INFORMATION PROCESSING ON THE TIME-FREQUENCY PLANE

Selin Aviyente

Department of Electrical and Computer Engineering,
Michigan State University, East Lansing, MI
e-mail:aviyente@egr.msu.edu

ABSTRACT

Time-frequency analysis is a major tool in representing the energy distribution of time-varying signals. There has been a lot of research on various properties of these representations. However, there is a general lack of quantitative analysis in describing the amount of information encoded into a time-frequency distribution. Recently, entropy based measures have been applied to the time-frequency plane to quantify the information content of signals. This paper aims to extend this approach to include other information theoretic measures such as the divergence measures to quantify how time-frequency distributions discriminate signals in an information theoretic framework. Different distance measures, such as Kullback-Leibler distance, Rényi distance, and Jensen difference based measures will be adapted to the time-frequency plane. The robustness of different distance measures under an additive perturbation model will be derived. The performance of different distance measures in quantifying the differences in information between signals will be demonstrated. Finally, the proposed distance measures will be applied on a set of event related brain potentials to discriminate different subject groups.

1. INTRODUCTION

Time-frequency distributions (TFDs) are used for representing the energy distribution of time-varying signals simultaneously in time and frequency. Despite their wide use in areas such as detection and classification of signals, their capacity in representing information has not been evaluated quantitatively. This paper aims at addressing this issue by using information-theoretic distance measures.

Distance measures between statistical models have been widely used in signal processing applications. Using entropy based distance functionals is a well-known discrimination method in signal processing. These functionals are known as divergence measures and are applied directly on statistical models describing the signals. Measures of divergence between two probability distributions are used to associate, cluster, classify, compress, and restore signals, images and patterns, in many applications. Many different measures of divergence have been constructed and characterized [1, 2].

Recently, there has been interest in applying information measures such as entropy on time-frequency representations to measure the signal complexity [3, 4]. This approach has been useful in quantifying the information content of time-varying signals. In this paper, we will extend this approach further so that the difference between signals can be quantified in an information theoretic sense. We will adopt some well-known divergence measures from probability theory and define them for time-frequency distributions. The major properties of these distance measures will be discussed and their robustness against possible perturbations to the signal model will be analyzed.

In Section 2, some well-known information-theoretic distance measures will be reviewed and defined for time-frequency distributions. A local curvature analysis for the divergence measures will be introduced to determine the robustness of a given divergence measure. Section 3 will present results that illustrate the performance of different divergence measures under an additive perturbation model. These results will be compared to the theoretical ones derived in Section 2. An application of the proposed distance measures in classifying a set of event related brain potentials will also be presented. Finally, Section 4 will discuss the major contributions of the paper and suggest some applications of the proposed method.

2. INFORMATION THEORETIC DISTANCE MEASURES ON THE TIME-FREQUENCY PLANE

A time-frequency distribution, $C(t, f)$, from Cohen's class can be expressed as [1] [5]:

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

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 $\phi(\theta, 0) = \phi(0, \tau) = 1 \;\; \forall\tau, \theta$.

$$\begin{aligned} \int\int C(t,f)\,dt\,df &= \int |s(t)|^2\,dt = \int |S(f)|^2\,df \\ \int C(t,f)\,df &= |s(t)|^2 \;\;, \int C(t,f)\,dt = |S(f)|^2. \end{aligned} \tag{2}$$

The formulas given above evoke an analogy between a TFD and the probability density function (pdf) of a two-dimensional random variable. Entropy is the measure of information for a given probability density function. Similarly, one can apply entropy and other information theoretic measures to time-frequency distributions to

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

quantify the information of a signal. The main difference between TFDs and pdfs is that TFDs are not always positive. Therefore, in this paper the analysis focuses on spectrograms since they are always positive. Another important point is that distributions have to be normalized by their energy before applying any information theoretic measures on them.

2.1. Distance Measures

The most general class of distance measures is known as Csiszar's f-divergence which includes some well-known measures like Hellinger distance, Kullback-Leibler divergence and Rényi divergence [2]. The divergence between two probability density functions, p_1 and p_2 for this class of distance measures can be expressed as:

$$d(p_1, p_2) = g\left[E_1\left[f\left(\frac{p_2}{p_1}\right)\right]\right], \tag{3}$$

where f is a continuous convex function, g is an increasing function and E_1 is the expectation operator with respect to p_1. The distance measures and their properties for time-frequency distributions are given below.

