MEASURING TIME-FREQUENCY INFORMATION CONTENT USING THE RÉNYI ENTROPIES

Richard Baraniuk,[◇] *Patrick Flandrin,*[□] *and Olivier Michel* [□] *

[◇] Department of Electrical and Computer Engineering
Rice University
P.O. Box 1892, Houston, TX 77251–1892, USA
E-mail: richb@rice.edu, Fax: (713) 524–5237

[□] Laboratoire de Physique (URA 1325 CNRS)
Ecole Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France
E-mail: flandrin@physique.ens-lyon.fr
omichel@physique.ens-lyon.fr

Submitted to *IEEE Transactions on Information Theory*, September 1995
Index Terms: Time-frequency analysis, Wigner distribution, Rényi entropies, complexity

Abstract

Of the many measures proposed for estimating signal information content and complexity on the time-frequency plane, entropy measures borrowed from probability theory show the greatest promise. When applied to a time-frequency representation from Cohen's class or the affine class, the generalized Rényi entropies conform closely to the visually based notion of complexity that we use when inspecting time-frequency images. In addition, these measures possess several interesting and useful properties, such as accounting and cross-component and transformation invariances, that make them natural for time-frequency analysis. This paper comprises a detailed study of the properties and several potential applications of the Rényi entropies, with emphasis on establishing a firm mathematical foundation for quadratic time-frequency representations.

*This work was supported by the National Science Foundation, grant MIP–9457438, the Office of Naval Research, grant N00014–95–1–0849, the Texas Advanced Technology Program, grant 003604–002, the NATO Collaborative Research Programme, grant GRG–950202, and URA 1325 CNRS.

Portions of this work have been presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing in Adelaide, Australia, May 1994, and at the IEEE International Symposium on Information Theory in Whistler, Canada, September 1995.

1 Introduction

The term *component* is ubiquitous in the signal processing literature. Intuitively, a component is a concentration of energy in some domain, but this notion is difficult to translate into a quantitative concept [1–3]. In fact, the concept of a signal component may never be clearly defined.

The use and abuse of this term is particularly severe in the literature on time-frequency analysis. Time-frequency representations (TFRs) generalize the concept of the time and frequency domains to a joint time-frequency function $C_s(t,f)$ that indicates how the frequency content of signals change over time [4, 5]. Common themes in the literature include the suppression of TFR cross-components, the concentration and resolution of auto-components, and the property that the time-varying spectral analysis of TFRs separate signal components such as parallel chirps that overlap in both time and frequency. Moreover, the quality of particular TFRs is very often judged based on subjective criteria related to the components of the signal being analyzed.

In this paper, rather than address the question "what is a component?" directly, we will investigate a class of quantitative measures of deterministic signal *complexity* and *information content*. While they do not yield direct answers regarding the locations and shapes of components, these measures are intimately related to the concept of a signal component, the connection being the intuitively reasonable supposition that signals of high complexity (and therefore high information content) must be constructed from large numbers of elementary components. Viable measures include moments and entropies of the signal represented in the time, frequency, or time-frequency domains.

Moment-based measures, such as the time-bandwidth product and its generalizations to second-order time-frequency moments [4–7] have found wide application, but unfortunately measure neither signal complexity nor information content [1, 2]. To demonstrate, consider a signal comprised of two components of compact support, and note that while the time-bandwidth product increases without bound with separation, signal complexity clearly does not rise once the components become disjoint.

A more promising approach to complexity based on entropy functionals exploits the powerful analogy between signal energy densities and probability densities [1].[1] Just as the instantaneous and spectral amplitudes $|s(t)|^2$ and $|S(f)|^2$ behave as unidimensional densities of signal energy in time and frequency, TFRs try very hard to act as bidimensional energy densities in time-frequency. In particular, there exist TFRs whose marginal properties parallel those of probability densities:

$$\int C_s(t,f)\,df \;=\; |s(t)|^2, \qquad \int C_s(t,f)\,dt \;=\; |S(f)|^2, \tag{1}$$

$$\iint C_s(t,f)\,dt\,df \;=\; \int |s(t)|^2\,dt \;\equiv\; \|s\|_2^2. \tag{2}$$

The quadratic TFRs of the large and useful Cohen's class can be obtained as the convolution [4]

$$C_s(t,f) \;=\; \iint W_s(u,v)\,\Phi(t-u,\,f-v)\,du\,dv \;\equiv\; (W_s * \Phi)(t,f) \tag{3}$$

[1] With the exception of Section 4.3, we will deal only with deterministic signals in this paper. Do not confuse the Rényi entropy of the *time-frequency* distribution of a signal with the Rényi entropy of its *probability* distribution.

of a kernel function Φ with the Wigner distribution W_s of the signal

$$W_s(t,f) \;=\; \int s\left(t+\frac{\tau}{2}\right) s^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\tau f}\, d\tau. \tag{4}$$

The probabilistic analogy evoked by (1) and (2) suggests the classical Shannon entropy

$$H(C_s) \;=\; -\iint C_s(t,f)\, \log_2 C_s(t,f)\, dt\, df \tag{5}$$

as a natural candidate for estimating the complexity of a signal through its TFR: The peaky TFRs of signals comprised of small numbers of elementary components would yield small entropy values, while the diffuse TFRs of more complicated signals would yield large entropy values. Unfortunately, however, the negative values taken on by most TFRs (including all fixed-kernel Cohen's class TFRs satisfying (1)) prohibit the application of the Shannon entropy due to the logarithm in (5).

In [1], Williams, Brown, and Hero sidestepped the negativity issue by employing the generalized entropies of Rényi [8]

$$H_\alpha(C_s) \;=\; \frac{1}{1-\alpha} \log_2 \iint C_s^\alpha(t,f)\, dt\, df, \tag{6}$$

shown here applied to a normalized TFR. Parameterized by $\alpha > 0$, this class of information measures is obtained simply by relaxing the mean value property of the Shannon entropy from an arithmetic to a geometric mean. (Shannon entropy appears as $\alpha \to 1$.) In several empirical studies, Williams, Brown, and Hero found that in addition to appearing immune to the negative TFR values that invalidate the Shannon approach, the third-order Rényi entropy seemed to measure signal complexity. Figure 1 repeats the principal experiment of [1]: The third-order Rényi entropy $H_3(W_s)$ of the Wigner distribution of the sum $s(t) = g(t)+g(t+\Delta t)$ of two lowpass Gaussian pulses is plotted versus the separation distance Δt. (At $\Delta t = 0$, the two pulses coincide and therefore, because of the assumed energy renormalization, have the same information content as a solitary pulse.) The time-bandwidth product of $s(t)$ is also plotted. It is clear from the Figure that, unlike the time-bandwidth product, which grows without bound with Δt, the Rényi entropy saturates exactly one bit above the value $H_3(W_g) = -0.208$.[2] Similar results hold for three separated copies of $g(t)$ ($\log_2 3$ bits information gain), four copies (2 bits information gain), and so on. To summarize, independent of the definition of signal component, the Rényi entropy indicates a "doubling of complexity" in $s(t)$ as Δt moves from 0 to ∞.

This paper comprises a detailed study of the properties and some potential applications of the promising Rényi time-frequency information measures (6), with emphasis on establishing a firm mathematical foundation for quadratic TFRs. In Section 2, after reviewing the development of these measures, we prove that the Rényi entropies of integer orders $\alpha > 1$ are defined for essentially all Cohen's class TFRs, including even those distributions taking locally negative values. In Section 3, we investigate the five key properties of the Rényi entropies that make them particularly well suited to time-frequency analysis [6, 7]:

[2] Readers should not be alarmed by negative Rényi entropy values. Even the Shannon entropy takes on negative values for certain distributions in the continuous-variable case.

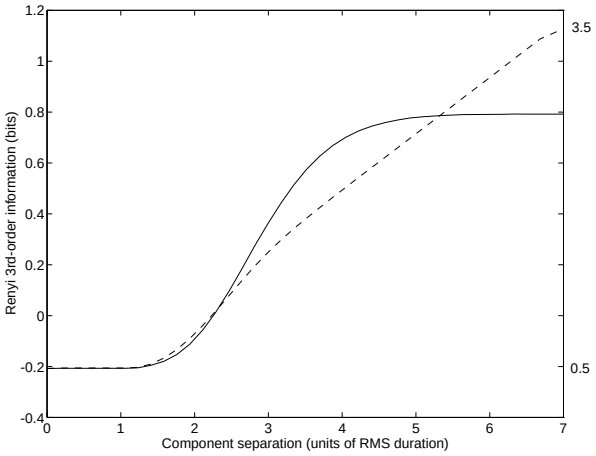


Figure 1: *Solid curve: Third-order Rényi entropy $H_3(W_s)$ of the Wigner distribution of the sum $s(t) = g(t) + g(t + \Delta t)$ of two lowpass Gaussian pulses plotted versus separation distance Δt normalized to units of RMS time width. Dotted curve: Time-bandwidth product of $s(t)$.*

1. $H_\alpha(C_s)$ counts the "number of components" in a broad class of multicomponent signals.

2. For odd orders $\alpha > 1$, $H_\alpha(C_s)$ is asymptotically invariant to TFR cross-components and therefore does not count them.

3. $H_\alpha(C_s)$ exhibits extreme sensitivity to phase differences between closely spaced components. (This sensitivity can be reduced through smoothing in time-frequency.)

4. The range of $H_\alpha(C_s)$ values is bounded above and below. For the Wigner distribution, a single Gaussian pulse attains the lower bound, while "deterministic white noise" nears the upper bound.

5. The values of $H_\alpha(C_s)$ are invariant to time and frequency shifts of the signal. Certain TFRs provide an additional invariance to scale changes, while the Wigner distribution boasts complete invariance to affine transformations on the time-frequency plane. For more general invariances, the Rényi theory extends easily to encompass not only the TFRs of the affine class [9] but also the generalized representations of the unitarily equivalent Cohen's and affine classes [10].

In Section 4, we discuss the role of these measures in adaptive time-frequency analysis, introduce the notion of Rényi dimension, and apply these measures to random signals. We close with a discussion and conclusions.

