INFORMATION BOUNDS FOR RANDOM SIGNALS IN TIME-FREQUENCY PLANE

Selin Aviyente and William J. Williams

Department of Electrical Engineering and Computer Science,
The University of Michigan, Ann Arbor, MI 48109-2122
e-mail:saviyent@eecs.umich.edu

ABSTRACT

Rényi entropy has been proposed as one of the methods for measuring signal information content and complexity on the time-frequency plane by several authors [1, 2]. It provides a quantitative measure for the uncertainty of the signal. All of the previous work concerning Rényi entropy in the time-frequency plane has focused on determining the number of signal components in a given deterministic signal. In this paper, we are going to discuss the behaviour of Rényi entropy when the signal is random, more specifically white complex Gaussian noise. We are going to present the bounds on the expected value of Rényi entropy and discuss ways to minimize the uncertainty by deriving conditions on the time-frequency kernel. The performance of minimum entropy kernels in determining the number of signal elements will be demonstrated. Finally, some possible applications of Rényi entropy for signal detection will be discussed.

1. INTRODUCTION

The main tool in measuring the information content or the uncertainty of a given probability distribution is the entropy function. Williams et al. have extended measures of information from probability theory to the time-frequency plane by treating the time-frequency distributions (TFDs) as density functions and have introduced Rényi entropy as an appropriate information measure [1]. The basic signal component with the lowest entropy in the time-frequency plane is known to be the gabor logon, Gaussian enveloped complex exponential signal. Rényi entropy has been shown to be an effective indicator of the number of basic signal components, gabor logons, in a given signal. The properties of Rényi entropy have been discussed in detail by Flandrin et al. and its application to component counting have been illustrated [2, 3]. Apart from quantifying the complexity of the signal, the entropy function also provides a way of evaluating the performance of different kernels. Previously, iterative methods for designing optimal kernels based on Rényi entropy have been considered for deterministic signals [4].

All of the previous work described above has been concentrated on analyzing the complexity of deterministic signals. In this paper, we focus on Rényi entropy for TFD of random signals. In many real life applications, signals are corrupted by noise. Therefore, it's important to quantify the behaviour of the entropy for noise. First, we review the basic definition of Rényi entropy for TFDs. Then in Section 3, we consider Rényi entropy for random signals and derive the bounds on the expected value of third order Rényi entropy for white Gaussian noise. These bounds give us an idea about how to assess randomness of a given signal and provide a means of detecting signals in noise. The dependence of the entropy on the kernel function is illustrated and some possible ways of designing kernels to minimize the uncertainty of the time-frequency distribution are proposed. The application of minimum entropy kernels for separating gabor logons is illustrated and compared with other TFDs in Section 4. Also applications of Rényi entropy for signal detection are discussed using the derived bounds. Finally, the possible extensions of the results presented in this paper are discussed in the conclusions.

2. RÉNYI ENTROPY FOR TIME-FREQUENCY DISTRIBUTIONS

The uncertainty of signals are studied indirectly through their time-frequency distributions, which represent the energy distribution of a signal as a function of both time and frequency. A time-frequency distribution, $C(t,f)$ from Cohen's class can be expressed as [1] [5]:

$$C(t,f) = \int\int\int \phi(\theta,\tau) s(u+\frac{\tau}{2}) s^*(u-\frac{\tau}{2}) e^{j2\pi(\theta u - \theta t - \tau f)} du d\theta d\tau \quad (1)$$

where the function $\phi(\theta,\tau)$ is the kernel function and s is the signal. The kernel completely determines the properties of its corresponding TFD. Some of the most desired properties of TFDs are the energy preservation and the marginals. They are given as follows and are satisfied when $\phi(\theta,0) = \phi(0,\tau) = 1 \;\; \forall \tau,\theta$.

$$\int\int C(t,f) dt df = \int |s(t)|^2 dt$$
$$\int C(t,f) df = |s(t)|^2 \quad , \quad \int C(t,f) dt = |S(f)|^2 \quad (2)$$

The formulas given above evoke an analogy between a TFD and the probability density function (pdf) of a two-dimensional random variable. In order to have the TFD behave like a pdf, one needs to normalize it properly. For

[1] All integrals are from $-\infty$ to ∞ unless otherwise stated.

this reason, in this paper before applying any entropy measure on the time-frequency plane we first normalize it, i.e. $C_{normalized}(t,f) = \frac{C(t,f)}{\int\int C(t,f)dtdf}$. Another main difference between TFDs and probability density functions is the non-positivity. Most Cohen's class TFDs are nonpositive and therefore cannot be interpreted strictly as densities of signal energy. Therefore, one should be careful while interpreting the results.

