BLIND SOURCE SEPARATION BASED ON THE FRACTIONAL FOURIER TRANSFORM

Sharon Karako-Eilon[(1)], Arie Yeredor[(1)], David Mendlovic[(2)]

[(1)]Department of Electrical Engineering - Systems
[(2)]Department of Electrical Engineering - Physical Electronics
Tel-Aviv University, Tel-Aviv, 69978, ISRAEL
e-mail: <sharon, arie, mend>@eng.tau.ac.il

ABSTRACT

Different approaches have been suggested in recent years to the blind source separation problem, in which a set of signals is recovered out of its instantaneous linear mixture. Many widely-used algorithms are based on second-order statistics, and some of these algorithms are based on time-frequency analysis. In this paper we set a general framework for this family of second-order statistics based algorithms, and identify some of these algorithms as special cases in that framework. We further suggest a new algorithm that is based on the fractional Fourier transform (FRT), and is suited to handle non-stationary signals. The FRT is a tool widely used in time-frequency analysis and therefore takes a considerable place in the signal-processing field.

In contrast to other blind source separation algorithms suited for the non-stationary case, our algorithm has two major advantages: it does not require the assumption that the signals' powers vary over time, and it does not require a pre-processing stage for selecting the points in the time-frequency plane to be considered.

We demonstrate the performance of the algorithm using simulation results.

1. INTRODUCTION

Many approaches have been suggested in recent years to the Blind Source Separation (BSS) problem. Among them, some of the most popular are algorithms based on Second-Order Statistics (SOS). This approach is preferable to other BSS techniques since it generally involves a more statistically-stable estimation comparing to the estimation of higher order statistics, such as cumulants.

However, SOS-based approaches are not applicable when the source signals are stationary with similar spectra. Nevertheless, when the sources are non-stationary, or stationary with different spectra (or non-stationary with different time-frequency structures), SOS-based separation is possible. Moreover, while the common assumption used in BSS is that the sources are statistically independent, a weaker assumption, namely that the sources are mutually uncorrelated, is sufficient for the SOS-based methods.

Algorithms for the case of stationary sources are, e.g., the Second Order Blind Identification algorithm (SOBI algorithm, by Belouchrani *et al.* [1]) or improvements thereof for the Gaussian case such as the Weights-Adjusted SOBI (WASOBI) algorithm [2]. For non-stationary sources, some commonly used algorithms are, e.g., Cardoso and Pham's likelihood-based algorithm for the Gaussian case [3] and a family of algorithms based on time-frequency (T-F) analysis [4-7], that do not employ any assumption on the statistical distribution of the sources. All of the above algorithms use joint diagonalization (or joint anti-diagonalization) of estimated "generalized correlation" matrices, a notion on which we shall further elaborate in section 2.

Fractional Fourier Transform (FRT) is one of the tools used in the T-F analysis field [9]. In the past decade the FRT has attracted the attention of researchers in the optics and signal processing community, being a generalization of the classical Fourier transform. In this paper we propose to use the FRT in the framework of SOS-based BSS.

The paper is structured as follows: This introduction is followed in section 2 by a presentation of a general framework for SOS-based BSS, which is closely related to the framework introduced in [8]. Additionally, we identify the existing algorithms mentioned above as special cases of that framework. A brief overview of the fractional Fourier transform and its properties is provided in section 3. We then proceed in section 4 to introduce a special family of SOS-based BSS algorithms, involving the FRT operation. Such algorithms are suitable for the case of non-stationary signals, exploiting their T-F structures without involving complex T-F analysis. Consequently, they can accommodate models in which the non-stationary

sources have constant power over time, in contrast to the methods proposed in [3]. We present our proposed BSS algorithm using the FRT as a special case of the general framework. In section 5 we show some experimental simulation results, demonstrating the performance of our algorithm. Possible improvements to our algorithm are proposed in section 6, and conclusions are summarized in section 7.