2 Rényi Entropy

2.1 Rényi entropy of a probability density

In the 1960s, Rényi introduced an alternative axiomatic derivation of entropy based on *incomplete* probability densities $p = \{p_1, p_2, \ldots p_n\}$ whose total probabilities sum to $w(p) \equiv \sum_i p_i \leq 1$ [8]. He observed that the Shannon entropy $H(p) = -\sum_i p_i \log_2 p_i$ uniquely satisfies the axioms of symmetry, continuity, normalization, additivity, and, in addition, the mean value condition

$$H(p \cup q) \;=\; \frac{w(p)\,H(p) \;+\; w(q)\,H(q)}{w(p) \;+\; w(q)}. \tag{7}$$

Here p and q are any two incomplete densities such that $w(p) + w(q) \leq 1$, and $p \cup q$ signifies the composite density $\{p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_m\}$.

Extending the arithmetic mean in (7) to a generalized mean yields generalized entropies closely resembling Shannon's. Considering the generalized mean value condition

$$H^{\mathrm{R}}(p \cup q) \;=\; m^{-1}\!\left(\frac{w(p)\,m[H(p)] \;+\; w(q)\,m[H(q)]}{w(p) \;+\; w(q)}\right), \tag{8}$$

with m a continuous monotone function, Rényi demonstrated that just two functions m are compatible with the other four axioms. The first, $m_1(x) \;=\; ax \;+\; b$, yields the arithmetic mean (7) and the Shannon entropy. The second,

$$m_2(x) \;=\; 2^{(\alpha-1)x}, \qquad\qquad \alpha > 0, \quad \alpha \neq 1,$$

yields the functional

$$H^{\mathrm{R}}_{\alpha}(p) \;=\; \frac{1}{1-\alpha}\log_2 \frac{\sum_i p_i^{\alpha}}{\sum_i p_i} \tag{9}$$

now known as the Rényi entropy of order α. Since $\lim_{\alpha\to 1} m_2 = m_1$, the Shannon entropy can be recovered as $\lim_{\alpha\to 1} H^{\mathrm{R}}_{\alpha} = H$. Extension of $H^{\mathrm{R}}_{\alpha}(p)$ to continuous-valued bivariate densities $P(x,y)$ is straightforward:

$$H^{\mathrm{R}}_{\alpha}(P) \;=\; \frac{1}{1-\alpha}\log_2 \frac{\iint P^{\alpha}(x,y)\,dx\,dy}{\iint P(x,y)\,dx\,dy}. \tag{10}$$

We emphasize that as the passage from the Shannon entropy to the class of Rényi entropies involves only the relaxation of the mean value property from an arithmetic to an exponential mean, H^{R}_{α} behaves much like H.

2.2 Rényi entropy of a time-frequency representation

The central theme of this paper is the application of entropy measures to TFRs to measure the complexity and information content of nonstationary signals indirectly via the time-frequency plane. Our primary TFR tools of choice lie in Cohen's class [4], which can be expressed as in (3) as the convolution between the Wigner distribution and a kernel Φ. The kernel and its inverse Fourier transform ϕ completely determine the properties of the corresponding TFR. For example, a fixed-kernel TFR possesses the energy preservation property (2) provided $\phi(0,0) = 1$ and the marginal

properties (1) provided $\phi(\theta,0) = \phi(0,\tau) = 1 \ \forall\, \theta,\tau$. We assume without loss of generality that $\phi(0,0) = 1$; our results extend easily to other cases. Besides the Wigner distribution, examples of Cohen's class TFRs include the spectrogram (ϕ = ambiguity function of the time-reversed window function) and the Choi-Williams exponential distribution ($\phi(\theta,\tau) = e^{-\theta^2\tau^2/\beta}$) [11].

The analogy between TFRs and bidimensional probability densities discussed in the Introduction breaks down at at least two key points. First, because of the freedom of choice of kernel function, the TFR of a given signal is nonunique, with many different distributions "explaining" the same data. Second, and more pertinent to the present discussion, most Cohen's class TFRs are nonpositive and, therefore, cannot be interpreted strictly as densities of signal energy.[3] These locally negative values will clearly play havoc with the logarithm in the Shannon entropy (5).

While the Rényi entropies (6) appear intriguing and encouraging for time-frequency application [1], it has remained an open question whether in general these measures can cope with the locally negative values of Cohen's class TFRs. Noninteger orders α yield complex $C_s^\alpha(t,f)$ values and so appear of limited utility. Even integer orders pose no such hazards, since the integral of the positive function $C_s^\alpha(t,f)$ remains positive. Odd integer orders require a more adroit analysis [6,7]:[4]

Theorem 1 (Existence of Rényi entropy for Cohen's class) *Let $C_s(t,f)$ be a Cohen's class TFR having kernel $\Phi \in L^1(\mathbb{R}^2)$.*[5] *Then for all integers $\alpha \geq 1$ and for all $s \in L^2(\mathbb{R})$*

$$\iint C_s^\alpha(t,f)\,dt\,df \;>\; 0.$$

We will assume throughout this paper that $\Phi \in L^1(\mathbb{R}^2)$. While guaranteeing the existence of $C_s(t,f)$, this mild condition merely constrains ϕ to be bounded.

Although it affirms existence, the Theorem offers no guidance on the appropriate choice of entropy order other than integer $\alpha \geq 2$. For the purpose of measuring information content, the first feasible value, $\alpha = 2$, can be ruled out, since for the important Wigner distribution (and indeed for any unitary TFR) we have $\iint W_s^2(t,f)\,dt\,df = \|s\|_2^4$ by Moyal's formula [4], and thus $H_2(W_s) = 0$ for all unit energy signals. The next possible choice, $\alpha = 3$, was observed by Williams, Brown, and Hero to yield a well-defined, useful information measure [1]. We will concentrate mainly on odd α, and particularly on $\alpha = 3$, in the following, but will find even α useful later in Section 4.1.

We close this section with some important notes on normalization. In their experiments, Williams, Brown, and Hero actually employed not H_α^{R} from (10), but a pre-normalized version equivalent to normalizing the signal energy *before* computing the entropy:

$$H_\alpha(C_s) \;=\; \frac{1}{1-\alpha}\log_2 \iint \left(\frac{C_s(t,f)}{\iint C_s(u,v)\,du\,dv}\right)^\alpha dt\,df. \tag{11}$$

[3] While there do exist nonquadratic classes of positive TFRs that satisfy (2) and (1) [4], we will consider only quadratic TFRs in this paper.

[4] All proofs are furnished in the Appendix.

[5] The $L^p(\mathbb{R}^n)$ spaces contain all functions having finite p norm, $\|g\|_p^p \equiv \int |g(\underline{x})|^p d\underline{x}$, $p \geq 1$.

The two measures are related by

$$H_\alpha^{\mathrm{R}}(C_s) \;=\; H_\alpha(C_s) \;-\; \log_2 \|s\|_2^2, \tag{12}$$

and thus $H_\alpha^{\mathrm{R}}(C_s)$ varies with the signal energy. Since an information measure should be invariant to the energy of the signal being analyzed, we will adhere strictly to the definition (11) for the duration of this paper. Discretization of this measure for use with computer generated, discrete TFRs yields

$$H_\alpha(C_s[n,k]) \;=\; \frac{1}{1-\alpha} \log_2 \sum_n \sum_k \left(\frac{C_s[n,k]}{\sum_{n'} \sum_{k'} C_s[n',k']} \right)^\alpha \;+\; \log_2 \delta_t \delta_f,$$

where δ_t and δ_f denote the sample spacings in time and frequency. The frequency spacing constant is computed as $\delta_f = \frac{F}{K}$, given K uniform frequency samples spanning the frequency range of F Hz/sample. For both continuous and discrete TFRs, operation in (t,ω) coordinates, with radial frequency $\omega = 2\pi f$ rad/s, introduces an offset: $H_\alpha[C_s(t,\omega)] = H_\alpha[C_s(t,f)] + \log_2 2\pi$. Sang and Williams explore an alternative magnitude normalization of the Rényi entropy in [12].

3 Properties of the Rényi Time-Frequency Information Measure

We now conduct a detailed analysis of the properties of the Rényi entropy that make it a powerful tool for studying the information content of nonstationary deterministic signals.

3.1 Component counting and cross-component invariance

If TFRs were "quasi-linear" — such that each signal component contributed essentially separately to the overall time-frequency representation with no intervening cross-components — then the analogy between TFRs and probability density functions would predict an additive or counting behavior from the Rényi entropy.

To gain more intuition into this most fundamental property of H_α, imagine applying this measure first to an ideal, quasi-linear TFR $I_s(t,f)$ of an arbitrary compactly supported signal s. Denote the support of s by $\mathrm{supp}(s)$, and form the two-component signal $s + \mathcal{T}s$, where $(\mathcal{T}s)(t) = s(t-\Delta t)$ represents translation by time Δt. Assuming that $\Delta t > \mathrm{supp}(s)$, the distribution is given by

$$I_{s+\mathcal{T}s}(t,f) \;=\; I_s(t,f) \;+\; I_s(t-\Delta t, f). \tag{13}$$

Since $I_s(t,f)$ is compactly supported in the time direction, we can appeal to the analogy between the right hand side of (13) and the composite probability distribution $p \cup q$ in (8) to compute $H_\alpha(I_{s+\mathcal{T}s})$. In particular, substituting (12) into (8) with $m(x) = 2^{(\alpha-1)x}$ and employing the facts $H_\alpha(I_{\mathcal{T}s}) = H_\alpha(I_s)$ and $\|s + \mathcal{T}s\|_2^2 = 2\,\|s\|_2^2$, some simple algebra yields

$$H_\alpha(I_{s+\mathcal{T}s}) \;=\; H_\alpha(I_s) \;+\; 1. \tag{14}$$

In words, the two-component signal $s + \mathcal{T}s$ contains exactly one bit more information than the one-component signal s.[6] The saturation levels of the entropy curve in Figure 1 display precisely this behavior.

While this simple analysis provides considerable insight into the counting behavior of H_α, it does not take into account the nonideal, nonlinear behavior of the quadratic TFRs of Cohen's class. In particular, we have ignored the presence of cross-components in these distributions [4,5], which violate the linearity assumption underlying (13). We will broaden our analysis to encompass actual TFRs in two stages.

