{\jmath\omega_1 t}$ and $s_2(t)=e^{\jmath\omega_2 t}$. This signal is taken for consideration due to the fact that the majority of signals can be represented as a sum of sinusoids during a short period of time. For the sake of simplicity, we use two sinusoids. The ambiguity function of the observed signal can be written as

$$\begin{aligned}A(\theta,\tau)=&e^{\jmath\omega_1\tau}\delta(\theta)+e^{\jmath\omega_2\tau}\delta(\theta)\\&+e^{\jmath(\omega_1+\omega_2)\frac{\tau}{2}}[\delta(\omega_1-\omega_2-\theta)+\delta(\omega_2-\omega_1-\theta)].\end{aligned} \tag{3}$$

The first two terms in the above equation represent the signal auto-terms which are located along the entire τ axis. The third term denotes the signal cross-terms which reside away from $\theta=0$ for $\omega_1\neq\omega_2$. In order to obtain the RID, we use the radial Gaussian kernel. This kernel is defined as

$$C(\theta,\tau)=e^{-\frac{\theta^2+\tau^2}{2\sigma^2}}, \tag{4}$$

where σ^2 is the kernel variance. Three distinct cases can occur when changing the kernel variance. This is illustrated in Fig. 1. The upper row shows RIDs, while the bottom one shows plots of a time slice taken from the corresponding RIDs. The variances considered are $\sigma_1^2=1/6$, $\sigma_2^2=50$ and $\sigma_3^2=250$.

We can notice that the first case (Fig. 1 (a)), which corresponds to the smallest kernel variance, depicts distorted auto-terms and shows lack of term concentrations and the lowest sparsity. Therefore, we can expect that neither sparsity or concentration measure will favor this specific variance. The auto-terms in the third case are more concentrated than in the second case. Concentration is defined by the bandwidth measured by drop in amplitude by $e\approx 2.7182$ times. Even though the signal TFR in the third case is more concentrated, it contains an undesirable cross-term which reduces TFR sparsity. Intuitively, the second case should be sparser than the third case, since the latter contains an additional component. This paradigm is made clear in the following analysis.

Since sparsity is commonly defined as the number of non-zero components, it is important to compute both the auto and cross-terms bandwidths. The RIDs auto and cross-terms properties, including their amplitude and bandwidth, are well studied in the literature [10]. In this paper, these properties are observed from the sparsity perspective. Based on (3) and (4), the auto-terms in the time-frequency plane are expressed as

$$\begin{aligned}AT(t,\omega)=&\int_{-\infty}^{\infty}(e^{\jmath\omega_1\tau}+e^{\jmath\omega_2\tau})e^{-\frac{\tau^2}{2\sigma^2}}e^{-\jmath\omega\tau}d\tau\\=&\sigma\sqrt{2\pi}e^{-\frac{\sigma^2}{2}(\omega-\omega_1)^2}+\sigma\sqrt{2\pi}e^{-\frac{\sigma^2}{2}(\omega-\omega_2)^2}.\end{aligned} \tag{5}$$

The bandwidth of the first auto-term is computed using the following relation

$$\frac{\sigma\sqrt{2\pi}}{e}=\sigma\sqrt{2\pi}e^{-\frac{\sigma^2}{2}(\omega-\omega_1)^2}. \tag{6}$$

Based on (6), we can compute the bandwidth for both auto-terms B_{AT} as follows:

$$\omega=\omega_1\pm\frac{\sqrt{2}}{\sigma}\Rightarrow B_{AT}=4\frac{\sqrt{2}}{\sigma}. \tag{7}$$

Similarly, the expression for the cross-term is given by

$$\begin{aligned}CT(t,\omega)=&\int_{-\infty}^{\infty}e^{\jmath\frac{\tau}{2}(\omega_1+\omega_2)}e^{-\frac{\tau^2+(\omega_1-\omega_2)^2}{2\sigma^2}}e^{-\jmath(\omega_1-\omega_2)t}e^{-\jmath\omega\tau}d\tau\\&+\int_{-\infty}^{\infty}e^{\jmath\frac{\tau}{2}(\omega_1+\omega_2)}e^{-\frac{\tau^2+(\omega_1-\omega_2)^2}{2\sigma^2}}e^{-\jmath(\omega_2-\omega_1)t}e^{-\jmath\omega\tau}d\tau\\=&\sigma\sqrt{2\pi}e^{-\frac{\sigma^2}{2}(\omega-\frac{\omega_1+\omega_2}{2})^2}e^{-\frac{(\omega_1-\omega_2)^2}{2\sigma^2}}\cos(\omega_1-\omega_2)t.\end{aligned} \tag{8}$$

