REGIONALLY OPTIMISED KERNELS FOR TIME-FREQUENCY DISTRIBUTIONS

Mark J. Coates, Christophe Molina and William J. Fitzgerald

Signal Processing and Communications Laboratory,
Cambridge University Engineering Department,
Trumpington Street, Cambridge CB2 1PZ, U.K.
e-mail mjc40@eng.cam.ac.uk

ABSTRACT

Ideally, kernels used to generate bilinear time-frequency distributions (TFD) should be signal-dependent, and optimised independently at every location in the time-frequency (TF) plane. This poses an extremely severe computational burden. A compromise is proposed in this paper: time-varying kernels are optimised for specific regions in the time-frequency plane. The regions, designed to isolate separate components comprising the signal, are determined by modelling the TFD using a finite mixture model of Gaussian distributions. The parameters of the model are estimated using a combination of the expectation-maximisation algorithm and functional merging. The regional optimisation provides improved separation and resolution of closely-spaced components when compared to methods using a solely time-varying kernel, without incurring an overwhelming computational expense.

1. INTRODUCTION

The time-frequency distributions (TFD) of Cohen's bilinear class [2] are generated by smoothing the Wigner-Ville distribution using a kernel designed to reduce cross-components. The choice of kernel is critical to the appearance and quality of the TFD. A fixed kernel cannot achieve a good representation, as defined by minimal smearing of auto-components and strong suppression of cross-component interference, for every type of signal encountered. To achieve satisfactory performance, the kernel must be dependent on the analysed signal. Recognising this, Baraniuk and Jones [4] developed a signal-dependent kernel optimisation method for a time-varying kernel, which performs well for signals comprised of time-separable components, but suffers when components with different time-frequency behaviour overlap in time.

Jones and Parks proposed an adaptive TFD [5], which maximised a measure of time-frequency concentration at each location in the plane by adapting the parameters of the windows inherent in the spectrogram. The method produces impressive time-frequency distributions, but the computational cost is extremely high. The TFD developed in [6] utilises a separable smoothing kernel comprised of two rectangular functions, whose shape is adapted at each location to reduce interference. The resulting distributions are good, but the use of a separable rectangular kernel is restrictive, leading to smearing of for some signals. The cost of optimising the kernel at each individual location is again high.

If the time-frequency components comprising the signal could be isolated and extracted, a time-varying kernel could be regionally optimised for each extracted component. This is the method proposed in this paper. Commencing with an initial approximate TFD, an iterative process is applied to identify and isolate the components comprising the signal. The number of components in the distribution and their locations and shapes are determined by modelling the TFD using a finite mixture model of Gaussian distributions. The number of Gaussians in the mixture model and the associated parameters are determined through a combination of the EM algorithm and the functional merging technique [7]. The model is used to define the time-frequency regions occupied by the components using a Bayesian classification approach. Non-separable filters are designed using the method in [3] to extract the components from the signal. The optimisation method proposed in [4], which poses a relatively minor computational burden, is used to procure the time-varying kernel for each isolated component.

The following section of the paper defines the time frequency distributions produced and reviews the kernel optimisation problem formulated in [1]. The third section contains an outline of the algorithm used to generate the TFD. The fourth section of the paper discusses the method used to generate the finite mixture model. In the fifth section, a comparison is made between the TFDs obtained using the algorithm presented in this paper and those obtained using the time-varying kernel of [4]. Conclusions follow in the final section.

2. DEFINITIONS AND PROBLEM STATEMENT

A time-frequency distribution which belongs to Cohen's bilinear class can be expressed as the Fourier transform of the

0-7803-4428-6/98 $10.00 © 1998 IEEE

product of the ambiguity function (AF) of the signal, $x(t)$, and a smoothing filter:

$$P(t,\omega) = \frac{1}{4\pi^2}\int\int A(\theta,\tau)\Phi(\theta,\tau)e^{j\theta t - j\tau\omega}d\theta d\tau \quad (1)$$

where $A(\theta,\tau)$ is the symmetrical AF, defined as:

$$A(\theta,\tau) \equiv \int x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)e^{j\theta t}dt \quad (2)$$

