REAL-TIME TIME-FREQUENCY BASED BLIND SOURCE SEPARATION

Scott Rickard, Radu Balan, Justinian Rosca

Siemens Corporate Research
Princeton, NJ 08540
{scott.rickard,radu.balan,justinian.rosca}@scr.siemens.com

ABSTRACT

We present a real-time version of the DUET algorithm for the blind separation of any number of sources using only two mixtures. The method applies when sources are W-disjoint orthogonal, that is, when the supports of the windowed Fourier transform of any two signals in the mixture are disjoint sets, an assumption which is justified in the Appendix. The online algorithm is a Maximum Likelihood (ML) based gradient search method that is used to track the mixing parameters. The estimates of the mixing parameters are then used to partition the time-frequency representation of the mixtures to recover the original sources. The technique is valid even in the case when the number of sources is larger than the number of mixtures.

The method was tested on mixtures generated from different voices and noises recorded from varying angles in both anechoic and echoic rooms. In total, over 1500 mixtures were tested. The average SNR gain of the demixing was 15 dB for anechoic room mixtures and 5 dB for echoic office mixtures. The algorithm runs 5 times faster than real time on a 750MHz laptop computer. Sample sound files can be found here:

http://www.princeton.edu/~srickard/bss.html

1. INTRODUCTION

In [1] a blind source separation technique was introduced that allows the separation of an arbitrary number of sources from just two mixtures provided the time-frequency representations of sources do not overlap. The key observation in the technique is that, for mixtures of such sources, each time-frequency point depends on at most one source and its associated mixing parameters. In anechoic environments, it is possible to extract the estimates of the mixing parameters from the ratio of the time-frequency representations of the mixtures. These estimates cluster around the true mixing parameters and, identifying the clusters, one can partition the time-frequency representation of the mixtures to produce the time-frequency representations of the original sources.

The original DUET algorithm involved creating a two-dimensional (weighted) histogram of the relative amplitude and delay estimates, finding the peaks in the histogram, and then associating each time-frequency point in the mixture with one peak. The original implementation of the method was offline and passed through the data twice; one time to create the histogram and a second time to demix. In this paper, we present an online version of the DUET algorithm which avoids the need for the creation of the histogram, which in turn avoids the computational load of updating the histogram and the tricky issue of finding and tracking peaks. The online DUET advantages are,

- online (5 times faster than real time),
- 15 dB average separation for anechoic mixtures,
- 5 dB average separation for echoic mixtures, and
- can demix >2 sources from 2 mixtures.

In Section 2 we define the time delay mixing model, explain the concept of W-disjoint orthogonality, and describe the mixing parameter tracking procedure. In Section 3 we describe the demixing method used for each algorithm. Section 4 describes the mixing tests and gives detailed demixing results. Justification for the W-disjoint orthogonality assumption can be found in Appendix A, and Appendix B contains the ML objective function derivation.

2. MIXING PARAMETER ESTIMATION

2.1. Source mixing

Consider measurements of a pair of sensors where only the direct path is present. In this case, without loss of generality, we can absorb the attenuation and delay parameters of the first mixture, $x_1(t)$, into the definition of the sources. The two mixtures can thus be expressed as,

$$x_1(t) = \sum_{j=1}^{N} s_j(t), \tag{1}$$

$$x_2(t) = \sum_{j=1}^{N} a_j s_j(t-\delta_j), \tag{2}$$

Presented at the *ICA2001 Conference, December 9-13, 2001, San Diego, CA.*

where N is the number of sources, δ_j is the arrival delay between the sensors resulting from the angle of arrival, and a_j is a relative attenuation factor corresponding to the ratio of the attenuations of the paths between sources and sensors. We use Δ to denote the maximal possible delay between sensors, and thus, $|\delta_j| \le \Delta, \forall j$.