2. THE GENERAL FRAMEWORK

The basic noiseless model commonly used in most BSS approaches is:

$$\underline{x}(t) = A\underline{s}(t) \qquad t = 1,2,...T$$

where $\underline{s}(t) = [s_1(t),...s_K(t)]^T$ is the vector of K zero-mean mutually uncorrelated source signals (assumed real-valued for simplicity), and $\underline{x}(t) = [x_1(t),...x_K(t)]^T$ is the observations vector obtained from a K-dimensional sensors array. $A \in \Re^{KxK}$ is a non-singular unknown matrix usually referred to as the *"mixing matrix"*. For convenience, we define the following matrix relation:

$$X = A \cdot S$$

where $X = [\underline{x}(1)\underline{x}(2)...\underline{x}(T)]$ is the observation matrix and $S = [\underline{s}(1)\underline{s}(2)...\underline{s}(T)]$ is the sources matrix.

Let P denote any (arbitrary) TxT matrix, and define $\Lambda_S^{(P)} \triangleq E\{SPS^T\}$. By virtue of the sources being mutually uncorrelated, $\Lambda_S^{(P)}$ is diagonal $\forall P$. Further defining $\widetilde{R}^{(P)} \triangleq E\{XPX^T\}$ as the observations' *"Generalized Correlation"* matrix with respect to the matrix P, we have

$$\widetilde{R}^{(P)} \triangleq E\{XPX^T\} = E\{ASPS^TA^T\} = A \cdot \Lambda_S^{(P)} \cdot A^T \quad (1)$$

Thus, for any given set of matrices $P_1,...,P_L$, the matrix A can, under some mild assumptions on the matrices $\Lambda_S^{(P_1)},...\Lambda_S^{(P_L)}$, be extracted from the (exact) joint diagonalization of the set $\widetilde{R}^{(P_1)},\widetilde{R}^{(P_2)},...\widetilde{R}^{(P_L)}$, up to arbitrary scaling and permutation on its columns. However, in practical situations the generalized correlation matrices $\widetilde{R}^{(P_1)},\widetilde{R}^{(P_2)},...\widetilde{R}^{(P_L)}$ are usually unknown. Nevertheless, for certain choices of matrices $P_1,...,P_L$ these generalized correlation matrices can be estimated from the data, leading to an estimate of A from the approximate joint diagonalization of these estimated matrices.

The following special cases can be easily identified:

1. Suppose that matrix P_l is all zeros, except for the $(l-1)th$ sub-diagonal, which contains the element $\frac{1}{T-l+1}$ throughout, namely:

$$P_l(i,j) = \begin{cases} 0, & i-j \neq l-1 \\ \dfrac{1}{T-l+1}, & i-j = l-1 \end{cases}$$

for any $l = 1,2,...L$, $1 \le i,j \le T$. Consequently, if the source signals are stationary, then $\widetilde{R}^{(P_l)} = R_X[l]$, where $R_X[l] = E\{\underline{x}(t+l)\underline{x}^T(t)\}$ is the classical correlation matrix of the observations. Moreover, $\hat{\widetilde{R}}_X^{(P_l)} = XP_lX^T$ is an unbiased consistent estimate of $\widetilde{R}^{(P_l)}$. Therefore, joint diagonalization of these estimated generalized correlation matrices leads to the SOBI algorithm.

2. Suppose that the interval $[1,T]$ is divided into L subintervals T_l, each containing $\|T_l\|$ samples, and let P_l be diagonal with:

$$P_l(j,j) = \begin{cases} 0, & j \notin T_l \\ 1/\|T_l\|, & j \in T_l \end{cases}$$

for $l = 1,2,...L$, $1 \le i,j \le T$. Consequently, if the source signals are non-stationary but with constant variance profiles (powers) in each subinterval, then $\widetilde{R}^{(P_l)}$ is the classical correlation matrix (at lag zero) of the l-th interval of X. As in the previous case, $\hat{\widetilde{R}}_X^{(P_l)} = XP_lX^T$ is an unbiased consistent estimate of $\widetilde{R}^{(P_l)}$. Therefore, approximate joint diagonalization of these estimated generalized correlation matrices leads to Pham and Cardoso's "Block Gaussian Likelihood" algorithm [3].