The well-known Shannon entropy for TFDs can be written as:

$$H(C) = -\int\int C(t,f) log_2 C(t,f) dtdf \quad (3)$$

Since the TFDs are nonpositive in some regions, this definition will not give finite entropy results. For this reason, Rényi entropy has been introduced as a more appropriate way of measuring time-frequency uncertainty. The αth order Rényi entropy is defined as:

$$R_\alpha(C) = \frac{1}{1-\alpha} log_2 \int\int C^\alpha(t,f) dtdf \quad (4)$$

One can easily see that the Shannon entropy is recovered as the limit of R_α as $\alpha \to 1$. As the passage from Shannon to Rényi entropy involves only the relaxation of the mean value property of entropy from an arithmetic to an exponential mean, R_α behaves much like H [6]. In particular, these functionals can be interpreted as inverse measures of concentration. In this paper, all of the analysis is done for third order Rényi entropy since it has been proved that $\alpha = 3$ is the smallest integer value to yield a well-defined, useful information measure, i.e. $\int\int C^\alpha(t,f)dtdf \geq 0$ and the entropy is not equal to zero [2] [2].

3. ANALYSIS OF RÉNYI ENTROPY FOR TFD OF WHITE NOISE

3.1. Bounds on Rényi Entropy

Rényi entropy for TFDs in discrete-time domain is defined as:

$$R_\alpha(C) = \frac{1}{1-\alpha} log_2 \sum_{k=-K}^{K} \sum_{n=-N}^{N} \left(\frac{C(n,k)}{\sum_{k=-K}^{K}\sum_{n=-N}^{N} C(n,k)}\right)^\alpha \quad (5)$$

where n and k are variables for discrete-time and discrete-frequency respectively, and α is the order of Rényi entropy.

In this section, bounds on the expected value of Rényi entropy for time-frequency distribution of white complex Gaussian noise will be derived. The bounds will be derived for general TFDs with arbitrary size and the order for Rényi entropy will be fixed as three.

It is known that the TFD of white noise should approach to a constant surface in the mean value, since the spectrum for white noise is flat. This is the most uncertain TFD that can be achieved, therefore Rényi entropy for this uniform distribution provides the upper bound on the expected value of Rényi entropy. The third order Rényi entropy for a uniformly distributed two dimensional random vector is given as follows:

$$\begin{aligned} R_3(p_{uniform}) &= -\frac{1}{2} log_2 \sum_{n=0}^{2N}\sum_{k=0}^{2K} \frac{1}{(2N+1)^3(2K+1)^3} \\ &= log_2(2N+1)(2K+1) \end{aligned} \quad (6)$$

where $p_{uniform}$ is the uniform pdf, $2N+1$ is the number of time points and $2K+1$ is the number of frequency points. Therefore,

$$E[R_3(C)] < log_2(2N+1)(2K+1) \quad (7)$$

for all $C(t,f)$. To derive the lower bound for the expected value, we have to make use of Jensen's inequality [7].

Theorem 3.1 *(Jensen's Inequality) If g(.) is a concave function, then $E[g(X)] \leq g(E[X])$, where equality occurs if and only if g is a strictly concave function.*

Since the logarithm is a concave function, we can exchange the order of the expectation operator and the logarithm function to obtain the following inequality.[3]

$$\begin{aligned} E[R_3(C)] &= -\frac{1}{2} E[log_2 \frac{\sum_k\sum_n C^3(n,k)}{(\sum_k\sum_n C(n,k))^3}] \\ &= -\frac{1}{2} E[log_2 \sum_k\sum_n C^3(n,k) - 3 log_2 \sum_k\sum_n C(n,k)] \\ &\geq -\frac{1}{2} log_2[\sum_k\sum_n E[C^3(n,k)]] + \frac{3}{2} E[log_2 \sum_k\sum_n C(n,k)] \end{aligned} \quad (8)$$

To compute the first part of this expression, we need to evaluate $E[C^3(n,k)]$ for white noise. This computation requires the sixth order joint moment for white, complex Gaussian process. For this computation, the well known moment generating theorem is used [8].

