JOINT ANTI-DIAGONALIZATION FOR BLIND SOURCE SEPARATION.

A. Belouchrani[1], *K. Abed-Meraim*[2], *M. G. Amin*[3], *A. M. Zoubir*[4]

[1] Elec. Eng. Dept, Ecole Nationale Polytechnique, EL Harrach 16200, Algiers, Algeria.
[2] Sig. & Image Proc. Dept., Telecom Paris (ENST), Cedex 13, France.
[3] Dept of Elec. and Comp. Eng. Villanova University, Villanova PA 19085 USA.
[4] Australian Telecomm. Research Institute and School of Elec. and Comp. Eng.
Curtin University of Technology, GPO Box U 1987, Perth 6485, Western Australia.

ABSTRACT

We address the problem of blind source separation of non-stationary signals of which only instantaneous linear mixtures are observed. A blind source separation approach exploiting both auto-terms and cross-terms of the time-frequency (TF) distributions of the sources is considered. The approach is based on the simultaneous diagonalization and anti-diagonalization of spatial TF distribution matrices made up of, respectively, auto-terms and cross-terms. Numerical simulations are provided to demonstrate the effectiveness of the proposed approach and compare its performances with existing TF-based methods.

1. INTRODUCTION

Blind source separation (BSS) consists of recovering a set of signals of which only instantaneous linear mixtures are observed. Source separation algorithms are based on the main assumption of mutual independence of the respective source signals. Various techniques have been proposed, including the separation by maximum likelihood [3], separation by decorrelation and rotation [1, 4], separation by neural networks [2], separation by contrast function [10], and separation by information-theoretic criteria [11].

For non stationary source signals, blind source separation methods based on time-frequency distributions have been introduced in [5, 6]. These methods consider only auto-terms of the signal time-frequency distributions (TFDs), and exploit the diagonal structure of the source TFD matrices. In this paper, we perfrom BSS of FM signals impinging on an antenna array using both the TFDs of the source signals as well as their cross TFDs. This is achieved by exploiting the anti-diagonal structure of the source TFD matrices, evaluated at the cross-term TF points. Moreover, we propose an automatic selection procedure to decide, with no a priori knowledge about the sources, whether a considered TF point corresponds to an auto-term or a cross-term.

As a consequence, in comparison with the method in [5], the proposed technique is more robust to noise and TF point selection errors and improves the quality of source separation.

2. PROBLEM FORMULATION

Consider m sensors receiving an instantaneous linear mixture of signals emitted from $n \leq m$ sources. The $m \times 1$ vector $\mathbf{x}(t)$ denotes the output of the sensors at time instant t which may be corrupted by additive noise $\mathbf{n}(t)$. Hence,

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{1}$$

where the $m \times n$ matrix $\mathbf{A}$ is called the 'mixing matrix'. The n source signals are collected in a $n \times 1$ vector denoted $\mathbf{s}(t)$ which is referred to as the source signal vector. The sources are assumed to have different structures and localization properties in the time-frequency domain. The mixing matrix $\mathbf{A}$ is full column rank but is otherwise unknown. In contrast to traditional parametric methods, no specific structure of the mixture matrix is assumed.

It is well known that the blind source separation is a technique for the identification of the mixing matrix and/or the recovering of the source signals up to a fixed permutation and some complex factors.

3. SPATIAL TIME-FREQUENCY DISTRIBUTIONS

The discrete-time form of Cohen's class of Time-Frequency Distributions (TFD) for a signal $x(t)$ is given by [8]

$$D_{xx}(t,f) = \sum_{l,m=-\infty}^{\infty} \phi(m,l) x(t+m+l) x^*(t+m-l) e^{-j4\pi fl} \tag{2}$$

where t and f represent the time index and the frequency index, respectively. The kernel $\phi(m,l)$ characterizes the distribution and is a function of both the time and lag variables. The cross-TFD of two signals $x_1(t)$ and $x_2(t)$ is defined by

$$D_{x_1x_2}(t,f) = \sum_{l,m=-\infty}^{\infty} \phi(m,l) x_1(t+m+l) x_2^*(t+m-l) e^{-j4\pi fl} \tag{3}$$