2.2. Source Assumptions

We call two functions $s_1(t)$ and $s_2(t)$ **W-disjoint orthogonal** if, for a given windowing function $W(t)$, the supports of the windowed Fourier transforms of $s_1(t)$ and $s_2(t)$ are disjoint. The windowed Fourier transform of $s_j(t)$ is defined,

$$F^W(s_j(\cdot))(\omega,\tau) = \int_{-\infty}^{\infty} W(t-\tau)s_j(t)e^{-i\omega t}dt, \quad (3)$$

which we will refer to as $S_j(\omega,\tau)$ where appropriate. The W-disjoint orthogonality assumption can be stated concisely,

$$S_1(\omega,\tau)S_2(\omega,\tau) = 0, \ \forall \omega,\tau. \quad (4)$$

In Appendix A, we introduce the notion of approximate W-disjoint orthogonality. When $W(t) \equiv 1$, we use the following property of the Fourier transform,

$$F^W(s_j(\cdot - \delta))(\omega,\tau) = e^{-i\omega\delta}F^W(s_j(\cdot))(\omega,\tau). \quad (5)$$

We will assume that (5) holds for all δ, $|\delta| \le \Delta$, even when $W(t)$ has finite support[2].

2.3. Amplitude-Delay Estimation

Using the above assumptions, we can write the model from (1) and (2) for the case with two array elements as,

$$\begin{bmatrix} X_1(\omega,\tau) \\ X_2(\omega,\tau) \end{bmatrix} = \begin{bmatrix} 1 & \ldots & 1 \\ a_1 e^{-i\omega\delta_1} & \ldots & a_N e^{-i\omega\delta_N} \end{bmatrix} \begin{bmatrix} S_1(\omega,\tau) \\ \vdots \\ S_N(\omega,\tau) \end{bmatrix} \quad (6)$$

For W-disjoint orthogonal sources, we note that at most one of the N sources will be non-zero for a given (ω,τ), thus,

$$\begin{bmatrix} X_1(\omega,\tau) \\ X_2(\omega,\tau) \end{bmatrix} = \begin{bmatrix} 1 \\ a_j e^{-i\omega\delta_j} \end{bmatrix} S_j(\omega,\tau), \qquad \text{for some } j. \quad (7)$$

The original DUET algorithm estimated the mixing parameters by analyzing the ratio of $X_1(\omega,\tau)$ and $X_2(\omega,\tau)$. In light of (7), it is clear that mixing parameter estimates can be obtained via,

$$(\hat{a}(\omega,\tau),\hat{\delta}(\omega,\tau)) = \left(\left| \frac{X_2(\omega,\tau)}{X_1(\omega,\tau)} \right|, \frac{1}{\omega}\mathrm{Im}\left\{ \ln\left(\frac{X_1(\omega,\tau)}{X_2(\omega,\tau)} \right) \right\} \right). \quad (8)$$

The original DUET algorithm constructed a 2-D histogram of amplitude-delay estimates and looked at the number and location of the peaks in the histogram to determine the number of sources and their mixing parameters. See [1, 3] for details.

2.4. ML Mixing Parameter Gradient Search

For the online algorithm, we take a different approach. First, note that,

$$|X_1(\omega,\tau)a_j e^{-i\omega\delta_j} - X_2(\omega,\tau)|^2 = 0, \quad (9)$$

if source j is the active source at time-frequency (ω,τ). Moreover, defining,

$$\rho(a_j,\delta_j,\omega,\tau) \doteq \frac{1}{1+a_j^2}|X_1(\omega,\tau)a_j e^{-i\omega\delta_j} - X_2(\omega,\tau)|^2, \quad (10)$$

we can see that,

$$\sum_\omega \min(\rho(a_1,\delta_1,\omega,\tau),\ldots,\rho(a_N,\delta_N,\omega,\tau)) = 0, \quad (11)$$

because at least one ρ term will be zero at each frequency. In Appendix B, it is shown that the maximum likelihood estimates of the mixing parameters satisfy,

$$\min_{a_1,\delta_1,\ldots,a_N,\delta_N} \sum_\omega \min(\rho_1,\ldots,\rho_N), \quad (12)$$