Theorem 3.2 *The joint moment for N random variables is given by:*

$$E[X_1X_2\ldots X_N] = j^{-N}\frac{\partial^N}{\partial\omega_1\partial\omega_2\ldots\partial\omega_N}\Phi(\omega_1,\omega_2,\ldots,\omega_N)|_{\omega_1=\ldots=\omega_N=0} \quad (9)$$

where $\Phi(\omega_1,\omega_2,\ldots,\omega_N)$ is the joint characteristic function.

Using the moment theorem, the sixth order joint moment for Gaussian random variables is given as:

$$\begin{aligned} E[X_1X_2X_3X_4X_5X_6] = & K_{12}K_{34}K_{56} + K_{12}K_{35}K_{46} + K_{12}K_{36}K_{45} \\ & + K_{13}K_{24}K_{56} + K_{13}K_{25}K_{46} + K_{13}K_{26}K_{45} \\ & + K_{14}K_{23}K_{56} + K_{14}K_{26}K_{35} + K_{14}K_{25}K_{36} \\ & + K_{15}K_{23}K_{46} + K_{15}K_{24}K_{36} + K_{15}K_{26}K_{34} \\ & + K_{16}K_{25}K_{34} + K_{16}K_{24}K_{35} + K_{16}K_{23}K_{45} \end{aligned} \quad (10)$$

where K_{ij} is the covariance between the i and jth random variables. Since the noise is assumed to be white, complex Gaussian, the only nonzero correlations are the ones

[2] Although there are some pathological signals for which this is not true, for the purposes of this paper this inequality is always true.

[3] $\sum_n = \sum_{n=-N}^{N}$ and $\sum_k = \sum_{k=-K}^{K}$ unless otherwise specified.

between the real and the complex components of the noise. Substituting this expression inside $\sum_n \sum_k E[C^3(n,k)]$ gives:

$$\sum_n \sum_k E[C^3(n,k)] = 3\sigma^6(2K+1)(2\pi)(2N+1)\sum_n \sum_{\tau=-2N}^{2N} |\Psi(n,\tau)|^2 + \sigma^6(2K+1)(2\pi)^3(2N+1)^3 + 2\sigma^6(2K+1)(2\pi)(2N+1) \quad (11)$$

where σ is the variance of the noise, (2N+1) is the number of time points, (2K+1) is the number of frequency points and $\Psi(n,\tau)$ is the time-frequency kernel in the time-lag domain.

The second part of equation 8 depends on computing the expected value for $log_2[\sum_n \sum_k C(n,k)]$. This expression can be simplified by noticing that $\sum_n \sum_k C(n,k) = (2\pi)(2K+1)\sum_n |x(n)|^2$. Since $x(n)$ is white Gaussian noise with zero mean and variance σ, after proper normalization it is seen that $\sum_n |x(n)|^2$ can be written as a χ^2 random variable with 2N+1 degrees of freedom. The logarithmic χ^2 is a known random variable with tabulated expected values [7]. After this evaluation equation 8 can be rewritten as follows:

$$E[R_3(C(n,k))] \geq \frac{1}{2}log_2\frac{(2K+1)^2(2\pi)^2(2N+1)^2}{3\sum_n \sum_{\tau=-2N}^{2N} |\Psi(n,\tau)|^2 + (2\pi)^2(2N+1)^2 + 2} - \frac{3}{2(\ln 2)(2N+1)} - \frac{1}{2(\ln 2)(2N+1)^2} \quad (12)$$

As it can be seen, this equation is independent of the variance of the noise. This is an expected result since Rényi entropy is invariant under the scaling of the amplitude. The entropy depends on the time-frequency kernel used and the dimensions of the distribution. It is also seen that Wigner distribution will give the smallest entropy among the well-known time-frequency distributions since the entropy is inversely proportional to the sum of the square of the kernel.