3. For the T-F analysis based algorithms, the matrix $P_{t,f}$ is:

$$P_{t,f}(i,j) = \phi\left(\frac{i+j}{2} - t, \frac{1-j}{2}\right)e^{-j2\pi f(i-j)} \quad (2)$$

for any (t,f) point in the T-F plane, where ϕ is some smoothing kernel (see [4]). In this case, assuming the observations are real-valued, we note that $\hat{R}_X^{P_{t,f}} = XP_{t,f}X^T$ is exactly the matrix $D_{XX}(t,f)$ defined in the BSS algorithm based on T-F analysis in [4]. To observe this relation, note that the (i,j) element of the matrix $D_{XX}(t,f)$ defined in [4] is:

$$[D_{XX}(t,f)](i,j) = \\ = \sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\phi(l,m)x_i(t+m+l)x_j(t+m-l)e^{-j4\pi fl} \quad (3)$$

In practical situations, we do not have all the observations' samples in the range $[-\infty,\infty]$, and therefore the summation in the above formula is truncated to a finite sum, where the convention is that the unavailable observations outside the given interval $[1,T]$ are zeros. By substituting $l=\frac{p-q}{2}, m=\frac{p+q}{2}-t$ into eq. (3) we get:

$$[D_{XX}(t,f)](i,j) = \\ = \sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty}\phi\left(\frac{p+q}{2}-t,\frac{p-q}{2}\right)e^{-j2\pi f(p-q)}x_i(p)x_j(q)$$

In this equivalent formula the same convention holds for observations outside the given interval, hence we can rewrite it:

$$[D_{XX}(t,f)](i,j) = \\ = \sum_{p=1}^{T}\sum_{q=1}^{T}\phi\left(\frac{p+q}{2}-t,\frac{p-q}{2}\right)e^{-j2\pi f(p-q)}x_i(p)x_j(q)$$

If we choose the matrices $P_{t,f}$ as defined in eq. (2), we get:

$$[D_{XX}(t,f)](i,j) = \sum_{p=1}^{T}\sum_{q=1}^{T}P(p,q)x_i(p)x_j(q)$$

which is the (i,j) element of the matrix $XP_{t,f}X$. Hence, approximate joint diagonalization of these estimated generalized correlation matrices leads to the T-F based algorithm presented in [4] and, with some modifications, to a similar algorithm presented in [6,7].

3. FRACTIONAL FOURIER TRANSFORM (FRT)

The FRT is a generalization of the conventional Fourier transform and has a history in mathematical physics and in digital signal processing (The interested reader is referred to [9] for a comprehensive overview). Basically, it is a one-parameter subclass of the class of linear canonical transforms. This parameter is called the *fractional order* of the transform, usually denoted as *p*. Due to some of the FRT properties, we can usually limit the discussion to the range $0 \le p \le 1$. When $p=0$, the FRT coincides with the identity operator, and when $p=1$ it coincides with the Fourier operator. Therefore, denoting by $f_p(u)$ the FRT of order p of a signal $f(x)$, the free variable u can be interpreted as some hybrid time/frequency variable: when $p=0$ it is a time variable, and when $p=1$ it is a frequency variable. As *p* takes values from 0 to 1, the interpretation of u changes gradually from "time" to "frequency", reflecting temporal changes in the frequency content of the transformed signal.

The FRT has several equivalent definitions. One of them is a kernel-based integral transformation of the form:

$$f_p(u) = F^p[f](x) = \\ = C_\alpha\int f(x)\exp\left[i\pi\frac{u^2+x^2}{\tan\alpha}-2i\pi\frac{ux}{\sin\alpha}\right]dx \quad (4)$$

where $\alpha=\frac{p\pi}{2}$ and:

$$C_\alpha = \frac{\exp\left[-i\left(\frac{\pi\cdot sign(\alpha)}{4}-\frac{\alpha}{2}\right)\right]}{|\sin\alpha|^{1/2}}$$

is constant given the fractional order.

Based on the Wigner chart, an alternative interpretation for the FRT has been proposed and consolidated [10]: The *p*th order fractional Fourier transform corresponds to the projection on the horizontal axis (in the T-F plane) of the signal's rotated Wigner distribution. The rotation is in the clockwise direction by an angle of α:

$$|f_p(x)|^2 = \int W_{f_p}(x,\nu)d\nu$$
$$W_{f_p}(x,\nu) = W_f(x\cos\alpha-\nu\sin\alpha, x\sin\alpha+\nu\cos\alpha)$$

where W_f is the Wigner transform. It can be shown [9] that both definitions are equivalent.

