

LINEAR AND QUADRATIC METHODS FOR POSITIVE TIME-FREQUENCY DISTRIBUTIONS

James W. Pitton

MathSoft
1700 Westlake Ave N, Suite 500
Seattle, WA USA 98109
yojimbo@statsci.com

ABSTRACT

This paper presents a new foundation for positive time-frequency distributions (TFDs). Based on an integral equation formulation of nonstationary systems, a positive TFD can be constructed from a decomposition of a signal over an orthonormal basis. This basis function definition of a positive TFD is used to obtain a relationship between the Wigner distribution and the positive TFD. The results are then generalized to derive positive joint distributions over arbitrary variables, following the approach of Baraniuk and Jones [1]. This general theory provides a common foundation for the two approaches of computing time-frequency representations: those based on linear decompositions of the signal (*e.g.*, best basis methods) and those based on a quadratic, or *bilinear*, functional of the signal (*i.e.*, Cohen's bilinear class).

1. INTRODUCTION

There have been two distinct classes of time-frequency analysis. Linear time-frequency analysis involves decomposing the signal over a set of basis functions to obtain a time-frequency representation. This class of methods includes the short-time Fourier transform, wavelets, and the adaptive basis decomposition methods. Linear methods are useful in such applications as signal compression, denoising, and reconstruction. Quadratic methods of time-frequency analysis transform a second-order function of the signal, *i.e.* its time-varying autocorrelation, to obtain a representation of the signal energy distributed over time and frequency. These methods all fall within the framework of Cohen's class of bilinear time-frequency representations¹ [4]. The squared-magnitude of the linear representations also fall within Cohen's class or its extensions.

All of these disparate methods are used to describe the time-varying structure of a signal. However, there is no consistent means of measuring the accuracy of the resulting representation, or comparing different representations. A goal of time-frequency research should be to discover a common foundation for time-frequency analysis that allows comparison of different methods of analysis, and en-

¹While these methods use quadratic functions of the signal to obtain the distributions, the resulting distribution is not necessarily quadratic, *i.e.* when signal-dependent estimation methods are used.

ables cross-fertilization of techniques. Towards this end, this paper presents a new foundation for the theory of positive TFDs, based on eigenfunction decompositions of linear integral equations [7]. It is shown that a positive distribution can be obtained from an eigenfunction decomposition of a signal. Using this definition of the TFD, an integral equation relating the Wigner distribution of the signal to the positive TFD is derived. Thus, TFDs may be obtained from either linear or quadratic signal representations. It is also shown that the approach can be used to obtain representations for positive distributions in other signal domains, *e.g.*, time-scale. Linear and quadratic methods can then be used in these domains as well to compute positive TFDs.

2. POSITIVE TIME-FREQUENCY DISTRIBUTIONS

A positive time-frequency distribution, or TFD, provides a measure of the instantaneous energy of a signal at a particular time and frequency. It is everywhere nonnegative, providing a meaningful estimate of the signal energy, and it yields the correct univariate marginal distributions in time and frequency (the instantaneous energy and the energy spectral density). Thus, it satisfies three fundamental properties of distributions:

$$P(t, \omega) \geq 0, \quad (1)$$

$$\int P(t, \omega) d\omega = |s(t)|^2, \quad (2)$$

$$\int P(t, \omega) dt = |S(\omega)|^2, \quad (3)$$

where $S(\omega)$ denotes the Fourier transform of the finite energy signal $s(t)$, and all integrals are from $-\infty$ to ∞ . TFDs satisfying these three properties are also known as Cohen-Posch TFDs, after the researchers who first showed that such distributions exist for all signals [5].

Positive TFDs satisfying the marginals can be computed using constrained optimization, *i.e.* finding the distribution that minimizes some cost function subject to a set of constraints. The first method for generating positive TFDs was presented in [9, 10]. TFDs were obtained by minimizing the cross-entropy to a prior distribution subject to a set of linear constraints. Spectrograms or combinations

of spectrograms were used as prior distributions, and positivity and the marginals were used as constraints. Positive TFDs have also been computed using maximum entropy estimation, which uses a uniform distribution as the prior. In [13, 14], a convolutional relationship between a spectrogram and a positive TFD was derived, and maximum entropy deconvolution was used to compute the TFD.