First, consider the Wigner distribution (4) of the compactly supported, two-component signal $s + \mathcal{T}s$:

$$W_{s+\mathcal{T}s}(t,f) \;=\; W_s(t,f) \;+\; X_{s,\mathcal{T}s}(t,f) \;+\; W_{\mathcal{T}s}(t,f).$$

The term $X_{s,\mathcal{T}s}(t,f)$, called the *cross-component* between s and $\mathcal{T}s$, is derived from the cross Wigner distribution [4,5]

$$X_{s,\mathcal{T}s}(t,f) \;=\; 2\,\mathrm{Re}\int s\left(t+\frac{\tau}{2}\right)(\mathcal{T}s)^*\left(t-\frac{\tau}{2}\right)e^{-j2\pi\tau f}\,d\tau. \tag{15}$$

In general, the Rényi entropy

$$H_\alpha(W_{s+\mathcal{T}s}) \;=\; \frac{1}{1-\alpha}\log_2\frac{1}{\|s+\mathcal{T}s\|_2^{2\alpha}}\iint\left[W_s(t,f)+X_{s,\mathcal{T}s}(t,f)+W_{\mathcal{T}s}(t,f)\right]^\alpha dt\,df. \tag{16}$$

involves a complicated polynomial in W_s, $X_{s,\mathcal{T}s}$, and $W_{\mathcal{T}s}$. However, due to the compact support of s, for separations $\Delta t > 2\,\mathrm{supp}(s)$, these terms lie disjoint in the time-frequency plane, and a tremendous simplification results:

$$\begin{aligned}
H_\alpha(W_{s+\mathcal{T}s}) \;&=\; \frac{1}{1-\alpha}\log_2\frac{1}{2^\alpha\|s\|_2^{2\alpha}}\iint\left[W_s^\alpha(t,f)+X_{s,\mathcal{T}s}^\alpha(t,f)+W_{\mathcal{T}s}^\alpha(t,f)\right]dt\,df \\
&=\; \frac{1}{1-\alpha}\log_2\frac{1}{2^\alpha\|s\|_2^{2\alpha}}\left[\iint W_s^\alpha(t,f)\,dt\,df \;+\; \iint W_{\mathcal{T}s}^\alpha(t,f)\,dt\,df\right] \\
&=\; H_\alpha(W_s) \;+\; 1,
\end{aligned}$$

provided

$$\iint X_{s,\mathcal{T}s}^\alpha(t,f)\,dt\,df \;=\; 0. \tag{17}$$

While this is obviously not the case for α even, the oscillatory structure of $X_{s,\mathcal{T}s}$ [4,5] cancels under integration with odd powers for Δt sufficiently large. We prove the following in the Appendix as a special case of Theorem 3.

Proposition 2 *For a signal $s \in L^2(\mathbb{R})$ of compact support, fix odd $\alpha \geq 3$ and $\Delta t > \alpha\,\mathrm{supp}(s)$. Then (17) holds and thus $H_\alpha(W_{s+\mathcal{T}s}) = H_\alpha(W_s) + 1$.*

[6] Note that the post-normalized entropy H_α^{R} from (10) exhibits the invariance $H_\alpha^{\mathrm{R}}(I_{s+\mathcal{T}s}) \;=\; H_\alpha^{\mathrm{R}}(I_s)$, since the energies of s and $s+\mathcal{T}s$ differ by a factor of two.

The linear growth of the separation condition $\Delta t > \alpha\,\mathrm{supp}(s)$ recommends the first of the odd integers $\alpha \geq 3$, namely $\alpha = 3$, as the best order for information analysis with the Wigner distribution. Numerical considerations (stability in the face of quantization errors) also justify small α values. Using the rotation-invariant properties of the Wigner distribution [5, 13], Proposition 2 can be easily extended from signals of compact time support to signals whose Wigner distributions are supported on a strip in the time-frequency plane.

By induction, this argument also extends to N components and $\log_2 N$ bits of information gain provided *all* components (auto and cross) become sufficiently disjoint in the time-frequency plane. While it appears relatively mild, this condition is not satisfied by the simple three-component signal $s(t+\Delta t)+s(t)+s(t-\Delta t)$, for example, because the cross-component between $s(t+\Delta t)$ and $s(t-\Delta t)$ will overlap with the auto-component of $s(t)$ for all values of Δt.

We now extend our counting results completely to include all Cohen's class TFRs and arbitrary finite energy signals. For noncompactly supported signals, the auto- and cross-components in the Cohen's class analog to (16) will always overlap to some degree, so we should expect only asymptotic expressions. Define the time-frequency displacement operator

$$(\mathcal{D}s)(t) \;\equiv\; e^{j2\pi t\,\Delta f}\; s(t-\Delta t)$$

that translates signals by the distance $|\mathcal{D}| \equiv \sqrt{(\Delta t)^2+(\Delta f)^2}$ in the time-frequency plane. The following is the key result of this section [6, 7].

Theorem 3 (Component counting) *For any Cohen's class TFR $C_s(t,f)$, any odd $\alpha \geq 3$, and any $s \in L^2(\mathbb{R})$*

$$\lim_{|\mathcal{D}|\to\infty} H_\alpha(C_{s+\mathcal{D}s}) \;=\; H_\alpha(C_s) \;+\; 1.$$

Theorem 3 implies also that the information in the cross-components of $C_{s+\mathcal{D}s}$ must decay to zero asymptotically.

Corollary 4 (Asymptotic cross-component invariance) *For any Cohen's class TFR $C_s(t,f)$, any odd $\alpha \geq 3$, and any $s \in L^2(\mathbb{R})$*

$$\lim_{|\mathcal{D}|\to\infty} \iint \Big[C^\alpha_{s+\mathcal{D}s}(t,f) \;-\; C^\alpha_s(t,f) \;-\; C^\alpha_{\mathcal{D}s}(t,f)\Big]\,dt\,df \;=\; 0.$$

3.2 Amplitude and phase sensitivity

The results of the experiment illustrated in Figure 1 and analyzed in the previous section are very appealing, but are also incomplete, because we introduced no amplitude or phase differences between the two signal components. Amplitude discrepancies alter the asymptotic saturation levels of the Rényi entropy; phase offsets induce strong oscillations between these levels.

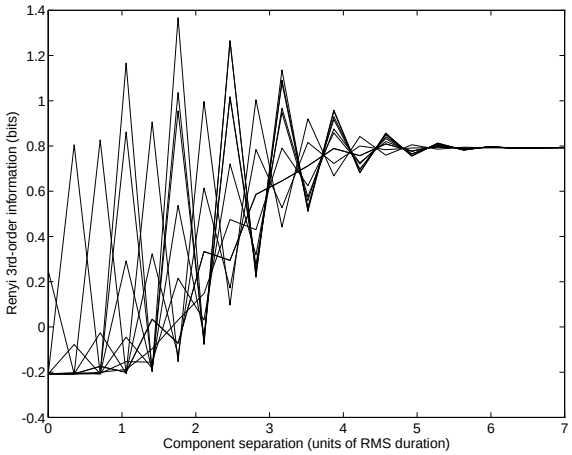


Figure 2: *Third-order Rényi entropy $H_3(W_s)$ of the Wigner distribution of the sum $g(t)\cos(\pi t/6) + g(t + \Delta t)\cos(\pi(t + \Delta t)/6 + \psi)$ of two modulated and phase-shifted Gaussian pulses plotted versus separation distance Δt for several values of phase ψ between 0 and π rad. Comparison with Figure 1 illustrates the sensitivity of $H_3(W_s)$ to relative phase.*

To study the amplitude sensitivity of the Rényi entropy, consider the signal $s + k\mathcal{D}s$, with k a real scaling factor. An analysis similar to that of Section 3.1 yields [14]

$$\lim_{|\mathcal{D}|\to\infty} H_\alpha(C_{s+k\mathcal{D}s}) \;=\; H_\alpha(C_s) \;+\; H_\alpha^{\mathrm{R}}(p_k),$$

with $p_k = \left\{\frac{1}{1+k^2}, \frac{k^2}{1+k^2}\right\}$ and H_α^{R} the discrete Rényi entropy of (9). $H_\alpha^{\mathrm{R}}(p_k)$ is a continuous function of k bounded by 0 and 1 and maximized by $k = 1$. The obvious conclusion that equal amplitudes maximize the complexity of signals composed of multiple identical components appears quite reasonable, for smaller components are dominated by larger ones and therefore carry less information.

In the region between saturation levels (where the TFR auto- and cross-components overlap and the assumptions of Section 3.1 fail to hold), the relative phase between components controls the value of the Rényi entropy. Figure 2 extends the experiment from Figure 1 by illustrating a more complete set of curves of the $H_3(W_s)$ entropy for the modulated and phase-shifted signal $g(t)\cos(\pi t/6) + g(t + \Delta t)\cos(\pi(t + \Delta t)/6 + \psi)$, with g a lowpass Gaussian pulse. Each curve corresponds to a different relative phase angle ψ between 0 and π rad. It is apparent from the curves that while phase changes do not affect the saturation levels of the information measure, they allow many possible trajectories between the two levels, including even trajectories where an "overestimation" (noted in [1]) of information content occurs.

The phase sensitivity of the $H_3(W_s)$ measure for closely spaced components is quite reasonable, given the sensitivity of the signals themselves to relative phase. For example, Figure 3 shows the composite signals and their respective Wigner distributions for a fixed offset Δt and relative phases $\psi = 0$ and $\psi = \frac{\pi}{2}$ rad. The difference in appearance is striking; clearly the components in the signal in (a) are more separated than those in (c). Accordingly, the $H_3(W_s)$ entropies for the two signals differ widely: from 1.31 to 0.31 bits, respectively.