3.2. Minimum Entropy Kernels

The lower bound presented in the previous section suggests a way of reducing the entropy of the TFD for white noise. Since the entropy is inversely proportional to the sum of the square of the kernel function in the time-lag domain, in order to minimize the entropy we need to maximize the energy of the kernel. This maximization problem can be formulated with the constraint that the distribution satisfies the frequency marginal and the time support properties.

$$maximize \sum_{\tau=-2N}^{2N} \sum_{n=-N}^{N} \Psi^2(n,\tau) \; subject\ to \sum_n \Psi(n,\tau) = 1 \;\; |\tau| \leq 2N \quad (13)$$

This is very similar to the problem formulation that Hearon and Amin present in their paper on minimum variance time-frequency kernels [9]. For minimizing the variance of the TFD for white noise, the energy of the kernel should be minimized. The minimum variance kernel problem is exactly the opposite of the problem we have presented above. Therefore, the minimum variance kernel implies maximum entropy for white Gaussian noise. The reason beyond this inverse relationship between variance and entropy is that minimum variance for the TFD of white noise suggests a spectrum which is very close to the flat spectrum, and flat spectrum corresponds to maximum entropy as discussed before.

Unlike the minimum variance problem, the minimum entropy problem does not have a unique solution unless some additional constraints are imposed. If the kernel's elements are constrained to be positive and symmetric, then the unique solution is the Wigner distribution. This is an expected result, since the Wigner distribution possesses the highest resolution on the time-frequency plane when compared to other TFDs.

In order to get different kernels with lower entropy we can relax the condition that the elements of the kernel are nonnegative. This gives a large class of kernels with negative values and the entropy can be made as small as desired since there is no upper bound on the energy of the kernel. To reduce the size of this class of kernels we suggest a symmetric structure with time-support property and the requirement that it is close to the Wigner distribution. This requirement results in kernels of the type depicted in Figure 1, where N is an arbitrary positive integer which controls the entropy of the distribution by increasing the energy of the kernel. In the next section, we are going to consider some implications of this type of kernels.

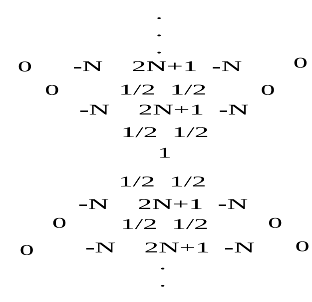

Figure 1: Structure for a symmetric minimum entropy kernel

4. RESULTS

In this section, we are going to show how the bounds derived in the previous section are related to the actual results obtained through simulations. The relation will be shown through simulations of white noise and computing the average entropy of the TFD of these simulation results. For a Born-Jordan kernel, the simulation averages and the bounds are given for different sizes of the kernel in the following table.

Size of TFD	*Upper Bound*	*Lower Bound*	*Simulation*
33 × 65	11.07	5.95	6.96
65 × 129	13.03	6.97	8.26
129 × 257	15.02	7.98	9.64

As it can be seen from the above table, the actual entropy tends to lie closer to the lower bound. This suggests the possible usage of the lower bound for detecting signals in noise. The entropy for the TFD of signal plus noise is always less than the entropy for the noise by itself. Using this fact, a gabor logon with noise is generated at different signal-to-noise ratios (SNR), and the lower bound for the entropy of the noise is used as a threshold to detect whether the logon is present or not. This simulation illustrates that

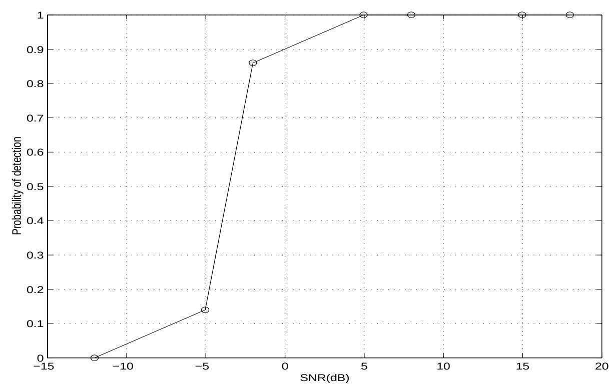

Figure 2: Probability of detection versus SNR for a gabor logon with different noise levels