Following are some of the fundamental properties of the fractional Fourier transform:

- Linearity:
$$F^P\{c_1f_0(u)+c_2g_0(u)\} = c_1F^P\{f_0(u)\}+c_2F^P\{g_0(u)\}$$
- Inverse:
$$(F^p)^{-1} = F^{-p}$$
- Unitarity:
$$(F^p)^{-1} = (F^p)^H$$
- Generalized Parseval's theorem:
$$\int f_0(x)g_0^*(x)dx = \int f_p(u)g_p^*(u)du$$

We shall note also that a discrete counterpart of the above transformation exists [9], known as the Discrete Fractional Fourier Transform (DFRT) [9]. Despite some difficulties in obtaining a closed-form definition of the discrete transform, an algorithm for computation of the kernel matrix was presented by Ozaktas *et al.* [11]. This algorithm, along with a fast algorithm for digital

computation of an approximation of the continuous transformation [12], enables us to use the FRT tool in our discrete framework.

Denoting the kernel matrix of the DFRT of order p by F^P, we can use the following notation for the discrete transform:

$$\underline{f}_p = F^P \cdot \underline{f}$$

where $\underline{f}$ and $\underline{f}_p$ denote the input series (vector) and the transformed series (vector), respectively.

4. BSS BASED ON FRT

In this section, in order to simplify our notations, we assume the signals are arranged as:

$S = \left[\underline{s}_1\ \underline{s}_2 \ldots \underline{s}_K\right]$, where $\underline{s}_i = \left[s_i(1)s_i(2)\ldots s_i(T)\right]^T$

and

$X = \left[\underline{x}_1\ \underline{x}_2 \ldots \underline{x}_K\right]$, where $\underline{x}_i = \left[x_i(1)x_i(2)\ldots x_i(T)\right]^T$

Hence, the modified (actually transposed) model is

$$X = SA^T$$

and the generalized correlation matrices are

$$\widetilde{R}^{(P_l)} = E\left\{X^T P_l X\right\}$$

Our selection of P_l is the following: we choose L values of the fractional order in the range $[0,1]$, denoted as $p_1, p_2, \ldots, p_L$. Denoting $D_l \triangleq F^{p_l}$, the set of matrices P_l is defined by

$$P_l \triangleq D_l^H \cdot W_l \cdot D_l$$

where $W_l = diag\left(w_l[1]w_l[2]\ldots w_l[T]\right)$ and $w_l[1]\ldots w_l[T]$ is a windowing series, containing a window of length T_l positioned differently for each l, and $diag(\cdot)$ is in the vector-to-matrix sense in MATLAB® notation. As a result of this selection, the generalized correlation matrices are:

$$\widetilde{R}^{(P_l)} = E\left\{X^T P_l X\right\} = E\left\{X^T D_l^H W_l D_l X\right\} \quad (5)$$

Denoting $Y^{p_l} = F^{p_l} \cdot X$ as the FRT of the observations, we can rewrite eq. (5):

$$\widetilde{R}^{(P_l)} = E\left\{\left(Y^{P_l}\right)^H W_l Y^{P_l}\right\} \quad (6)$$

As usual, these matrices are generally unknown. We therefore need to estimate them, e.g., using

$$\hat{\widetilde{R}}^{(P_l)} = \left(Y^{P_l}\right)^H W_l Y^{P_l} = \\ = \sum_k w_l[k] \cdot \left(Y^{P_l}[k]\right)^* \cdot Y^{P_l}[k] \quad (7)$$

where the summation in eq. (7) needs only be computed over indices k in which the window $w_l[k]$ is non-zero.

In general, such an estimate may be quite inaccurate, since it basically estimates the mean using only one realization of the signals. However, the actual accuracy of such an estimate depends on the statistical-spectral properties of the signals, and, in a sense, on their ergodicity - in exactly the same way as other algorithms like [1], [3], [4] estimate different expectations out of one realization.

