A SIMPLE CLOSED-FORM SOLUTION FOR OVERDETERMINED BLIND SEPARATION OF LOCALLY SPARSE QUASI-STATIONARY SOURCES

Xiao Fu and Wing-Kin Ma

Department of Electronic Engineering, The Chinese University of Hong Kong,
Shatin, N.T., Hong Kong

E-mail: xfu@ee.cuhk.edu.hk, wkma@ieee.org

ABSTRACT

We consider the scenario of an unknown overdetermined instantaneous mixture of quasi-stationary sources. Blind source separation (BSS) under this scenario has drawn much attention, motivated by applications such as speech and audio separation. The ideas in the existing BSS works often focus on exploiting the time-varying statistics characteristics of quasi-stationary sources, through various kinds of formulations and optimization methods. In this paper, we are interested in further assuming that the sources exhibit some form of local sparsity, which is generally satisfied in speech. By exploiting this additional assumption, we show that there is a simple closed-form solution for the BSS problem. Simulation results based on real speech show that the proposed closed-form algorithm is computationally much lower than some existing BSS algorithms, while delivering a promising mean-square-error performance.

Index Terms— blind source separation, quasi-stationary sources, sparsity

1. INTRODUCTION

This paper focuses on the scope of blind source separation (BSS) of quasi-stationary signals (BSS-QSS) [1]. This class of BSS techniques assumes that sources exhibit time-varying second order statistics (SOSs) characteristics, and exploit such phenomenon to blindly identify the unknown mixing system, or its inverse. Speech and audio sources, for instance, are found to show significant variations of their statistical natures in time. In fact, an important application of BSS-QSS is speech separation in microphone array systems [2, 3]. BSS-QSS can be treated using either the parallel factor analysis (PARAFAC) framework [4] or the joint diagonalization (JD) framework [1,5], both of which are recognized as key techniques for BSS, with many interesting results. In this paper, we take a different approach, where we incorporate one more assumption with the source characteristics, namely, local sparsity, and take advantage of it to develop a simple alternative to BSS-QSS.

It is important for us to define local sparsity in the context of this paper. We assume that among a collection of SOSs estimated locally in time, there are particular time instants in which the SOSs are dominated by one source. However, we do not know where these locally dominant SOSs are, and the SOSs in the other time instants are composed of multiple source components. We will call this assumption the *local dominance* assumption in this paper, to distinguish it from some other sparsity assumptions, such as those based on sparse signal representation and L_1 minimization [6]. For speech signals, which often contain many unvoiced segments between utterances, local dominance is generally an appropriate assumption. It is interesting to note that local dominance is considered in several BSS frameworks. In BSS using time-frequency distribution (TFD), where the sources of interest are monocomponent or multicomponent signals (e.g., a chirp), sources are assumed to be totally disjoint in the time frequency domain [7, 8]; see [9] for an extension to the partially disjoint case. Exploitation of such disjointness, or local dominance, is found to be helpful in BSS, even in the underdetermined case. In BSS of non-negative signals, which recently finds meaningful applications in biomedical imagery and hyperspectral remote sensing [10, 11], there has been growing interest in using the local dominance assumption, together with signal non-negativity, to solve the BSS problems there. In particular, one major category of BSS methods in hyperspectral remote sensing relies on the so-called pure-pixel assumption [11], which is the same as local dominance.

The main contribution of this paper lies in exploiting the local dominance assumption, together with the underlying problem structures in BSS-QSS, to come up with an algebraically simple BSS-QSS algorithm. To be specific, we consider an overdetermined instantaneous mixture of locally dominant quasi-stationary signals. In this scenario, where one is permitted to apply prewhitening to orthonormalize the mixing system, we observe an interesting algebraic property with the local SOSs. That property can be turned to a criterion to identify locally dominant SOSs, consequently leading to the proposed BSS-QSS algorithm. The criterion mentioned above is easy to compute, as will be seen soon. Moreover, in our algorithm development, we address a modeling error issue caused by source correlations. A simple projection process is proposed to mitigate the modeling errors, and the effectiveness of this process is supported by analysis. We will demonstrate by simulations that the proposed algorithm is much more effective than the clustering-based methodology commonly used in TFD-BSS [9], in terms of both estimation performance and computational times.