The TFDs generated by the algorithm proposed in this paper are members of an extension of the bilinear class. Both the smoothing kernel and the ambiguity function are regionally localized by dividing the time-frequency plane into $n+1$ regions. The first n regions correspond to components identified in the signal; the final region covers the remainder of the plane. The regionally localized ambiguity function, for $(t,\omega) \in R_k$, obtained by performing the time-frequency filtering process of [3], is given by:

$$A(k;\theta,\tau) \equiv \int_{t\in R_k} x_k\left(t+\frac{\tau}{2}\right)x_k^*\left(t-\frac{\tau}{2}\right)e^{j\theta t}dt \quad (3)$$

where

$$x_k(t) = (\mathbf{S_k}x)(t) = \sum_{v=1}^{N_k}\langle x, s_v\rangle s_v(t). \quad (4)$$

Here $x_k(t)$ is the orthogonal projection of $x(t) \in \mathcal{L}_2(R)$ on $\mathcal{S}_k$, a linear signal space covering R_k energetically and with little energy outside the region. $\mathbf{S}_k$ is an orthogonal projection operator onto $\mathcal{S}_k$, and $\{s_v(t)\}_{v=1}^{N_k}$ is an orthonormal basis of $\mathcal{S}_k$, where N_k is the dimension of the space.

The kernels developed within the algorithm have Gaussian radial cross-section:

$$\Phi(\theta,\tau) = \exp\left(-\frac{\theta^2+\tau^2}{2\sigma^2(\psi)}\right) \quad (5)$$

where $r = \sqrt{\theta^2+\tau^2}$. The spread function, σ, controls the spread of the Gaussian at radial angle $\psi = \arctan(\tau/\theta)$. Such a kernel provides more flexibility in shape than a conventional Gaussian, whilst maintaining sufficient computational ease. The kernels are optimal in the sense that they are the solution to the optimisation problem of [1], which proposes constraints and a performance index designed to suppress cross-components whilst passing auto-components with as little distortion as possible. The problem is formulated in the ambiguity plane, exploiting its property of separation of auto- and cross-components, and can be stated for the radially Gaussian kernel as [4]:

$$\Phi_{opt} = \arg\max_{\Phi}\int_0^{2\pi}\int_0^{\infty}|A(r,\psi)\Phi(r,\psi)|^2\, r dr d\theta \quad (6)$$

subject to:

$$\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{\infty}|\Phi(r,\psi)|^2\, r dr d\theta \le \alpha \quad (7)$$

where α is a parameter controlling the kernel volume. The optimisation is performed at every time location using a region and time dependent ambiguity function $A(k,t;\theta,\tau)$.

The regional optimisation algorithm decomposes the TF plane into a number of regions, each representing a component of the signal. The portion of the signal within that region can then be extracted, and optimisation of a local kernel performed. The algorithm commences with an initial estimate of the distribution, and iteratively refines the distribution via the regional optimisation process until the number of components detected in the signal stabilises.

3. OUTLINE OF ALGORITHM

The first task performed by the algorithm is the detection and isolation of the major components of the signal. An initial TFD is required to determine a reasonably accurate estimate of the number of components and their approximate locations and shapes. It is important that cross-components be suppressed as much as possible in the initial distribution to prevent the masking of genuine components and the false inclusion of cross-components in the model generation. Auto-component smearing in the initial representation is not important, because after component identificaton they will be re-processed, and the smearing removed. The TFD generated using the method of [4], with a very small volume parameter, is sufficiently accurate, and is quickly obtained.

The number of distinct components in the distribution must be estimated, and the locations and shapes of the components approximated. A finite mixture model is developed to approximate the energy distribution. The mixture model is a weighted sum of Gaussian distributions, with the shape of each Gaussian governed by a full covariance matrix. The manner in which the parameters of the finite mixture model are estimated is discussed in the following section. At this stage, an over-estimation of the number of components is preferable to an under-estimation. An over-estimation can occur when cross-components are included in the model or when over-fitting occurs. In the first instance, the fault will be removed in the next iteration, because once the TF region corresponding to a cross-component is isolated, the apparent energy in the region will disappear. The clarification of the distribution encourages merging of over-fitted components.

The finite mixture model is used to divide the TF plane into regions corresponding to the components. The division is accomplished according to a Bayesian classification approach, with equal priors associated with each component. Each location in the TF plane is assigned to a component

based on the MAP probability, with a threshold on the minimum inclusion probability. The choice of threshold does not affect the performance of the algorithm, provided it is small enough such that all locations significantly contributing to a distribution's energy are included.