As we gradually increase the kernel variance, the cross-term bandwidth can be computed using the same threshold as for the auto-terms, i.e., $\sigma\sqrt{2\pi}/e$,

$$B_{CT}=2\frac{\sqrt{2}}{\sigma}\sqrt{1-\frac{(\omega_1-\omega_2)^2}{2\sigma^2}+\ln[\cos(\omega_1-\omega_2)t]}. \tag{9}$$

It is clear from the above analysis that the bandwidths of both types of time-frequency terms change with the kernel variance. These bandwidths can be used to measure sparsity, since they provide the number of occupied frequency bins. In this respect, in order to cast the second case of the variance to be sparser than the third case, the following condition should be satisfied:

$$B_{AT}(\sigma_2)<B_{AT}(\sigma_3)+B_{CT}(\sigma_3). \tag{10}$$

It can be noticed that the fulfilment of this condition depends not only on the used kernel variances, but also on the signal components. Measuring the bandwidth, or equivalently, the number of occupied frequency bins, corresponds to the ℓ_0 norm. However, it is well-known that this sparsity measure is not suitable in the presence of noise. Another norm which is commonly used is ℓ_1 norm. Based on the previous analysis, unfortunately there are no indications that ℓ_1 norm is sensitive to the cross-terms since the sum of the absolute values does not discriminate between signal auto and cross-terms. This paradox confirms that even though sparsity can be a useful indicator when searching for the optimal RID, an appropriate metric is required when measuring sparsity in the time-frequency plane.

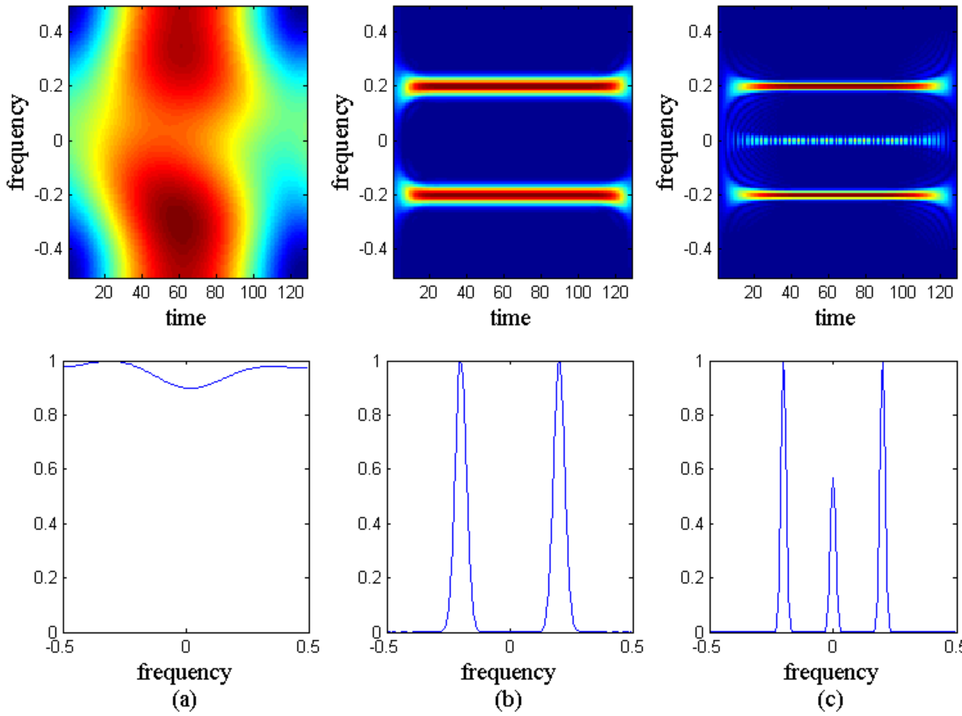

Fig. 1. RIDs and corresponding time slices at *t*=64. Three cases in the RID kernel adjustment: (a) Distorted auto-terms, (b) Cross-terms free TFR, (c) TFR which contains cross-terms. Plots are normalized.

