Fast Approximations to Positive Time-Frequency Distributions, with Applications

Patrick J. Loughlin[1], James W. Pitton[2] and Blake Hannaford[3]

[1]Dept. of Electrical Engineering
University of Pittsburgh
Pittsburgh, PA 15261

[2]AT&T Bell Laboratories
Speech Research Dept.
Murray Hill, NJ 07974

[3]Dept. of Electrical Engineering
University of Washington
Seattle, WA 98195

Abstract

We present a general approach to approximating positive time-frequency distributions (TFDs) through nonlinear combinations of spectrograms. Closed-form solutions for the combinations are obtained via optimization of entropy functionals subject to an energy constraint. We apply two such combinations to generating approximate TFDs for whale sounds and speech. Through these applications, it can be seen that these methods give results superior to that achieved with individual spectrograms, and remarkably close to the positive TFDs obtained via computationally-intensive methods.

1. Introduction

Joint distributions, particularly time-frequency distributions (TFDs), continue to be an important tool in signal analysis, especially where nonstationary signals are involved (*e.g.*, biomedical signals). Among the many possible distributions, the positive time-frequency distributions of Cohen (with Zaparovanny [5] and Posch [4]) have received renewed interest [9], [12], [13], [14], [15], [16]. Finch and Groblicki have shown that the Cohen-Posch-Zaparovanny formulation includes all positive distributions with the correct marginals [8]. These distributions are particularly appealing from a theoretical viewpoint, as they satisfy the marginal conditions *and* they are everywhere nonnegative, unlike the well-known bilinear distributions [4].

From a practical viewpoint, the utility of positive TFDs has recently been demonstrated in the analysis of speech [15], rotating machines [13], and heart sounds [1]. Methods for constructing positive TFDs have been developed by Loughlin *et al.* [12], [13], and Pitton *et al.* [15], [16]. The methods are similar in that they involve the optimization of an entropy criterion, subject to integral constraints. Fonollosa and Nikias recently proposed additional constraints [9].

A potential limitation in the practical application of these methods for generating positive TFDs is that, despite iterative algorithms, these methods can become computationally prohibitive, especially for large data records (*e.g.*, thousands of data samples). Fast approximations may suffice in many cases. Cunningham and Williams [6] have investigated sums of spectrograms as fast approximations to bilinear TFDs. Frazier and Boashash have considered "*ad hoc*" [10] combinations of smoothed Wigner distributions to obtain nonnegative TFDs. In this paper, we explore some optimal spectrogram combinations for approximating the positive TFDs Cohen, Posch and Zaparovanny [4], [5].

2. Background

All time-frequency distributions can be obtained from Cohen's general formulation [2], [3]:

$$C(t,f) = \iiint s(u+\tau/2)\, s^*(u-\tau/2) \times \phi(\eta,\tau;t,f,s(t))\, e^{j2\pi(\eta(u-t)-f\tau)}\, du\, d\tau\, d\eta \qquad (1)$$

where ϕ is a kernel that determines the particular distribution. As Cohen noted [2], [3], the kernel can be functionally dependent on time, frequency and the signal, as explicitly denoted. The particular case of signal-independent kernels corresponds to the subclass of "bilinear distributions," which have been extensively studied (see [3] and references therein).

The positive, marginal-satisfying TFDs of Cohen, Posch and Zaparovanny [4], [5] can be written in terms of the "bilinear" formulation above, with signal-dependent kernels. A more tractable formulation for the positive TFDs is [3], [4]

$$P(t,f;\Omega) = |s(t)|^2 |S(f)|^2 \Omega(u(t), v(f); s(t)) \qquad (2)$$

where $\Omega(u,v)$ is a nonnegative, unit area function that can be functionally dependent on the signal. As shown by Cohen [3], it is mathematically simple to choose $\Omega(u,v)$ such that the distribution is positive and the marginals are satisfied.