Expressions (2) and (3) are used to define the following data *spatial time-frequency distribution* (STFD) *matrix*,

$$\mathbf{D_{xx}}(t,f) = \sum_{l,m=-\infty}^{\infty} \phi(m,l) \mathbf{x}(t+m+l) \mathbf{x}^H(t+m-l) e^{-j4\pi fl} \tag{4}$$

where $[\mathbf{D_{xx}}(t,f)]_{ij} = D_{x_ix_j}(t,f)$, for $i,j = 1,\cdots,m$ and the superscript ^{H} denotes the transpose conjugate operator.

Under the linear data model (1) and assuming a noise-free environment, the STFD matrix takes the following structure:

$$\mathbf{D_{xx}}(t,f) = \mathbf{A}\mathbf{D_{ss}}(t,f)\mathbf{A}^H \tag{5}$$

where $\mathbf{D_{ss}}(t,f)$ is the signal TFD matrix whose entries are the auto- and cross-TFDs of the sources.

Auto-source TFD: We define the auto-source TFD by

$$\mathbf{D}^a_{\mathbf{ss}}(t,f) = \mathbf{D_{ss}}(t,f) \text{ for auto-term TF points} \tag{6}$$

THE WORK BY M. AMIN IS SPONSORED BY ONR, GRANT # N00014-98-0176

Since the off-diagonal elements of $\mathbf{D_{ss}}(t,f)$ are cross-terms, the auto-source TFD matrix is quasi[1] diagonal for each TF point that corresponds to a true power concentration, i.e. a source auto-term.

Cross-source TFD: We define the Cross-source TFD by

$$\mathbf{D}^c_{\mathbf{ss}}(t,f) = \mathbf{D_{ss}}(t,f) \text{ for cross-term TF points} \quad (7)$$

Since the diagonal elements of $\mathbf{D_{ss}}(t,f)$ are auto-terms, the cross-source TFD matrix is quasi anti-diagonal (i.e. its diagonal entries are close to zero) for each TF point that corresponds to a cross-term.

4. PROPOSED ALGORITHM

Let $\mathbf{W}$ denote an $m \times n$ matrix, such that $(\mathbf{WA})(\mathbf{WA})^H = \mathbf{UU}^H = \mathbf{I}$, i.e. $\mathbf{WA}$ is an $n \times n$ unitary matrix ($\mathbf{W}$ is referred to as the whitening matrix, since it whitens the signal part of the observations). Pre- and post-multiplying the TFD-matrices $\mathbf{D_{xx}}(t,f)$ by $\mathbf{W}$, lead to the *whitened TFD-matrices*, defined as:

$$\underline{\mathbf{D}}_{\mathbf{xx}}(t,f) = \mathbf{W}\mathbf{D_{xx}}(t,f)\mathbf{W}^H \quad (8)$$

From the definition of $\mathbf{W}$ and Eq.(5), we can express $\underline{\mathbf{D}}_{\mathbf{xx}}(t,f)$ as

$$\underline{\mathbf{D}}_{\mathbf{xx}}(t,f) = \mathbf{U}\mathbf{D_{ss}}(t,f)\mathbf{U}^H \quad (9)$$

Joint Diagonalization (JD): By selecting auto-term TF points, the data auto-source TFD will have the following structure,

$$\underline{\mathbf{D}}^a_{\mathbf{xx}}(t,f) = \mathbf{U}\mathbf{D}^a_{\mathbf{ss}}(t,f)\mathbf{U}^H \quad (10)$$

where $\mathbf{D}^a_{\mathbf{ss}}(t,f)$ is diagonal. The missing unitary matrix $\mathbf{U}$ is retrieved (up to permutation and phase shifts) by Joint Diagonalization (JD) of a combined set $\{\underline{\mathbf{D}}^a_{\mathbf{xx}}(t_i,f_i)|i = 1,\cdots,p\}$ of p auto-source TFD matrices. The incorporation of several auto-term TF points in the JD reduces the likelihood of having degenerate eigenvalues.