B. Measures for determining the optimal RID

The above discussion, although based on a simple example and specific kernel, serves to strengthen the general argument of the importance of having a proper time-frequency performance measure. In essence, we require a measure which can distinguish the three aforementioned critical cases decided by the kernel. The task of finding the optimal RID can be described as searching for the TFR which is both well-concentrated and sparse, but has reduced cross-terms. In [11], the authors defined several attributes which should be held by any sparsity measure. They proved that, among the examined measures, the Gini index and pq-mean satisfy all desirable properties.

In what follows, we examine several well-known sparsity and concentration measures. For a signal $\mathbf{s} \in C^N$, we observe the discrete RID $\mathbf{X}$. More specifically, we use the absolute form of the normalized $\mathbf{X}$ or its vectorized version $\mathbf{x}$.

ℓ_0 and ℓ_1 norm

As previously mentioned, the ℓ_0 norm is commonly used to describe sparsity. The ℓ_0 norm for vector $\mathbf{x}$ is defined as

$$\ell_0(\mathbf{x}) = card(\mathbf{x} \neq 0). \tag{11}$$

It is possible to use the threshold to detect the number of components which are above the threshold value, but the determination of that threshold is itself a difficult task. The ℓ_1 norm can be used as an alterative to the ℓ_0 norm. This norm is defined as the sum of the absolute values in vector $\mathbf{x}$.

TFR kurtosis

TFR kurtosis is one of the well-known measures used in time-frequency analysis [3]. It is defined as

$$k(\mathbf{X}) = \frac{\sum_n \sum_k X^4(n,k)}{(\sum_n \sum_k X^2(n,k))^2}. \tag{12}$$

Even though this measure satisfies few desirable properties for measuring sparsity and concentration, it has one disadvantage. Namely, in the case of a multicomponent signal, it favors a single highly localized component over less, but similarly localized components in the time-frequency domain. Thus, it is expected that this norm would fail in the presence of cross-terms.

Renyi entropy

Another approach of determining the quality of the TFR is using the entropy measures [12]. Renyi entropy of third order is a traditionally used measure in time-frequency analysis, and it is computed as

$$H_\alpha(\mathbf{X}) = \frac{1}{1-\alpha} log_2 \sum_n \sum_k X(n,k)^\alpha, \tag{13}$$

where $\alpha = 3$.

pq-mean

pq-mean is a sparsity measure formulated as

$$pq(\mathbf{x}) = -\frac{(\frac{1}{N^2}\sum_{m=1}^{N^2} x^p(m))^{1/p}}{(\frac{1}{N^2}\sum_{m=1}^{N^2} x^q(m))^{1/q}}, \tag{14}$$

where $p < q$. In this paper, we use $p = 1, q = 2$.

Gini index

The Gini index is a sparsity measure [13], [14], which for sorted form of vector $\mathbf{x}, x(1) \leq ... \leq x(N^2)$ is defined as

$$GI(\mathbf{x}) = 1 - 2 \sum_{m=1}^{N^2} \frac{x(m)}{\|x\|_1} \frac{N^2 - m + 0.5}{N^2}. \tag{15}$$

As the number of non-zero values of the signal decreases, the value of the Gini index increases. The pq-mean and the Gini index have shown to satisfy several desirable properties. One of the advantages of the Gini index is its normalization so it assumes the values in the range between 0 and 1 for any given vector. This gives a useful interpretation of the sparsity and enables a more efficient monitoring of the change in sparsity over kernel parameters. An alternative definition of the Gini index is based on the relative mean difference [15], i.e.,

$$GI(\mathbf{x}) = \frac{\sum_{m=1}^{N^2} \sum_{n=1}^{N^2} ||x(m)| - |x(n)||}{2N\|\mathbf{x}\|_1}. \tag{16}$$

This definition shows another advantage of using the Gini index. Namely, the Gini index takes into account the energy distribution of the nonzero coefficients. This property can be useful in detecting cross-terms since these terms can reduce the difference between two coefficients in vector $\mathbf{x}$.