Least-squares estimation has also been used to compute positive TFDs. In [17, 8], a positive TFD was computed that minimized the L_2 norm of the distance to a Wigner distribution. In [15], an alternating projections procedure was used to minimize the squared distance to a reduced interference distribution subject to positivity and the marginals. In [12], a general approach was presented for using linearly-constrained quadratic programming to compute positive TFDs minimizing the weighted squared-error to a set of constraints derived from a statistical formulation of positive TFDs.

3. LINEAR DECOMPOSITIONS AND POSITIVE TFDs

A comprehensive theory for time-frequency distributions must address both deterministic and stochastic signals. For the stochastic case, a TFD should satisfy a stochastic form of the marginals:

$$\int P(t, \omega) d\omega = E|s(t)|^2, \quad (4)$$

$$\int P(t, \omega) dt = E|S(\omega)|^2. \quad (5)$$

The TFD may then be derived from a linear time-varying (LTV) filter model. Define $s(t)$ as the output of a white-noise-driven LTV filter:

$$s(t) = \int h(t, \tau) e(\tau) d\tau = \int h(t, t - \tau) e(t - \tau) d\tau. \quad (6)$$

Using a definition originally due to Kailath, the Fourier transform of $h(t, \tau)$ is defined as:

$$H(t, \omega) \stackrel{\text{def}}{=} \int h(t, t - \tau) e^{-j\omega\tau} d\tau. \quad (7)$$

This formulation is purely for notational convenience, letting $H(t, \omega)$ be interpreted as a time-varying weight of a complex exponential input, consistent with linear time-invariant filters. Next, let $e(t)$ be bandlimited Gaussian white noise with bandwidth much greater than that of the filter $h(t, \tau)$:

$$e(t) = \int e^{j\omega t} dZ(\omega), \quad (8)$$

where $dZ(\omega)$ is an orthogonal process with unit variance, gives

$$s(t) = \int H(t, \omega) e^{j\omega t} dZ(\omega). \quad (9)$$

Consistent with stationary spectral estimation, the time-varying spectral density² is defined by:

$$P(t, \omega) = |H(t, \omega)|^2. \quad (10)$$

²This form subsumes Priestley's evolutionary spectrum if $H(t, \omega)$ is restricted to be slowly varying.

Equation (6) models $s(t)$ as an integral transform of $e(t)$. $s(t)$ may equivalently be modeled as the solution to a nonhomogeneous linear differential equation with time-varying coefficients. Every linear differential equation has a corresponding integral equation specified by a unique kernel $h(t, \tau)$ [7]. Thus, any nonstationary signal which is the solution to a time-varying linear system may be uniquely represented by the kernel of its integral equation.

For any linear system, there exists a corresponding adjoint system whose kernel $h(t, \tau)$ is symmetric [7]. Thus, for any signal $s(t)$, there exists a corresponding symmetric kernel. According to Mercer's Theorem, a symmetric kernel can be expanded in a series

$$h(t, \tau) = \sum_i \frac{1}{\lambda_i} \phi_i(t) \phi_i^*(\tau), \quad (11)$$

where $\{\phi_i\}$ are the eigenfunctions and $\{\lambda_i\}$ are the eigenvalues of the expansion. Using (7) gives:

$$H(t, \omega) = \sum_i \frac{1}{\lambda_i} \phi_i(t) \Phi_i^*(\omega) e^{-j\omega t}, \quad (12)$$

where $\Phi_i(\omega)$ is the Fourier transform of $\phi_i(t)$. A positive TFD derived from the filter transform is then given by:

$$P(t, \omega) = |H(t, \omega)|^2 = \left| \sum_i \frac{1}{\lambda_i} \phi_i(t) \Phi_i^*(\omega) \right|^2. \quad (13)$$

