# A NEW SIGNAL ADAPTIVE APPROACH TO POSITIVE TIME-FREQUENCY DISTRIBUTIONS WITH SUPPRESSED INTERFERENCE TERMS

Robert M. Nickel, Tzu-Hsien Sang and William J. Williams

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

# ABSTRACT

Quadratic time varying-spectral analysis methods<sup>1</sup> that achieve a high resolution jointly in time and frequency generally suffer from interference terms that obscure the true location of the auto components in the resulting time-frequency representation. Unfortunately, as of now, there is no general mathematical model available for an exact distinction between cross-terms and auto-terms. Consequently an attempt to suppress interferences can only rely on a few qualitative properties which are commonly associated with cross terms. Most of the reduced interference distributions that have been developed so far exploit the fact that cross terms tend to oscillate and can hence be suppressed by a properly chosen twodimensional low pass filter (see [1] and [4]). Besides the fact that cross-terms oscillate, they are also known to be responsible for all negative density values of a time-frequency distribution. Non of the currently existing methods addresses this characteristic. In this paper we introduce an entirely new approach that achieves a significant interference reduction by specifically exploiting the negative density structure of cross-terms.

## **1. PRELIMINARIES**

A fundamental tool for time-frequency analysis with high joint time-frequency resolution is the Wigner distribution (WD) (see [3]). Its discrete-time form can be approximated numerically by computing windowed FFTs from a sampled local ACF<sup>2</sup> ( $R_{xx}$ ):

$$W_{xx}(n,k) = \frac{1}{2L+1} \sum_{m=-L}^{L} h(m) R_{xx}(n,m) e^{-j2\pi \frac{mk}{2L+1}}$$
(1)

The window function h(m) is assumed to be symmetric, real, positive definite and normalized to h(0) = 1. Furthermore, we restrict the attention to Wigner distributions that are zero for |n| > M.

Despite many desirable properties, the most predominant drawback of the Wigner distribution can be found in its interferenceterms and its negative energy density values. The key idea to what is currently state-of-the-art in interference-term suppression was developed by Choi and Williams [3]. They realized that since cross-terms tend to oscillate it is possible to suppress interferences with a fixed, two dimensional low-pass filter, which is called the kernel of the distribution. Besides being low-pass, the kernel might also satisfy additional constraints in order to maintain most of the Wigner distributions desirable properties (reduced interference distribution (RID), see [4]). Baraniuk and Jones extended the RID concept from a fixed kernel into a signal adaptive kernel [1].

A mathematical framework that encompasses positive timefrequency distributions (TFDs) that satisfy the marginals was first formulated by Cohen and Posch. This framework, however, does not provide a concrete construction procedure [3]. A successful construction procedure for positive TFDs that satisfy the marginals was introduced by Loughlin, Pitton, and Atlas [6]. Their approach is based on the minimization of the cross-entropy (MCE) between the resulting positive time-frequency distribution and a positive prior distribution, which is usually a spectrogram.

# 2. THE NEW APPROACH

Within this section we introduce a new approach for time-frequency analysis that readily generates distributions that are positive, satisfy the marginals, and have suppressed interference-terms. The key idea is easily explained with an example. Consider two Gaussian logons s(n) and y(n) which each have a non-negative Wigner distribution. Assume that the two logons occur at different times and different frequencies and define x(n) = s(n) + y(n). The Wigner distribution of the signal x(n) is shown in figure 1.



Figure 1: The Wigner distribution  $W_{xx}(n, k)$  of two Gaussian logons. The cross-terms of the distribution and its negative energy density values are clearly visible in the center of the time-frequency plane.

There are 'three' noteworthy components in this representation : the two auto-components in the front and back of the plot and the cross-terms in the center of the time-frequency plane. Besides

This research has been supported in part by the Office of Naval Research, ONR contract no. N00014-97-1-0072

<sup>&</sup>lt;sup>1</sup>Like the well-known Wigner-Ville Distribution for example [3].