The regions generated by the classification procedure act as indicator functions. Based on these indicator functions, non-separable time-frequency filters are designed as in [3]. The generated filters are applied to the signal to isolate the components in the model. The time-dependent kernel acting within each component's region is optimised, and applied within the indicated region to update the TFD.

This process enhances the TFD clarity and the concentration of components. Smearing is reduced because kernel volume values can be determined based on component size. In subsequent iterations, the number of components in the model will reduce or stabilise, as the effects of over-fitting and inclusion of cross-components are removed. When the number of components stabilises, the algorithm terminates.

4. THE FINITE MIXTURE MODEL AND FUNCTIONAL MERGING

A key task in the regional optimisation algorithm is the identification of the components and the time-frequency spaces they occupy. The task is approached by adapting a Gaussian mixture model to approximate the number of modes in the underlying multi-modal distribution. The Gaussian mixture is defined as a linear combination of Gaussian distributions:

$$F_N(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i f_i(\mathbf{x}|\mu_i, \phi_i), \tag{8}$$

where N is the number of Gaussians $f_i(\mathbf{x})$, λ_i are the mixing weights, μ_i the means, and ϕ_i the covariance matrices.

The model is initialised by setting N to a value significantly larger than the number of components expected. The value of N does not affect the model produced, but provides an upper bound for the distribution complexity. The Gaussians are initialised with equal covariance matrices. To ensure a good coverage of the initial distribution, the Gaussian means are assigned progressively to the next highest energy level in the distribution which lies at least a specified distance from any previously assigned mean. The algorithm is robust to the specified separation, but the value should be large enough to allow the assignment of means to all components. This process is repeated until N means have been assigned, or no points exist which satisfy the minimum distance constraint. In the latter case, N is reduced to the number of assigned Gaussians. The expectation-maximisation (EM) algorithm is applied to adapt the model parameters to achieve a good approximation of the distribution.

At this stage, the number of modes in the distribution is over-approximated, and a better approximation must be obtained. This is achieved through functional merging [7], a technique based on the principle that Gaussians in the mixture model which are used to model the same mode should be closer together than Gaussians modelling separate modes. The arc-cosine distance is used as the distance measure because it has a closed form expression for Gaussian distributions. The distance is equal to the angle between two pdfs $f(\cdot)$ and $g(\cdot)$, and is defined in a Hilbert space $\mathcal{H}$ as:

$$\Omega = \arccos\left(\frac{\langle f, g\rangle}{\|f\|_2 . \|g\|_2}\right) \tag{9}$$

where $< \cdot, \cdot >$ is the inner product and $\| \cdot \|_2$ is the L_2-norm.

Functional merging operates by repeatedly determining whether the two closest Gaussians in the current model can be merged into a single Gaussian. The merging decision is made by comparing to a specified threshold the arc-cosine distance between the mixture of the two Gaussians and the new single Gaussian. The parameters of the new Gaussian are chosen to minimise this distance:

$$\begin{aligned}
\lambda_{new} &= \sum_{i=1}^{2} \lambda_i \\
\mu_{new} &= \sum_{i=1}^{2} \lambda_i . \mu_i \\
\phi_{new} &= \sum_{i=1}^{2} \lambda_i (\phi_i + (\mu_i - \mu_{new})(\mu_i - \mu_{new})^T).
\end{aligned}$$

The choice of threshold critically affects the number of Gaussians in the mixture model after merging has been performed. Bearing in mind that an over-estimation of components is preferable, an upper limit on the threshold can be determined by considering the maximum separation between components for which merging is acceptable. In the range of thresholds up to the upper limit, the number of Gaussians in the merged model should be stable for the largest range of thresholds when the number of modes has been correctly identified. This reasoning is based on the principle that the range begins when all the intra-mode components have been merged, and ends when extra-mode components begin to be merged. Figure 2 indicates how the choice of threshold is made, displaying how the number of components in the mixture model varies depending on the threshold. The threshold is determined as a member of the largest range of thresholds over which the number of Gaussians is stable (in this case two). The cases of no merging and a single Gaussian are not considered because more than one component is assumed to be present in the TFD, and over-estimation should have occurred initially.