The joint diagonalization [4] of a set $\{\mathbf{M}_k|k = 1..p\}$ of $p\ \times$ matrices is defined as the maximization of the JD criterion:

$$C(\mathbf{V}) \stackrel{\text{def}}{=} \sum_{k=1}^{p}\sum_{i=1}^{n} |\mathbf{v}_i^H \mathbf{M}_k \mathbf{v}_i|^2 \quad (11)$$

over the set of unitary matrices $\mathbf{V} = [\mathbf{v}_1,\cdots,\mathbf{v}_n]$. An efficient joint approximate diagonalization algorithm exists in [4] and it is a generalization of the Jacobi technique [9] for the exact diagonalization of a single normal matrix.

Joint Anti-Diagonalization (JAD): By selecting cross-term TF points, the data cross-source TFD will have the following structure,

$$\underline{\mathbf{D}}^c_{\mathbf{xx}}(t,f) = \mathbf{U}\mathbf{D}^c_{\mathbf{ss}}(t,f)\mathbf{U}^H \quad (12)$$

where $\mathbf{D}^c_{\mathbf{ss}}(t,f)$ is anti-diagonal. The missing unitary matrix $\mathbf{U}$ is 'uniquely' (i.e. up to permutation and phase shifts) retrieved by Joint Anti-Diagonalization (JAD) of a combined set $\{\underline{\mathbf{D}}^c_{\mathbf{xx}}(t_i,f_i)|i = 1,\cdots,q\}$ of q cross-source TFD matrices.

The joint anti-diagonalization is explained by first noting that the problem of anti-diagonalization of a single $n \times n$ matrix $\mathbf{N}$ is equivalent[2] to the maximization of the criterion

$$C(\mathbf{N},\mathbf{V}) \stackrel{\text{def}}{=} -\sum_{i=1}^{n} |\mathbf{v}_i^H \mathbf{N} \mathbf{v}_i|^2 \quad (13)$$

over the set of unitary matrices $\mathbf{V} = [\mathbf{v}_1,\cdots,\mathbf{v}_n]$. Hence, JAD of a set $\{\mathbf{N}_k|k = 1..q\}$ of q $n \times n$ matrices is defined as the maximization of the JAD criterion:

$$C(\mathbf{V}) \stackrel{\text{def}}{=} \sum_{k=1}^{q} C(\mathbf{N}_k,\mathbf{V}) = -\sum_{k=1}^{q}\sum_{i=1}^{n} |\mathbf{v}_i^H \mathbf{N}_k \mathbf{v}_i|^2 \quad (14)$$

under the same unitary constraint. A Jacobi-like algorithm has been derived for the maximization of the JAD criterion (14).

Combined JD/JAD algorithm: The Combined joint diagonalization and joint anti-diagonalization of two sets $\{\mathbf{M}_k|k = 1..p\}$ and $\{\mathbf{N}_k|k = 1..q\}$ of $n \times n$ matrices is defined as the maximization of the JD/JAD criterion:

$$C(\mathbf{V}) \stackrel{\text{def}}{=} \sum_{i=1}^{n}\left(\sum_{k=1}^{p} |\mathbf{v}_i^H \mathbf{M}_k \mathbf{v}_i|^2 - \sum_{k=1}^{q} |\mathbf{v}_i^H \mathbf{N}_k \mathbf{v}_i|^2\right) \quad (15)$$

over the set of unitary matrices $\mathbf{V} = [\mathbf{v}_1,\cdots,\mathbf{v}_n]$. A Jacobi-like algorithm has been derived[3] for the maximization of the JD/JAD criterion (15).