1. **Kullback-Leibler divergence**: The most common distance measure used for probability distributions is the Kullback-Leibler divergence measure. This measure can be adapted to the time-frequency distributions as follows:

$$K(C_1, C_2) = \int\int C_1(t,f) \log \frac{C_1(t,f)}{C_2(t,f)} dt\, df. \tag{4}$$

This measure belongs to the class of Csiszar's f-divergence with $f(x) = -\log x$, and $g(x) = x$. $0 \leq K(C_1, C_2) \leq \infty$, the first equality holds if and only if $C_1 = C_2$ and the second equality holds if and only if $Supp\ C_1 \bigcap Supp\ = \emptyset$. This is not a symmetric distance measure but can easily be symmetrized by taking the average of $K(C_1, C_2)$ and $K(C_2, C_1)$.

2. **Rényi Divergence**: Rényi divergence is a generalized formulation of Kullback-Leibler divergence and can be expressed as:

$$D_\alpha(C_1, C_2) = \frac{1}{\alpha - 1} \log \int\int C_1^\alpha(t,f) C_2^{1-\alpha}(t,f) dt\, df. \tag{5}$$

where $\alpha \in [0, 1]$ is the order of Rényi divergence. This measure converges to Kullback-Leibler distance as $\alpha \to 1$. It is also a member of Csiszar's f-divergence with $f(x) = x^{1-\alpha}$, and $g(x) = \frac{1}{\alpha-1}\log(x)$. $0 \leq D_\alpha(C_1, C_2) \leq \infty$, the first equality holds if and only if $C_1 = C_2$ and the second if and only if $Supp\ C_1 \bigcap Supp\ C_2 = \emptyset$.

3. **Jensen-Shannon Divergence**: One common approach for constructing divergence measures is to apply Jensen inequality on the entropy functional. For time-frequency distributions, Jensen-Shannon divergence can be defined as:

$$J(C_1, C_2) = H\left(\frac{C_1 + C_2}{2}\right) - \frac{H(C_1) + H(C_2)}{2}. \tag{6}$$

This distance measure is always positive since

$$H\left(\frac{C_1 + C_2}{2}\right) \geq \frac{H(C_1)}{2} + \frac{H(C_2)}{2} \tag{7}$$

by concavity of H. It is equal to zero when $C_1 = C_2$ and is a symmetric divergence measure. Unlike the Kullback-Leibler divergence, Jensen-Shannon distance does not diverge when the two distributions are disjoint.

4. **Jensen-Rényi Divergence**: The Rényi entropy is derived from the same set of axioms as the Shannon entropy, the only difference being the employment of a more general exponential mean instead of the arithmetic mean in the derivation. This realization inspires the modification of Jensen-Shannon divergence from an arithmetic to a geometric mean, and the following quantity is obtained for two positive TFDs C_1 and C_2.

$$J_1(C_1, C_2) = H_\alpha(\sqrt{C_1 C_2}) - \frac{H_\alpha(C_1) + H_\alpha(C_2)}{2}, \tag{8}$$

where $(\sqrt{C_1 C_2})(t,f) = \sqrt{C_1(t,f) C_2(t,f)}$. This quantity is obviously null when $C_1 = C_2$. The positivity of this quantity can be proven using the Cauchy-Schwartz inequality.

$$\left|\int\int [C_1(t,f) C_2(t,f)]^{\alpha/2}\ dt\, df\right|^2 \leq \int\int C_1^\alpha(t,f) dt\, df \int\int C_2^\alpha(t,f)\, dt\, df, \tag{9}$$

and since the log function is monotonically increasing, for $\alpha > 1$

$$\frac{1}{1-\alpha} \quad \log\left|\int\int [C_1(t,f) C_2(t,f)]^{\alpha/2}\ dt\, df\right|^2 \geq \frac{1}{1-\alpha} \quad \left[\log \int\int C_1^\alpha(t,f)\, dt\, df + \log \int\int C_2^\alpha(t,f)\, dt\, df\right]. \tag{10}$$

Thus $H_\alpha(\sqrt{C_1 C_2}) \geq \frac{H_\alpha(C_1) + H_\alpha(C_2)}{2}$, which proves that the distance measure is always positive.