Since the interference pattern generated by the cross-component encodes inter-component phase

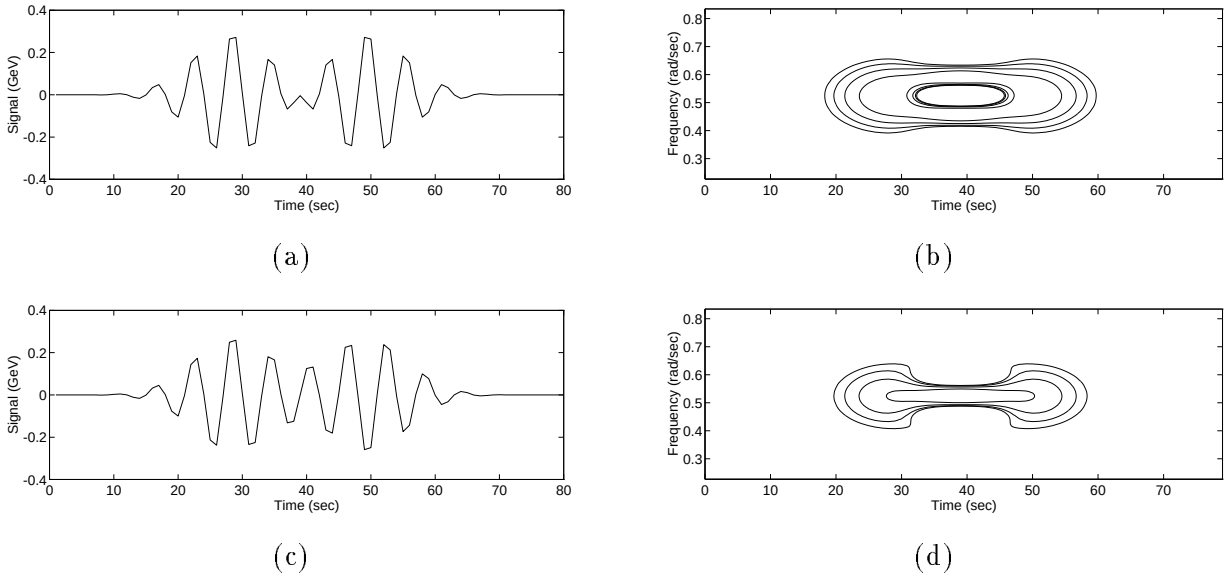


Figure 3: *Signals and positive parts of the Wigner distributions from the experiment illustrated in Figure 2 corresponding to a single fixed Δt and two different phases. (a) Signal and (b) Wigner distribution given $\psi = 0$; $H_3(W_s) = 1.31$ bits. (c) Signal and (d) Wigner distribution given $\psi = \frac{\pi}{2}$; $H_3(W_s) = 0.31$ bits.*

information, signals with low information content ("almost mono-component signals") must exhibit mainly *constructive interference* in the sense of [5]. Relative phase fades from importance after components become disjoint.

We can obtain further insight regarding the region between saturation levels via an analytic expansion for the entropy of the Wigner distribution of a sum of modulated Gaussian pulses of the form $g(t) = (\pi\sigma^2)^{-1/4} \exp(-t^2/2\sigma^2)$. Define the time-frequency-phase displacement operator $(\mathcal{D}'g)(t) \equiv e^{j(2\pi t\,\Delta f + \psi)}\, g(t - \Delta t)$; then the Rényi entropy of the displaced sum $g + \mathcal{D}'g$ can be expanded as [14]

$$H_\alpha(W_{g+\mathcal{D}'g}) \;\approx\; H_\alpha(W_g) \;+\; 1 \;+\; \frac{1}{(1-\alpha)\log_e 2}\left[\, c(\alpha)\,\varepsilon \;-\; \frac{1}{2}\,c^2(\alpha)\,\varepsilon^2 \;+\; \alpha\cos(\gamma)\,\varepsilon^\alpha \;+\; o(\varepsilon^3)\,\right],$$

with normalized time-frequency distance $d^2 = (2\pi\Delta f\,\sigma)^2 \;+\; \Delta t^2/\sigma^2$, absolute phase $\gamma = \psi + 2\pi\Delta f\,(t + \Delta t/2)$, $H = H_\alpha(W_g)$, $c(\alpha) = \frac{1}{\alpha}\,2^{(\alpha-1)H}\binom{\alpha}{\frac{\alpha-1}{2}}\cos(\gamma)$, and $\varepsilon = \exp(-d/4\alpha)$. Valid for $\varepsilon \ll 1$, this expression conforms closely to experimental results and quantifies the effects of separation d, phase ψ, and order α on the Rényi entropy:

1. As $d \to \infty$, the third term decays to zero, and we have the asymptotic, "counting" result of Section 3.1. Clearly, then, the interesting action takes place for small d.

2. For small d, the phase offset ψ, through γ, imposes an oscillatory structure on the entropy value. This behavior is evident in Figure 2.

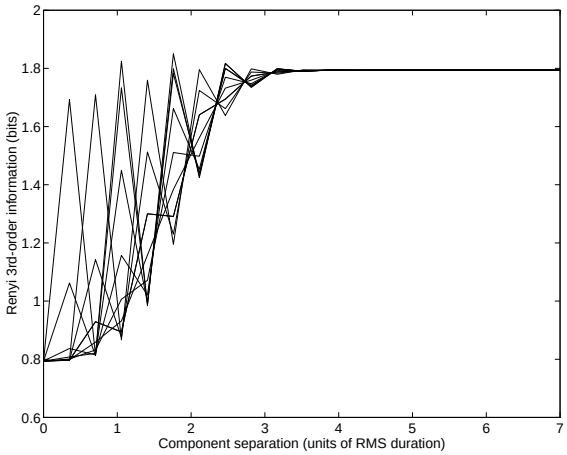


Figure 4: *Third-order Rényi entropy $H_3(C_s)$ of the matched window spectrogram for the same signals utilized in Figure 2. The reduced sensitivity of the spectrogram to relative phase results in swifter saturation with reduced overshoot.*

3. The coefficient ε controlling the amplitude of the third term decays more slowly for larger values of α. As a consequence, larger values of α lead to larger departures from the asymptotic values of H_α. This property should be expected; large values of α emphasize the larger values of the TFR,[7] which for the Wigner distribution occur on the cross-components.

This analysis remains relevant for any signals that can be exponentially bounded in some direction in the time-frequency plane.

3.3 Effects of smoothing

TFRs based on lowpass kernels lead to more robust Rényi information estimates, since smoothing suppresses the Wigner cross-components that carry the inter-component phase information. To illustrate, we repeat in Figure 4 the experiment of Figure 2 using a matched window spectrogram TFR rather than the Wigner distribution. While the spectrogram information estimate remains somewhat phase sensitive, it climbs more swiftly to the saturation level and with a reduced overshoot compared to the Wigner distribution estimate. In general, the ascent to saturation accelerates with increasing order α once the cross-components are smoothed to the same peak level as the auto-components. (The opposite holds for the Wigner distribution, because Wigner cross-components can tower over Wigner auto-components by up to a factor of two).

The price paid for the more robust information estimates derived from smoothed TFRs is a signal-dependent bias of entropy levels compared to those derived from the Wigner distribution, with the amount of bias increasing with the amount of smoothing. This bias has proved difficult to quantify, since the convolution in (3) and the power and integral in (6) do not permute in any simple fashion. In the special case of the matched window spectrogram applied to a sum of Gaussian signal components, a direct computation finds a one-bit bias in asymptotic information compared to that

[7] In fact, we have that $\lim_{\alpha\to\infty} H_\alpha(C_s) = \max_{(t,f)} C_s(t,f)$.

estimated using the Wigner distribution (compare Figure 1 with Figure 4). Despite the introduction of systematic bias, some amount of smoothing appears crucial for obtaining accurate information estimates for complicated multicomponent signals with overlapping auto- and cross-components.

3.4 Bounds on signal information content

Simple to derive, lower and upper bounds on the Rényi entropy correspond to the "peakiest" and "flattest" Cohen's class TFRs.

Theorem 5 (Lower bound on information content for Cohen's class) *For any Cohen's class TFR $C_s(t,f)$, any $\alpha \geq 1$, and any $s \in L^2(\mathbb{R})$*

$$H_\alpha(C_s) \;\geq\; \frac{\log_2 \alpha}{\alpha - 1} \;-\; 1 \;-\; \frac{\alpha}{\alpha - 1} \log_2 \|\Phi\|_1. \tag{18}$$

In particular, for the Wigner distribution,

$$H_\alpha(W_s) \;\geq\; \frac{\log_2 \alpha}{\alpha - 1} \;-\; 1, \tag{19}$$

with equality if and only if s is a Gaussian.

Compare the lower saturation level in Figure 1 with the theoretical bound $H_\alpha(W_g) = \frac{1}{2}\log_2 3 - 1 \approx -0.208$ for the Gaussian.

Theorem 5 can be interpreted as a more powerful version of the well-known time-frequency uncertainty principle [15]; more powerful because it takes the entire time-frequency plane into account rather than just the marginal distributions $|s(t)|^2$ and $|S(f)|^2$. This result also generalizes the inequality of Hirschman that relates the classical principle to the Shannon entropy of the marginals, as [16]

$$H_1(|s|^2) \;+\; H_1(|S|^2) \;\geq\; 1 \;-\; \log_2 e$$

with equality if and only if s is Gaussian. Note that Theorem 5 marks the third breakdown of the analogy between probability density functions and TFRs (the first two being nonpositivity and nonuniqueness), since the delta function probability density minimizes the Rényi entropies of all orders.

While a finite upper bound on signal complexity cannot exist in general (since new components can always be appended to a signal to increase its complexity), an approximate upper bound can be derived for signals constrained to lie in a region $\mathcal{R}$ in the time-frequency plane. In this case, the maximizing signal will be "white" such that its TFR is uniformly distributed over $\mathcal{R}$.

Proposition 6 (Upper bound on information content) *Let $C_s(t,f)$ be a Cohen's class TFR and let $\alpha \geq 1$. Then for signals concentrated in the time-frequency region $\mathcal{R}$*

$$H_\alpha(C_s) \;<\; \frac{1}{1-\alpha} \log_2 \iint_{\mathcal{R}} \mathrm{area}(\mathcal{R})^{-\alpha}\, dt\, df \;=\; \log_2 \mathrm{area}(\mathcal{R}).$$

3.5 Information invariant signal transformations and the affine class

An *information invariant* signal transformation $\mathcal{M}$ leaves the Rényi entropy measure unchanged, with $H_\alpha(C_{\mathcal{M}s}) = H_\alpha(C_s)$ [1]. Distributions information invariant to such a transformation $\mathcal{M}$, provided it displaces the center of gravity of the signal in time-frequency, admit a useful generalization of Theorem 3 to $\lim_{|\mathcal{M}|\to\infty} H_\alpha(C_{s+\mathcal{M}s}) = H_\alpha(C_s) + 1$.