This expansion satisfies the stochastic marginals defined above. To see this, expand $h(t, \tau)$ in (6):

$$s(t) = \sum_i \frac{1}{\lambda_i} \phi_i(t) \int \phi_i^*(\tau) e(\tau) d\tau = \sum_i \frac{1}{\lambda_i} \phi_i(t) \chi_i. \quad (14)$$

χ_i is a zero-mean Gaussian random variable with unit variance. The variance of $s(t)$ is given by

$$E|s(t)|^2 = \sum_i \left| \frac{1}{\lambda_i} \phi_i(t) \right|^2. \quad (15)$$

The spectrum of $s(t)$ is given by

$$E|S(\omega)|^2 = \sum_i \left| \frac{1}{\lambda_i} \Phi_i(\omega) \right|^2. \quad (16)$$

It is readily shown that $P(t, \omega)$ satisfies these marginals. Thus, $P(t, \omega)$ is a valid TFD for the stochastic signal $s(t)$.

When $s(t)$ is a deterministic signal, the kernel formulation can still be used. $s(t)$ is now a solution to a homogeneous integral equation:

$$\phi(t) = \int h(t, \tau) \phi(\tau) d\tau. \quad (17)$$

The kernel of the integral equation is a consistent representation for both deterministic and stochastic signals, corresponding to homogeneous and nonhomogeneous systems, respectively. As before, there exists an adjoint system with an associated symmetric kernel, described by the series expansion in (11). The signal is then given by a sum of the eigenfunctions of the kernel $h(t, \tau)$:

$$s(t) = \sum_i \frac{1}{\lambda_i} \phi_i(t), \quad (18)$$

where the eigenvalues now form the coefficients of the expansion. Thus, any nonstationary signal which is the solution to a time-varying linear system may be uniquely represented by the eigenfunctions of the kernel of the defining integral equation. Methods such as *best basis* [6], *basis pursuit* [2], and *matching pursuit* [11] can be used to obtain a decomposition of the signal.

Given a decomposition of the signal as in (18), a positive TFD is can be defined by the same expansion used for stochastic signals:

$$P(t, \omega) = \left| \sum_i \frac{1}{\lambda_i} \phi_i(t) \Phi_i^*(\omega) \right|^2. \quad (19)$$

This distribution does not precisely satisfy the deterministic time and frequency marginals. The marginals obtained from this distribution are:

$$\int P(t, \omega) d\omega = \sum_i \left| \frac{1}{\lambda_i} \phi_i(t) \right|^2. \quad (20)$$

$$\int P(t, \omega) dt = \sum_i \left| \frac{1}{\lambda_i} \Phi_i(\omega) \right|^2. \quad (21)$$

These marginals do not exhibit the interaction between the components of the signal, as observed in the true marginals. Rather, they are identical to those obtained in the stochastic case. As such, the TFD is a distribution of the energy of the individual signal components in time-frequency, not the time-frequency energy of the signal itself.

4. QUADRATIC METHODS FOR POSITIVE TFDs

The basis decomposition approaches are essentially non-parametric spline approximations of the signal or TFD. The difficulty in a basis function approach to signal representations is finding a good basis for a given signal. The decomposition methods mentioned above maintain a library of candidate basis functions from which a unique basis is selected according to some optimality criterion. However, the library of functions does not always contain an adequate representation of the signal. This is particularly true with complicated signals such as speech, or when the signal is noisy or stochastic and subject to random phase variations. Nonetheless, the candidate basis function approach can be very useful, particularly when the decomposition can be concisely parameterized.