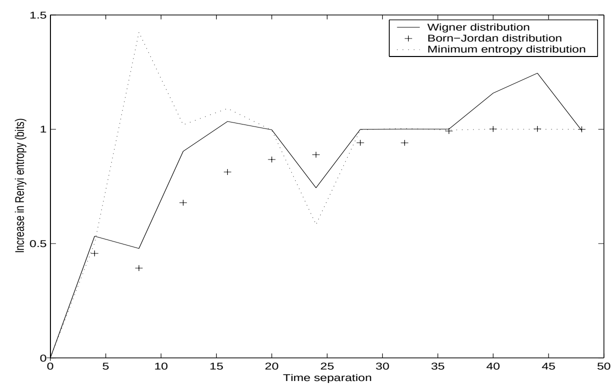

Figure 3: R_3 information gain for Wigner, Born-Jordan and Minimum Entropy kernels with respect to the time separation between the two gabor logons

Rényi entropy is a reliable measure for detecting signals in noise even when the SNR is low. [Figure 2]

Apart from the direct usage of the bounds for detection purposes, the formula derived for the lower bound brings up interesting questions about minimizing entropy. The minimum entropy kernel results presented in the previous section suggest that the entropy can be made as small as desired at the expense of introducing negative terms in the kernel. The minimum entropy kernels have a high resolution in the time-frequency plane but they cause negative sidelobes to occur due to the negative terms in the kernel. The high resolution property of these kernels make them a natural candidate for detecting the number of signal components. For example, the well-known counting components problem can be addressed with this new class of kernels. The performance of different kernels for two gabor logons at different time separations is presented in Figure 3. As it can be seen from the figure, the minimum entropy kernel reaches the 1 bit gain level, i.e. detects the second component, before the other two distributions. It is superior to Wigner distribution due to increased resolution. The minimum entropy kernel produces more unstable behaviour compared to the other two distributions, especially when the gabor logons are very close to each other, since the distribution is more negative than the others. This is a natural result, since we know that minimum entropy kernels have high variance and thereby yield TFDs that are not very stable.

5. CONCLUSIONS

In this paper, we have extended the previous work in using Rényi entropy for measuring the complexity of deterministic signals to measuring the uncertainty of random signals, or more specifically white Gaussian noise. We have presented a quantitative analysis for the expected value of Rényi entropy for white noise and derived the bounds on this quantity. These bounds show how random signals behave in general and suggest an easy and direct way of detecting signals in white noise. The proposed detection scheme is shown to be effective at SNR values as low as -2 dB. Apart from detecting signals in noise, the lower bound derived in this paper is shown to be useful in interpreting and constructing minimum entropy kernels. The class of minimum entropy kernels are shown to be more effective at separating gabor logon components compared to the existing distributions.

The work presented in this paper can be extended to different signal classes, and some preliminary results suggest that Rényi entropy can be used as an effective measure for separating between different classes of real-time signals. Also, the class of minimum entropy kernels discussed here can be expanded by modifying the constraints on the time-frequency distribution.

6. REFERENCES

[1] W. J. Williams, M. Brown, and A. Hero, "Uncertainty, Information, and Time-Frequency Distributions," *Proc. SPIE Advanced Signal Processing Algorithms, Architectures, and Implementations,* vol. 1566, pp. 144-156, 1991.

[2] P. Flandrin, R. G. Baraniuk, and O. Michel, "Time-Frequency Complexity and Information," *Proc. ICASSP'94,* vol. 3, pp. 329-332, 1994.

[3] O. Michel, R. G. Baraniuk, and P. Flandrin, "Time-Frequency Based Distance and Divergence Measures," *Proc. Time-Frequency and Time-Scale Analysis,* pp. 64-67, 1994.

[4] T. Sang, W. J. Williams, "Rényi Information and Signal-Dependent Optimal Kernel Design," *Proc. ICASSP'95,* vol. 2, pp. 997-1000, 1995.

[5] L. Cohen, *Time-Frequency Analysis,* Prentice Hall, 1995.

[6] A. Rényi, "On Measures of Entropy and Information," *Proc. 4th Berkeley Symp. Math., Stat., Prob.,* vol. 1, pp. 547-561, 1961.

[7] M. Abramowitz, *Handbook of mathematical functions with formulas, graphs, and mathematical tables,* U.S. Govt. Print. Off., 1964.

[8] H. Stark, J. W. Woods, *Probability, Random Processes, and Estimation Theory for Enginers,* Prentice Hall, 1994.

[9] S. B. Hearon, M. G. Amin, "Minimum-variance time-frequency distribution kernels," *IEEE Trans. on Signal Processing*, vol. 43, no. 5, pp. 1258-1262, May 1995.