The transformations leaving the Rényi entropy invariant correspond to those that do not change the value of the integral in (11). For Cohen's class TFRs, the invariance properties of three nested kernel classes are simple to quantify. All fixed-kernel TFRs are information invariant to time and frequency shifts. Product-kernel TFRs, with kernels of the form $\Phi(t,f) = \Phi(tf)$, are in addition invariant to scale changes of the form $s(t) \mapsto |a|^{-1/2}\, s(t/a)$. The Wigner distribution is the lone fixed-kernel TFR information invariant to time and frequency shifts, scale changes, and the modulation and convolution by linear chirp functions that realize shears in the time-frequency plane. It is not coincidental that these same five operations leave invariant the form of the (minimum information) Gaussian signal [13].

The *affine class* provides additional TFRs information invariant to time shifts and scale changes [5, 9, 17]. Affine class TFRs are obtained from the affine smoothing

$$\Omega_s(t,f) \;=\; \iint W_s(u,v)\,\Pi\Big(f(t-u),\, v/f\Big)\,du\,dv \;\equiv\; (W_s \mathbin{@} \Pi)(t,f) \tag{20}$$

of the Wigner distribution of the signal with a kernel function Π.[8] Since given proper normalization of the kernel we have $\iint \Omega_s(t,f)\,dt\,df = \|s\|_2^2$, the Rényi entropy of an affine class TFR can be defined exactly as in (11). The resulting time-scale information measure $H_\alpha(\Omega_s)$ shares all of the properties discussed above in the context of Cohen's class, except with time and frequency shifts replaced by time shifts and scale changes. To set this theory on a firm foundation, we need only ensure that $H_\alpha(\Omega_s)$ is well defined.

Theorem 7 (Existence of Rényi entropy for the affine class) *Let $\Omega_s(t,f)$ be an affine class TFR with kernel such that $\Pi \in L^1(\mathbb{R}^2)$, $\frac{1}{f}\Pi(t,f) \in L^1(\mathbb{R}^2)$, and $\iint \Pi(t,f)\,dt\,df > 0$. Then for all integers $\alpha \geq 1$ and for all $s \in L^2(\mathbb{R})$*

$$\iint \Omega_s^\alpha(t,f)\,dt\,df \;>\; 0.$$

The condition $\frac{1}{f}\Pi(t,f) \in L^1$ reduces to $\Pi(t,0) = 0\ \forall t$ for continuous kernels and ensures that the affine smoothing (20) is defined. Since the kernel generating the scalogram (the squared magnitude of the continuous wavelet transform) corresponds to the Wigner distribution W_ψ of the wavelet function ψ, this condition also generalizes the now classical "wavelet admissibility condition" [5]; in particular, we have

$$\iint \left|\frac{1}{f}\,\Pi(t,f)\right| dt\,df \;=\; \iint \left|\frac{1}{f}\,W_\psi(t,f)\right| dt\,df \;\geq\; \iint W_\psi(t,f)\,\frac{dt\,df}{|f|} \;=\; \int |\Psi(f)|^2\,\frac{df}{|f|}.$$

[8] In order to emphasize the similarity of (20) to (3), we have reparameterized the original time-scale formulation of (20) from [9] in terms of time-frequency coordinates by setting scale $a = f_0/f$, with $f_0 = 1$ Hz.

The lower bound on Rényi entropy for affine class TFRs incorporates this prerequisite as well.

Theorem 8 (Lower bound on information content for the affine class) *Let $\Omega_s(t,f)$ be an affine class TFR with kernel such that $\Pi \in L^1(\mathbb{R}^2)$ and $\frac{1}{f}\Pi(t,f) \in L^1(\mathbb{R}^2)$. Then for all integers $\alpha \geq 1$ and for all $s \in L^2(\mathbb{R})$*

$$H_\alpha(\Omega_s) \;\geq\; \frac{\log_2 \alpha}{\alpha - 1} \;-\; 1 \;-\; \frac{\alpha}{\alpha - 1} \log_2 \left\| \frac{1}{f}\, \Pi(t,f) \right\|_1 .$$

For information invariances different from time and frequency shifts, scale changes, and chirp modulations and convolutions, we must look beyond Cohen's class and the affine class. Fortunately, all the above results extend easily to the recently developed unitarily equivalent Cohen's and affine classes [10]. The TFRs in these new classes are information invariant to generalized time-frequency shifts and time-scale changes.

4 Selected Applications

The foregoing properties of the Rényi entropies make these new information and complexity measures particularly appropriate for time-frequency analysis. In this section, we briefly discuss three areas of past and potential application.

4.1 Information-based performance measures

The Rényi entropies make excellent measures of the information extraction performance of TFRs. By the analogy to probability density functions, minimizing the complexity or information in a particular TFR is equivalent to maximizing its concentration, peakiness, and, therefore, resolution [18]. Optimization of a TFR (through its kernel) with respect to an information measure yields a high performance "information optimal" TFR that changes its form to best match the signal at hand [12,19].

Many of the optimal-kernel TFRs in the literature have been based either implicitly or explicitly on information measures. As noted by Williams and Sang [12,19], the performance index common to the 1/0 [20], radially Gaussian [21], and running radially Gaussian kernel [22] optimization formulations can be rewritten using Parseval's theorem as

$$\iint |(W_s * \Phi)(t,f)|^2 \, dt \, df \;=\; 2^{-H_2(W_s * \Phi)}. \tag{21}$$

Since the second-order Rényi entropy squares the TFR, it remains sensitive to cross-components and hence can be considered as a measure of their information content [1]. Thus, maximizing (21) over a class of lowpass smoothing kernels ϕ simultaneously minimizes the information in the cross-components of the optimal-kernel TFR. Maximizing the $\|C_s\|_4^4 \,/\, \|C_s\|_2^4$ concentration ratio of [23,24] can also be viewed in information theoretic terms, since this is equivalent to minimizing the differential entropy $3H_4(C_s) - 2H_2(C_s)$.

Differential performance measures formed with odd and even order entropies also prove interesting [19]. For example, the differential measure

$$H_3(C_s) \; - \; \beta H_2(C_s), \qquad 0 \leq \beta \leq 1 \tag{22}$$

exploits the fact that odd and even order entropies decouple to some degree the information content in the auto- and cross-components in a TFR. Minimizing this measure balances *(i)* maximizing the information in the auto-components (by keeping them peaky through less smoothing) with *(ii)* minimizing the information in the cross-components (by flattening them through more smoothing). For the special choice $\beta = \frac{3}{4}$, minimizing (22) is equivalent to maximizing the concentration ratio $\|C_s\|_3^6 \, / \, \|C_s\|_2^6$, making it an interesting alternative to the $\|C_s\|_4^4 \, / \, \|C_s\|_2^4$ measure used in [23, 24].

Figure 5 explores the effect of time-frequency smoothing on (22) as a function of the parameter β. Forming a signal s from two well-separated Gaussian pulses, we smooth the Wigner distribution of s with a Gaussian kernel of increasing volume to generate a series of smoothed TFRs C_s. In the Figure, the smoothing parameter ρ corresponds to the normalized degree of smoothing, with $\rho = 0$ leaving the Wigner distribution untouched and $\rho = 1$ generating the matched window spectrogram. We plot $H_3(C_s) - \beta H_2(C_s)$ versus the smoothing parameter for several values of β ranging between 0 and 1. From the Figure, it is clear that β controls the tradeoff between measuring auto-component concentration and measuring cross-component suppression: Small β favors auto-component concentration, and (22) is minimized by very little smoothing — no smoothing at all for the extreme $\beta = 0$ case. On the other hand, large β favors cross-component suppression, and (22) is minimized only after considerable smoothing. In fact, we see (but have not proved) that as $\beta \to 1$, C_s at the minimum tends to the spectrogram. More complicated signals will exhibit local minima.

4.2 Rényi dimensions

Based on the counting property of the Rényi entropy (Section 3.1), we can define a *Rényi dimension* $D_\alpha(C_s)$ of a signal s in terms of its TFR C_s and a basic building block function b [6, 7]

$$D_\alpha(C_s) \; = \; 2^{H_\alpha(C_s) \, - \, H_\alpha(C_b)}.$$

This dimension attempts to indicate — relative to a highly redundant set of building blocks obtained from b by all possible translations and modulations — the number of blocks required to "cover" the TFR of s. For the Wigner TFR, a Gaussian is the natural choice for the building block function, since it has minimum intrinsic information and leads to an always positive dimension. A similar dimension can be defined for affine class TFRs.

By permitting redundant time-frequency building blocks, the Rényi time-frequency dimension generalizes the concepts of the number of "independent degrees of freedom" and number of "independent coherent states" that have proved useful in signal analysis and quantum physics [25, p. 23]. Desirable invariance properties result from this redundancy: Cohen's class Rényi dimension estimates remain invariant under time and frequency shifts in the signal, while affine class estimates

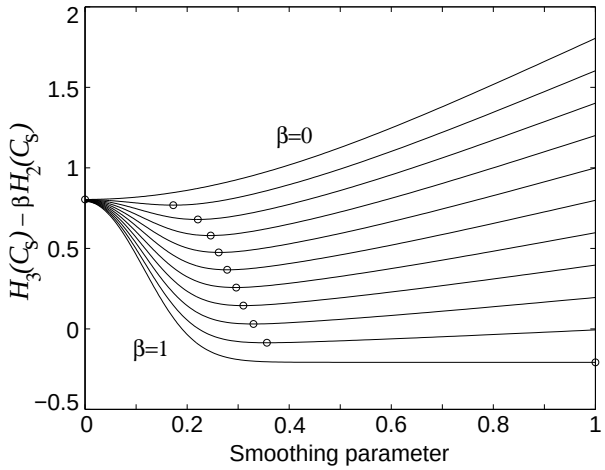


Figure 5: *The effect of time-frequency smoothing on (22) as a function of the parameter β. For a signal consisting of two well separated Gaussian pulses, we form the TFR $C_s = W_s * \Phi_\rho$, where $\Phi_\rho(t,f) = \frac{2}{\rho^2}\, e^{-2\pi(\theta^2+\tau^2)/\rho^2}$. The parameter ρ controls the degree of time-frequency smoothing: $\rho = 0$ generates the Wigner distribution; $\rho = 1$ generates the matched window spectrogram. We plot $H_3(C_s) - \beta H_2(C_s)$ versus the smoothing parameter ρ for the eleven values $\beta = 0, 0.1, 0.2, \ldots, 1$. (The upper curve corresponds to $\beta = 0$, while the lower curve corresponds to $\beta = 1$.) The minumum point of each curve, marked by a circle, corresponds to an "information optimal TFR."*

remain invariant under time and scale changes. Alternative dimensions that measure signal complexity with respect to an orthonormal basis of (wavelet or Gabor) functions (see [26], for example) cannot share these invariances without carrying out an optimization over all "nice" bases [2].

