Introduction

Many signals that exist in nature are inherently non-stationary or transient like Bat and Whale echoes, evoked-potentials, RADAR clutters. There exist many methods to analyze such signals including wavelet-transforms, short-time Fourier transform, adaptive time-frequency distributions, Matching Pursuits etc.. The signal representations which adapt the the tilings in the time-frequency (t-f) plane perform prominently in terms of achieving higher resolution in time and frequency simultaneously. Atomic decompositions [1, 2, 3] is one such class of algorithms which attempt to approximate the signal as a weighed combination of the atoms from a chosen dictionary. The atoms in the dictionary should contain all possible tilings that best suit the inner structure of the signal under analysis. Hence, some *a priori* information about the signal under analysis is needed if one were working with a finite vocabulary. Another alternative is to use, if an efficient method for choosing an atom exists, an infinite size vocabulary. Chirplets, windowed linear frequency modulated signals, offer a very flexible set of tilings like shearing, translation and rotation. Gaussian chirplets are particularly attractive in terms of computational tractability in addition to offering optimum time-frequency resolution (attain minimum time-bandwidth product). Numerous algorithms have been proposed to decompose a signal as a weighed combination of chirplets. However, most of the methods proposed in the literature use a divide and conquer approach, i.e., repeatedly decompose the residue until the residue is of insignificant energy. It should be noted that chirplets are not orthogonal and hence such decompositions are not unique. Also, there is no formal mechanism to estimate the number of components in the model and efficiency of the estimator in terms Cramer-Rao bounds were not provided (except in single component cases).

In the present work, we jointly estimate the parameters of these chirlpets. We assume that the number of components is unknown. We model the computed (specified) the spectrogram of the signal to be analyzed (synthesized) as a mixture of normal densities (Gausssian Mixture model - GMM). The mixture model parameters are estimated using the Expectize-Maximize (EM) algorithm. Functional merging is used to estimate the number of components in the GMM. We obtain the chirplets parameters (except for the phase) by the method of moments. These parameters are refined by Fisher-Scoring method. We obtain the Cramer-Rao bounds for all the parameters. If the EM algorithms produces a closer-to-global maximum, the parameters estimated would be optimal.

Chirpet Decompostiion

Employing chirplets to decompose a signal unifies the t-f tilings such as rotation, shearing, scaling and shifting. Hence, we proceed with the following assumption that any signal can be represented as a weighed sum of Gaussian chirplets given

by:

$$y(t) = s(t) + w(t) \tag{1a}$$

$$s(t) = \sum_{k=1}^{K} (a_k + jb_k) g_k(t) \tag{1b}$$

where

$$g_{k;t_c,\omega_c,\beta,d^2}(t) = \left(\frac{1}{\pi d^2}\right)^{\frac{1}{4}} \exp\left(-\frac{(t-t_c)^2}{4d^2} + j\frac{\beta(t-t_c)^2}{2} + j\omega_c(t-t_c)\right) \tag{2}$$

$t_c, \omega_c, \beta, d^2$ represent the location in time, location in frequency, chirp rate and the duration parameter of the chirplet and $w(t)$ is additive complex white Gaussian noise with unknown variance. Further

$$\|g_{t_c,\omega_c,\beta,d^2}\| = 1 \tag{3}$$

to make them unit energy signals.

Time-frequency modeling

We model the t-f plane as a mixture of normal pdfs [4] and estimate the parameters of the chirplets . It requires interpreting the t-f representation as an arbitrary bivariate pdf which would be modeled as GMM. It may appear strange that we are tying to view the t-f space as a mixture of normal pdfs by applying a random process modeling technique to a deterministic problem. However, we are using it just as a means to accomplish the task of decomposing the signal into chirplets. The initial choice of the t-f representation affects the model parameter estimation. We observe that the spectrogram of a Gaussian chirplet and normal density have structural similarities. We will utilize this fact to derive a set of mapping rules. Let $T(t, f)$ be the spectrogram of the signal under consideration. Also assume that t and f axis, representing time and frequency respectively, are sampled sufficiently smoothly on an $N \times M$ grid and let $\theta_{mn} = [t(n), f(m)]^T$

$$f(\theta_{mn}; \Phi) = \sum_{k=1}^{K} \lambda_k f_k(\theta_{mn}; \phi_k) \tag{4a}$$

$$T(\theta_{mn}) \approx f(\theta_{mn}; \Phi) \tag{4b}$$

where

- K is the total number of densities

- θ_{mn} is the mnth sample in the $N \times M$ grid

2

- $\phi_k = [\mu_k, \Sigma_k]$ represent the mean vector and covariance matrix of the kth density