Selection procedure: The success of the JD or JAD of TFD matrices in determining the unitary matrix $\mathbf{U}$ depends strongly on the correct selection of the auto-term and cross-term points. A simulation example is given in Section 6 to emphasize this point. Therefore, it is crucial to have a selection procedure that is able to distinguish between auto-term and cross-term points based only on the TFD matrices of the observations. Here, we propose a selection approach that exploits the anti-diagonal structure of the cross-source TFD matrices. More precisely, we have

$$\begin{aligned}\text{Trace}(\underline{\mathbf{D}}^c_{\mathbf{xx}}(t,f)) &= \text{Trace}(\mathbf{U}\mathbf{D}^c_{\mathbf{ss}}(t,f)\mathbf{U}^H) \\ &= \text{Trace}(\mathbf{D}^c_{\mathbf{ss}}(t,f)) \approx 0.\end{aligned}$$

Based on this observation, we derive the following testing procedure:

$$\text{if } \frac{\text{Trace}(\underline{\mathbf{D}}_{\mathbf{xx}}(t,f))}{\text{norm}(\underline{\mathbf{D}}_{\mathbf{xx}}(t,f))} < \epsilon \longrightarrow \text{ decide that } (t,f) \text{ is a cross-term}$$

$$\text{if } \frac{\text{Trace}(\underline{\mathbf{D}}_{\mathbf{xx}}(t,f))}{\text{norm}(\underline{\mathbf{D}}_{\mathbf{xx}}(t,f))} > \epsilon \longrightarrow \text{ decide that } (t,f) \text{ is an auto-term}$$

where ϵ is a 'small' positive real scalar. The correct choice of the value of ϵ is still under investigation. An ad-hoc value ($\epsilon = 0.1$) has been used in our simulation experiment.

Identification Procedure: Equations (5-15) constitute the proposed blind source separation approach which is summarized by the following steps:

- Determine the whitening matrix $\hat{\mathbf{W}}$ from the eigen decomposition of an estimate of the covariance matrix of the data (see [5] for more details).
- Compute the TF distribution of the array output according to (4).
- Select a set of TF points (usually corresponding to the high amplitude points of the signal TF transform) then distinguish between auto-term and cross-term points using the above selection procedure.
- Determine the unitary matrix $\hat{\mathbf{U}}$ by maximizing the JD/JAD criterion applied to the whitened TFD matrices computed at the selected TF points.

[1]Because of the finite window effect, the sidelobes from cross terms would often prevent a full diagonal structure of the above matrix.

[2]This is due to the fact that the Frobenius norm of a matrix is constant under unitary transform, i.e. norm($\mathbf{N}$) =norm($\mathbf{V}^H\mathbf{NV}$).

[3]Details of the JAD and the JD/JAD algorithms are omitted here due to space limitation.

- Obtain an estimate of the mixture matrix $\hat{\mathbf{A}}$ as $\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#}\hat{\mathbf{U}}$, where the superscript $\#$ denotes the pseudo-inverse, and an estimate of the source signals $\hat{\mathbf{s}}(t)$ as $\hat{\mathbf{s}}(t) = \hat{\mathbf{U}}^H\mathbf{W}\mathbf{x}(t)$.

5. DISCUSSION

We present below several comments to obtain more insight into the proposed blind source separation (BSS) method:

1) In practice, the cross-source TFD matrices will not be purely anti-diagonal. This is because some auto-terms, through their side lobes or main lobes, will intrude over the cross-term regions. The cross-terms will be however the dominant components. This situation is similar to the earlier work on joint diagonalization of TFD matrices selecting auto-term points [5], where the source auto-TFD matrices are not purely diagonal because of cross-term intrusion. This impairment is mitigated by the joint approximation property of the JD/JAD algorithm and its robustness.

2) In contrast to the previously proposed Time-Frequency Separation (TFS) approach [5], the new proposed algorithm, allows selecting TF points in both auto-term and cross-term regions, as both regions provide separate key information about the signals, and in turn provides improved separation performance (see simulation example in Section 6).

3) The cross-term issues rise in both the TF and ambiguity domain [7] based BSS. Therefore the proposed blind separation method can be applied to both domains.