For the simplest signals, composed of disjoint, equal-amplitude copies of one basic function, the Rényi dimension simply counts the number of components. As the relative amplitudes of these components change, however, the dimension estimate will also change, as some components begin to dominate others.

Finally, we note that there exists a tantalizing connection between the Rényi entropy and the generalized *fractal dimension* [27]

$$d_\alpha \;=\; \lim \frac{\frac{1}{\alpha-1} \log \sum_i p_i^\alpha(a)}{\log a}. \tag{23}$$

Here, $p_i(a)$ denotes the area of the scale-invariant, fractal object inside a box of area a placed at position i, and the limit is taken either as $a \to 0$ or as $a \to \infty$. Clearly, for a fixed box size a, d_α corresponds to the Rényi entropy of the distribution $p_i(a)$. This inspires a second, *multiresolution Rényi dimension* based on the scale-invariant TFRs of the affine class

$$d_\alpha(\Omega_s) \;=\; \lim \frac{H_\alpha(\Omega_s(a))}{\log_2 a}. \tag{24}$$

Here, $\Omega_s(a)$ is an affine class TFR (20) obtained from the Wigner distribution via an affine smoothing of degree a, where small/large a corresponds to small/large amounts of smoothing. We can

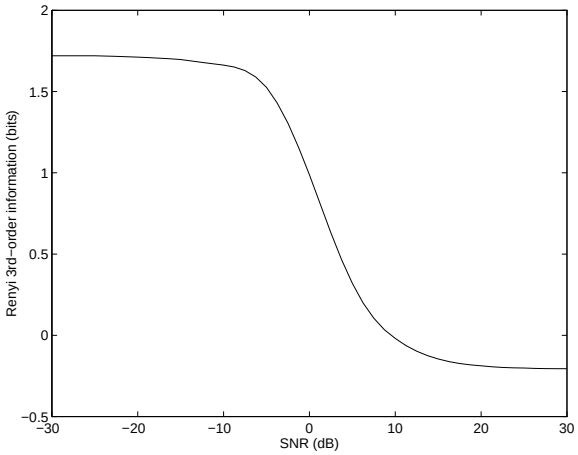


Figure 6: *Third-order Rényi entropy $H_3(W_{g+n})$ of the Wigner distribution of a Gaussian pulse g immersed in white Gaussian noise n plotted versus signal-to-noise ratio (SNR). To obtain the curve, we averaged the $H_3(W_{g+n})$ values of 200 experiments (post-averaging). Averaging the W_{g+n} before computing H_3 (pre-averaging) would cause the upper saturation level to approach the information upper bound of Theorem 6.*

draw a parallel between (23) and (24) as $a \to \infty$: Just as the increasing box size in (23) "zooms out" to view the global structure of a fractal with ever decreasing resolution, the increasing degree of smoothing in (24) "zooms out" to view the global structure of a TFR with ever decreasing resolution.

4.3 Random signals

While up to now we have considered only deterministic signals, time-frequency information estimates are also very useful for random signals. However, care must be taken not to confuse the Rényi entropy of the *time-frequency* distribution of a random signal with the Rényi entropy of the *probability* distribution of the signal. We illustrate with a simple example.

The $H_3(W_s)$ time-frequency information estimate provides an intriguing alternative to the signal-to-noise ratio (SNR) for signals embedded in additive noise [6,7]. For example, Figure 6 illustrates the relationship between the two for a single Gaussian pulse in additive white Gaussian noise. Interestingly, the sigmoidal characteristic of the information measure behaves more like our eyes and ears than the SNR: For high SNRs, it indicates that there is virtually only signal present, whereas for low (negative) SNRs, it indicates that there is virtually only noise present. Furthermore, the 0 dB SNR point (the point of equal signal and noise energies) occurs roughly midway between the two information extremes.

5 Conclusions

Taking off where Williams, Brown, and Hero left off in [1], this paper has studied a most interesting class of new signal analysis tools — the Rényi entropies. The key properties of these functions,

namely their existence, accounting, and cross-component and transformation invariances, make them the measures of choice for estimating the complexity of deterministic signals through TFRs. While simple to apply, the Rényi entropies provide often deep insights into the structure of the time-frequency plane. For instance, a simple lower bound on the entropy of the Wigner distribution yields a new time-frequency uncertainty principle (Theorem 5) based on the entire time-frequency plane as a whole rather than on the time and frequency domains separately.

The explorations of Section 4 into TFR performance measures, Rényi dimensions, and random signals merely scratch the surface of potential applications of the Rényi entropies in time-frequency analysis. Worthy of pursuit seems the extension of our results to TFRs outside the quadratic Cohen's and affine classes. The positive TFRs of the Cohen-Posch class [4], for example, would allow the unrestricted use of the Shannon entropy. Moreover, an axiomatic derivation of the "ideal" time-frequency complexity measure along the lines of Rényi's work in probability theory [8] could yield other entropies meriting investigation.

In information theory, entropies form the basis for distance and divergence measures between probability densities. In time-frequency analysis, analogous measures between TFRs would find immediate application in detection and classification problems. Unfortunately, the Rényi entropy complicates the formation of distances, because it is neither a concave nor a convex function for $\alpha \neq 1$. Although the bulk of the work lies ahead, some progress has been made in this direction recently [14]. By considering only positive TFRs (smoothed spectrograms in Cohen's class), we defined in [14] a quasi-Jensen difference

$$J_\alpha(C_1, C_2) \;=\; H_\alpha\left(\sqrt{C_1\,C_2}\right) \;-\; \frac{H_\alpha(C_1) \;+\; H_\alpha(C_2)}{2},$$

that measures the distance between the TFRs C_1 and C_2 of two different signals s_1 and s_2. (Here, $\sqrt{C_1\,C_2}(t,f) \equiv \sqrt{C_1(t,f)\,C_2(t,f)}$.) Currently, we are evaluating the potential of this measure for problems in nonparametric and blind transient detection.

In short, the surprisingly good fit of the Rényi entropies with Cohen's class and affine class TFRs make them powerful tools for time-frequency analysis theory and applications.

Appendix: Proofs

Preliminaries: The *ambiguity function* of a signal s is defined as [4,5]

$$A_s(\theta,\tau) \;=\; \int s\left(u+\frac{\tau}{2}\right) s^*\left(u-\frac{\tau}{2}\right) e^{j2\pi\theta t}\,du$$

and corresponds to the two-dimensional Fourier transform of the Wigner distribution (4).

In the proofs that follow, we use extensively the following three lemmata. The first two are classical, the last one relatively recent [28]. All three can be found in [28], where the first is sharpened even further. We will omit R and R^2 in the following except where the domain is not obvious from context. (Lemmata 1 and 2 apply to both one- and two-dimensional functions.)

Lemma 1 (Young) *Let $1/p+1/q = 1/r+1$ with $1 \le p,q,r \le \infty$. Then, when $g \in L^p$ and $h \in L^q$, $g*h \in L^r$ and $\|g*h\|_r \;\le\; \|g\|_p\,\|h\|_q$. In particular, if $h \in L^1$, then $\|g*h\|_p \;\le\; \|g\|_p\,\|h\|_1$.*

Lemma 2 (Hausdorff-Young) *Let $1/q + 1/p = 1$ with $2 \le p \le \infty$ (and thus $1 \le q \le 2$). If $g \in L^q$, then its Fourier transform $G \in L^p$, with $\|G\|_p \le \|g\|_q$.*

Lemma 3 (Lieb) *Let $p \ge 2$ and assume that $s \in L^2$. Then $\|W_s\|_p^p \;\le\; (2^{p-1}/p)\,\|s\|_2^{2p}$ with equality if and only if s is Gaussian. A similar bound holds for the ambiguity function A_s (with constant $2/p$).*

A result similar to Lemma 3 holds for the cross Wigner distribution [28]. In short, if $s \in L^2(\mathsf{R})$, then W_s, $X_{s,\mathcal{D}s}$, and A_s inhabit *every* space $L^p(\mathsf{R}^2)$, from $2 \le p \le \infty$.

In addition, we will assume that these functions lie in $L^1(\mathsf{R}^2)$ as well. While there exist signals $s \in L^2(\mathsf{R})$ such that this is not the case (a rectangular pulse, for example), such signals can be uniformly approximated by continuous functions whose Wigner distributions do lie in $L^1(\mathsf{R}^2)$ [29]. Thus, many of our results hold rigorously only in the limit of this approximation, which we will not perform explicitly.

Proof of Theorem 1: First, note that since $W_s \in L^\alpha$ and $\Phi \in L^1$, Lemma 1 yields $C_s \in L^\alpha$, and hence the integral makes sense.