As with the bilinear distributions and the choice of ϕ, there are an infinite number of choices of Ω for a given signal. The known constraints, *i.e.*, positivity, marginals, conditional moments (*e.g.*, instantaneous frequency), are not sufficient to uniquely fix the choice of Ω (or ϕ for that matter). The problem of determining a distribution given conditions that do not uniquely fix the distribution is common in many fields (*e.g.*, chemistry, astronomy). One method of solution is maximum entropy, or more generally, minimum cross-entropy, estimation. This approach has recently been applied to generate positive TFDs for any signal [12],[13],[15],[16]. The method is straightforward in concept and implementation: make an initial educated guess $P_0(t,f) > 0$ at the unknown distribution, and then find the distribution $P(t,f)$ that minimizes the cross-entropy

$$\Delta H(P, P_0) = \iint P(t,f) \log \frac{P(t,f)}{P_0(t,f)}\, dt\, df \qquad (3)$$

subject to the marginal constraints and possibly others ([13],[15]). The solution can be obtained iteratively. If only the marginal constraints are imposed, the iterations are

$$P^{(k+1)}(t,f) = P^{(k)}(t,f) \frac{|S(f)|^2}{P^{(k)}(f)} \qquad (4a)$$

$$P^{(k+2)}(t,f) = P^{(k+1)}(t,f) \frac{|s(t)|^2}{P^{(k+1)}(t)} \qquad (4b)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

where $k = 2i, i \in \{0, 1, 2, ...\}$ and

$$P^{(0)}(t,f) = P_0(t,f) \quad (4c)$$

$$p^{(k+1)}(t) = \int P^{(k+1)}(t,f)\, df \quad (4d)$$

$$p^{(k)}(f) = \int P^{(k)}(t,f)\, dt \quad (4e)$$

and $S(f)$ is the Fourier transform of the signal $s(t)$.

The initial guess, or "prior" $P_0(t,f)$, can be a spectrogram, for example, or a combination of spectrograms where each spectrogram is calculated with a different, unit-energy window [13]. In this paper, we examine different combinations of spectrograms and show that sometimes the combined prior is good enough to get a quick sense of the underlying distribution.

3. Optimal Combinations

We examine three combinations, each of which optimizes an entropy-based measure, subject to an energy normalization constraint. Let $\Pi_1(t,f)$, $\Pi_2(t,f)$ and $\Pi_3(t,f)$ be the functions that are closest to a set of N spectrograms $\{S_i(t,f) : i \in (1, .., N), N \geq 2\}$ in minimum mean cross-entropy, minimax cross-entropy, and minimum mean Itakura-Saito distance, respectively. Furthermore, let the total energy of each $\Pi_i(t,f)$ equal that of the given signal. Mathematically,

$$\Pi_1(t,f) = \underset{P_0(t,f)}{argmin} \frac{1}{N} \sum_i \Delta H(P_0, S_i) \quad (5)$$

$$\Pi_2(t,f) = \underset{P_0(t,f)}{argmin} \max_i \Delta H(P_0, S_i) \quad (6)$$

$$\Pi_3(t,f) = \underset{P_0(t,f)}{argmin} \frac{1}{N} \sum_i d_{IS}(P_0, S_i) \quad (7)$$

subject to the energy constraint

$$\|\Pi_i(t,f)\|_1 = \iint \Pi_i(t,f)\, dt df$$

$$= \int |s(t)|^2 dt = \int |S(f)|^2 df = E, \quad (8)$$

where

$$d_{IS}(P_0, S_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{P_0(t,f)}{S_i(t,f)} - \log \frac{P_0(t,f)}{S_i(t,f)} - 1 \right) dt df \quad (9)$$

is the Itakura-Saito distance measure [11], which has been previously used to combine estimates of the evolutionary spectrum [7]. Solving each of these constrained optimizations via the method of Lagrange multipliers yields [14][1]

$$\Pi_1(t,f) = \frac{E}{\left\| \prod_i \sqrt[N]{S_i(t,f)} \right\|_1} \prod_i \sqrt[N]{S_i(t,f)} \quad (10)$$

$$\Pi_2(t,f) = \frac{E}{\left\| \min_i S_i(t,f) \right\|_1} \min_i S_i(t,f) \quad (11)$$

$$\Pi_3(t,f) = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{S_i(t,f)} + \lambda \right)^{-1} \quad (12)$$

where $\| \ \|_1$ denotes the L_1 norm, and in the last equation, λ is a Lagrange multiplier that must be chosen to ensure that the energy constraint is met (*i.e.*, eq. (8)). Unlike the other two combinations, the Itakura-Saito combination may not be nonnegative, since λ may take on negative values, which could drive the solution below zero. We focus our attention on applying the nonnegative combinations in equations (10) and (11).

