CLASS-DEPENDENT, DISCRETE TIME-FREQUENCY DISTRIBUTIONS VIA OPERATOR THEORY

Jack McLaughlin, James Droppo and Les Atlas

Interactive Systems Design Laboratory, Department of Electrical Engineering, Box 352500
University of Washington, Seattle, WA 98195-2500
jackm | jdroppo | les @isdl.ee.washington.edu

ABSTRACT

We propose a property for kernel design which results in distributions for each of two classes of signals which maximally separates their energies in the time-frequency plane. Such maximally separated distributions may result in improved classification because the signal representation is optimized to accentuate the differences in signal classes. This is not the case with other time-frequency kernels which are optimized based upon some criteria unrelated to the classification task. Using our operator theory formulation for time-frequency representations, our "maximal separation" criteria takes on a very easily solved form. Analysis of the solution in both the time-frequency and ambiguity planes is given along with an example on discrete signals.

1. INTRODUCTION

A single discrete signal is associated with an essentially infinite set of quadratic time-frequency distributions (TFDs) through appropriate choices for the kernel function. Useful kernels can be selected by incorporating into the kernel design properties that are desired in the end distribution.

Kernel design for quite a number of "desired properties" has been researched, most notably design for suppression of distracting cross terms [1][2][3]. By making the kernel signal-dependent, a wider variety of properties can be obtained in the TFD. For example, *proper* distributions (distributions which are nonnegative and have physically meaningful marginal distributions) can be obtained [4][5][6]. These methods begin with a prior distribution that is not proper, and iteratively calculate that proper distribution which is nearest, by some measure, to the original distribution. A method for computing proper, signal dependent distributions without a prior has been developed [7], but the marginal constraints must be altered to achieve this. Though some of these TFDs may offer advantages in classification of certain types of signals, they cannot hope to offer improved signal discrimination for all signals because discrimination is not one of the goals of the kernel design procedure.

We present a kernel design procedure in which signal discrimination (in the time-frequency plane) is the *only* goal. Using our operator theory formulation for TFDs, we are able to easily develop a closed-form solution for an optimally discriminating kernel which is not signal dependent, but *signal class* dependent. Use of TFDs for signal classification [8] and detection [9][10] (a similar problem) has been researched previously, but from the point of view of discovering which, if any, of the existing TFDs might succeed with certain signal types. The idea of custom designing kernels has been explored [11], but not with an eye to classification.

We begin with a discussion of operator theory and our application of this to time-frequency analysis. We then discuss our kernel design procedure from both the time-frequency perspective and the ambiguity plane perspective which provides valuable insight into the nature of our operator theory formulation as well as our kernel design process. We conclude with a simple example.

2. OPERATOR THEORY

Just as the discrete wavelet transform does not need to be viewed in the context of the continuous version, discrete-time discrete-frequency distributions do not need to be viewed simply as sampled versions of their continuous counterparts [12][13]. Here, we present a means of directly connecting a discrete-time discrete-frequency distribution with a discrete, finite-length input signal through the use of a discrete version of operator theory [14]. Though operator theory has been invoked in the generation of discrete-time discrete-frequency TFDs in the past [15], it has not been used in this context. With this in hand, we develop a wholly new expression for discrete-time, discrete-frequency distributions. A more detailed discussion is given in [12].

We first define two operators, $\boldsymbol{L}$ and $\boldsymbol{K}$ such that

$$\sum_n x^*[n]\,\boldsymbol{K}\{x[n]\} = E\{X[k]\}$$

and

$$\sum_n x^*[n]\,\boldsymbol{L}\{x[n]\} = E\{x[n]\}$$

where $E\{\}$ is the expectation operation. The operator $\boldsymbol{L}$ is just the sequence {-(N-1)/2, ... 0, ... (N-1)/2} which can be written as a diagonal matrix with this sequence along the diagonal, and $\boldsymbol{K}$ is the inverse discrete Fourier transform of the sequence {-(N-1)/2, ... 0, ... (N-1)/2}. $\boldsymbol{K}$ is circulant when written in matrix form as it performs a circular convolution as its method of operation.

The discrete TFD can be computed from these operators by first calculating the characteristic function, $M[\theta, m]$, and then performing a two-dimensional Fourier transform:

$$M[\theta, m] = \sum_n x^*[n]\,\phi[\theta, m]\, e^{\frac{j2\pi\theta \boldsymbol{L}}{N}}\, e^{\frac{j2\pi m \boldsymbol{K}}{N}}\, x[n]$$

Here, we make use of the correspondence rule, which maps permutations of the terms in the exponent to a kernel function $\phi[\theta,m]$, which is a scalar function of θ and m. Expressing the above equation in matrix notation (and using H to indicate the conjugate transpose), we obtain (after the 2-D Fourier transform) the TFD.

$$P[n, k] = x^H(\Phi_{(n,k)} .* \boldsymbol{F})\,\boldsymbol{F}^H x \qquad (1)$$

$\boldsymbol{F}$ is the unitary Fourier transform matrix ($\boldsymbol{F} = \boldsymbol{KLK}^{\mathrm{H}}$), and the .* operation indicates an element-by-element multiplication. Φ is a function of n and k only in that for different values of these variables, the Φ matrix has its rows and/or columns circularly shifted. This results in a kernel which is effectively circularly convolved with some other matrix to yield the TFD.