For $\alpha = 1$ or an even integer, the result is obvious. Assuming now that $\alpha > 1$ is an odd integer, we have

$$\begin{aligned} 0 \;&\le\; \iint C_s^{\alpha+1}(t,f)\,dt\,df \;=\; \iint (W_s*\Phi)(t,f)\,C_s^\alpha(t,f)\,dt\,df \\ &=\; \iint W_s(t,f)\,(C_s^\alpha*\Phi)(t,f)\,dt\,df, \end{aligned}$$

and so we can interpret the function $C_s^\alpha * \Phi$ as a *Wigner weight symbol* [29]. As shown by Janssen in [29], any Wigner weight symbol $K \in L^1 \cap L^2$ admits a decomposition of the form $K(t,f) =$

$\sum_{k=1}^{\infty} c_k\, W_{b_k}(t,f)$, with $\{b_k(t)\}$ an orthonormal basis for $L^2(\mathbb{R})$ and each $c_k \geq 0$. It follows that $\iint K(t,f)\,dt\,df = \sum_{k=1}^{\infty} c_k > 0$, from which we find

$$0 \;<\; \iint (C_s^\alpha * \Phi)(t,f)\,dt\,df \;=\; \phi(0,0) \iint C_s^\alpha(t,f)\,dt\,df.$$

Since we have assumed $\phi(0,0) > 0$, the result follows provided $C_s^\alpha * \Phi \in L^1 \cap L^2$. Applying Lemma 1 again, we obtain $\|C_s^\alpha * \Phi\|_1 \leq \|C_s^\alpha\|_1 \|\Phi\|_1 = \|C_s\|_\alpha^\alpha \|\Phi\|_1 < \infty$ and $\|C_s^\alpha * \Phi\|_2 \leq \|C_s^\alpha\|_2 \|\Phi\|_1 = \|C_s\|_{2\alpha}^{2\alpha} \|\Phi\|_1 < \infty$, and thus the result. The less restrictive assumption $\phi(0,0) \geq 0$ yields a similar inequality in the Theorem. □

Proof of Theorem 3: To keep the notation simple, we first prove the result for the Wigner distribution. Let p denote any integer $1 \leq p \leq \infty$ and let c denote a generic bounded constant.

Setting $C_s = W_s$ for now, the integrand of (16) can be expanded as

$$W_{s+\mathcal{D}s}^\alpha(t,f) \;=\; W_s^\alpha(t,f) + W_{\mathcal{D}s}^\alpha(t,f) + X_{s,\mathcal{D}s}^\alpha(t,f) + \sum_{\alpha_i} c_{\alpha_1,\alpha_2,\alpha_3} W_s^{\alpha_1}(t,f)\, X_{s,\mathcal{D}s}^{\alpha_2}\, W_{\mathcal{D}s}^{\alpha_3}(t,f), \quad (25)$$

where the sum ranges over all positive integer α_i such that $\alpha_1 + \alpha_2 + \alpha_3 = \alpha$. The first two terms yield the desired result, using the same simple arguments employed at the beginning of Section 3.1. We will now show that, in the limit as $|\mathcal{D}| \to \infty$, the third and fourth terms integrate to zero. Before we proceed, note that the cross-component $X_{s,\mathcal{D}s}$ can be written as [5, p. 240]

$$X_{s,\mathcal{D}s}(t,f) \;=\; 2\, W_s(t - \Delta t/2, f - \Delta f/2) \cos[2\pi(t\Delta f - f\Delta t + \Delta t\Delta f/2)]. \quad (26)$$

Fourth term of (25): Let I_0 denote the integral of a generic term in the sum, and assume for now that $\alpha_1, \alpha_2 > 0$. Using the fact that $W_{\mathcal{D}s} \in L^\infty$ to pull it out of the integral, we have

$$|I_0| \;\leq\; I_1 \;=\; |c_{\alpha_1,\alpha_2,\alpha_3}|\, \|W_{\mathcal{D}s}\|_\infty^{\alpha_3} \iint |W_s(t,f)|^{\alpha_1}\, |X_{s,\mathcal{D}s}(t,f)|^{\alpha_2}\,dt\,df. \quad (27)$$

Now, using (26) plus the fact that $|\cos x| \leq 1$, we have the convolution

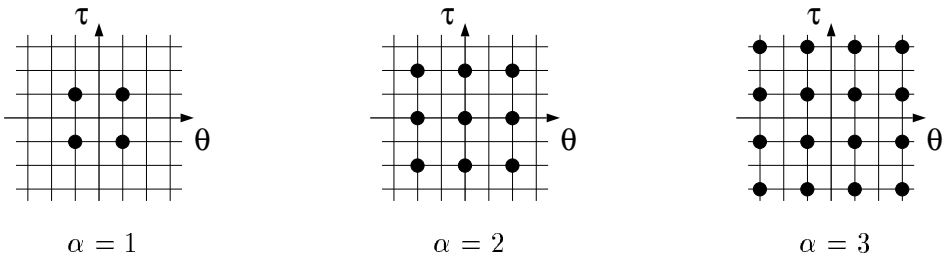
$$I_1 \;\leq\; I_2 \;=\; c \iint |W_s(t,f)|^{\alpha_1}\, |W_s(t - \Delta t/2,\, f - \Delta f/2)|^{\alpha_2}\,dt\,df.$$

Since $W_s \in L^{\alpha_1} \cap L^{\alpha_2}$, by Lemma 1, $I_2(\Delta t, \Delta f) \in L^1$. Therefore, using the fact that L^1 functions must decay at infinity, we have that $\lim |I_0| = \lim I_2 = 0$ as Δt, Δf, or both diverge. Proof for the other cases ($\alpha_1 = 0$, $\alpha_2 = 0$) proceeds in exactly the same fashion, with the obvious modifications.

Third term of (25): Denote by J_0 the integral of the cross-component $X_{s,\mathcal{D}s}^\alpha$. Substituting (26) yields

$$J_0 \;=\; 2^\alpha \iint W_s^\alpha(t - \Delta t/2, f - \Delta f/2)\, \cos^\alpha[2\pi(t\Delta f - f\Delta t + \Delta t\Delta f/2)]\,dt\,df. \quad (28)$$

Since $W_s^\alpha \in L^1 \cap L^2$, the Fourier transform of the integrand exists, and formally J_0 corresponds to the convolution of the Fourier transforms of the W_s^α and $\cos^\alpha$ terms evaluated at the origin. Denote these transforms by $B_\alpha(\theta,\tau)$ and $Z_\alpha(\theta,\tau)$, respectively, so that $J_0 = (B_\alpha * Z_\alpha)(0,0)$.



$\alpha = 1$ $\alpha = 2$ $\alpha = 3$

Figure 7: *The impulse array $Z_\alpha(\theta, \tau)$ for $\alpha = 1, 2, 3$. Round blobs denote impulses; the grid is spaced in units of Δt in the τ direction and in units of Δf in the θ direction. Z_2 results from convolving Z_1 with itself. Z_3 results from convolving Z_2 by Z_1.*

The term B_α corresponds to the α-fold convolution of a phase-shifted version of the ambiguity function A_s of the signal. Note that since A_s is symmetrical in the (θ, τ) plane and reaches its maximum at the origin, so does B_α. Furthermore, since $A_s \in L^p$, by Lemma 1, $B_\alpha \in L^p$ as well.

The term Z_α corresponds to the α-fold convolution of the Fourier transform Z_1 of the cosine term in (26). This transform is easily worked out formally to be four impulses at the corners of a rectangle:

$$Z_1(\theta, \tau) \;=\; k\,[\delta(\theta - \Delta f, \tau - \Delta t) \;+\; \delta(\theta - \Delta f, \tau + \Delta t) \;+\; \delta(\theta + \Delta f, \tau - \Delta t) \;+\; \delta(\theta + \Delta f, \tau + \Delta t)],$$

where k is a complex constant. Since the array of impulses in Z_1 is square, the convolution $Z_\alpha = Z_1 * \ldots * Z_1 = Z_{\alpha-1} * Z_1$ also results in square arrays. In particular, Z_α consists of an impulse at each point $(m\Delta t,\, n\Delta f)$, where $m, n = 0, \pm 2, \ldots, \pm\alpha$ for even α and $m, n = \pm 1, \pm 3, \ldots, \pm\alpha$ for odd α. The maximum weighting on any impulse in Z_α is $\alpha^2\, k$. Note that for even α, Z_α has an impulse at the origin, while for odd α, Z_α has no impulse at the origin. See Figure 7.

Computing $(B_\alpha * Z_\alpha)(0, 0)$ is equivalent to summing the B_α values lying at the impulse locations of the array. Thus, for odd α, we can bound J_0 by

$$|J_0| \;\leq\; \alpha^2\, |k| \sum_{m,n = \pm 1, \pm 3, \ldots, \pm\alpha} |B(m\Delta t,\, n\Delta f)|.$$

Since $B_\alpha \in L^1$, each term in this sum falls to zero with $|\mathcal{D}| \to \infty$. Thus the entire sum $|J_0|$ does likewise, completing the proof of both Theorem 3 and Corollary 4 in the case of the Wigner distribution.

For the proof of Proposition 2: Assuming s is compactly supported, as soon as $T = \Delta t > \alpha\, \mathrm{supp}(s)$, the third and fourth terms of (25) immediately become zero.

Extension to Cohen's class: Simply make the substitutions $C_s = W_s * \Phi$ for W_s and $Y_{s,\mathcal{D}s} = X_{s,\mathcal{D}s} * \Phi$ for $X_{s,\mathcal{D}s}$ in the above proof. Lemma 1 coupled with the hypothesis $\Phi \in L^1$ yields $C_s,\, Y_{s,\mathcal{D}s} \in L^p$, and thus, the proof for the term corresponding to I_0 in (27) proceeds exactly as above. The term corresponding to J_0 in (28) equals the value at the origin of the convolution

$$[(A_s * Z_1)\phi] \;*\; \ldots \;*\; [(A_s * Z_1)\phi] \;=\; [(A_s\phi) * \ldots * (A_s\phi)] * [Z_1 * \ldots * Z_1],$$

so we redefine B_α as the α-fold convolution of the weighted ambiguity function $A_s\phi$. Since the hypothesis $\Phi \in L^1$ implies that $\phi \in L^\infty$ by Lemma 2, we have $A_s\phi \in L^p$. Thus this convolution is defined, with the result $B_\alpha \in L^1$. The remainder of the proof remains unchanged. □

From the above proof we can gain valuable intuition on how to choose α for maximum cross-component invariance. For the Wigner distribution, smaller α appear preferable: As α increases, the support of the function B_α grows in the (θ, τ) plane, and thus the integral of the third term in (25) falls off more slowly. This broadening can be counteracted by choosing a different TFR having a rapidly decreasing kernel $\phi(\theta, \tau)$. In fact, once ϕ supplies sufficient smoothing to make the auto- and cross-components of equal height over the time-frequency plane, we reach a threshold point, after which it is actually advantageous to choose large α, to emphasize the auto-components over the cross-components. This effect explains why the third-order Rényi entropy of the spectrogram (with ϕ = the ambiguity function of the analysis window) saturates more swiftly in Figure 4 than that of the Wigner distribution in Figure 1.