5. RESULTS

Figure 1 compares the time-frequency distributions of a synthetic signal comprised of two components, a chirp and a

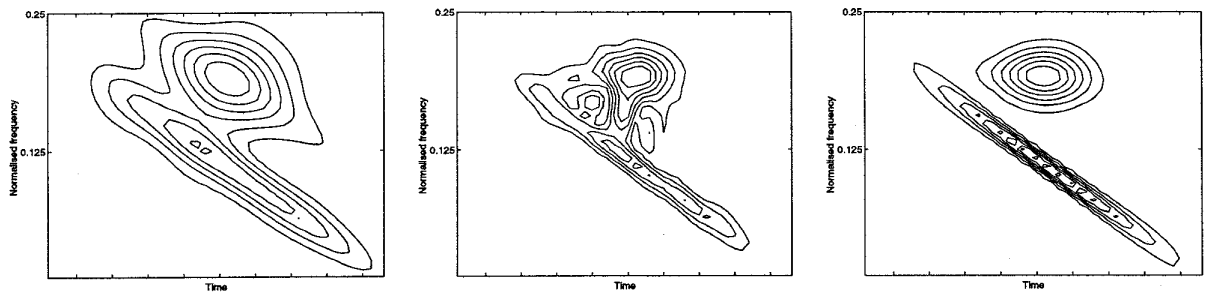

Figure 1: Time-frequency distributions for signal composed of time-overlapping chirp and Gaussian. (a) Adaptive optimal kernel (small kernel $\alpha = 0.75$) (b) Adaptive optimal kernel (large kernel $\alpha = 1.8$) (c) The regionally optimised distribution

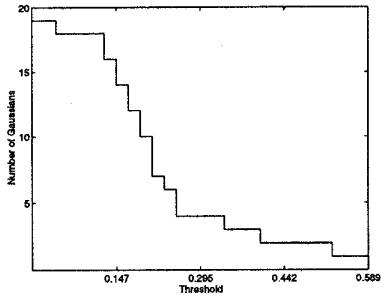

Figure 2: Variation of number of Gaussians in the mixture model generated after functional merging as the threshold is changed for the synthetic signal in Section 5.

Gaussian, which overlap in time and are closely separated. The comparison is made between two distributions generated using the time-adaptive kernel of [4] and that generated using the regional optimisation technique. The time-adaptive kernel does not have the flexibility to change shape for the two components, and the resultant kernel is a compromise between the optimal kernels for each component. As a result, cross-components are observed for larger kernel volumes, and distortion for smaller volumes. Over the entire range of reasonable volumes, distortion, cross-components or a combination are observed. The regional optimisation of the kernel allows different kernels to be applied to the two components, providing clear separation and displaying the true time-frequency behaviour of the signal.

6. CONCLUSION

This paper has introduced a technique for the generation of time-frequency distributions using kernels optimised over specific regions in the time-frequency plane. The technique includes a method of determining the number of components present in the analysed signal, and the time-frequency regions the components occupy. Whenever signal components overlap in time and have different time-frequency behaviour, the resultant distributions provide better resolution and separation of components than distributions generated using purely time-adaptive kernels. The computational expense is less than that of techniques which adapt the kernel at all locations in the time-frequency plane.

7. REFERENCES

[1] R.G. Baraniuk and D.L. Jones. A signal-dependent time-frequency representation: optimal kernel design. *IEEE Trans. Signal Processing*, 41:1589–1602, 1993.

[2] L. Cohen. Time-frequency distributions - a review. *Proc. IEEE*, 77:941–981, 1989.

[3] F. Hlawatsch and W. Kozek. Time-frequency projection filters and time-frequency signal expansions. *IEEE Trans. Signal Processing*, 42:3321–3334, 1994.

[4] D.L Jones and R.G. Baraniuk. Adaptive optimal-kernel time-frequency representation. *IEEE Trans. Signal Processing*, 43:2361–2371, 1995.

[5] D.L Jones and T.W. Parks. A high-resolution data-adaptive time-frequency representation. *IEEE Trans. ASSP*, 38:2127–2135, 1990.

[6] H. Oehlmann and D. Brie. The reduced-interference local wigner-ville distribution. In *Proc. 1997 IEEE ICASSP.*, pages 3645–3648, 1997.

[7] C.M. Stow, A.C.T. Kennington, C. Molina, and W.J. Fitzgerald. Experimental issues of functional merging on probability density estimation. In *Proc. 1997 IEE Int. Conf. Artificial Neural Networks*, pages 1–3–1–8, 1997.