While (1) may seem to be simply a rewriting of the conventional expression for a discrete TFD, we must remark here that whereas the conventional expression generates a TFD from the convolution of a discrete approximation of the Wigner distribution of the signal with the kernel, our expression generates a TFD from the *circular* convolution of the kernel with the Rihaczek distribution [16] of the signal.

3. CLASS-DEPENDENT DISTRIBUTIONS

Time-Frequency Plane

We wish to find a kernel such that the TFDs P_1 and P_2 for each of two signal classes are maximally separated:

$$\max_{\Phi}\left\{\sum_{n,k}\left|P_1[n,k] - P_2[n,k]\right|^2\right\} \qquad (2)$$

(For the moment, we assume for simplicity that there is only one representative signal for each class.)

We begin our simplifications by noting that for a single ordered pair (n,k) and a given input signal x, the right-hand side of (1) is a linear combination of the elements of Φ. Let us write the coefficients of this linear combination as the column vector x_1. This vector contains the elements of the discrete Rihaczek distribution. As n and k change, the rows and columns of Φ shift, but this shift can also be captured by rearranging the elements of x_1. So, by reshaping the Φ matrix into a vector $\overline{\Phi}$, (1) can be rewritten as

$$P = \boldsymbol{X}^T\overline{\Phi} \quad \text{where} \quad \boldsymbol{X} = \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix} \qquad (3)$$

The x_1, ..., x_N are column vectors containing the same elements, but in different orders. The elements of the distribution $P[n,k]$ are now all contained in the vector P.

Having done this, we can rewrite (2) as

$$\max\{\overline{\Phi}^H(\boldsymbol{X}_1 - \boldsymbol{X}_2)^H(\boldsymbol{X}_1 - \boldsymbol{X}_2)\overline{\Phi}\} \qquad (4)$$

The expression in brackets is of the form $y^H\boldsymbol{B}y$ which appears commonly in matrix algebra. It is maximized when y is the eigenvector corresponding to the largest eigenvalue of the matrix $\boldsymbol{B}$. Thus, we can calculate the kernel that produces maximal separation by performing an eigenvalue decomposition on $(\boldsymbol{X}_1-\boldsymbol{X}_2)^H(\boldsymbol{X}_1-\boldsymbol{X}_2)$.

Ambiguity Plane

For a signal of length N, the $\boldsymbol{X}$ matrix as given in (3) is of size N^2xN^2. For a signal of any useful length, performing an eigenvalue decomposition on a matrix of such size (even calculating the single eigenvector associated with the largest eigenvalue) is a lengthy process at best. Fortunately, we can circumvent this computation completely while obtaining the decomposition exactly by merely viewing the problem from the ambiguity plane, the two-dimensional Fourier transform of the time-frequency plane.

In the time-frequency plane, a TFD is computed from a kernel via a 2-D circular convolution, therefore the same TFD can be computed in the ambiguity plane via an element by element multiplication of the 2-D Fourier transforms of the matrices involved. Transforming (3) to the ambiguity plane then, we have:

$$A = \boldsymbol{Y}\Psi$$

A may be obtained from P by rewriting P as a matrix, taking the 2-D discrete Fourier transform and then revectorizing the result. The same method transforms Φ into Ψ, and the first column of $\boldsymbol{X}$ in (2) into the diagonal of $\boldsymbol{Y}$. $\boldsymbol{Y}$ is a strictly diagonal matrix, whose diagonal contains the elements of the 2-D discrete Fourier transform of the Rihaczek distribution of the signal. (4) can now be written in the ambiguity domain as:

$$\max\{\Psi^H (\boldsymbol{Y}_1 - \boldsymbol{Y}_2)^H (\boldsymbol{Y}_1 - \boldsymbol{Y}_2)\Psi\} \qquad (5)$$

As $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are diagonal matrices, the eigenvalue decomposition is trivial in this domain.

Looking at the kernel design in the ambiguity plane can also give us insight into what is actually being done. The kernel accentuates regions of maximum absolute difference (in the ambiguity function) of the Rihaczek distributions of the signals.