<sup>&</sup>lt;sup>2</sup>The ACF is assumed to be computed from an oversampled or sincinterpolated signal in order to avoid aliasing.

tackling these cross-terms by exploiting their oscillating nature we can also attempt to identify them based on the 'negative-parts' of the distribution. It is easy to show that the Wigner distribution of the signal x(n) can be expressed in terms of two auto-Wigner distributions and two cross-Wigner distributions.

$$W_{xx}(n,k) = \underbrace{W_{ss}(n,k) + W_{yy}(n,k)}_{\text{auto-terms}} + \underbrace{W_{sy}(n,k) + W_{ys}(n,k)}_{\text{cross-terms}}$$

Since we know from Cohen and Posch ([3]) that there is at least one distribution that is manifestly positive and satisfies the marginals, we can always rearrange the energy density of the Wigner distribution in such a way that

$$W_{xx}(n,k) = \underbrace{W_{ss}(n,k) + W_{yy}(n,k) + \bar{\chi}_{xx}(n,k)}_{\hat{=}C_{xx}^{P}(n,k)} + \chi_{xx}(n,k)$$

 $C_{xx}^P(n,k)$  is a positive time-frequency distribution that satisfies the marginals.  $\chi_{xx}(n,k)$  is consequently the part of the crossterms that is responsible for the negative energy density values of  $W_{xx}(n,k)$ . The key to the new approach is to construct an estimate for  $\chi_{xx}(n,k)$  based on the negative parts of the distribution. A positive distribution  $C_{xx}^P(n,k)$  with reduced interference-terms can thus be constructed by subtracting the cross-terms estimate from the Wigner distribution.

$$C_{xx}^{P}(n,k) = W_{xx}(n,k) - \chi_{xx}(n,k)$$
 (2)

Note that by definition the rows and columns of  $\chi_{xx}(n, k)$  must sum to zero. Furthermore, we want  $\chi_{xx}(n, k)$  to 'absorbs' the negative parts of  $W_{xx}(n, k)$ , i.e.  $\chi_{xx}(n, k) = W_{xx}(n, k)$  for all (n, k)for which  $W_{xx}(n, k) < 0$ . We can try to find a  $\chi_{xx}(n, k)$  that satisfies these conditions in a minimum norm sense. Moreover, we might want to utilize general 'qualitative' knowledge about the location of the cross-terms by designing a special signal adaptive norm that takes such considerations into account. In case of the proposed algorithm a special signal dependent mask function  $P(n, k) > 0 \forall (n, k)$  serves to incorporate this kind of knowledge. These considerations provide the motivation for the following iterative procedure. The resluting time-frequency distribution will be referred to as an iterated projections distribution (IPD).

### IPD Algorithm

• Initialize the iteration with a TFD that satisfies the marginals<sup>3</sup>:

$$W_0(n,k) = W_{xx}(n,k)$$
 (3)

- Iterate the following steps :
  - 1. Take the negative part of the current iteration

$$W_i^{-}(n,k) = \frac{1}{2} (W_i(n,k) - |W_i(n,k)|)$$
(4)

2. Obtain an estimate  $\chi_i(n, k)$  based on a mask function P(n, k) > 0 by minimizing

$$\sum_{n=-M}^{M} \sum_{k=-L}^{L} \frac{\left|W_{i}^{-}(n,k) - \chi_{i}(n,k)\right|^{2}}{P(n,k)}$$
(5)

 $\sum_{n=-M}^{M} \chi_i(n,k) = 0 \;\forall\; k$ 

subject to

$$\sum_{k=-L}^{L} \chi_i(n,k) = 0 \ \forall \ n \quad (6)$$

<sup>3</sup>With a Wigner distribution for example.

3. Subtract the estimate  $\chi_i(n, k)$  from  $W_i(n, k)$ 

$$W_{i+1}(n,k) = W_i(n,k) - \chi_i(n,k)$$
(7)

It can be shown that the proposed algorithm establishes a sequence of successive projections onto intersecting convex polyhedra and thus converges linearly and pointwise to a function<sup>4</sup>

$$C_{xx}^{P}(n,k) = \lim_{i \to \infty} W_{i}(n,k)$$
(8)

that is non-negative and satisfies the marginals. The fact that the distribution  $C_{xx}^{P}(n, k)$  has suppressed interference-terms follows by construction. The result of the IPD Algorithm, when applied to the initial example of two Gaussian logons, can be seen in figure 2.