Finally, we obtain an estimate of the matrix A by approximate joint diagonalization of the estimated generalized matrices. To this end, several approaches can be taken. A commonly used approach is to impose "hard-whitening" on one of the generalized correlation matrices (usually, but not necessarily, the standard correlation matrix), and to transform accordingly all the generalized correlation matrices (see [1] for a detailed description of this stage). The next stage is to find a unitary matrix that attains the "best" (in the Least Squares (LS) sense) approximate joint diagonalization of the whitened set of generalized correlation matrices (e.g. by using [13]). An alternative way is to find a non-orthogonal matrix that attains the "best" (LS) approximate joint diagonalization of the original generalized correlation matrices (e.g. by using the AC-DC algorithm presented in [14]).

Note that the different location of the window for each different value of the fractional order is meant to capture different regions in the T-F structure of the signals. For example, if we had no windowing at all, then we would get identical matrices for every l. This is due to the fact that the FRT is an energy-conserving transform, so that $\sum_{k=1}^{T}\left|Y^{p_l}[k]\right|^2$ would be constant for all fractional orders.

The major advantage of this algorithm over other T-F analysis-based algorithms is the fact that we don't have to find and apply a sophisticated procedure on the data in order to select the desirable (t, f) points in the T-F plane (e.g. points corresponding to "cross-terms" and "auto-terms" of the data [4-6]). On the other hand, this algorithm requires the selection of fractional order values and the windows' locations, but the performance is expected to be relatively more robust with respect to arbitrary selection of these parameters, since they have more of a "global" rather than "local" effect.

5. SIMULATION RESULTS

In our simulation we used only $K = 2$ source signals and observations signals, and computed only $L = 2$ generalized correlation matrices, for simplicity.

The source signals are comprised of two subintervals of length $\frac{T}{2}$, where each subinterval is a stationary Gaussian

Moving Average (MA) process of the order $q=1$. The 1st and 2nd subintervals of $s_1(t)$ are MA(1) processes generated by passing a zero-mean white Gaussian noise with variance $\sigma^2=1$ through the systems $1+z^{-1}$ and $1-z^{-1}$, respectively. The 1st and 2nd subintervals of $s_2(t)$ are MA(1) processes generated by passing a zero-mean white Gaussian noise with variance $\sigma^2=1$ through the systems $1+0.5z^{-1}$ and $1-0.6z^{-1}$, respectively. Note that the power is constant over time in $s_1(t)$ but just nearly constant in $s_2(t)$. Nevertheless, the overall signals are non-stationary.

The mixing matrix was $A=\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}$.

The selected fractional order values were $p_1=0$ and $p_2=0.5$. We used a rectangular window $w_l[k]$, of length $T/2$, covering $t=T/2+1,T/2+2,...T$ for both $l=1,2$.

For reference, we compared to the performance of the "Block Gaussian Likelihood" (BGL) method [3], using the same data. For this algorithm we windowed the two consecutive length-$T/2$ subintervals, and computed two generalized (conventional in this case) correlation matrices, one for each subinterval. Note that the generalized correlation matrix $\widetilde{R}^{P_1}$ in the FRT algorithm is thus identical to $\widetilde{R}^{P_2}$ in the BGL algorithm.

Since we had only two generalized correlation matrices, $\widetilde{R}^{P_1}$ and $\widetilde{R}^{P_2}$, the matrix $\hat{A}$ was obtained by the exact joint diagonalization thereof, i.e. $\hat{A}$ is composed of the eigenvectors of the matrix $\widetilde{R}^{P_1}\cdot\left(\widetilde{R}^{P_2}\right)^{-1}$.

Performance was measured in terms of residual Interference-to-Signal Ratio (ISR) (see, e.g., [1], [4]) computed from the overall mixing-demixing matrix $\hat{A}^{-1}\cdot A$. The ISR was computed separately for $s_1(t)$ and $s_2(t)$ (denoted as ISR_1 and ISR_2, respectively), after resolving the permutation ambiguity. Results are shown in figure 1, in terms of averaged ISR for different values of T. For each different signals' length we averaged 200 Monte-Carlo experiments.

As could be expected, the advantage of our algorithm over the "Block Gaussian Likelihood" algorithm is quite evident in this example. The improved performance dwells mostly on the fact that the variance profiles of both sources are nearly similar, whereas the main difference lies in their rich temporal-spectral content, which can be captured by the FRT, by exploiting their T-F signatures in the fractional domain.