Define a prototype basis function:

$$\phi_i(t) = \frac{1}{\sqrt{E_i}} e^{-\frac{(t-t_i)^2}{2\sigma_i^2}} e^{j(\omega_i t + \beta_i t^2/2)} \text{rect} \frac{t - \bar{t}_i}{T_i}. \quad (22)$$

$\phi_i(t)$ is parameterized by six variables: t_i , \bar{t}_i , T_i , σ_i , ω_i , and β_i . A seventh variable, E_i , normalizes the energy of the eigenfunction. This family of functions offers much more flexibility than the traditional Gabor basis of truncated Gaussians. The *rect* function allows any region of the Gaussian pulse to be selected, providing a wider range of envelopes for ϕ_i . Appropriate choice of parameters will approximate *rect* functions, ramps, and decaying exponentials. The resulting basis will not form an orthonormal set;

however, such a basis is readily obtained, *e.g.* via Gram-Schmidt. Note that the complex exponentials e^{st} also do not form an orthonormal basis: once a linearly-independent subset has been determined for a given linear system, an orthogonal basis can be obtained.

To relate the linear and quadratic approaches to time-frequency analysis, the quadratic representations must be described in terms of the linear decompositions. The Wigner distribution of a signal $s(t)$ is:

$$W(t, \omega) = \int s(t + \tau/2) s^*(t - \tau/2) e^{-j\omega\tau}. \quad (23)$$

$s(t)$ can be decomposed over the space defined by ϕ_i :

$$s(t) = \sum_i \frac{\lambda_i}{\lambda_i} \phi_i(t). \quad (24)$$

The stochastic formulation of the signal is used to obtain its second-order statistical properties.

Replacing $\phi(t)$ in (24) by its form in (22), and plugging into (23), it can be shown that the Wigner distribution of the signal is given by:

$$E[W(t, \omega)] = \int H(t, \omega_1) H^*(t, \omega_2) \frac{\sin[(2\omega - \omega_1 - \omega_2)T_{t,\omega}]}{(2\omega - \omega_1 - \omega_2)} d\omega_1 d\omega_2. \quad (25)$$

where $H(t, \omega)$ is given by (12), and $T_{t,\omega}$ is related to T_i and \bar{t}_i in the expansion. $T_{t,\omega}$ parameterizes a signal-dependent smoothing window, akin to Dirichlet's kernel. $T_{t,\omega}$ must be estimated from the data, suggesting an iterative procedure for solving (25). Such methods are currently being investigated. Note that no restrictions on the smoothness of ϕ , such as assumptions of "local stationarity" or "slowly-varying processes," have been made: $\phi(t)$ can be frequency-modulated and discontinuous at its endpoints. (25) is solved for $H(t, \omega)$, which is then used to obtain $P(t, \omega)$, if desired. The importance of this form for $P(t, \omega)$ is that no explicit decomposition of the signal needs to be performed; no library of basis functions needs to be maintained.

5. JOINT POSITIVE DISTRIBUTIONS IN ARBITRARY VARIABLES

In section 3, a positive TFD was obtained by computing the Fourier transform of the time-varying filter $h(t, \tau)$. The filter may likewise be transformed to yield other joint distributions. For example, a positive time-scale distribution may be obtained by computing a scale transform [3] of the decomposition:

$$P(t, c) = \left| \sum_i \frac{1}{\lambda_i} \phi_i(t) \Phi_i^*(c) \right|^2, \quad (26)$$

where $\Phi_i(c) = \frac{1}{\sqrt{2\pi}} \int \phi_i(t) \frac{e^{-jc \ln t}}{\sqrt{t}} dt.$

Positive time-scale distributions have also been computed using cross-entropy minimization [16].

This approach may be extended to obtain positive distributions in arbitrary variables. Following the notation of [4], given two variables a and b with associated Hermitian operators \mathcal{A} and \mathcal{B} , a positive distribution in (a, b) is given by

$$P(a, b) = \left| \sum_i \frac{1}{\lambda_i} F_i(a) F_i^*(b) \right|^2, \quad (27)$$

where $F_i(a)$ and $F_i(b)$ are the \mathcal{A} and \mathcal{B} transforms, respectively, of $\phi_i(t)$. This distribution satisfies the a and b marginals of the signal, $|F(a)|^2$ and $|F(b)|^2$.