4. EXAMPLE

Two 128-point discrete signals were used to develop an optimally discriminating kernel. Figure 1 shows diagrammatically the two signals used. The first is a real chirp running from a normalized frequency of 0.2 to 0.8 Nyquist. The other signal, again a real chirp, runs from 0.3 to 0.7.

Figure 2 shows the magnitude of the difference of the Rihaczek distributions of these two signals in the ambiguity plane. The points of maximum difference are circled. The kernel, also plotted in the ambiguity plane and appearing in the lower half of the figure, displays large values precisely at the points where the differences are the greatest. It should be noted here that there are in fact eight equal and maximum eigenvalues for this example. Because all linear combinations of eigenvectors corresponding to the same eigenvalue are themselves eigenvectors, we're free to choose any weighting of these eight points for our kernel. All satisfy our optimality criterion.

Figure 3 shows the TFDs of the two signals resulting from the application of our kernel. Note that there is little of the expected time and/or frequency structure. This is because the kernel design criteria make no attempt to include this. The kernel is geared strictly to achieving time-frequency separation. In this regard, we can see that it succeeds; where the TFD for chirp 1 has large amounts of energy, the TFD for chirp 2 has little resulting in minimal overlap and maximal separation.

5. DISCUSSION

Using the concepts of operator theory, we've been able to forge a direct connection between a discrete, finite-length input signal, and it's discrete-time, discrete-frequency TFD. With our approach, we represent each TFD as the 2-D circular convolution of a kernel with the Rihaczek distribution of the signal.

It is important to note that the kernel we obtain for optimal separation maximizes the time-frequency difference *given* the original distribution (the Rihaczek). If the two signal classes have very dissimilar Rihaczek TFDs, then our method will find very little room for improvement.

Can a base distribution other than the Rihaczek be used to form the $\boldsymbol{X}$ matrices in (4)? Yes. Since one TFD can be derived from any other with application of the appropriate transforming kernel, any TFD may serve as an initial, base distribution in our method. The optimal discriminating kernel will vary with the base TFD chosen, however, due to the varying amounts of time-frequency similarity between the signal classes.

Lastly, throughout this paper, we have assumed a single representative signal for each class, but this constraint is not necessary. One way of incorporating multiple examples of each class in (2) is to average all the individual TFDs for class one and class two resulting in a representative P_1 and P_2 as is done in [8]. (This is tantamount to averaging the $\boldsymbol{X}$ matrices in (4).) More effective means of combining multiple examples may be possible.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Boeing Commercial Airplane Group, the Office of Naval Research and Daniel H. Wagner Associates for their financial support of this work, and Siva Bala Narayanan for his input during the early stages of this research.

REFERENCES

[1] H.I. Choi and W.J. Williams, "Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. ASSP-37, no. 6, pp. 862-871, 1989.

[2] Loughlin P., J. Pitton and L. Atlas, "Bilinear Time-Frequency Representations: New Insights and Properties," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 750-767, 1993.

[3] Y. Zhao, L.E. Atlas, R.J. Marks II, "The Use of Cone-shaped Kernels for Generalized Time-frequency Representations of Nonstationary Signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.38, no.7, pp. 1084-91,July 1990.

[4] P. Loughlin, J. Pitton and L. Atlas, "An information-theoretic approach to positive time-frequency distributions," *Proc. ICASSP 92*, vol. V, pp. 125-128, 1992.

[5] J. Pitton, "Applications of Positive Time-Frequency Distributions to Speech Processing," *Trans. on Speech and Audio Processing*, vol.2, no. 4, pp. 544-566, Oct., 1994.

[6] T. Sang, W. Williams and J. O'Neill, "An Algorithm for Positive Time-Frequency Distributions", *Proc. of the Int. Sym. of Time-Frequency and Time-Scale Analysis*, p. 165, June, 1996.

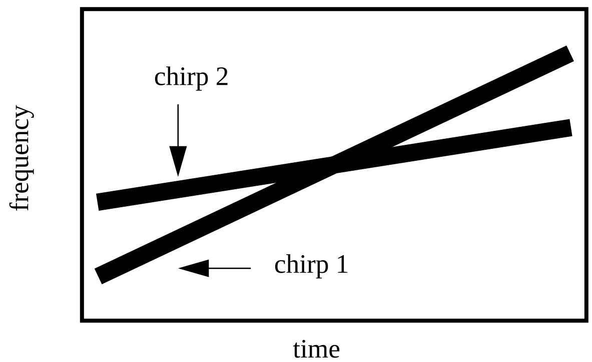

Figure 1: Diagrammatic representation of the two chirps tested.

[7] J. Fonollosa and C. Nikias, "A New Positive Time-Frequency Distribution," *Proc. ICASSP*, vol. 4, pp. 301-304, 1994.