where we have used ρ_j as shorthand for $\rho(a_j,\delta_j,\omega,\tau)$. We perform gradient descent with (12) as the objective function to learn the mixing parameters. In order to avoid the discontinuous nature of the minimum function, we approximate it smoothly as follows,

$$\begin{aligned} \min(\rho_1,\rho_2) &= \frac{\rho_1+\rho_2-|\rho_1-\rho_2|}{2} && (13) \\ &\approx \frac{\rho_1+\rho_2-\phi(\rho_1-\rho_2)}{2} && (14) \\ &= \frac{-1}{\lambda}\ln(e^{-\lambda\rho_1}+e^{-\lambda\rho_2}), && (15) \end{aligned}$$

where,

$$\phi(x) = \int_0^x \frac{1-e^{-\lambda t}}{1+e^{-\lambda t}}dt = x + \frac{2}{\lambda}\ln(1+e^{-\lambda}). \quad (16)$$

Generalizing (15), the smooth ML objective function is,

$$J(\tau) = \min_{a_1,\delta_1,\ldots,a_N,\delta_N} \sum_\omega -\frac{1}{\lambda}\ln(e^{-\lambda\rho_1}+\cdots+e^{-\lambda\rho_N}), \quad (17)$$

which has partials,

$$\frac{\partial J(\tau)}{\partial \delta_j} = \sum_{\omega} \frac{e^{-\lambda \rho_j}}{\sum_k e^{-\lambda \rho_k}} \frac{-2\omega a_j}{1+a_j^2} \mathrm{Im}\{X_1(\omega,\tau)\overline{X_2(\omega,\tau)}e^{-i\omega\delta_j}\}, \tag{18}$$

and,

$$\frac{\partial J(\tau)}{\partial a_j} = \sum_{\omega} \frac{e^{-\lambda \rho_j}}{\sum_k e^{-\lambda \rho_k}} \frac{2}{(1+a_j^2)^2} \cdot \\ (((a_j^2-1)\mathrm{Re}\{X_1(\omega,\tau)\overline{X_2(\omega,\tau)}e^{-i\omega\delta_j}\} \\ +a_j(|X_1(\omega,\tau)|^2+|X_2(\omega,\tau)|^2)). \tag{19}$$

We assume we know the number of sources we are searching for and initialize an amplitude and delay estimate pair to random values for each source. The estimates $(a_j[k], \delta_j[k])$ for the current time $\tau_k = k\tau_\Delta$ (where τ_Δ is the time separating adjacent time windows) are updated based on the previous estimate and the current gradient as follows,

$$a_j[k] = a_j[k-1] - \beta\alpha_j[k]\frac{\partial J(\tau_k)}{\partial a_j}, \tag{20}$$

$$\delta_j[k] = \delta_j[k-1] - \beta\alpha_j[k]\frac{\partial J(\tau_k)}{\partial \delta_j}, \tag{21}$$

where β is a learning rate constant and $\alpha_j[k]$ is a time and mixing parameter dependent learning rate for time index k for estimate j. In practice, we have found it helpful to adjust the learning rate depending on the amount of mixture energy recently explained by the current estimate. We define,

$$q_j[k] = \sum_{\omega} \frac{e^{-\lambda\rho(a_j,\delta_j,\omega,\tau_k)}}{\sum_l e^{-\lambda\rho(a_l,\delta_l,\omega,\tau_k)}} |X_1(\omega,\tau_k)|\,|X_2(\omega,\tau_k)|, \tag{22}$$

and update the parameter dependent learning rate as follows,

$$\alpha_j[k] = \frac{q_j[k]}{\sum_{m=0}^{k} \gamma^{k-m} q_j[m]}, \tag{23}$$

where γ is a forgetting factor.

3. DEMIXING

In order to demix the jth source, we construct a time-frequency mask based on the ML parameter estimator (see (33) in Appendix B),

$$\Omega_j(\omega,\tau) = \begin{cases} 1 & \rho(a_j,\delta_j,\omega,\tau) \le \rho(a_m,\delta_m,\omega,\tau)\ \forall m \ne j \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

The estimate for the time-frequency representation of the jth source is,

$$\hat{S}_j(\omega,\tau) = \Omega_j(\omega,\tau)X_1(\omega,\tau). \tag{25}$$

We then reconstruct the source using the appropriate dual window function[4]. In this way, we demix all the sources by partitioning the time-frequency representation of one of the mixtures. Note that because the method does not invert the mixing matrix, it can demix all sources even when the number of sources is greater than the number of mixtures ($N > M$).