IV. Simulation Results

In this section, we evaluate the measures described previously for obtaining the optimal RID. Radial Gaussian kernel is used as a parameterized kernel. For each measure, we use the following range of variances 1:10:1000.

Example 1: We observe a signal composed of a chirp and a sinusoid. Fig. 2 shows plots of different sparsity measures. These plots depict how the sparsity of a noiseless TFR changes for different variances. The TFRs, corresponding to the optimal points in Fig. 2, are shown in Fig. 3. We can notice that the ℓ_0 norm (Fig. 3(a)) provides the most efficient solution, thus underscoring the general argument that the sparsity, i.e., the number of occupied frequency bins, is considerably altered when the cross-terms are included. However, it should be noted that this is a noiseless signal. Among other measures, the Gini index provided the TFR with the most suppressed cross-terms (Fig. 3(f)).

Example 2: In this example, a signal consisting of two components with 3-degree polynomial phase is considered (SNR=10dB). Results are shown in Fig. 4. The sensitivity of the ℓ_0 norm, even with an applied threshold, demonstrates the inefficiency of this measure (Fig. 4 (a)). The solutions which most resemble the optimal case are provided by pq-mean and Gini index. It can be noticed from Figs. 4, 5 that in the noisy case, sparsity measures, such as pq-mean and Gini index, tend to choose smaller variance, and hence, provide TFR with more suppressed cross-terms. This behaviour is attributed to the fact that, by increasing the kernel extent in the ambiguity domain, the noise power becomes higher which significantly affects the TFR sparsity. These results confirm that the Gini index can be successfully used for determining a sparse RID.

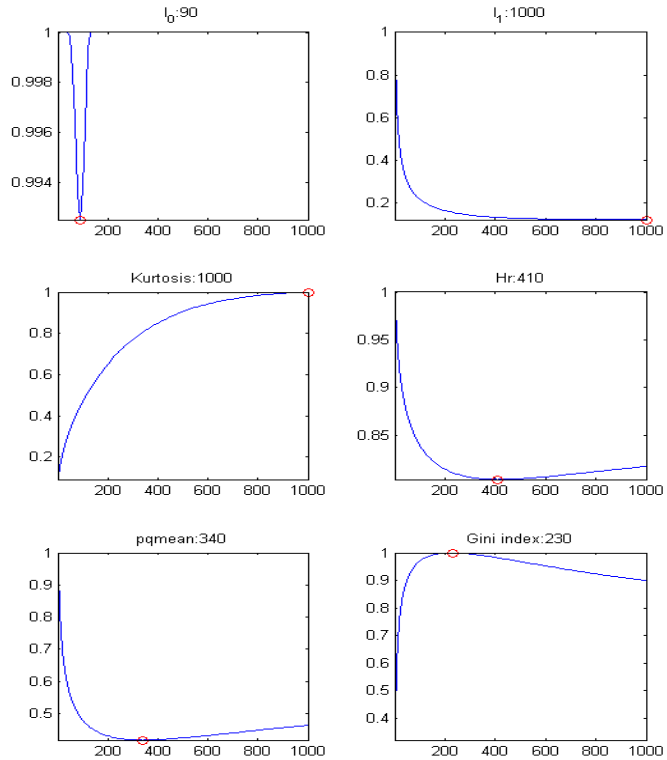

Fig. 2. Different measures applied to the TFRs. x axis represents the variance of the radial Gaussian kernel. The optimal points, according to the applied measures, are denoted by red circles and their variance value is also depicted. The measures values are normalized to be in the range [0,1], except the Gini index which already provides results in this range.