4) The smoothing kernel reduces the cross terms by re-distributing them across the t-f domain, rather than being concentrated at specific points where they can be confused with true energy. This re-distribution process will place some of these terms on the top of the autoterms, rendering the TFD matrix, constructed from autoterms, non-diagonal. So, in many cases, the Wigner-Ville distribution is more robust than other distributions.

5) The JAD algorithm provides an estimate of the unitary matrix $\mathbf{U}$ and cross-source TFD matrices $\mathbf{D}^c_{\mathbf{ss}}(t, f)$. A necessary condition for the uniqueness of the solution is that the number of equations is greater than the total number of unknown parameters. This leads to the condition $q \geq n - 1$, where q is the number of the $n \times n$ matrices to be processed by the JAD algorithm. Note that for the JD algorithm, we need only $p \geq 1$ as a necessary condition. A detailed study on the identifiability of the problem will be given elsewhere.

6. SIMULATION

First experiment: We consider a uniform linear array of $m = 3$ sensors having half wavelength spacing and receiving signals from $n = 2$ sources in the presence of white Gaussian noise. The sources arrive from different directions $\phi_1 = 10$ and $\phi_2 = 20$ degrees. The emitted signals are two chirps. The kernel used for the computation of the TFDs is the Wigner-Ville kernel. Eight TFD matrices are considered.

The performance is characterized in terms of signal rejection. The mean rejection level is defined as

$$\mathcal{I} \stackrel{\text{def}}{=} \sum_{q \neq p} E|\hat{\mathbf{A}}^{\#}\mathbf{A}_{pq}|^2 \qquad (16)$$

We compare in Figure 1 the performance of the TFS algorithm proposed in [5] and the proposed algorithm for a signal-to-noise ratio (SNR) in the range [5 - 20 dB]. The mean rejection levels are evaluated here over 100 Monte Carlo runs with 1024 samples. It turns out that, in this case, the new algorithm performs slightly better than the TFS algorithm.

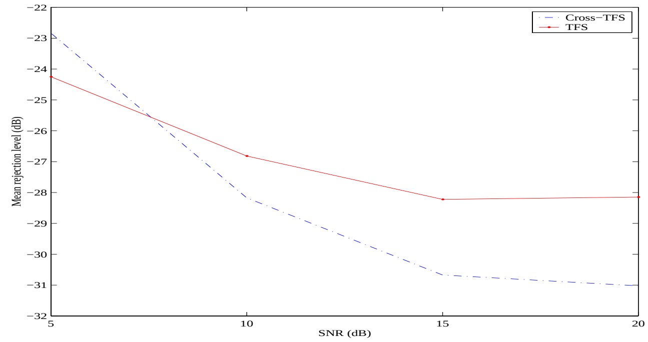

Figure 1: Mean rejection level vs SNR.

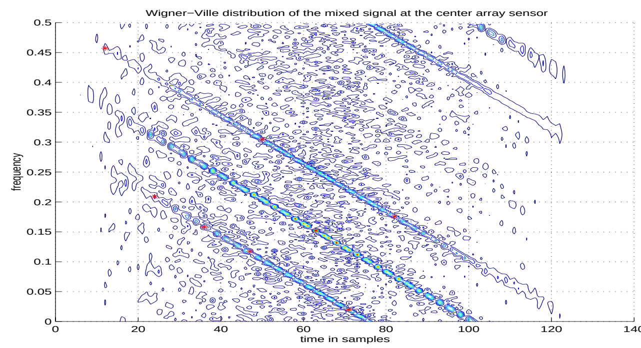

Figure 2: WVD of two mixed signals at 0 dB SNR.