Figure 2: The positive iterated projections distribution  $C_{zz}^{P}(n,k)$  for two Gaussian logons. The distribution is almost free of any artifacts.

#### 2.1. Mask Functions

The mask function P(n, k) in the IPD Algorithm is a flexible tool to incorporate additional 'qualitative' knowledge about the nature and location of cross-terms in the initial time-frequency distribution. If we do not want to use such additional information we may simply use a unity mask :  $P(n, k) = 1 \forall (n, k)$ . It can be shown that the unity mask case corresponds to the least squares approach for the construction of positive time-frequency distributions, which was introduced by Sang, Williams and O'Neill in 1996 (see [7]). The presented algorithm can thus be viewed as a generalization of this concept.

Usually we would be interested in utilizing additional prior information. The way P(n, k) will influence the result of the IPD Algorithm is by assigning different weights to different locations of the error-term  $|W_i^-(n, k) - \chi_i(n, k)|^2$ . Whenever P(n, k) is small we want to have a small deviation between  $W_i^-(n, k)$  and  $\chi_i(n, k)$  and whenever P(n, k) is large we want to encourage a large deviation between  $W_i^-(n, k)$  and  $\chi_i(n, k)$ . This implies that P(n, k) should be large at those index-pairs (n, k) for which  $W_i$ has cross-terms and is <u>not</u> negative. P(n, k) should be small everywhere else. A mask function that proved to be very successful

<sup>&</sup>lt;sup>4</sup>It can be shown that the IPD Algorithm constitutes a special case of the theory presented in [2], theorem 5.7, page 393. Note that we are operating in a finite dimensional Hilbert space.

in capturing these conditions is a combination of a 'distribution' mask  $(P_D)$  with a 'vicinity' mask  $(P_V)$ :

$$P(n,k) = P_D(n,k) \cdot P_V(n,k)$$
(9)

with 
$$P_D(n,k) = |W_{xx}(n,k)| + W_{xx}(n,k) + \epsilon$$
 (10)

$$P_{V}(n,k) = \begin{cases} -\sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} W_{xx}^{-}(n,k) e^{-\frac{(n-p)^{2} + (k-q)^{2}}{2\alpha^{2}}} \\ \forall n, k \text{ s.t. } W_{xx}(n,k) \ge 0 \\ \epsilon & \forall n, k \text{ s.t. } W_{xx}(n,k) < 0 \end{cases}$$
  
where  $W_{xx}^{-}(n,k) = \frac{1}{2} (W_{xx}(n,k) - |W_{xx}(n,k)|)$  (11)

The 'distribution' mask provides information about the location of large positive terms in the distribution. The 'vicinity' mask incorporates 'qualitative' knowledge about the probable location of cross-terms in the distribution. It assumes that cross-terms are usually locally concentrated and thus most likely in the vicinity of strong negative density values. The parameter  $\alpha$  determines how far the positive parts of the cross-terms are expected to be spread away from their corresponding negative parts.

It can be verified by inspection that P(n, k) satisfies the desired properties that were mentioned above. Note, that it is seems necessary to add in a small constant  $\epsilon$ , which ensures that P(n, k)remains strictly larger than zero. This is required for equation (5) to be well defined. However, one can show that the resulting system of equations (see section 3) remains solvable and that the convergence of the algorithm remains guaranteed even if a significant number of elements of P(n, k) is identically equal to zero<sup>5</sup>. Generically we can thus pick  $\epsilon = 0$  for any practical application.