4. TESTS AND SUMMARY

We tested the method on mixtures created in both an anechoic room and an echoic office environment. The algorithm used parameters $\beta = 0.02, \gamma = .95, \lambda = 10$ and a Hamming window of size 512 samples (with adjacent windows separated by 128 samples) in all the tests. For all tests, the method ran more than 5 times faster than real time.

For the anechoic test, the setup is pictured in Figure 1. Separate recordings at 16kHz were made of six speech files (4 female, 2 male) taken from the TIMIT database played from a loudspeaker placed at the X marks in the figure. Pairwise mixtures were then created from all possible voice/angle combinations, excluding same voice and same angle combinations, yielding a total of 630 mixtures ($630 = 6 \times 5 \times 7 \times 6/2$).

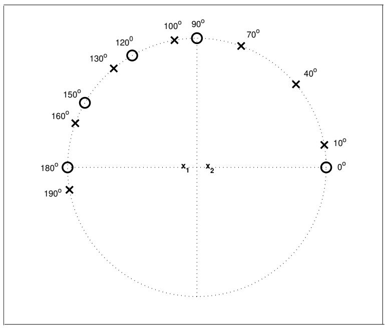

Fig. 1. Experimental setup. Microphones are separated by ~1.75 cm centered along the 180^o to 0^o line. The X's show the source locations used in the anechoic tests. The O's show the locations of the sources in the echoic tests.

The SNR gains of the demixtures were calculated as follows. Denote the contribution of source j on microphone k as $S_{jk}(\omega,\tau)$. Thus we have,

$$X_1(\omega,\tau) = S_{11}(\omega,\tau) + S_{21}(\omega,\tau) \tag{26}$$

$$X_2(\omega,\tau) = S_{12}(\omega,\tau) + S_{22}(\omega,\tau) \tag{27}$$

As we do not know the permutation of the demixing, we calculate the SNR gain conservatively,

$$\mathrm{SNR}_1 = \max\left(10\log\frac{||\Omega_1 S_{11}||^2}{||\Omega_1 S_{21}||^2}, 10\log\frac{||\Omega_2 S_{12}||^2}{||\Omega_2 S_{22}||^2}\right) - \max\left(10\log\frac{||S_{11}||^2}{||S_{21}||^2}, 10\log\frac{||S_{12}||^2}{||S_{22}||^2}\right)$$

$$\mathrm{SNR}_2 = -\min\left(10\log\frac{||\Omega_1 S_{11}||^2}{||\Omega_1 S_{21}||^2}, 10\log\frac{||\Omega_2 S_{12}||^2}{||\Omega_2 S_{22}||^2}\right) + \min\left(10\log\frac{||S_{11}||^2}{||S_{21}||^2}, 10\log\frac{||S_{12}||^2}{||S_{22}||^2}\right)$$

In order to give the method time to learn the mixing parameters, the SNR results do not include the first half second of data.

Figure 2 shows the average SNR gain results for each angle difference. For example, the 60 degree difference results average all the 10-70, 40-100, 70-130, 100-160, and 130-190 results. Each bar shows the maximum SNR gain, one standard deviation above the mean, the mean (which is labeled), one standard deviation below the mean, and the minimum SNR gain over all the tests (both SNR_1 and SNR_2 are included in the averages). The separation results improve as the angle difference increases. Figure 3 details the 30 degree difference results by angle comparison, averaging 30 tests per angle comparison. The performance is a function of the delay. That is, the worst performance is achieved for the smallest delay (corresponding to the 160-190 mixtures), the second worst performance is achieved for the second smallest delay (corresponding to the 10-40 mixtures), and so forth.