Second experiment: In this experiment, we consider two chirp signals ($n = 2$), depicted by

$$\begin{aligned} s_1(t) &= \exp(-j0.004\pi t^2) \\ s_2(t) &= \exp(-j0.004\pi t^2 - j\pi 0.4t), \end{aligned}$$

impinging on an array of m=5 sensors at 30 and 60 degrees. White Gaussian noise was added, leading to an SNR of 0dB. The Wigner-Ville distribution (WVD) of the mixture at the middle sensor is depicted in Figure 2. From Figure 2, we selected eight arbitrary TF points, among which only one was a cross-term. Using the algorithm based on JD only, suggested in [5], we obtain the estimated signals, described by WVDs, shown in Figure 3. The figure clearly shows that the algorithm had failed. An estimate of the mean rejection level was as high as 3 dB. However, if we apply the proposed method from Section 4, the results are more promising, leading to Figure 4 with a signal rejection level estimate of -26 dB. One may

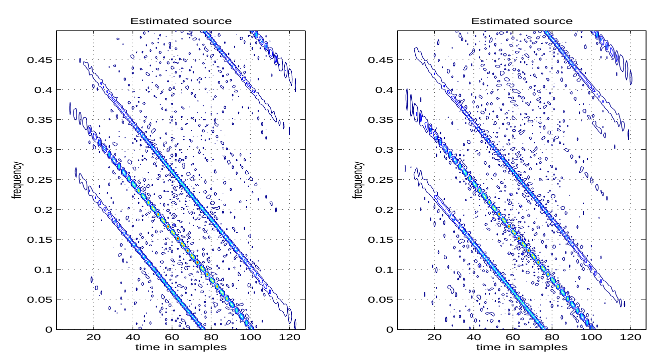

Figure 3: WVDs of the two chirps using JD with seven auto-terms and one cross-term.

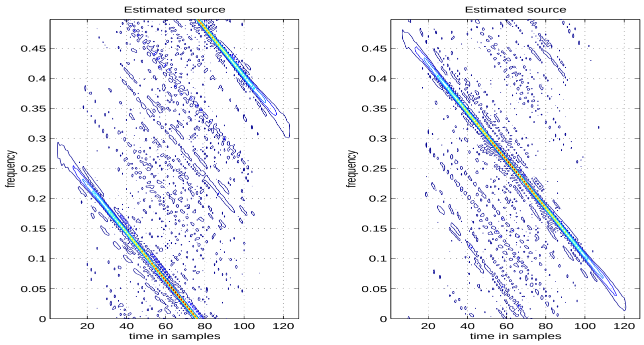

Figure 4: WVDs of the two chirps using JD/JAD with seven auto-terms and one cross-term.

suggest to remove the cross-term, identified with the method suggested in Section 4, and run a JD algorithm based on the auto-terms only. The result of this approach is depicted in Figure 5. Although visually not noticeable, this approach does not perform as well as the JD/JAD method as the signal mean rejection level estimate is higher by circa 2 dB (-24dB). This loss of mean rejection level can be more severe in other situations.

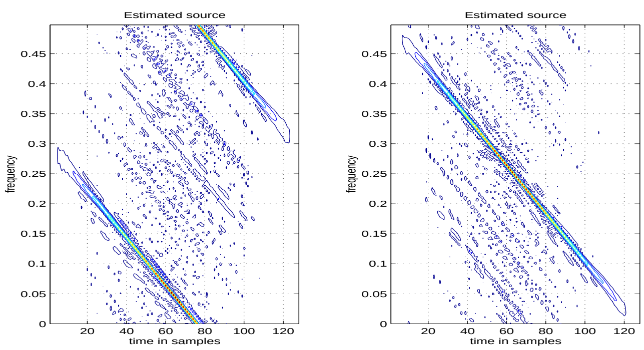

Figure 5: WVDs of the two chirps using JD with the seven auto-terms from above.

Third experiment: Here, we use three sources signals at 20 dB SNR (the third source being at 90 degrees). The number of antenna elements is again $m = 5$. The WVD of the mixture at the middle sensor is depicted in Figure 6. Six TF points are considered, among which five are cross-terms. The procedure described above