#### 3. PROJECTION COMPUTATION

In this section we describe the solution for the constraint optimization problem that is implied in step 2 of the IPD iterations. The derivation of the following equations has to be omitted due to the brevity of the paper. It is based on a Lagrange multiplier ( $\lambda_n$  and  $\sigma_k$ ) strategy that leads to a linear, positive definite system of equations. We will use the following vector notation :

$$w_{i}(n) = \sum_{k=-L}^{L} W_{i}^{-}(n,k) \qquad w_{f}(k) = \sum_{n=-M}^{M} W_{i}^{-}(n,k)$$
(12)

$$Q_{n} = \sum_{k=-L}^{L} P(n,k) \qquad R_{k}|_{k \neq L} = \sum_{n=-M}^{M} P(n,k) \quad (13)$$

$$\mathbf{t} = \begin{bmatrix} w_{t}(-M) \\ \vdots \\ w_{t}(M) \end{bmatrix} \qquad \lambda = \begin{bmatrix} \lambda_{-M} \\ \vdots \\ \lambda_{M} \end{bmatrix}$$
(14)

$$\mathbf{f} = \begin{bmatrix} w_f(-L) \\ \vdots \end{bmatrix} \qquad \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{-L} \\ \vdots \end{bmatrix}$$
(15)

$$\begin{bmatrix} \cdot \\ w_{f(L-1)} \end{bmatrix} \begin{bmatrix} \cdot \\ \sigma_{L-1} \end{bmatrix}$$

$$\begin{pmatrix} \lceil Q_{-M} \rceil \end{pmatrix} \qquad \begin{pmatrix} \lceil R_{-L} \rceil \end{pmatrix}$$

$$\mathbf{Q} = \operatorname{diag}\left( \begin{bmatrix} \mathbf{Q} - \mathbf{M} \\ \vdots \\ \mathbf{Q}_{\mathbf{M}} \end{bmatrix} \right) \quad \mathbf{R} = \operatorname{diag}\left( \begin{bmatrix} \mathbf{R} - \mathbf{L} \\ \vdots \\ \mathbf{R}_{L-1} \end{bmatrix} \right) \tag{16}$$

<sup>5</sup>The details have to be omitted due to the brevity of the paper.

(diag denotes a diagonal matrix)

$$\mathbf{\Omega} = \begin{bmatrix} P_{(-M,-L)} & \cdots & P_{(-M,L-1)} \\ \vdots & P_{(0,0)} & \vdots \\ P_{(M,-L)} & \cdots & P_{(M,L-1)} \end{bmatrix}$$
(17)

The matrix  $\mathbf{H}$  is a lower triangular matrix that results form the Cholesky decomposition of the following term :

$$(\mathbf{R} - \mathbf{\Omega}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Omega}) = \mathbf{H} \mathbf{H}^{\mathrm{T}}$$
(18)

Using these definitions one can compute the projection result  $\chi_i$ as follows: first compute vector  $\vartheta$  from  $\vartheta = \mathbf{f} - \Omega^T \mathbf{Q}^{-1} \mathbf{t}$ , second solve for vector  $\nu$  via back-substitution of  $\mathbf{H} \ \nu = \vartheta$ , now back-substitute for  $\sigma$  in  $\mathbf{H}^T \sigma = \nu$ , then compute  $\lambda$  from  $\lambda = \mathbf{Q}^{-1}(\mathbf{t} - \Omega \sigma)$ , and finally obtain the result from

$$\chi_i(n,k) = W_i^{-}(n,k) - P(n,k)(\lambda_n + \sigma_k) \quad \forall \ k \neq L; n$$
  
$$\chi_i(n,L) = W_i^{-}(n,L) - P(n,L)\lambda_n \qquad \forall \ n$$

For the special case of a unity mask P(n, k) = 1 we obtain the following simple solution :

$$\chi_{i}(n,k) = W_{i}^{-}(n,k) - \frac{1}{K}w_{t}(n) - \frac{1}{N}w_{f}(k) + \frac{\Delta E}{NK} \quad (19)$$
with  $\Delta E = \sum_{n=-M}^{M} w_{t}(n) = \sum_{k=-L}^{L} w_{f}(k)$ 

It is worth mentioning that this solution implicitly establishes a fast computation procedure for the LMS method proposed by Sang, Williams and O'Neill [7].