The linear and quadratic methods can be treated under a common framework again by considering the case when a and b are related to time and frequency by a unitary transform \mathcal{U} :

$$A = \mathcal{U}^\dagger T \mathcal{U}, \text{ and } B = \mathcal{U}^\dagger W \mathcal{U}. \quad (28)$$

In this case, a and b are said to be “unitarily equivalent” to t and f . Unitary equivalence of time–frequency representations was first addressed in [1]. Using the authors’ approach, a \mathcal{U} -Wigner distribution is obtained by computing the Wigner distribution of the unitary-transformed signal $\mathcal{U}s$. A positive distribution $P(a, b)$ is obtained by solving (25) in the (a, b) domain. The (a, b) distribution may be mapped back to the time–frequency plane by using the method described in [1].

6. DISCUSSION

A commonly heard criticism of time–frequency analysis is that the problem is inherently ill-posed, as N points of data can not be used to estimate N^2 values in time–frequency. A basis function definition for time–frequency distributions refutes this criticism. The TFD is constructed from a finite set of eigenvalues and eigenvectors, which are mapped to the time–frequency plane by (13). The number of eigenvalues will be less than or equal to N . In any case, this criticism was never well-founded. Computing a TFD does not entail estimating N^2 independent points of data, no more than constructing a Gaussian probability density function requires estimating an infinite number of values of the function. Both functions are distributions, dependent on a finite set of parameters which must be estimated from the available data. Equation (13) is one method of constructing the distribution from that finite set.

Acknowledgements

The author is grateful to Dr. Don Percival for many helpful discussions on these and related topics.

7. REFERENCES

- [1] R. Baraniuk and D. Jones. Unitary equivalence: a new twist on signal processing. *IEEE Transactions on Signal Processing*, 43(1), 1995.
- [2] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. Technical report, Stanford University, 1996.
- [3] L. Cohen. The scalè representation. *IEEE Transactions on Acoustics Speech and Signal Processing*, 41(12), 1993.
- [4] L. Cohen. *Time-Frequency Analysis*. Prentice-Hall, 1995.
- [5] L. Cohen and T. Posch. Positive time-frequency distributions. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33(1):31–38, 1985.
- [6] R. Coifman and V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, 1992.
- [7] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. John Wiley and Sons, Inc., N.Y., 1953.
- [8] S. Kikkawa and H. Tohyama. A least squares method for generating positive time-frequency distributions. Unpublished manuscript.
- [9] P. Loughlin, J. Pitton, and L. Atlas. An information-theoretic approach to positive time-frequency distributions. In *ICASSP-92*, volume V, pages 125–128, 1992.
- [10] P. Loughlin, J. Pitton, and L. Atlas. Construction of positive time-frequency distributions. *IEEE Transactions on Signal Processing*, 42(10):2697–2705, 1994.
- [11] S. Mallat and Z. Zhang. Matching pursuits with time frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12), December 1993.
- [12] J. Pitton. Positive time-frequency distributions via quadratic programming. *Multidimensional Systems and Signal Processing*, 1996. In review.
- [13] J. Pitton, L. Atlas, and P. Loughlin. Applications of positive time-frequency distributions to speech processing. *IEEE Trans. on Speech and Audio Proc.*, 2(4):554–566, 1994.
- [14] J. Pitton, P. Loughlin, and L. Atlas. Positive time-frequency distributions via maximum entropy deconvolution of the evolutionary spectrum. In *ICASSP-93*, volume IV, pages 436–439, 1993.
- [15] T. Sang, W. Williams, and J. O’Neill. An algorithm for positive time-frequency distributions. In *Proc. IEEE-SP Symp. Time-Freq. / Time-Scale Analysis*, pages 165–168, 1996.
- [16] B. Tacer and P. Loughlin. Time-scale energy density functions. *IEEE Transactions on Signal Processing*, 44(5):1310–1314, 1996.
- [17] H. Tohyama, S. Kikkawa, and K. Ohara. On an optimum positive time-frequency distribution by the least squares method (in Japanese). *The Transactions of the Institute of Electronics, Information and Communication Engineers*, J75-A1(3):661–663, 1992.