2. PROBLEM FORMULATION

We follow a standard BSS-QSS formulation [1], wherein observed signals are linear instantaneous mixtures of sources

$$\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{a}_k s_k(t) = \mathbf{A}\mathbf{s}(t), \qquad t = 1, 2, \ldots, \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^N$ is the observed signal vector, $s_k(t)$ is the kth source signal, $\mathbf{s}(t) = [\, s_1(t), \ldots, s_K(t) \,]^T \in \mathbb{R}^K$, $\mathbf{a}_k \in \mathbb{R}^N$ is the system response vector of the kth source, $\mathbf{A} = [\, \mathbf{a}_1, \ldots, \mathbf{a}_K \,] \in$

This work is supported by a General Research Fund of Hong Kong Research Grant Council (CUHK415509).

$\mathbb{R}^{N\times K}$, and N and K are the number of sensors and sources, respectively. The source signals $s_k(t)$ are assumed to be zero-mean wide-sense quasi-stationary with frame length L; i.e.,

$$\mathrm{E}\{|s_k(t)|^2\} = d_k[m], \quad \text{for any } (m-1)L+1 \le t \le mL, \quad (2)$$

where m is the local time frame index. This means that the source second order statistics remain constant within local time intervals, and can change from one time frame to another. Let us define

$$\mathbf{R}[m] = \mathrm{E}\{\mathbf{x}(t)\mathbf{x}^T(t)\}, \quad \text{for any } (m-1)L+1 \le t \le mL,$$

to be the local covariance of the observed signal in frame m; such local covariances are, in practice, acquired by local averaging, like $\mathbf{R}[m] \approx \frac{1}{L}\sum_{t=(m-1)L+1}^{mL} \mathbf{x}(t)\mathbf{x}^T(t)$. By further assuming that $s_k(t)$ are mutually independent, one can deduce the following model

$$\mathbf{R}[m] = \sum_{k=1}^{K} d_k[m]\mathbf{a}_k\mathbf{a}_k^T, \quad m = 1,\ldots,M, \quad (3)$$

where M is the number of available frames.

Consider the following assumption, the local dominance assumption mentioned in the introduction:

A1) (*local dominance* [10]) For each source index k, there exists a time frame, indexed by m_k, such that $d_k[m_k] > 0$ and $d_\ell[m_k] = 0$ for all $\ell \neq k$.

Physically, A1) means that the sources exhibit some form of sparsity in the local covariance domain, such that there exist local time frames where only one source dominates. Hence, temporally sparse source signals may satisfy A1) well. Assumption A1) is also considered reasonable for speech sources, since speech signals often contain many unvoiced segments between utterances.

There is an intuitively simple way to exploit local dominance for blind identification of $\mathbf{A}$; the idea can be found in BSS-TFD [8, 9] and here we will call it the *clustering-based methodology* for convenience. Under A1), we have, at locally dominant points,

$$\mathbf{R}[m_k] = d_k[m_k]\mathbf{a}_k\mathbf{a}_k^T. \quad (4)$$

Hence, if we know where the locally dominant points are, then $\mathbf{a}_k$'s can be retrieved (up to a scaling factor) by obtaining the principal eigenvector of the locally dominant $\mathbf{R}[m]$. A practically working clustering-based algorithm generally follows the following steps: i) detect locally dominant points by evaluating the ranks of all $\mathbf{R}[m]$'s; ii) extract the principal eigenvector of each detected $\mathbf{R}[m]$; iii) apply a K-means clustering algorithm to the obtained principal eigenvectors to construct the mixing matrix $\mathbf{A}$.

In the next sections, we will explore a different approach for blind identification of $\mathbf{A}$.

3. THE PROPOSED ALGORITHM

The algorithm to be proposed also aims at finding the locally dominant points. However, we do not use rank as in the clustering-based methodology. Instead, we employ another idea that uses non-negativity of the local source variances $d_k[m]$, together with the locally dominant assumption A1), to come up with an alternative solution for identifying the locally dominant points. In addition, our algorithm is based on a successive search strategy and does not require clustering. The proposed algorithm is shown in Algorithm 1. As we can see, the arithmetic operations of this algorithm are simple, involving 2-norm computations and linear projections mostly.

Algorithm 1:

input : $\mathbf{R}[1],\ldots,\mathbf{R}[M]$;
$\mathbf{y}[m] = \mathrm{vec}(\mathbf{R}[m])$, $z[m] = \mathrm{Tr}(\mathbf{R}[m])$, $m = 1,\ldots,M$;
$\hat{\mathbf{h}}_1 = \mathbf{y}[\hat{m}_1]$, where $\hat{m}_1 \in \arg\max_{m=1,\ldots,M} \|\mathbf{y}[m]\|_2/z[m]$;
obtain $\hat{\mathbf{a}}_1$ as the principal eigenvector of $\mathrm{vec}^{-1}(\hat{\mathbf{h}}_1)$;
for $k = 2,\ldots,K$ **do**
  $\hat{\mathbf{h}}_k = \mathbf{y}[\hat{m}_k]$, where
  $\hat{m}_k \in \arg\max_{m=1,\ldots,M} \|\mathbf{P}^{\perp}_{\hat{\mathbf{H}}_{1:k-1}}\mathbf{y}[m]\|_2/z[m]$;
  obtain $\hat{\mathbf{a}}_k$ as the principal eigenvector of $\mathrm{vec}^{-1}(\hat{\mathbf{h}}_k)$.
end
output: $\hat{\mathbf{A}} = [\,\hat{\mathbf{a}}_1,\ldots,\hat{\mathbf{a}}_K\,]$.