Finally, we note that the exact same means can be applied without modification to demonstrate the cross-component invariance of the Rényi entropy for more general signals than $s + \mathcal{D}s$. In particular, for $\alpha \geq 3$ we have

$$\lim_{|\mathcal{D}|\to\infty} \iint W^\alpha_{s_1+\mathcal{D}s_2}(t,f)\,dt\,df \;=\; \iint W^\alpha_{s_1}(t,f)\,dt\,df \;+\; \iint W^\alpha_{s_2}(t,f)\,dt\,df,$$

where $s_1, s_2 \in L^2(\mathsf{IR})$ are any two signals.

Proof of Theorem 5: Using first Lemma 1 and then Lemma 3, we have for unit-energy s

$$\iint C^\alpha_s(t,f)\,dt\,df \;\leq\; \|C_s\|^\alpha_\alpha \;\leq\; \|W_s\|^\alpha_\alpha\,\|\Phi\|^\alpha_1 \;\leq\; \frac{2^{\alpha-1}}{\alpha}\,\|\Phi\|^\alpha_1.$$

Thus,

$$H_\alpha(C_s) \;\geq\; \frac{1}{1-\alpha}\log_2\frac{2^{\alpha-1}}{\alpha}\,\|\Phi\|^\alpha_1,$$

and (18) follows. The bound (19) for the Wigner distribution follows from the fact that $\|\Phi_{\mathrm{WD}}\|_1 = 1$. While Gaussian signals reach the bound (19) for the Wigner distribution (because of the second assertion of Lemma 3), the more general bound (18) may be unattainable for other Cohen's class TFRs. □

Proof of Theorem 7: Since the classical Young's theorem (Lemma 1) does not apply to the affine smoothing of (20), we begin by stating an analogue matched to the affine convolution

$$(g\#h)(c,d) \;\equiv\; \iint g(a,b)\,h\Big(b\,(c-a)\,,\,\frac{d}{b}\Big)\,da\,db \tag{29}$$

defined on the affine group. The following was obtained by specializing the general results of [30, p. 293–8] to the scalar affine group having group operation $(a,b)\circ(c,d) = (a + c/b, bd)$, $b, d > 0$, and left Haar measure $da\,db$. All integrals and norms in the following can be interpreted to run over the upper half-plane $\mathsf{IR}\times\mathsf{IR}_+$ to account for $b > 0$ in (a,b).

Lemma 4 *Let $1/p+1/q = 1/r+1$ with $1 \le p,q,r \le \infty$. Then, when $g \in L^p$ and $h \in L^q$, $g\#h \in L^r$ and $\|g\#h\|_r \;\le\; \|g\|_p \, \|h\|_q$. In particular, if $h \in L^1$, then $\|g\#h\|_p \;\le\; \|g\|_p \, \|h\|_1$.*

While the affine smoothing (20) is not a group convolution proper, the condition for existence and integrability of an affine class TFR follows immediately from this Lemma. Substituting $\lambda(tf, 1/f) \equiv \Pi(t,f)$ into (20) immediately yields the form (29) and the conclusion that $\Omega_s \in L^p$, $1 \le p \le \infty$, provided $W_s \in L^p$ and $\lambda \in L^1$. A change of variable converts the constraint on λ into a constraint on the original kernel Π

$$\lambda(t,f) \in L^1 \quad \Leftrightarrow \quad \frac{1}{f}\,\Pi(t,f) \in L^1.$$

Now onward with the proof. For $\alpha = 1$ and even integers, the result follows immediately. For odd integer $\alpha > 1$, we follow the proof of Theorem 1 and write

$$\begin{aligned} 0 \;\le\; \iint \Omega_s^{\alpha+1}(t,f)\,dt\,df \;&=\; \iint (W_s \mathbin{@} \Pi)(t,f)\,\Omega_s^{\alpha}(t,f)\,dt\,df \\ &=\; \iint W_s(u,v) \left\{ \iint \Omega_s^{\alpha}(t,f)\,\Pi\Big(f(u-t),\,\frac{v}{f}\Big)\,dt\,df \right\} du\,dv. \end{aligned}$$

Thus, the term in braces (call it $K(u,v)$) is a Wigner weight symbol provided $K \in L^1 \cap L^2$. This condition is easily checked from Lemma 4: The expression for K coincides with (29), $\Omega_s^{\alpha} \in L^1 \cap L^2$ as shown above, and $\Pi \in L^1$ by hypothesis. Wrapping up, since a Wigner weight symbol has a positive integral, we have

$$0 \;<\; \iint \Omega_s^{\alpha}(t,f) \left\{ \iint \Pi\Big(f(u-t),\,\frac{v}{f}\Big)\,du\,dv \right\} dt\,df \;=\; \iint \Omega_s^{\alpha}(t,f)\,dt\,df \left\{ \iint \Pi(u,v)\,du\,dv \right\}$$

and the result. □

Proof of Theorem 8: Using first Lemma 4 and then Lemma 3, we have for unit-energy s

$$\iint \Omega_s^{\alpha}(t,f)\,dt\,df \;\le\; \|\Omega_s\|_{\alpha}^{\alpha} \;\le\; \|W_s\|_{\alpha}^{\alpha} \left\| \frac{1}{f}\,\Pi(t,f) \right\|_1^{\alpha} \;\le\; \frac{2^{\alpha-1}}{\alpha} \left\| \frac{1}{f}\,\Pi(t,f) \right\|_1^{\alpha}.$$

Taking logartithms yields the result. □

Acknowledgements

The authors wish to thank Michèle Basseville, Paulo Gonçalvès, Alfred Hero III, Douglas Jones, Richard Orr, and William Williams for stimulating discussions at various stages of this work.

References

[1] W. J. Williams, M. L. Brown, and A. O. Hero, "Uncertainty, information, and time-frequency distributions," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1566, pp. 144–156, 1991.

[2] R. Orr, "Dimensionality of signal sets," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1565, pp. 435–446, 1991.

[3] L. Cohen, "What is a multicomponent signal?," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '92*, vol. V, pp. 113–116, 1992.

[4] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[5] P. Flandrin, *Temps-Fréquence*. Paris: Hermes, 1993.

[6] R. G. Baraniuk, P. Flandrin, and O. Michel, "Information and complexity on the time-frequency plane," in *14ème Colloque GRETSI*, (Juan Les Pins, France), pp. 359–362, 1993.

[7] P. Flandrin, R. G. Baraniuk, and O. Michel, "Time-Frequency complexity and information," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '94*, vol. III, pp. 329–332, 1994.

[8] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. and Prob.*, vol. 1, pp. 547–561, 1961.

[9] O. Rioul and P. Flandrin, "Time-scale energy distributions: A general class extending wavelet transforms," *IEEE Trans. Signal Processing*, vol. 40, pp. 1746–1757, July 1992.

[10] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Processing*, vol. 43, Oct. 1995.

[11] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 862–871, June 1989.

[12] T.-H. Sang and W. J. Williams, "Rényi information and signal-dependent optimal kernel design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '95*, vol. 2, pp. 997–1000, 1995.

[13] G. B. Folland, *Harmonic Analysis in Phase Space*. Princeton, NJ: Princeton University Press, 1989.

[14] O. Michel, R. G. Baraniuk, and P. Flandrin, "Time-frequency based distance and divergence measures," *IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 64–67, Oct. 1994.

[15] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pp. 429–457, 1946.

[16] W. Beckner, "Inequalities in Fourier analysis," *Annals Math.*, vol. 102, pp. 159–182, 1975.

[17] J. Bertrand and P. Bertrand, "A class of affine Wigner functions with extended covariance properties," *J. Math. Phys.*, vol. 33, pp. 2515–2527, July 1992.

[18] D. L. Jones and T. W. Parks, "A resolution comparison of several time-frequency representations," *IEEE Trans. Signal Processing*, vol. 40, pp. 413–420, Feb. 1992.

[19] W. J. Williams and T. H. Sang, "Adaptive RID kernels which minimize time-frequency uncertainty," *IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 96–99, Oct. 1994.

[20] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: Optimal kernel design," *IEEE Trans. Signal Processing*, vol. 41, pp. 1589–1602, Apr. 1993.

[21] R. G. Baraniuk and D. L. Jones, "A radially Gaussian, signal-dependent time-frequency representation," *Signal Processing*, vol. 32, pp. 263–284, June 1993.

[22] D. L. Jones and R. G. Baraniuk, "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Signal Processing*, vol. 43, Oct. 1995.

[23] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2127–2135, Dec. 1990.

[24] D. L. Jones and R. G. Baraniuk, "A simple scheme for adapting time-frequency representations," *IEEE Trans. Signal Processing*, vol. 42, pp. 3530–3535, Dec. 1994.

[25] I. Daubechies, *Ten Lectures on Wavelets.* New York: SIAM, 1992.

[26] R. Coifman and V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. IT–38, pp. 713–718, Mar. 1992.

[27] A. B. Chhabra, C. Meneveau, R. V. Jensen, and K. R. Sreenivasan, "Direct determiniation of the $f(\alpha)$ singularity spectrum and its application to fully developed turbulence," *Phys. Rev. A*, vol. 44, pp. 5284–5294, Nov. 1989.

[28] E. Lieb, "Integral bounds for radar ambiguity functions and the Wigner distribution," *J. Math. Phys.*, vol. 31, pp. 594–599, Mar. 1990.

[29] A. J. E. M. Janssen, "Positivity properties of phase-plane distribution functions," *J. Math. Phys.*, vol. 25, pp. 2240–2252, July 1984.

[30] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis*, vol. I. New York: Academic Press, 1963.