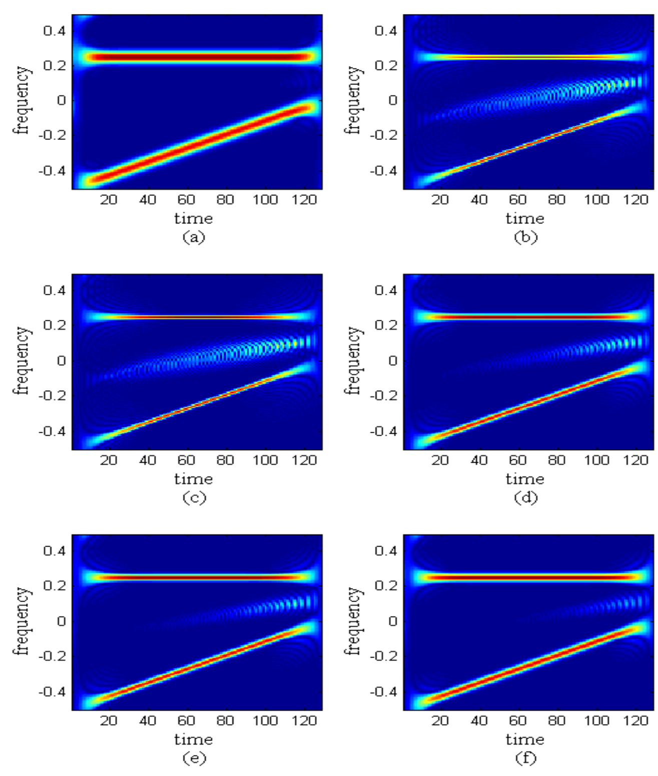

Fig. 3. TFRs corresponding to the optimal points according to the measures in Fig. 2 (a) ℓ_0 norm result, (b) ℓ_1 norm result, (c) Kurtosis, (d) Renyi entropy, (e) pq-mean ($p = 1, q = 2$), (f) Gini index.

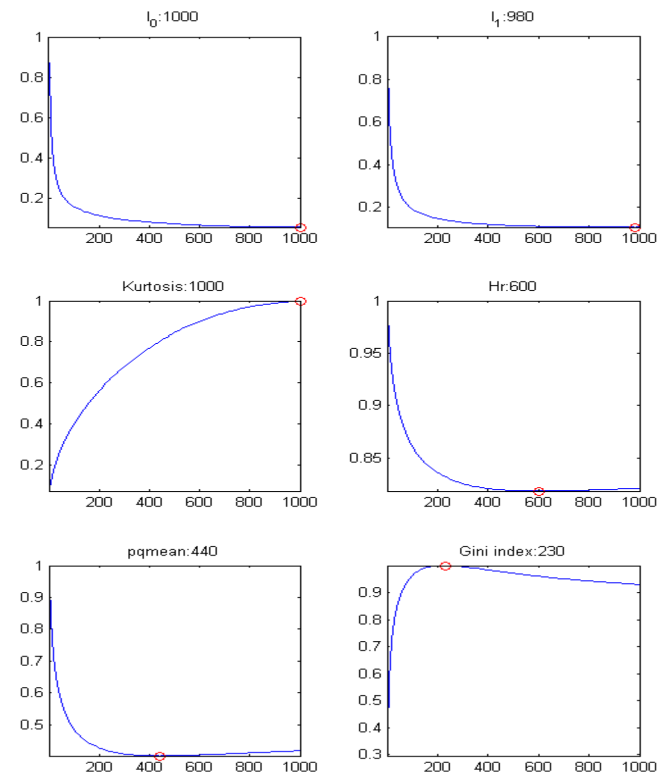

Fig. 4. Different measures applied to the TFRs of a signal consisting of two components with 3-degree polynomial phase.