2.2. Sensitivity Analysis

In order to investigate the effectiveness of a given divergence measure, we carry out a sensitivity analysis using the local curvature approach [6]. The local curvature can be interpreted as the instantaneous acceleration of the divergence measure when it takes off from zero as one signal starts to depart from the other. It can also be used as a measure of robustness for a given distance.

Under an additive perturbation model such as $s(t) = (1 - \epsilon)x(t) + \epsilon g(t)$, where $\epsilon \in [0, 1]$ controls the amount of perturbation and $g(t)$ is the actual perturbation, we can compute the local sensitivity or the robustness of the distance measures as:

$$\frac{\partial^2}{\partial \epsilon^2} D(C_{xx}(t,f), C_{ss}(t,f))\,|_{\epsilon=0} \tag{11}$$

where $D(\cdot,\cdot)$ is an information theoretic distance measure, $C_{xx}(t,f)$ and $C_{ss}(t,f)$ are the time-frequency distributions of the original and the perturbed signals respectively. Under an additive perturbation model, $s(t) = (1 - \epsilon)x(t) + \epsilon g(t)$, where $\epsilon \in [0, 1]$, the spectrogram of $s(t)$ will be the sum of the individual spectrograms and the cross-spectrogram as follows:

$$C_{ss}(t,f) = (1-\epsilon)^2 C_{xx}(t,f) + 2(1-\epsilon)C_{xg}(t,f) + \epsilon^2 C_{gg}(t,f), \tag{12}$$

where $C_{xg}(t,f)$ will refer to the real part of the cross-terms between $x(t)$ and $g(t)$. In this section, the local curvature of the distance measures will be formulated under this perturbation model.

- The sensitivity of the Kullback-Leibler divergence measure can be computed as follows:

$$\begin{aligned} \frac{\partial^2}{\partial\epsilon^2} & \quad K(C_{xx}(t,f), C_{ss}(t,f))\,|_{\epsilon=0} \\ = & \quad -4\int\int C_{xg}(t,f)\,dt\,df + 4\int\int \frac{C_{xg}^2(t,f)}{C_{xx}(t,f)} dt\,df. \end{aligned} \tag{13}$$

  As it can be seen from this equation, the robustness of the divergence measure depends on the energy of the cross-term, and the energy of the ratio of the cross-term to the original signal. In order to increase robustness of the distance measure, we need to minimize $4\int\int \frac{C_{xg}^2(t,f)}{C_{xx}(t,f)} dt\,df$, since in general $\int\int C_{xg}(t,f)dt\,df$ is negligible for normalized signals. Minimizing this quantity corresponds to minimizing the cross-terms which agrees with our intuition of reducing cross-terms for better discrimination between signals. This energy can be minimized by using reduced interference distributions. The robustness of the distance measure will also improve when $C_{gg}(t,f)$, the TFD of the perturbation signal, is well separated from $C_{xx}(t,f)$, the TFD of the original signal, on the time-frequency plane.

- For the Rényi divergence, the same measure can be computed as follows:

$$\begin{aligned} &\frac{\partial^2}{\partial\epsilon^2} D_\alpha(C_{xx}(t,f), C_{ss}(t,f))|_{\epsilon=0} \\ &= \frac{\partial^2}{\partial\epsilon^2}\left[\frac{1}{\alpha-1}\log_2\int\int C_{xx}^\alpha(t,f)C_{ss}^{1-\alpha}(t,f)dt\,df\right]_{\epsilon=0} \\ &= 4\alpha\int\int\frac{C_{xg}^2(t,f)}{C_{xx}(t,f)}\,dt\,df - 4\int\int C_{xg}(t,f)\,dt\,df \\ &+ 4(1-\alpha)\left(\int\int C_{xg}(t,f)\,dt\,df\right)^2 \end{aligned} \tag{14}$$

  where α is the order of the Rényi divergence. As $\alpha \to 1$, the above expression becomes equal to

$$-4\int\int C_{xg}(t,f)dt\,df + 4\int\int\frac{C_{xg}^2(t,f)}{C_{xx}(t,f)}dt\,df, \tag{15}$$

  which is the same as the sensitivity of Kullback-Leibler measure. This is an expected result since Kullback-Leibler distance measure is a special case of Rényi distance as $\alpha \to 1$. In general, the cross-terms oscillate between positive and negative values, thus one can show that,

$$\int\int C_{xg}^2(t,f)\,dt\,df > \left(\int\int C_{xg}(t,f)\,dt\,df\right)^2. \tag{16}$$

  Therefore, the dominant term in Rényi divergence will be $4\alpha\int\int\frac{C_{xg}^2(t,f)}{C_{xx}(t,f)}dt\,df$. For $0 < \alpha < 1$, the dominant term in Rényi divergence is less than the dominant term in Kullback-Leibler divergence. Therefore, Rényi divergence will be more robust compared to Kullback-Leibler divergence.