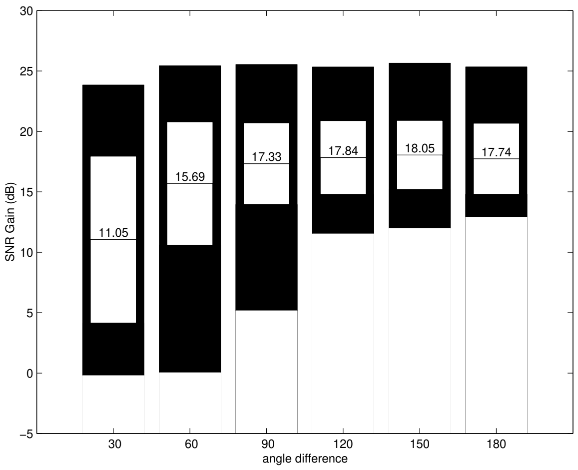

Fig. 2. Comparison of overall separation SNR gain by angle difference. Anechoic data.

Recordings were also made in an echoic office with reverberation time of $\sim$500 ms, that is, the impulse response of the room fell to -60 dB after 500 ms. For the echoic tests, the sources were placed at 0, 90, 120, 150, and 180 degrees

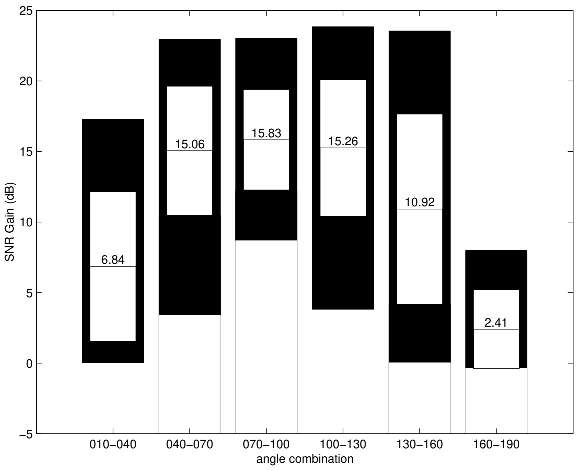

Fig. 3. Overall separation SNR gain by 30 degree angle pairing. Anechoic data.

(see the O's in Figure 1). Separation results for pairwise mixtures of voices (4 female, 4 male) are shown in Figure 5. Separation results for pairwise mixtures of voices (4 female, 4 male) and noises (line printer, copy machine, and vacuum cleaner) are shown in Figure 4. The results are considerably worse in the echoic case, which is not surprising as the method assumes anechoic mixing. However, the method does achieve 5 dB SNR gain on average and is real-time.

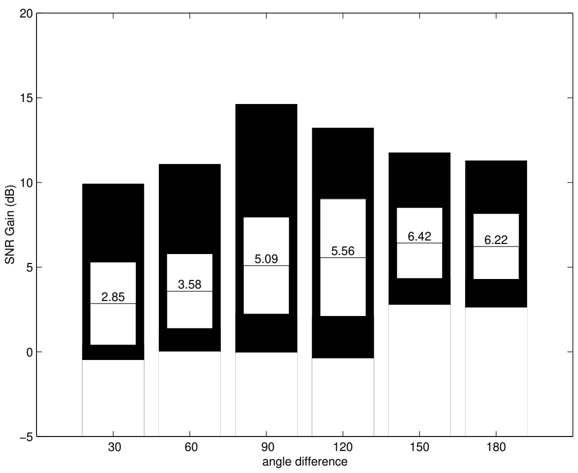

Fig. 4. Comparison of overall separation SNR gain by angle difference. Echoic office data. Voice vs. Noise.