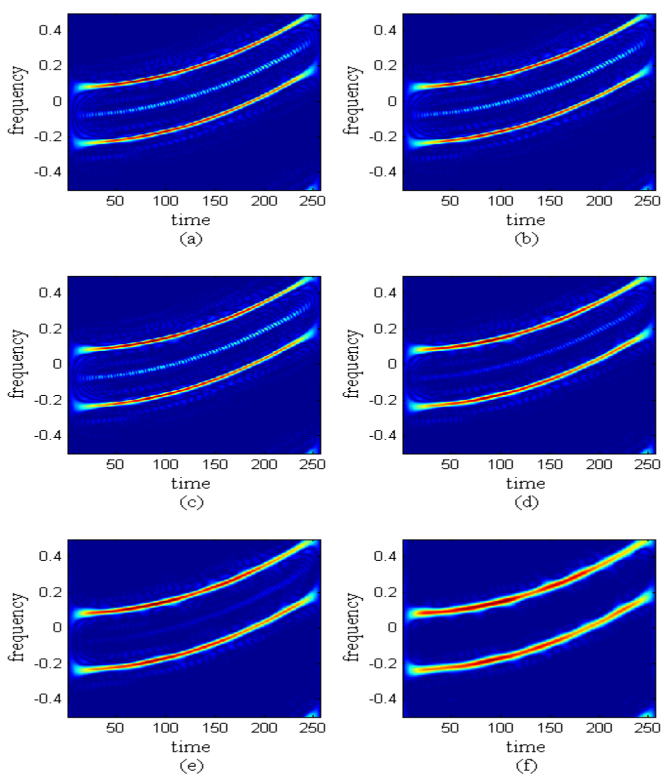

Fig. 5. TFRs corresponding to the optimal points according to the described measures when SNR=10dB: (a) ℓ_0 norm result, (b) ℓ_1 norm result, (c) Kurtosis, (d) Renyi entropy, (e) pq-mean ($p = 1, q = 2$), (f) Gini index. When computing ℓ_0 norm, we use threshold-based definition. Threshold is set to the $1/e$ of the maximum TFR value.

V. Conclusion

This paper examined sparsity in the time-frequency domain from a quadratic power distribution perspective. It was shown that one can improve time-frequency representations of non-stationary signals using sparsity, in lieu of concentration, as a performance measure in the time-frequency domain. However, the presence of undesirable distribution cross-terms, which are inherent in multicomponent signal TFR, renders ℓ_0 and ℓ_1 based sparsity measures inadequate and ineffective. Simulation results showed that, among the employed measures, the Gini sparsity index is considered most reliable for finding both sparse and cross-terms free time-frequency signal representation.

References

[1] L. Cohen, *Time-Frequency Analysis: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1995.

[2] B. Boashash, Ed., *Time-Frequency Signal Analysis and Processing. A Comprehensive Reference*, 1st ed. Oxford, U.K.: Elsevier, 2003.

[3] D. L. Jones and T. W. Parks, “A high resolution data-adaptive time-frequency representation,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 12, pp. 2127–2135, 1990.

[4] E. Sejdić, I. Djurović, and J. Jiang, “Time–frequency feature representation using energy concentration: An overview of recent advances,” *Digital Signal Processing*, vol. 19, no. 1, pp. 153–183, 2009.

[5] M. G. Amin and W. J. Williams, “High spectral resolution time-frequency distribution kernels,” *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2796–2804, 1998.

[6] L. Stanković, “A measure of some time-frequency distributions concentration,” *Signal Processing*, vol. 81, no. 3, pp. 621–631, 2001.

[7] P. Flandrin and P. Borgnat, “Time-frequency energy distributions meet compressed sensing,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 2974–2982, 2010.

[8] L. Stankovic, I. Orovic, S. Stankovic, and M. Amin, “Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency,” *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4562–4572, Sep. 2013.

[9] M. G. Amin, B. Jokanovic, and T. Dogaru, “Reconstruction of locally frequency sparse nonstationary signals from random samples,” in *Proc. Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, Sep. 2014, pp. 1–5.

[10] L. Stankovic, “Auto-term representation by the reduced interference distributions: a procedure for kernel design,” *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1557–1563, 1996.

[11] N. Hurley and S. Rickard, “Comparing measures of sparsity,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.

[12] R. G. Baraniuk, P. Flandrin, A. J. Janssen, and O. J. Michel, “Measuring time-frequency information content using the Rényi entropies,” *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1391–1409, 2001.

[13] H. Dalton, “The measurement of the inequality of incomes,” *The Economic Journal*, pp. 348–361, 1920.

[14] J. L. Gastwirth, “The estimation of the Lorenz curve and Gini index,” *The Review of Economics and Statistics*, pp. 306–316, 1972.

[15] D. Zonoobi, A. A. Kassim, and Y. V. Venkatesh, “Gini index as sparsity measure for signal reconstruction from compressive samples,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 927–932, 2011.