4. Examples

Figure 1 illustrates various positive time-frequency representations for a recorded whale sound. Figure 1a shows a positive (Cohen-Posch) density obtained via deconvolution [15]. Close examination of this figure reveals three closely-spaced frequency components in the mid-frequency range. Figures 1b and 1c illustrate wideband and narrowband spectrograms, respectively. The wideband spectrogram shows the temporal detail, while the narrowband spectrogram resolves the three components. Figures 1d and 1e show the combined spectrograms, which are better approximations to the joint density in figure 1a than is either spectrogram; figure 1d is the MMCE combination of the two spectrograms, and figure 1e is the minimax cross-entropy combination. Both approximations simultaneously resolve the temporal structure and all three frequency components.

Figure 2 compares the results of these techniques when applied to speech. Figure 2a illustrates a Cohen-Posch TFD for this signal, obtained via deconvolution. Figures 2b and 2c show wideband and narrowband spectrograms, respectively, of the signal. The wideband spectrogram resolves short-duration events such as plosives (*e.g.*, just after 5.3 seconds), and shows the time variation of the individual formants. The narrowband spectrogram resolves the individual harmonics of the fundamental frequency, but smears out the transient events. The combinations of equations (10) and (11) are shown in figures 2d and 2e. Both methods simultaneously resolve the temporal and spectral structure observed in wideband and narrowband spectrograms, providing representations similar to, though less sharp than, the TFD of figure 2a.

5. Discussion and Conclusion

With the advent of positive TFDs and general methods for constructing them, it is no longer necessary to compromise and accept negative-valued "energy" density functions that satisfy the marginals (*e.g.*, the Wigner distribution), or nonnegative densities that fail to satisfy them (*e.g.*, the spectrogram). The continued development of methods for generating or approximating positive TFDs in a quick and efficient manner opens up new avenues for analysis of nonstationary signals. Fast approximations are of particular benefit in the initial exploratory phases of applications where large amounts of data need to be processed.

1. The Itakura-Saito combination derived in [7] was an unconstrained optimization, normalized after the fact. Taking the energy normalization into account in the optimization itself yields eq. (12), which is readily derived by following the procedure in [14], using the Itakura-Saito distortion measure.