We concentrate on the overdetermined scenario; i.e, $N > K$. In this scenario, we may assume that

A2) The mixing matrix $\mathbf{A}$ is orthonormal.

By using prewhitening [13], a popularized preprocessing procedure in BSS, we can turn a general overdetermined mixing model (more precisely, with a full-rank $\mathbf{A}$) to an equivalent model whose $\mathbf{A}$ is orthonormal. Now, from the local covariance model (3), let us apply vectorization to obtain

$$\mathbf{y}[m] \triangleq \mathrm{vec}(\mathbf{R}_m) = \sum_{k=1}^{K} d_k[m]\mathrm{vec}(\mathbf{a}_k\mathbf{a}_k^T) = \mathbf{H}\mathbf{d}[m] \quad (5)$$

where $\mathbf{H} = [\,\mathbf{h}_1,\ldots,\mathbf{h}_K\,]$, $\mathbf{d}[m] = [\,d_1[m],\ldots,d_K[m]\,]^T$, and $\mathbf{h}_k = \mathrm{vec}(\mathbf{a}_k\mathbf{a}_k^T) = \mathbf{a}_k \otimes \mathbf{a}_k$ with $\otimes$ being the Kronecker product. It can be easily shown that $\mathbf{H}$ is orthonormal, as far as A2) holds. Hence, we have that

$$\|\mathbf{y}[m]\|_2 = \|\mathbf{d}[m]\|_2 \le \|\mathbf{d}[m]\|_1 \quad (6)$$

where $\|\cdot\|_2$ and $\|\cdot\|_1$ are the 2-norm and 1-norm, respectively. The inequality in (6) follows from the basic linear algebra result that $\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1$, with equality being satisfied if and only if $\mathbf{x}$ is a scaled unit vector[1] [12]. Moreover, we notice that

$$\|\mathbf{d}_1[m]\|_1 = \sum_{k=1}^{K} d_k[m] = \mathrm{Tr}(\mathbf{R}[m]) \quad (7)$$

where, in the first equality above, we use the fact that $d_k[m] \ge 0$; recall from Eq. (2) that $d_k[m]$ are modeled as local source variances and thus must be non-negative. As for the second equality in (7), it is obtained by using $\mathbf{a}_k^T\mathbf{a}_k = 1$, which is implied by A2). From (6)-(7), we conclude that

$$\frac{\|\mathbf{y}[m]\|_2}{\mathrm{Tr}(\mathbf{R}[m])} \le 1 \quad (8)$$

and equality holds if and only if $\mathbf{d}[m]$ is a scaled unit vector; i.e., $\mathbf{y}[m]$ is locally dominant, taking the form $\mathbf{y}[m] = d_k[m]\mathbf{h}_k$ for some k. As a consequence, any

$$\hat{m} \in \arg\max_{m=1,\ldots,M} \frac{\|\mathbf{y}[m]\|_2}{\mathrm{Tr}(\mathbf{R}[m])} \quad (9)$$

[1] To be specific, $\mathbf{x}$ takes the form $\mathbf{x} = \alpha\boldsymbol{e}_i$ for some α, i, where $\boldsymbol{e}_i$ is a unit vector with $[\boldsymbol{e}_i]_k = 1$ for $k = i$ and $[\boldsymbol{e}_i]_k = 0$ for $k \neq i$.

corresponds to a locally dominant point. This result can be extended to provide successive search of all $\mathbf{h}_k$. To describe it, suppose that the first $k-1$ columns of $\mathbf{H}$ have been obtained. Let $\mathbf{H}_{1:k-1} = [\,\mathbf{h}_1, \ldots, \mathbf{h}_{k-1}\,]$, and $\mathbf{P}_{\mathbf{X}}^{\perp} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{\dagger}\mathbf{X}^T$ be the orthogonal complement projector of its argument $\mathbf{X}$. Following the derivations above, we can show that

$$\hat{m} \in \arg \max_{m=1,\ldots,M} \frac{\|\mathbf{P}_{\mathbf{H}_{1:k-1}}^{\perp}\mathbf{y}[m]\|_2}{\mathrm{Tr}(\mathbf{R}[m])}, \tag{10}$$

corresponds a locally dominant point of $\mathbf{h}_k, \mathbf{h}_{k+1}, \ldots,$ or $\mathbf{h}_K$, but not the previously found. Based on these results, we obtain Algorithm 1.