Summary results for all three testing groups (anechoic, echoic voice vs. voice, and echoic voice vs. noise) are shown in the table. We have presented a real-time version of the DUET algorithm which uses gradient descent to learn the anechoic mixing parameters and then demixes by partitioning the time-frequency representations of the mixtures. We have also introduced a measure of W-disjoint orthogo-

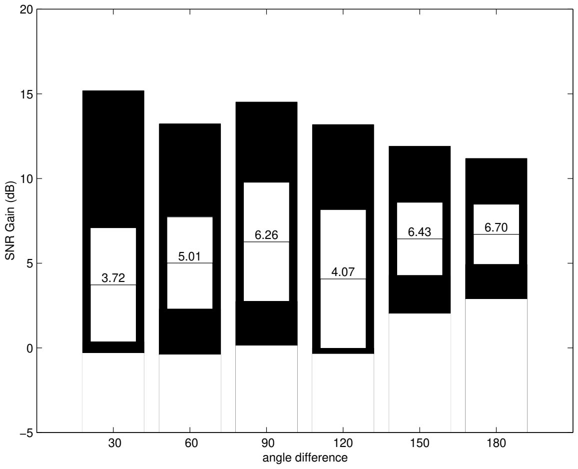

Fig. 5. Comparison of overall separation SNR gain by angle difference. Echoic office data. Two Voices.

	AVV	EVV	EVN
number of tests	630	560	480
mean SNR gain (dB)	15.31	5.09	4.41
std SNR gain (dB)	5.69	3.34	2.87
max SNR gain (dB)	25.65	15.18	14.61
min SNR gain (dB)	-0.21	-0.42	-0.50

Table 1. Results Summary. AVV = Anechoic Voice vs. Voice. EVV = Echoic Voice vs. Voice. EVN = Echoic Voice vs. Noise

nality and provided empirical evidence for the approximate W-disjoint orthogonality of speech signals.

A. W-DISJOINT ORTHOGONALITY OF SPEECH

Clearly, the W-disjoint orthogonality assumption is not exactly satisfied for our signals of interest. We introduce here a measure of W-disjoint orthogonality for a group sources and show that speech signals are indeed nearly W-disjoint orthogonal to each other. Consider the time-frequency mask,

$$\Phi_x^{12}(\omega,\tau) = \begin{cases} 1 & 20\log(|S_1(\omega,\tau)|/|S_2(\omega,\tau)|) > x \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

and the resulting energy ratio,

$$r(x) = ||\Phi_x^{12}(\omega,\tau)S_1(\omega,\tau)||^2/||S_1(\omega,\tau)||^2, \tag{29}$$

which measures the percentage of energy of source 1 for time-frequency points where it dominates source 2 by x dB. We propose $r(x)$ as a measure of W-disjoint orthogonality. For example, Figure 6 shows $r(x)$ averaged for pairs of sources used in the demixing tests. We can see from the graph that $r(3) > .9$ for all three, and thus say that the signals used in the tests were 90% W-disjoint orthogonal at 3 dB. If we can correctly map time-frequency points with 3 dB or more single source dominance to the correct corresponding output partition, we can recover the 90% of the energy of the original sources. The figure also demonstrates the W-disjoint orthogonality of six speech signals taken as a group and the fact that independent Gaussian white noise processes are less than 50% W-disjoint orthogonal at all levels.

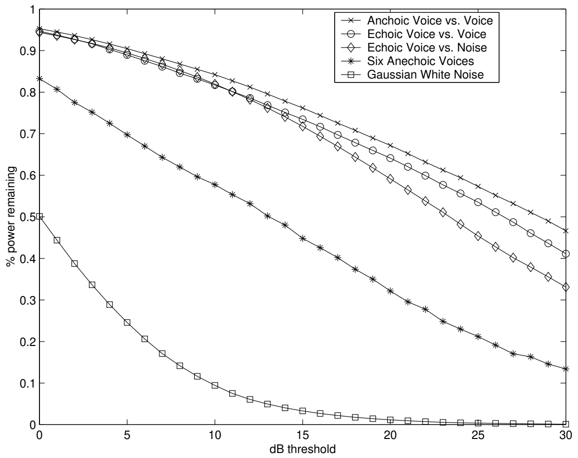

Fig. 6. W-Disjoint Orthogonality. The signals used in the tests were 90% W-disjoint orthogonal at 3 dB and more than 80% W-disjoint orthogonal at 10 dB. Comparing one source to the sum of five others, we still have more than 75% W-disjoint orthogonality at 3 dB. Independent Gaussian white noise processes are less than 50% W-disjoint orthogonal at all levels.