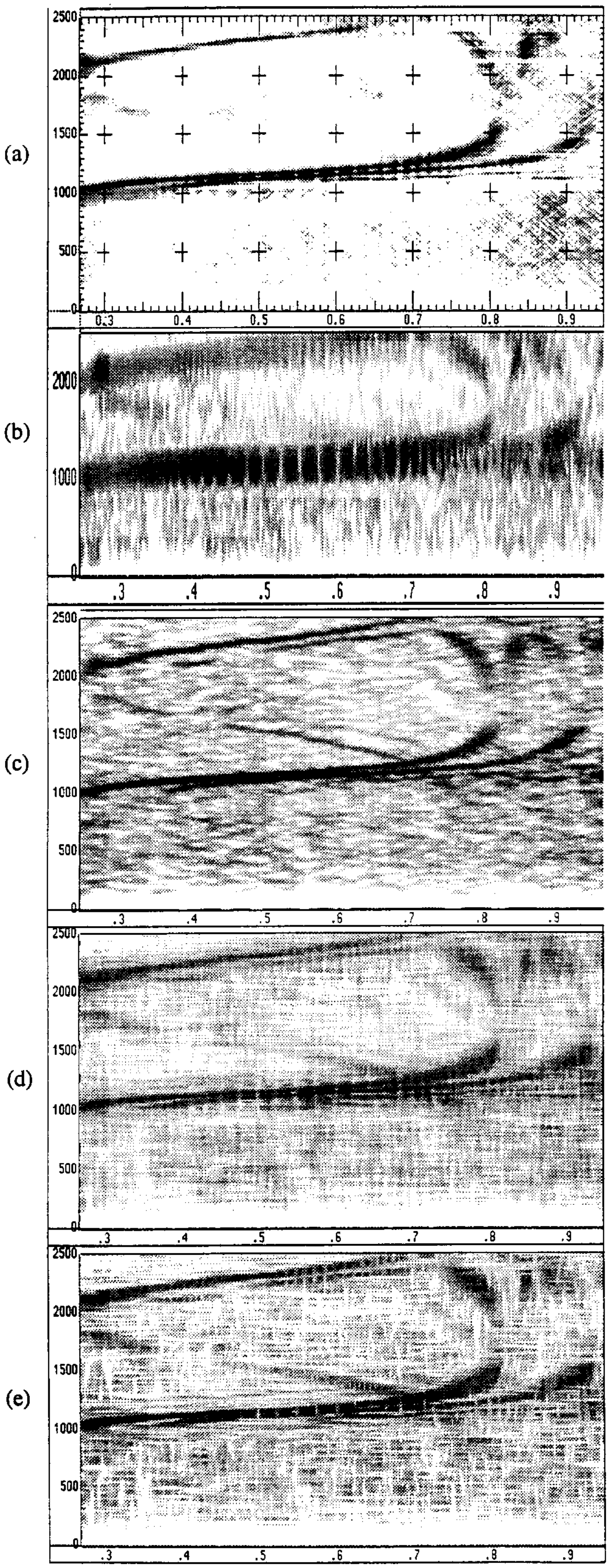

Figure 1: (a) Cohen-Posch TFD of a whale sound; (b) wideband spectrogram; (c) narrowband spectrogram; (d) combination 1 (product); (e) combination 2 (min).

Three methods for optimally combining spectrograms (in an entropy sense) to obtain an improved approximation of the joint time-frequency signal density were presented and explored. Examples demonstrate the effectiveness of the combinations. Regarding computational complexity, an $N \times M$ TFD combining P spectrograms requires approximately $(P\text{-}1)NM$ multiplications (by (10)) or comparisons (by (11)). Computing a Cohen-Posch TFD as per [13] requires at least $2NM$ multiplications *per iteration*; the deconvolution method used here is even more computationally intensive. We have found that on the order of 100 iterations yields acceptable results. The fast combinations presented here greatly reduce the computation time at a cost of resolution *vis-a-vis* the Cohen-Posch TFDs. The results, however, are clearly superior to the original spectrograms.

6. References

[1] J. Bulgrin, B. Rubal, T. Posch and J. Moody, "Comparison of binomial, ZAM and minimum cross-entropy time-frequency distributions of intracardiac heart sounds," *28th Asil. Conf. Sigs. Syst. Comps.*, Pacific Grove, CA, Oct. 31-Nov. 2, 1994.

[2] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, no. 5, pp. 781-786, 1966.

[3] L. Cohen, "Time-frequency distributions — A review," *Proc. IEEE,* vol. 77, no. 7, pp. 941-981, 1989.

[4] L. Cohen and T. Posch, "Positive time-frequency distribution functions," *IEEE Trans. ASSP,* vol. 33., no. 1, pp. 31-38, 1985.

[5] L. Cohen and Y. Zaparovanny, "Positive quantum joint distributions," *J. Math. Phys.*, vol. 21, pp. 794-796, 1980.

[6] G. Cunningham and W. Williams, "Kernel decomposition of time-frequency distributions," *IEEE Trans. Sig. Proc.* vol., 42, no. 6, pp. 1425-1442, 1994.

[7] C. Detka, P. Loughlin and A. El-Jaroudi, "On combining evolutionary spectral estimates," *IEEE Proc. SSAP'94*, to appear.

[8] P. Finch and R. Groblicki, "Bivariate probability densities with given margins," *Found. Phys.*, vol. 14, pp. 549-552, 1984.

[9] J. Fonollosa and C. Nikias, "A new positive time-frequency distribution," *IEEE Proc. ICASSP'94*, pp. IV-301-304, 1994.

[10] G. Frazer and B. Boashash, "Multiple view time-frequency distributions," *Proc. 27th Asilomar Conf. Sigs. Syst. and Comps.*, pp. 513-517, 1993.

[11] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electronics and Communications in Japan,* 53A: 36-43, 1970.

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

[13] P. Loughlin, J. Pitton and L. Atlas, "Construction of positive time-frequency distributions," *IEEE Trans. Sig. Proc.* vol., 42, no. 10, pp. 2697-2705, 1994.

[14] P. Loughlin, J. Pitton and B. Hannaford, "Approximating time-frequency density functions via optimal combinations of spectrograms," *IEEE Sig. Proc. Letters*, Dec. 94.

[15] J. Pitton, L. Atlas and P. Loughlin, "Applications of positive time-frequency distributions to speech processing," *IEEE Trans. Sp. Aud. Proc.*, vol. 2, no. 4, pp. 554-566, 1994.

[16] J. Pitton, P. Loughlin and L. Atlas, "Positive time-frequency distributions via maximum entropy deconvolution of the evolutionary spectrum," *IEEE Proc. ICASSP'93*, vol. IV, pp. 436-439, 1993.