- $f_k \sim \mathcal{N}(\mu_k, \Sigma_k)$

- $\Phi = [\phi_1 \ldots \phi_K, \lambda_1 \ldots \lambda_K]$

- $\{\lambda_k\}$ represent the mixing weights with constraint $\sum_{k=1}^{K} \lambda_k = 1$

We can estimate these parameters by the application of EM algorithm as:

$$\hat{\lambda}_{k,new} = \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_k(\theta_{mn}; \hat{\Phi}) \tag{5}$$

$$\hat{\mu}_{k,new} = \frac{1}{\hat{\lambda}_{k,new}} \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_k(\theta_{mn}; \hat{\Phi}) \theta_{mn} \tag{6}$$

$$\hat{\Sigma}_{k,new} = \frac{1}{\hat{\lambda}_{k,new}} \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_k(\theta_{mn}; \hat{\Phi})(\theta_{mn} - \hat{\mu}_k)^T(\theta_{mn} - \hat{\mu}_k) \tag{7}$$

where

$$\gamma_k(\theta_{mn}; \hat{\Phi}) = \frac{\hat{\lambda}_k f_k(\theta_{mn}; \hat{\theta}_k) T(\theta_{mn}; \Phi)}{f(\theta_{mn}; \hat{\Phi})} \tag{8}$$

In [4], time-frequency projection filters were used synthesize time-domain signals based on the regions identified by the individual densities in the mixture. However, we show that, if the initial t-f representation is chosen as spectrogram, one can immediately synthesize corresponding time-domain signals except for the phase. This is so because phase information is lost in the spectrogram decomposition. We map the parameters of a bivariate normal density to a chirpelt by the matching the moments. A Gaussian window with duration parameter α^2 is chosen in computing the spectrogram. The spectrogram of the a chirplet g_k is given by [5]:

$$T_{g_k}(t, f) = \frac{P(t - t_c)}{\sqrt{2\pi\sigma_{\omega/t}^2}} \exp\left(-\frac{(\omega - \omega_c - <\omega_t>)^2}{2\sigma_{<\omega_t>}^2}\right) \tag{9a}$$

$$P(t - t_c) = \left(\frac{1}{\pi(d^2 + \alpha^2)}\right)^{\frac{1}{2}} \exp\left(\frac{1}{\pi(d^2 + \alpha^2)} t^2\right) \tag{9b}$$

$$<\omega_t> = \left(\frac{1}{\pi(d^2 + \alpha^2)}\right) \beta(t - t_c) \tag{9c}$$

$$\tag{9d}$$

Soma Sekhar Dhavala

and a bivariate normal density is given as:

$$f_k(\theta) = \frac{1}{\sqrt{2\pi|\Sigma_k|}} \exp\left(-\frac{1}{2}(\mu_k - \theta)^T \Sigma_k^{-1}(\mu_k - \theta)\right) \tag{10a}$$

$$\Sigma_k = \begin{bmatrix} \sigma_{k1}^2 & \sigma_{k1}\sigma_{k2} \\ \sigma_{k1}\sigma_{k2} & \sigma_{k2}^2 \end{bmatrix} \tag{10b}$$

$$\mu_k = \begin{bmatrix} \mu_{k1} \\ \mu_{k2} \end{bmatrix} \tag{10c}$$

By matching the moments (method of moments) of the spectrogram (a time-frequency distribution)and the bivirate normal density given above, we obtain:

$$\frac{\alpha^2}{(d^2 + \alpha^2)}\beta = \rho\left(\frac{\sigma_1}{\sigma_2}\right) \tag{11a}$$

$$\frac{1}{(d^2 + \alpha^2)} = \frac{1}{2\sigma_1^2} \tag{11b}$$

$$\sigma_2^2(1 - \rho^2) = \frac{d^2 + \alpha^2}{2d^2\alpha^2} + \frac{\beta^2 d^2\alpha^2}{2(d^2 + \alpha^2)} \tag{11c}$$

$$\begin{bmatrix} \mu_{k1} \\ \mu_{k2} \end{bmatrix} = \begin{bmatrix} t_c \\ \omega_c \end{bmatrix} \tag{11d}$$

$$\tag{11e}$$

The parameters of the chirplet are obtained by solving the above equations simultaneously as:

$$d^2 = \frac{1 + (\alpha^2 \rho\, r)^2}{r^2(1 - \rho^2)} \tag{12a}$$

$$\beta = \frac{\rho\, r(d^2 + \alpha^2)}{d^2\alpha^2} \tag{12b}$$

$$\begin{bmatrix} \mu_{k1} \\ \mu_{k2} \end{bmatrix} = \begin{bmatrix} t_c \\ \omega_c \end{bmatrix} \tag{12c}$$

where $r = \sigma_1/\sigma_2$. It should be noted that we can not estimate the complex weight c_k since we loose phase information in spectrogram. We consider using quasi-Newton optimization algorithms to refine the parameters and also to estimate the amplitudes of the individual components. We use the estimates provided by the EM algorithms as initial estimates to the quasi-Newton algorithms. Besides providing faster convergence, if the initial estimates are carefully chosen, they also provide error covariances which could be used in designing time-frequency detectors.