4. CROSS-CORRELATION EFFECTS AND REMEDY

There are practical situations where source signals may exhibit cross correlations in some frames. Although such cross correlations are often weak and intermittent, the subsequent modeling errors can result in performance deterioration. In the following, we will suggest a procedure of mitigating the cross-correlation effects.

Assuming that $s_k(t)$ may be correlated at times, the local covariance model in (3) should be modified as

$$\mathbf{R}[m] = \mathbf{A}\mathbf{D}[m]\mathbf{A}^T \tag{11}$$

where $\mathbf{D}[m] = \mathrm{E}\{\mathbf{s}(t)\mathbf{s}^T(t)\}$ for $(m-1)L+1 \le t \le mL$, and $\mathbf{D}[m]$ may contain non-zero off-diagonal elements. Let $d_k[m] = [\mathbf{D}[m]]_{kk}$ as before. Also let $e_{k\ell}[m] = [\mathbf{D}[m]]_{k\ell}$, $k \neq \ell$, which represent the cross-correlation components. Consequently, the model of $\mathbf{y}[m]$ in (5) is replaced by

$$\mathbf{y}[m] = \mathbf{H}\mathbf{d}[m] + \mathbf{G}\mathbf{e}[m] \tag{12}$$

where $\mathbf{G} = [\,\mathbf{G}_1, \ldots, \mathbf{G}_{K-1}\,]$, $\mathbf{G}_k = [\,\mathbf{g}_{k,k+1}, \ldots \mathbf{g}_{k,K}\,]$, $\mathbf{g}_{k,\ell} = \mathbf{a}_k \otimes \mathbf{a}_\ell + \mathbf{a}_\ell \otimes \mathbf{a}_k$, $\mathbf{e}[m] = [\,\mathbf{e}_1^T[m], \ldots, \mathbf{e}_{K-1}^T[m]\,]^T$, and $\mathbf{e}_k[m] = [\,e_{k,k+1}[m], \ldots, e_{k,K}[m]\,]^T$.

Our rationale of mitigating the cross-correlations term is to project $\mathbf{y}[m]$ into a principal component subspace. Let

$$\mathbf{\Psi} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{y}[m]\mathbf{y}^T[m], \tag{13}$$

and consider its eigen-decomposition $\mathbf{\Psi} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is the (orthogonal) eigenvector matrix, and $\mathbf{\Lambda}$ is the (diagonal) eigenvalue matrix in which the diagonal elements or eigenvalues are arranged in descending order. We use the following projection process

$$\tilde{\mathbf{y}}[m] = \mathbf{U}_{1:K}\mathbf{U}_{1:K}^T\mathbf{y}[m] \tag{14}$$

to mitigate the undesired term $\mathbf{G}\mathbf{e}[m]$. The intuition is that the main term $\mathbf{H}\mathbf{d}[m]$ is often much stronger than the cross-correlations term $\mathbf{G}\mathbf{e}[m]$ in practice, and therefore $\mathbf{U}_{1:K}$, which contains the first K principal components of $\mathbf{\Psi}$, should be dominated by $\mathbf{H}\mathbf{d}[m]$.

By simulations, we found that the projection process in (14) can lead to significant performance improvements. Here, we intend to establish a theoretical justification by modeling $\mathbf{d}[m]$ and $\mathbf{e}[m]$ as random processes. Let us assume

A3) Each $d_k[m]$, $k = 1, \ldots, K$ is a wide-sense stationary (WSS) random process, each $e_{k,\ell}[m]$, $k = 1, \ldots, K-1$, $\ell = k+1, \ldots, K$ is a zero-mean WSS random process, and all $d_k[m]$ and $e_{k,\ell}[m]$ are statistically independent of one other.