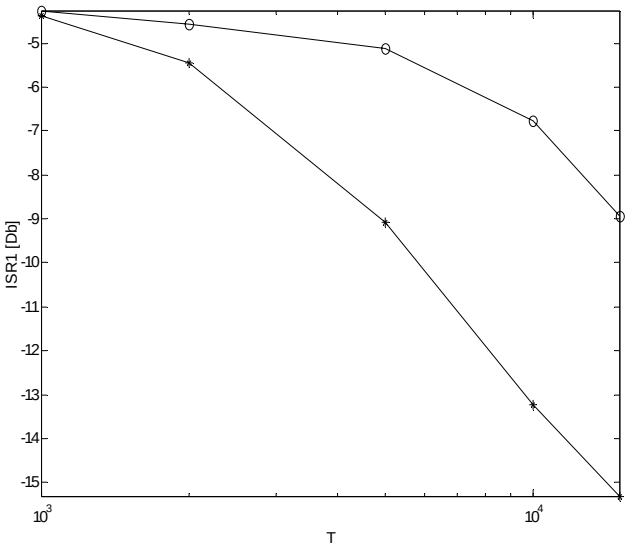

Fig. 1(a). ISR_1 versus the sample size (T)

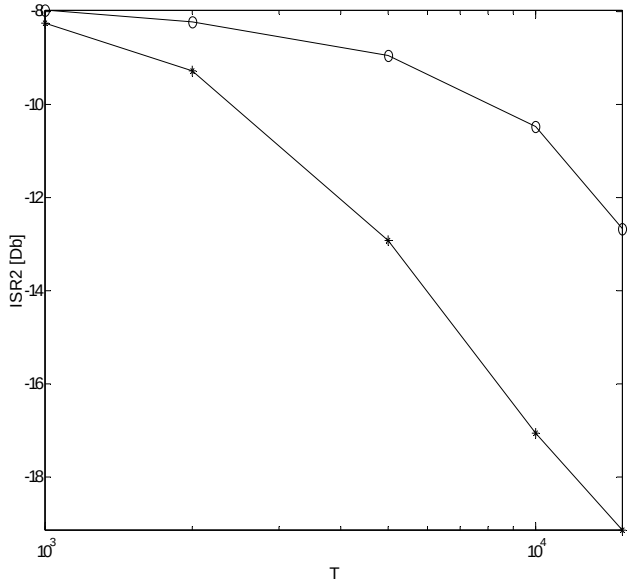

Fig. 1(b). ISR_2 versus the sample size (T)

6. POSSIBLE FURTHER IMPROVEMENT

When the sources are known to be MA sources of a maximum known order q, the estimation of $\widetilde{R}^{(P_l)}$ (and consequently of A) can be improved using the following strategy: Recall that (i,j) element of the matrix $\widetilde{R}^{(P_l)}$ is given by

$$\begin{aligned}&\tilde{R}^{(P_l)}(i,j)=\\&=E\left\{\underline{x}_i^{\,T}P_l\,\underline{x}_j\right\}=E\left\{Tr\left\{\underline{x}_i^{\,T}P_l\,\underline{x}_j\right\}\right\}=\\&=E\left\{Tr\left\{P_l\,\underline{x}_j\,\underline{x}_i^{\,T}\right\}\right\}=Tr\left\{P_l\cdot E\left\{\underline{x}_j\,\underline{x}_i^{\,T}\right\}\right\}=\\&=Tr\left\{P_l\cdot R_{x_jx_i}\right\},\end{aligned}\qquad(8)$$

where $Tr\{\cdot\}$ is the trace operator and $R_{x_jx_i}=E\left\{\underline{x}_j\,\underline{x}_i^{\,T}\right\}$ is the correlation matrix of the signals x_i, x_j.