4

Nested Fisher scoring

We rewrite the model in vector form as:

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{w} \tag{13a}$$

$$\boldsymbol{s} = \boldsymbol{G}\boldsymbol{\lambda} \tag{13b}$$

$$\boldsymbol{G} = [\boldsymbol{g}_1, j\boldsymbol{g}_1, \ldots \boldsymbol{g}_K, j\boldsymbol{g}_K] \tag{13c}$$

where

- $\boldsymbol{y}$: $N \times 1$ vector of measurements

- $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ complex AWGN with unknown variance σ^2

- $\boldsymbol{\theta}_k = [t_{c_k}, f_{c_k}, \alpha_k, \beta_k]^T$

- $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_K^T]^T$

- $\boldsymbol{\lambda} = [a_1, b_1, \ldots, a_K, b_K]^T$

- $\boldsymbol{\phi} = [\boldsymbol{\lambda}^T, \boldsymbol{\theta}^T]$

- $\boldsymbol{\gamma} = [\sigma^2, \boldsymbol{\phi}^T]^T$

- g_k is a function of $\boldsymbol{\theta}_k$

The log-likelihood is given as:

$$\ell(\boldsymbol{y}; \boldsymbol{\gamma}) = -N \log(\pi\sigma^2) - \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{s})^H(\boldsymbol{y} - \boldsymbol{s}) \tag{14}$$

and the Fisher's method of scoring produces the estimates using the following interation,

$$\hat{\boldsymbol{\gamma}}^{i+1} = \hat{\boldsymbol{\gamma}}^i + \boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}^{-1} 2\mathcal{R}e\left\{ \frac{\partial}{\partial \boldsymbol{\gamma}} \ell^H(\boldsymbol{y}; \boldsymbol{\gamma}) \right\}_{|\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}^i} \tag{15}$$

Using [6, Appendix 15.c], the Fisher Information matrix is given as:

$$\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}} = \begin{bmatrix} \mathcal{I}_{\sigma^2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mathcal{I}}_{\boldsymbol{\phi}} \end{bmatrix} \tag{16}$$

where

$$\mathcal{I}_{\sigma^2} = \frac{N}{\sigma^4} \text{ and } \boldsymbol{\mathcal{I}}_{\boldsymbol{\phi}} = \frac{2}{\sigma^2}\left[\frac{\partial \boldsymbol{s}^H}{\partial \boldsymbol{\phi}} \frac{\partial \boldsymbol{s}}{\partial \boldsymbol{\phi}} \right] \tag{17}$$

Ignoring the nuisance parameter σ^2,

$$\hat{\boldsymbol{\phi}}^{i+1} = \hat{\boldsymbol{\phi}}^i + \boldsymbol{\mathcal{I}}_{\boldsymbol{\phi}}^{-1} 2\mathcal{R}e\left\{ \frac{\partial}{\partial \boldsymbol{\phi}} \ell^H(\boldsymbol{y}; \boldsymbol{\gamma}) \right\}_{|\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}^i} \tag{18}$$

Since the parameters $\boldsymbol{\lambda}$ enter linearly, we could use nested-Fisher scoring [7]. This also becomes the separable non-linear least squares since the noise

is Gaussian. Fisher-scoring iteration on the concentrated log-likelihood is then given as:

$$\hat{\boldsymbol{\theta}}^{i+1} = \hat{\boldsymbol{\theta}}^{i} + \boldsymbol{\mathcal{I}'}_{\boldsymbol{\theta}}^{-1} \, 2\mathcal{R}e \left\{ \frac{\partial}{\partial \boldsymbol{\phi}} \ell^{H}(\boldsymbol{y}; \boldsymbol{\gamma}) \right\} \tag{19a}$$