B. ML ESTIMATION FOR DUET MODEL

Assume a mixing model of type (1)-(2) to which we add measurement noise:

$$X_1(\omega,\tau) = \sum_{j=1}^{N} q_j(\omega,\tau)S_j(\omega,\tau) + \nu_1(\omega,\tau) \tag{30}$$

$$X_2(\omega,\tau) = \sum_{j=1}^{N} a_j e^{-i\omega\delta_j} q_j(\omega,\tau)S_j(\omega,\tau) + \nu_2(\omega,\tau) \tag{31}$$

The ideal model (1)-(2) is obtained in the limit $\nu_1,\nu_2 \rightarrow 0$. In practice, we make the computations assuming the existence of such a noise, and then we pass to the limit. We assume the noise and source signals are Gaussian distributed and independent from one another, with zero mean and known variances:

$$\begin{bmatrix} \nu_1(\omega,\tau) \\ \nu_2(\omega,\tau) \end{bmatrix} \sim \mathcal{N}(0,\sigma^2 I_2)$$

$$S_j(\omega,\tau) \sim \mathcal{N}(0, \rho_j(\omega))$$

The Bernoulli random variables $q_j(\omega,\tau)$'s are NOT independent. To accommodate the *W-disjoint orthogonality* assumption, we require that for each (ω,τ) at most one of the $q_j(\omega,\tau)$'s can be unity, and all others must be zero. Thus the N-tuple $(q_1(\omega,\tau),\ldots,q_N(\omega,\tau))$ takes values only in the set

$$\mathcal{Q} = \{(0,0,\ldots,0),(1,0,\ldots,0),\ldots,(0,0,\ldots,1)\}$$

of cardinality $N+1$. We assume uniform priors for these R.V.'s.

The short-time stationarity implies different frequencies are decorrelated (and hence independent) from one another. We use this property in constructing the likelihood. The likelihood of parameters $(a_1,\delta_1,\ldots,a_N,\delta_N)$ given the data $(X_1(\omega,\tau),X_2(\omega,\tau))$ and spectral powers σ^2, $\rho_j(\omega)$ at a given τ, is given by conditioning with respect to $q_j(\omega,\tau)$'s by:

$$\begin{aligned} &L(a_1,\delta_1,\ldots,a_N,\delta_N;\tau) \\ &\quad := p(X_1(\cdot),X_2(\cdot)|a_1,\delta_1,\ldots,a_N,\delta_N;\tau,\sigma^2,\rho_j) \\ &= \prod_{\omega}\sum_{j=0}^{N} \frac{\exp\{-M\}}{\pi^2\det(\sigma^2 I_2+\rho_j(\omega)\Gamma_j(\omega))} p(q_j(\omega,\tau)=1) \end{aligned} \tag{32}$$

where:

$$M = \left[\overline{X_1(\omega,\tau)}\ \overline{X_2(\omega,\tau)}\right](\sigma^2 I_2+\rho_j(\omega)\Gamma_j(\omega))^{-1}\left[\begin{array}{c} X_1(\omega,\tau) \\ X_2(\omega,\tau)\end{array}\right]$$

and

$$\Gamma_j = \left[\begin{array}{c} 1 \\ a_j e^{-i\omega\delta_j}\end{array}\right]\left[\begin{array}{cc} 1 & a_j e^{i\omega\delta_j}\end{array}\right] = \left[\begin{array}{cc} 1 & a_j e^{i\omega\delta_j} \\ a_j e^{-i\omega\delta_j} & a_j^2\end{array}\right]$$

and we have defined $q_0(\omega,\tau) = 1-\sum_{k=1}^N q_k(\omega,\tau)$, $\rho_0(\omega) = 0$, and $\Gamma_0(\omega) = I_2$ for notational simplicity in (32) in dealing with the case when no source is active at a given (ω,τ).