[8] I. Vincent, C. Doncarli and E. Le Carpentier, "Non Stationary Signals Classification Using Time-Frequency Distributions," *Proc. of the Int. Sym. of Time-Frequency and Time-Scale Analysis*, p. 233-236, June, 1996.

[9] P. Flandrin, "A Time-Frequency Formulation of Optimum Detection," *IEEE Transactions on Signal Processing*, vol. 36, no. 9, Sept., 1988.

[10] S. Kay and F. Boudreaux-Bartels, "On the Optimality of the Wigner Distribution for Detection," *Proc. ICASSP 85*, pp. 1017-1020, 1985.

[11] M.G. Amin, G.T. Venkatesan and J.F. Carroll, "A Constrained Weighted Least Squares Approach for Time-Frequency Distribution Kernel Design," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, May, 1996.

[12] Siva Bala Narayanan, Jack McLaughlin, Les Atlas and James Droppo, "An Operator Theory Approach to Discrete Time-Frequency Distributions", *Proc. of the Int. Sym. of Time-Frequency and Time-Scale Analysis*, p. 521, June, 1996.

[13] J. O'Neill and W. Williams, "New Properties for Discrete, Bilinear Time-Frequency Distributions", *Proc. of the Int. Sym. of Time-Frequency and Time-Scale Analysis*, p. 505-508, June, 1996.

[14] L. Cohen, "A General Approach for Obtaining Joint Representations in Signal Analysis. I. Characteristic Function Operator Method." *IEEE Transactions on Signal Processing*, vol.44, no.5. pp. 1080-90, May 1996.

[15] M. Richman, T. Parks and R. Shenoy, "Discrete-Time, Discrete-Frequency Time-Frequency Representations," *Proc. ICASSP 95*, vol. 2, pp. 1029-32, 1995.

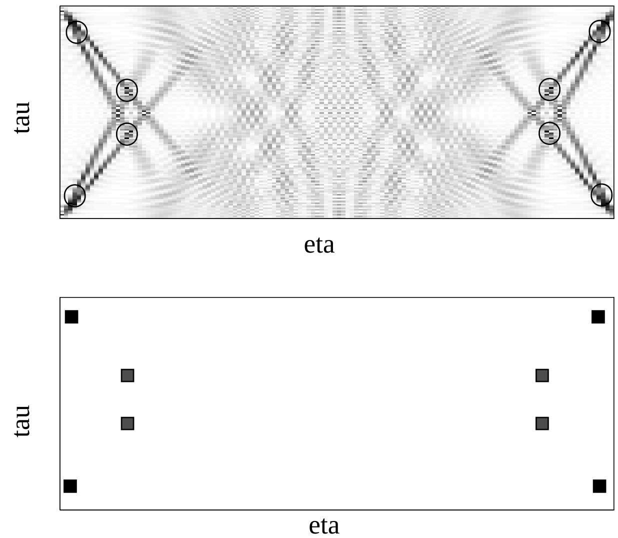

Figure 2: (top) Magnitude of the difference between the Rihaczek distributions of the two chirps, shown in the ambiguity plane. (bottom) The optimal discrimination kernel, also shown in the ambiguity plane. Note that tau and eta are the transforms of frequency and time respectively. Also, points in the kernel have been enlarged for clarity.

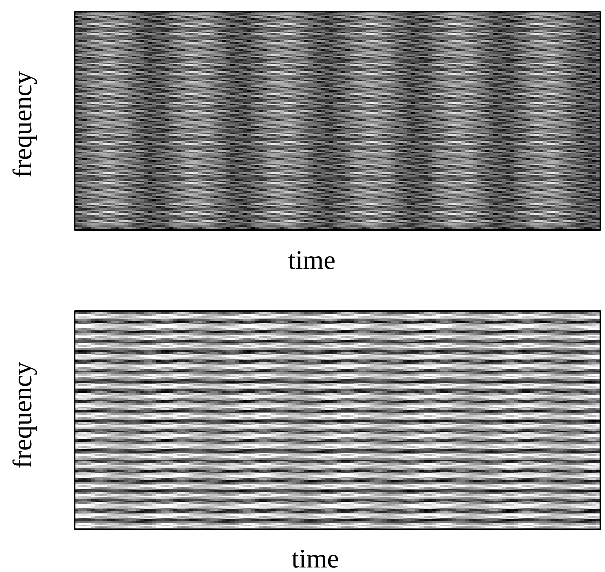

Figure 3: TFD magnitudes produced with the optimal discrimination kernel. (top) Chirp 1. (bottom) Chirp 2.

[16] A.W. Rihaczek, "Signal Energy distribution in Time and Frequency," *IEEE Trans. Info. Th.*, vol. IT-14, no. 3, pp. 369-374, 1968.