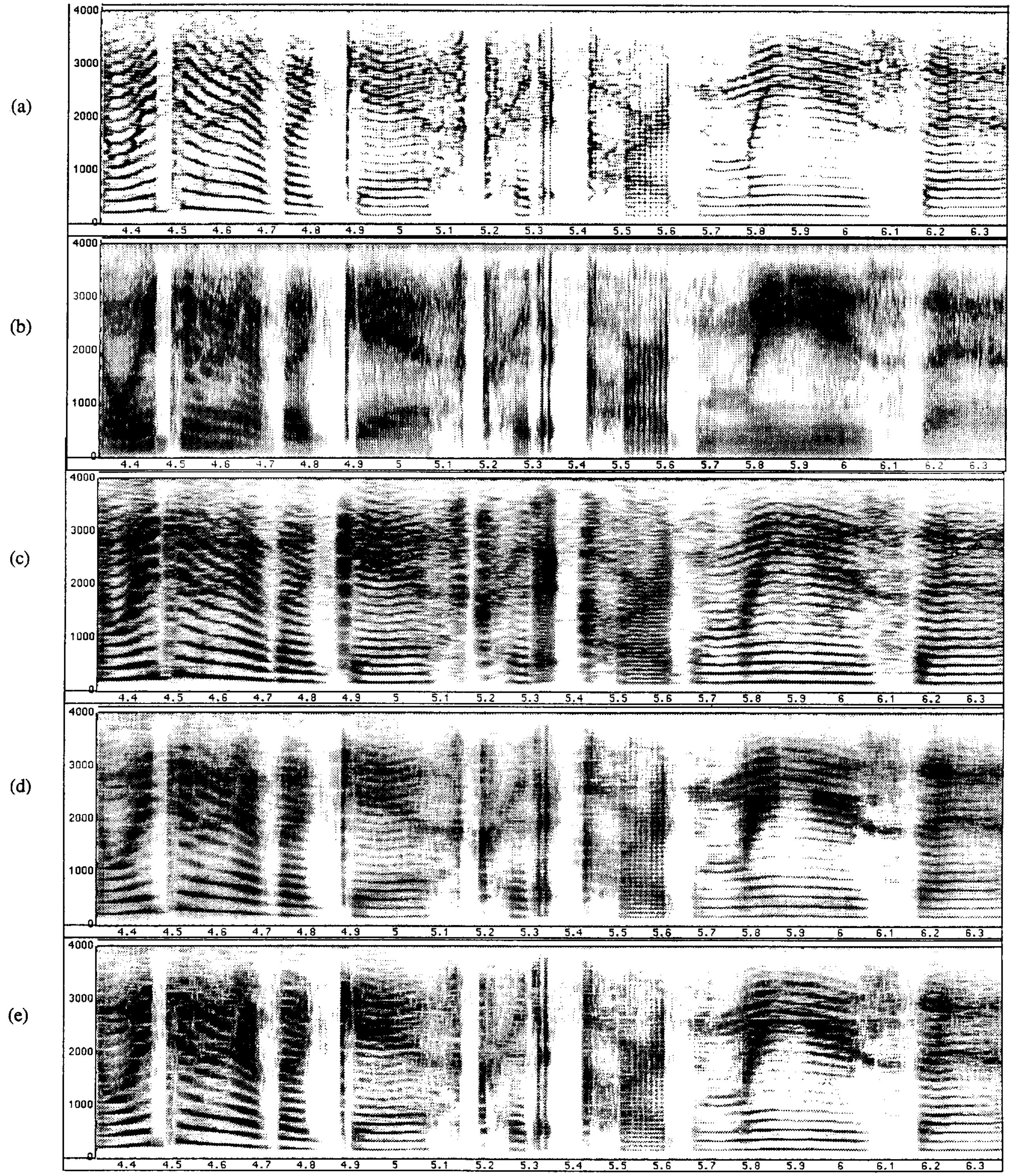

Figure 2: Positive time-frequency representations of a speech utterance, spoken by a female: (a) a positive (Cohen-Posch) TFD, satisfying both marginals; (b) wideband spectrogram; (c) narrowband spectrogram; (e) combination 1 (product); (f) combination 2 (min). The Cohen-Posch TFD and both "combograms" simultaneously resolve fine temporal structure, such as the two stop consonants near 5.3 seconds, and the individual harmonics of the fundamental frequency, unlike either spectrogram.