Therefore, an alternative way to estimate $\tilde{R}^{(P_l)}$ is by using some more elaborate estimate $\hat{R}_{x_jx_i}$ of $R_{x_jx_i}$ than just $\underline{x}_j\,\underline{x}_i^{T}$ in:

$$\hat{\tilde{R}}^{(P_l)}(i,j)=Tr\left\{P_l\cdot\hat{R}_{x_jx_i}\right\}.\qquad(9)$$

This point of view allows us to incorporate a-priori knowledge on the sources into the algorithm, and thus improve the estimation of the generalized correlation matrices. For example, in the case of MA sources, each observation signal x_i is a linear combination of uncorrelated source signals with finite correlation of length q (double-sided). Thus, the true correlation matrices $R_{x_jx_i}$ are all zeros except on their $2q+1$ main diagonals (the main diagonal and the $\pm 1,2,...q$ sub-diagonals). Hence, zeroing all corresponding elements in the matrix $\hat{R}_{x_jx_i}=\underline{x}_j\,\underline{x}_i^{T}$ is expected to significantly improve the estimation of $\tilde{R}^{(P_l)}$.

7. CONCLUSIONS

We presented a new algorithm for BSS, based on the T-F representation of the signals, using the FRT. This algorithm is part of a whole family of SOS-based algorithms that jointly diagonalize a set of generalized correlation matrices. Our choice of these generalized correlation matrices takes into account the different representations of the signals in the different fractional domains of the T-F plane.

This algorithm is suited to handle non-stationary signals without the need of selecting specific (t,f) points in the T-F plane in a pre-processing stage.

We have demonstrated the potential advantage of our proposed algorithm in a simulation example, and proposed further directions for improving the performance. Additionally, on-going research is aimed at optimizing the selection of fractional orders and window functions.

REFERENCES

[1] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics", *IEEE Trans. on Signal Processing*, vol.45(2), pp.434-444, Feb. 1997.

[2] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting", *IEEE Signal Processing Letters*, vol. 7(7), pp. 197-200, July 2000.

[3] D.-T. Pham, and J.-F. Cardoso, "Blind separation of instantaneous mixtures of non-stationary sources", *IEEE Trans. on Signal Processing*, vol. 49(9), pp. 1837-1848, Sep. 2001.

[4] A. Belouchrani, and M. G. Amin, "Blind source separation based on time-frequency signal representations", *IEEE Trans. on Signal Processing*, vol.46(11), pp.2888-2897, Nov. 1998.

[5] A. Belouchrani, K. Abed-Meraim, M. G. Amin, and A. M. Zoubir, "Joint anti-diagonalization for blind source separation", *Proc. ICASSP 2001*, Salt Lake City, May 2001.

[6] L. Giulieri, N. Thirion-Moreau, and P.-Y Arques, "Blind source separation using bilinear and quadratic time-frequency representations", *Proc. ICA 2001*, pp. 486-491, Dec. 2001.

[7] L. Giulieri, N. Thirion-Moreau, and P.-Y Arques, "Blind source separation based on bilinear time-frequency representations: A performance analysis", *Proc. ICASSP 2002*, pp. II 1649-1652, Orlando, May 2002.

[8] A. Leshem, "Blind source separation using bilinear forms", Proceedings of *IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 102-106, Caesarea, Israel, June 1999.

[9] H. M. Ozaktas, Z. Zalevsky, and M.A. Kutay, *The Fractional Fourier Transform*, Wiley, England, 2001.

[10] A.W. Lohmann, "Image Rotation, Wigner rotation, and the fractional order Fourier transform", *J Opt Soc Am A*, vol. 10, pp. 2181-2186, 1993.

[10] C. Candan, M.A. Kutay, and H.M. Ozaktas, "The Discrete Fractional Fourier Transform", *IEEE Trans. on Signal Processing*, vol.48, pp. 1329-1337, 2000.

[11] H.M. Ozaktas, O. Arikan, M.A. Kutay, and G. Bozdagn, "Digital computation of the fractional Fourier transform", *IEEE Trans. on Signal Processing*, vol.44, pp. 2141-2150, 1996.

[12] J.-F. Cardoso, and A Souloumiac, "Jacobi angles for simultaneous diagonalization", SIAM journal on Matrix Analysis and Applications, vol. 17(1), pp. 161-164, Jan. 1996.

[13] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation", *IEEE Trans. On Signal Processing*, vol. 50 no. 7 pp. 1545-1553, July 2002.