- For the Jensen-Shannon divergence, the local sensitivity under an additive perturbation model can be computed as:

$$\begin{aligned} &\frac{\partial^2}{\partial\epsilon^2} JS(C_{xx}(t,f), C_{ss}(t,f))|_{\epsilon=0} \\ &= 1 - 2\int\int C_{xg}(t,f)dt\,df + \int\int\frac{C_{xg}^2(t,f)}{C_{xx}(t,f)}dt\,df. \end{aligned} \tag{17}$$

  This result is comparable to the sensitivity of Kullback-Leibler measure. Depending on the magnitude of $\int\int\frac{C_{xg}^2}{C_{xx}}dtdf$, Jensen-Shannon divergence may be more or less robust than Kullback-Leibler distance.

- For Jensen-Rényi divergence, the local curvature under the additive perturbation model will depend on α as follows:

$$\begin{aligned} \frac{\partial^2}{\partial\epsilon^2} & \quad JR(C_{xx}(t,f), C_{ss}(t,f))|_{\epsilon=0} \\ = & \quad \frac{\alpha^2}{\alpha-1}\left[\frac{\int\int C_{xx}^\alpha + C_{xx}^{\alpha-2}C_{xg}^2 - 2C_{xx}^{\alpha-1}C_{xg}dt\,df}{\int\int C_{xx}^\alpha(t,f)dt\,df}\right] \\ + & \quad \frac{\alpha^2}{\alpha-1}\left[\frac{\left(\int\int -C_{xx}^\alpha + C_{xx}^{\alpha-1}C_{xg}dt\,df\right)^2}{\left(\int\int C_{xx}^\alpha(t,f)dt\,df\right)^2}\right], \end{aligned} \tag{18}$$

  where $\alpha > 1$.

3. RESULTS

In this section, the robustness of the different distance measures introduced above will be computed for an example signal both theoretically and experimentally.

Example 1: In this example, the change in the information-theoretic distance as the signal is perturbed will be explored. The original signal is a gabor logon, $x(t) = \exp(-(t-t_0)^2)\exp(-j\omega_0 t)$, centered at time $t_0 = 32$, normalized frequency $\omega_0 = 0.2$, and the perturbation signal is another gabor logon, $g(t) = \exp(-(t-t_1)^2)\exp(-j\omega_0 t)$, centered at $t_1 = 64, \omega_0 = 0.2$. The perturbed signal is $s(t) = (1-\epsilon)x(t) + \epsilon g(t)$, where $\epsilon \in [0,1]$. The distance between the time-frequency distributions of the perturbed signal and the original one is computed as ϵ goes from 0 to 1.

Figure 1 shows the comparison between the Kullback-Leibler and the Rényi divergences for different values of α. As $\alpha \to 1$, Rényi divergence gets closer to the Kullback-Leibler divergence as expected. The local curvature is computed for these distance measures theoretically and experimentally by computing the second derivative of the curve at $\epsilon = 0$. The results are summarized in Table 1. The theoretical and experimental results are close to each other. The deviations between the two are due to the fact that computing the local curvature from the experimental data depends on fitting a polynomial which is prone to error. It can be seen that as α decreases, the distance measure becomes more robust and we get smaller curvature values.

Figure 2 shows a similar comparison between the Jensen-Shannon and the Jensen-Rényi divergences for different values of

Order of Rényi Entropy	Simulation	Theoretical
1	2.4428	2.8854
0.9	2.3162	2.5969
0.5	1.7458	1.4427
0.2	1.0496	0.5771

Table 1. Local curvature of Rényi divergence for different α

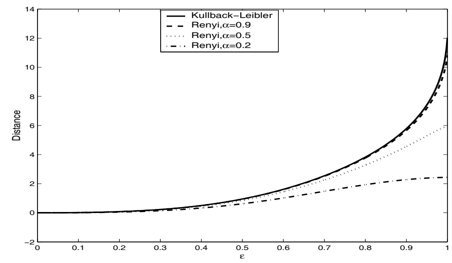

Fig. 1. Kullback-Leibler and Rényi divergence for $\alpha = 0.2, 0.5, 0.9$

α. The theoretical values of the local curvature for these distance measures are given in Table 2. For these distance measures, since the distance versus ϵ curves are flat close to $\epsilon = 0$, it is hard to fit polynomials accurately. Therefore, only the theoretical curvature values are presented.