# 4. EXAMPLES

In this section we demonstrate the performance of the iterated projections distribution (IPD) side by side with other signal adaptive distributions that are either positive or have suppressed crossterms. A commonly used signal for TFD performance demonstrations is a signal with a sinusoidal frequency modulation :

$$g_{sime}(n) = e^{-jm \cdot \sin(\omega_m n) - j\omega_0 n}$$



Figure 3: A signal adaptive distribution of  $s_{sine}(n)$  after Baraniuk and Jones [1].

Figure 3 displays the result for the signal adaptive distribution after Baraniuk and Jones [1]. The kernel volume was chosen to achieve a reasonable tradeoff between cross-term suppression and autoterm resolution. The signal adaptive approach is neither positive, nor does it satisfy the marginals (unless additional constraints are incorporated into the optimization procedure).Note that the axis systems below and aside of the contour plots represent the time and frequency marginal of the distribution (solid line) compared to the true instantaneous/spectral power of the signal (dotted line).



Figure 4: The MCE distribution after Loughlin, Pitton, and Atlas [6] (computed from a spectrogram).

The positive TFD that satisfies the marginals and is closest to the spectrogram, in a minimum cross-entropy sense, is shown in figure 4. It was computed according to the algorithm suggested by Loughlin, Pitton, and Atlas [6]. The joint time-frequency resolution is, as expected, not better than the one of the underlying prior.



Figure 5: The positive IPD of the signal  $s_{sime}(n)$  after 50 iterations.

Figure 5 shows the positive IPD of the signal  $s_{\text{sime}}(n)$ . The crosscomponents are strongly suppressed, and yet the auto-term resolution is still very high: jointly in time and frequency. Moreover, the distribution is positive and it satisfies the marginals.

## 5. CONCLUSIONS

A new approach to the reduction of interferences in discrete quadratic TFDs with a high joint time-frequency resolution was introduced. Instead of smoothing the time-frequency plane with a lowpass filter, the cross-components were estimated in parts based on the location of the negative energy density values of the initial distribution. The subtraction of these cross-term estimates from the initial TFD suggested an iterative procedure. This procedure was shown to be linearly convergent to a manifestly positive representation that satisfies the marginals.

An advantage of the described algorithm is that we can readily incorporate 'qualitative' prior knowledge about the structure of cross-terms and auto-terms into the resulting representation by means of a mask function. Examples for the construction of several mask functions were given.

It was demonstrated with examples that the proposed algorithm can improve upon existing methods for the reduction of crossterms, as well as improve upon methods for the design of positive TFDs that satisfy the marginals. A drawback can be found in the high computational complexity, even though a 'fast' computation method was developed.

Furthermore, it was shown that the approach can be viewed as the generalization of the method proposed by Sang, Williams and O'Neill [7]. Additionally, it is worth to mention that the very recent work on positive TFDs conducted by Loughlin and Emresoy seems closely related [5].

## 6. REFERENCES

- R.G. Baraniuk and D.L. Jones. A signal-dependent timefrequency representation: Optimal kernel design. *IEEE Transactions on Signal Processing*, 41(4):1589-1602, April 1993.
- [2] H.H. Bauschke and J.M. Borwein. On projection algorithms for solving complex feasibility problems. *SIAM Review*, 38(3):367-426, September 1996.
- [3] L. Cohen. *Time-Frequency Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1995.
- [4] J. Jeong and W.J. Williams. Kernel design for reduced interference distributions. *IEEE Transactions on Signal Process*ing, 40(2):402–412, February 1992.
- [5] P.J. Loughlin and M. Emresoy. Weighted least squares implementation of cohen-posch time-frequency distributions. *IEEE Transactions on Signal Processing*, in press.
- [6] P.J. Loughlin, J.W. Pitton, and L.E. Atlas. Construction of positive time-frequency distributions. *IEEE Transactions on Signal Processing*, 42(10):2697-2705, 1994.
- [7] H. Sang, W.J. Williams, and J. O'Neill. An algorithm for positive time-frequency distributions. Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, 37:165-168, June 18-21 1996. Paris, France.