$$\boldsymbol{\mathcal{I}'}_{\boldsymbol{\theta}} = \frac{2}{\sigma^2} \frac{\partial \boldsymbol{s}^H}{\partial \boldsymbol{\theta}} \left(\boldsymbol{I} - \frac{\partial \boldsymbol{s}^H}{\partial \boldsymbol{\lambda}} \left(\frac{\partial \boldsymbol{s}}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{s}^H}{\partial \boldsymbol{\lambda}} \right)^{-1} \frac{\partial \boldsymbol{s}^H}{\partial \boldsymbol{\lambda}} \right) \frac{\partial \boldsymbol{s}}{\partial \boldsymbol{\theta}} \tag{19b}$$

$$\frac{\partial}{\partial \boldsymbol{\phi}} \ell^{H}(\boldsymbol{y}; \boldsymbol{\gamma}) = \frac{2}{\sigma^2} \frac{\partial \boldsymbol{s}^H}{\partial \boldsymbol{\theta}} (\boldsymbol{y} - \boldsymbol{s}) \tag{19c}$$

The score vector and the Fisher information matrix are evaluated at:

$$\boldsymbol{\theta} \leftarrow \hat{\boldsymbol{\theta}}^{i} \tag{20a}$$

$$\boldsymbol{\lambda} \leftarrow (\boldsymbol{G}^H \boldsymbol{G})^{-1} \boldsymbol{G}^H \boldsymbol{y} \tag{20b}$$

Cramer-Rao bound

The Cramer-Rao lower bounds on the variance of the parameters are given as:

- $\mathrm{CRLB}_{\sigma^2} = \frac{\sigma^4}{N}$
- $\mathrm{CRLB}_{\boldsymbol{\lambda}} = \frac{1}{\sigma^2} \boldsymbol{G}^H \boldsymbol{G}$
- $\mathrm{CRLB}_{\boldsymbol{\theta}} = \boldsymbol{\mathcal{I}'}_{\boldsymbol{\theta}}^{-1}$

Gradients

Here analytical expressions for the gradients are given. The discrete version of the signal is given by:

$$g_k(n) = \left(\frac{1}{\pi d_k^2} \right)^{\frac{1}{4}} \exp \left(-\frac{(n - t_{c,k})^2}{4 d_k^2} + j \frac{\beta (n - t_{c,k})^2}{2} + j \omega_{c,k} (n - t_{c,k}) \right) \tag{21}$$

By noting that,

$$\frac{\partial s(n)}{\partial \boldsymbol{\theta}_k} = \frac{\partial g_k(n)}{\partial \boldsymbol{\theta}_k} \tag{22}$$

we calculate the gradients of a chirplet and drop the subscript k for simplicity

$$\frac{\partial g(n)}{\partial t_c} = g(n) \left(\frac{n - t_c}{4 d^2} - j\beta(n - t_c) - j\omega_c \right) \tag{23a}$$

$$\frac{\partial g(n)}{\partial \omega_c} = g(n) \left(j\omega(n - t_c) \right) \tag{23b}$$

$$\frac{\partial g(n)}{\partial d^2} = g(n) \left(\frac{(n - t_c)^2}{4 d^4} - \frac{1}{d^2} \right) \tag{23c}$$

$$\frac{\partial g(n)}{\partial \beta} = g(n) \left(\frac{j}{2} (n - t_c)^2 \right) \tag{23d}$$

We can use [8] to obtain closed-form expressions for the Cramer-Rao bound. This completes the estimation process.

References

[1] A. Bultan, "A four-parameter atomic decomposition of chirplets," *IEEE Trans. on Signal Processing*, vol. 41, pp. 731–745, 1999.

[2] J. C. O'Neill and P. Flandrin, "Chirp hunting," in *Proc. of the IEEE-SP International Symposium on Time-Frequency and Time-ScaleAanalysis*, 1998, pp. 425–428.

[3] R. N. Bracewell and D Mihovilovic, "Adaptive chirplet representation of signals on time-frequency plane," *Electronics Letters*, vol. 27, pp. 1159–1161, 1991.

[4] Mark Coates, *Time-frequency modeling*, Ph. D, Univ. of Cambridge, Signal Procesing and Communications Laboratory, Dept. of Engineering, August 1998.

[5] Leon Cohen, *Time-Frequency Analysis*, Printice-Hall, Englewood cliffs, New Jersey, 1995.

[6] Steven M. Kay, *Fundamentals of Statistical Signal Processing, Vol-1 Estimation theory*, Prentice-Hall PTR, 1993.

[7] Gordon K Smith, "Partitioned algorithms for maximum likelihood and other nonlinear estimation," *Statistics and Computing*, vol. 6, pp. 201–216, 1996.

[8] Kostas Traintafyllopoulos, "Moments and cumulants of the multivariate real and complex gaussian distributions," 2002, www.stats.bris.ac.uk/research/stats/pub/ResRept/2002.html.