Next the Matrix Inversion Lemma (or an explicit computation) gives:

$$\begin{aligned} -M = &-\frac{1}{\sigma^2}\frac{1}{\sigma^2+\rho_j(\omega)(1+a_j^2)}\cdot \\ &(\rho_j(\omega)|a_j e^{-i\omega\delta_j}X_1(\omega,\tau)-X_2(\omega,\tau)|^2 + \\ &\qquad \sigma^2(|X_1(\omega,\tau)|^2+|X_2(\omega,\tau)|^2)) \end{aligned}$$

and

$$\det(\sigma^2 I_2+\rho_j(\omega)\Gamma_j(\omega)) = \sigma^2(\sigma^2+\rho_j(\omega)(1+a_j^2))$$

Now we pass to the limit $\sigma\to 0$. The dominant terms from the previous two equations are:

$$-\frac{1}{\sigma^2}\frac{|a_j e^{-i\omega\delta_j}X_1(\omega,\tau)-X_2(\omega,\tau)|^2}{1+a_j^2}$$

and

$$\sigma^2\rho_j(\omega)(1+a_j^2)$$

Of the $N+1$ terms in each sum of (32), only one term is dominant, namely the one of the largest exponent. Assume $\pi:\omega\mapsto\{0,1,\ldots,N\}$ is the selection map defined by:

$$\pi(\omega) = k\ \ ,\ \ \text{if}\ \rho(a_k,\delta_k,\omega,\tau)\le\rho(a_j,\delta_j,\omega,\tau)\ \ \forall j\neq k \tag{33}$$

where:

$$\rho(a_0,\delta_0,\omega,\tau) = |X_1(\omega,\tau)|^2+|X_2(\omega,\tau)|^2$$

and for $k\in\{1,2,\ldots,N\}$:

$$\rho(a_k,\delta_k,\omega,\tau) = \frac{|a_j e^{-i\omega\delta_j}X_1(\omega,\tau)-X_2(\omega,\tau)|^2}{1+a_j^2}$$

Then the likelihood becomes:

$$\begin{aligned} &L(a_1,\delta_1,\ldots,a_N,\delta_N;\tau) = \\ &\quad \frac{C}{\sigma^{2M}}\prod_{k=0}^{N}\prod_{\omega\in\pi^{-1}(k)} t_k\exp\left\{-\frac{\rho(a_k,\delta_k,\omega,\tau)}{\sigma^2}\right\} \end{aligned} \tag{34}$$

with M, the number of frequencies and:

$$t_k = \left\{\begin{array}{ll} \frac{1}{\sigma^2} & k=0 \\ \frac{1}{\rho_k(\omega)(1+a_k^2)} & k\in\{1,2,\ldots,N\}\end{array}\right.$$

The dominant term in log-likelihood remains the exponent. Thus:

$$\log L \approx -\frac{1}{\sigma^2}\sum_{k=0}^{N}\sum_{\omega\in\pi^{-1}(k)}\rho(a_k,\delta_k,\omega,\tau) \tag{35}$$

and maximizing the log-likelihood is equivalent to the following (which is (12)):

$$\min_{a_1,\delta_1,\ldots,a_N,\delta_N}\sum_{\omega}\min(\rho(a_1,\delta_1,\omega,\tau),\cdots,\rho(a_N,\delta_N,\omega,\tau))$$

6. REFERENCES

[1] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind Separation of Disjoint Orthogonal Signals: Demixing N Sources from 2 Mixtures," in *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, June 2000, vol. 5, pp. 2985–88.

[2] R. Balan, J. Rosca, S. Rickard, and J. O'Ruanaidh, "The influence of windowing on time delay estimates," in *Proceedings of the 2000 CISS*, Princeton, NJ, March 15-17 2000.

[3] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind Separation of Disjoint Orthogonal Signals," *IEEE Trans. Signal Proc.*, 2000, Submitted.

[4] I. Daubechies, *Ten Lectures on Wavelets*, chapter 3, SIAM, Philadelphia, PA, 1992.