Divergence Measure	Local Curvature
JS	1.7214
JR(2)	1.4395
JR(3)	1.6212
JR(4)	1.9174

Table 2. Local curvature of Jensen-Shannon (JS) and Jensen-Rényi (JR) divergences for different α

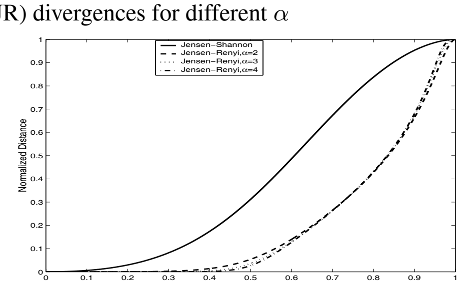

Fig. 2. Jensen-Shannon and Jensen-Rényi divergence $\alpha = 2, 3, 4$

Example 2: In this example, an application of distance measures for distinguishing between brain responses will be presented. The event-related potentials are collected during an experiment that aims at differentiating between the responses of two different groups of subjects, spider phobics and non-phobics, to subliminal presentations of spider stimulus. For purposes of comparison, event-related potentials were collected at two electrodes: Oz and Cz. The major phenomena that need to be investigated are the difference in responses for the two groups, the phobics and non-phobics, and the difference in responses at different electrodes. For each subject and each trial the time-frequency distribution of both the pre- and post-stimulus regions are computed. The distance between the pre-stimulus and the post-stimulus is obtained by applying the distance measure on the average time-frequency surface. A two-way analysis of variance (ANOVA) is used to explore the interactions between the two factors, i.e. the electrode and the phobic group. We are especially interested in the interaction between the phobia group and the electrode since that will give us information about how phobic stimulus is processed by different subjects at different parts of the brain. For this purpose, Jensen-Rényi distance with $\alpha = 3$ is used since it is a symmetric distance measure and has been shown to be robust. Using this distance measure, we observe a significant effect for the interaction between the phobia group and the electrode at the 5% significance level ($F(1, 14) = 8.073, p = 0.013, \eta^2 = 0.366$).

4. CONCLUSIONS

In this paper, the idea of discriminating probability models using divergence measures is extended to the time-frequency plane. Distance measures are defined for quantifying the difference between two signals on the time-frequency plane. The robustness of these measures is evaluated by performing a local curvature analysis under an additive perturbation model. It is observed that for all of the distance measures introduced in this paper robustness is inversely proportional to the energy of the cross-terms between the original signal and the perturbation. Therefore, using TFDs that minimize the energy of the cross-terms will increase the robustness of the distance measures. It is also observed that in general Rényi entropy based divergence measures are more robust than Shannon entropy based ones. The proposed distance measures are applied on an example signal under perturbation and the local curvature computed directly from the distance graphs are shown to agree with the theoretical ones. The distance measures are also shown to be effective in classifying event related brain potentials.

The distance measures introduced in this paper focused on positive distributions such as the spectrogram. However, most bilinear time-frequency distributions belonging to Cohen's class are non-positive. In that case, the divergence measures will not be well-defined. Therefore, the distance measures introduced here should be modified to account for the negativity in the distributions. The results presented here can also be extended for different perturbation models such as a multiplicative one.

5. REFERENCES

[1] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, no. 4, pp. 349–369, Dec. 1989.

[2] I. Csiszar, "Information-type distance measures and indirect observations," *Stud. Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[3] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Trans. on Info. Theory*, vol. 47, no. 4, pp. 1391–1409, May 2001.

[4] W. J. Williams, M. Brown, and A. Hero, "Uncertainty, information and time-frequency distributions," in *SPIE-Advanced Signal Processing Algorithms*, 1991, vol. 1556, pp. 144–156.

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

[6] D. K. Dey and L. R. Birmiwal, "Robust bayesian analysis using divergence measures," *Statist. Probab. Letters*, pp. 287–294, 1994.