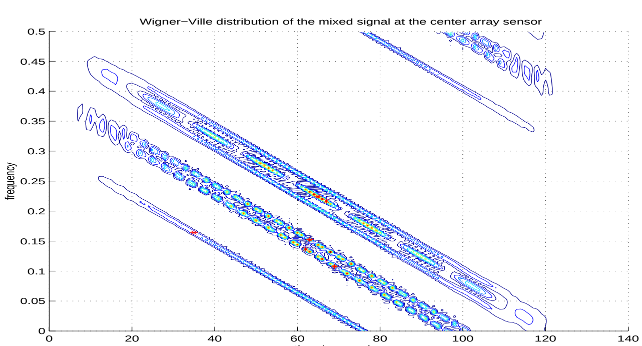

Figure 6: WVD of the mixture of three chirps at 20 dB SNR.

was used with $\epsilon = 0.1$ to identify the auto-terms and cross-terms and the JD/JAD criterion ran.The result is depicted in Figure 7. It is clearly seen that the method performs very well with a rejection mean level estimate of -28 dB.

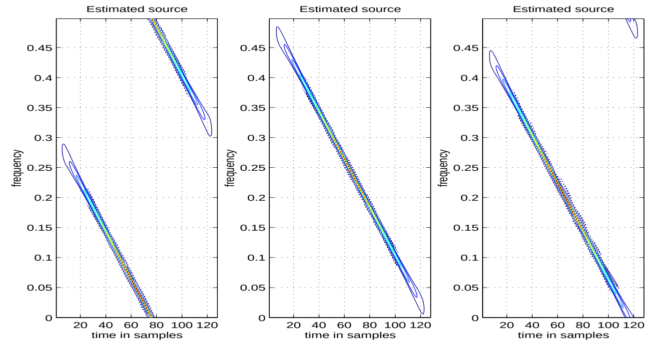

Figure 7: WVDs of the three chirps using JD/JAD with one auto-term and five cross-terms.

7. CONCLUSIONS

In this paper, the problem of blind separation of linear spatial mixtures of non-stationary source signals based on time-frequency distributions has been investigated. A solution based on the hybrid diagonalization / anti-diagonalization of a combined set of spatial time-frequency distribution matrices, selected in both the auto-term and cross-term regions, has been proposed. The identifiability problem as well as the problem of TF points selection have been discussed. Numerical simulations are provided to illustrate the effectiveness of the proposed approach.

8. REFERENCES

[1] P. Comon, “Independent component analysis. a new concept,” *Sig. Proc.*, vol. 36, pp. 287–314, 1994.

[2] A. Cichocki, W. Kasprzak, S. Amari, “Multi-layer neural networks with a local adaptive learning rule for blind separation of source signals,” *Proc. Nolta, Las Vegas, USA*, Dec. 1995.

[3] A. Belouchrani, J.-F. Cardoso, “Maximum likelihood source separation by the EM technique: Deterministic and stochastic implementation,” in *Proc. Nolta, Las Vegas, USA*, Dec. 1995.

[4] A. Belouchrani, K. Abed-Meraim, J.-F Cardoso, E. Moulines, “A blind source separation technique using second order statistics,” *IEEE Trans. on SP*, pp. 434–444, Feb. 1997.

[5] A. Belouchrani, M. G. Amin, “Blind source separation based on time-frequency signal representation,” *IEEE Trans. on SP*, pp. 2888–2898, Nov. 1998.

[6] Y. Zhang, M. G. Amin, “Signal averaging of time-frequency distributions for signal recovery in uniform linear arrays,” *IEEE Trans. on SP*, pp. 2892–2902, Oct. 2000.

[7] M. G. Amin, A. Belouchrani, Y. Zhang, “The spatial ambiguity function and its applications,” *IEEE Sig. Proc. Letters*, pp. 138–140, June 2000.

[8] L. Cohen, *Time-frequency analysis*. Prentice Hall, 1995.

[9] G.H. Golub, C.F. Van Loan, *Matrix computations*. The Johns Hopkins University Press, 1989.

[10] P. Comon and E. Moreau, “Improved contrast dedicated to blind separation in communications”, *Proc ICASSP, Munich*, pp. 3453–3456, 1997.

[11] J.F. Cardoso, “Infomax and maximum likelihood for blind source separation”, *IEEE Sig. Proc. Letters*, 4(4): pp. 112–114, Apr. 1997.