In particular, we model the local source variances and cross correlations as independent processes, which is arguably reasonable since their physical behaviors are supposed to be different. We show that

Proposition 1 *Suppose that* A2)-A3) *hold true, that* $M \to \infty$ *such that* $\mathbf{\Psi} = \mathrm{E}\{\mathbf{y}[m]\mathbf{y}^T[m]\}$*, and that*

$$\min_{k=1,\ldots,K} \mathrm{var}\{d_k[m]\} > \max_{k\neq\ell} \mathrm{E}\{|e_{k,\ell}[m]|^2\}. \tag{15}$$

Then, the projection process in (14) *completely eliminates the cross-correlations term and keeps the main term intact; i.e.,*

$$\tilde{\mathbf{y}}[m] = \mathbf{H}\mathbf{d}[m].$$

Proposition 1 implies that if the sources exhibit significant frame-wise power variations and the cross correlations are weak, then the projection process in (14) can eliminate the cross-correlations term very substantially for sufficiently large M. The proof of Proposition 1 is omitted here due to lack of space; the key insight behind the proof is that under A2), $[\,\mathbf{H}\ \mathbf{G}\,]$ is shown to be an orthogonal matrix. As a result, we can derive a sufficient condition [i.e., (15)] under which $\mathbf{U}_{1:K}$ is aligned to the subspace spanned by $\mathbf{H}$; such an alignment also assures that $\mathbf{U}_{1:K}\mathbf{U}_{1:K}^T\mathbf{G} = \mathbf{0}$.

5. SIMULATION RESULTS AND CONCLUSION

Simulations were conducted to demonstrate the performance of the proposed algorithm.

Simulation Settings: The sources are real speech. At each simulation trial, the sources are randomly chosen from a database of 32 recorded speech signals. They are sampled at 16kHz. The mixing matrix is also randomly generated at each trial. The frame length is set at $L = 200$. To get more frames, we employ 50%-overlapping local averaging; i.e., $\mathbf{R}[m] = \frac{1}{L}\sum_{0.5(m-1)L+1}^{0.5(m-1)L+L} \mathbf{x}(t)\mathbf{x}^T(t)$. The number of trials of the simulations is $1,000$.

Noisy observations are considered in the simulations. Specifically, we consider $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t)$, where $\mathbf{v}(t)$ is white Gaussian with zero mean and variance σ^2. Under such circumstances, the local covariances should be modeled as

$$\mathbf{R}[m] = \mathbf{A}\mathbf{D}[m]\mathbf{A}^T + \sigma^2\mathbf{I} \tag{16}$$

In the simulations we remove the noise term by a standard procedure, where we first estimate σ^2 by $\hat{\sigma}^2 = \min_{m=1,\ldots,M} \lambda_{\min}(\mathbf{R}[m])$, in which $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue, and then obtain $\tilde{\mathbf{R}}[m] = \mathbf{R}[m] - \hat{\sigma}^2\mathbf{I}$ as noise-removed local covariances.

We benchmark the proposed algorithm against the clustering-based algorithm mentioned in Section 2, and FFDIAG [5]. FFDIAG is a competitive BSS-QSS algorithm based on JD, which does not assume local dominance. We run all the algorithms using the same set of noise-removed local covariances $\{\tilde{\mathbf{R}}[m]\}_{m=1}^M$ described above.

Recall that the proposed algorithm requires prewhitening in practice. The pre-whitener is obtained from the averaged covariance $\frac{1}{M}\sum_{m=1}^M \tilde{\mathbf{R}}[m]$. The procedure is standard; see [13] for example.

Simulation Results: Fig. 1 shows the mean square errors (MSEs) of the estimated $\mathbf{A}$ yielded by the three algorithms, when $N = 6$, $K = 5$, and the total signal length $= 6$ sec. We can see that the proposed algorithm, with the projection process in Section 4, achieves better MSEs than the other algorithms for $\mathrm{SNR} \le 22$dB. For $\mathrm{SNR} > 22$dB, FFDIAG is better than the proposed algorithm by about 2dB. Moreover, there is a significant performance gap between the clustering-based algorithm and the proposed algorithm, despite the fact that they both utilize local dominance. Our numerical experience is that the clustering-based algorithm may have sensitivity issues with locally dominant points detection in the presence of noise and modeling errors.

Table 1: The MSEs and running times under various signal length. $K = 5$; $N = 6$; SNR=16dB.

Method \ Source duration		2 sec.	3 sec.	4 sec.	5 sec.	6 sec.
Proposed with projection	MSE (dB)	-31.0111	**-33.4356**	**-34.7539**	**-35.9755**	**-36.7884**
	Time (sec.)	**0.0021**	**0.0026**	**0.0033**	**0.0036**	**0.0042**
Clustering-based method	MSE (dB)	-14.7995	-16.5108	-17.2223	-18.2085	-18.8416
	Time (sec.)	0.0778	0.0934	0.1078	0.1233	0.1395
FFDiag	MSE (dB)	**-31.9315**	-33.1174	-33.7476	-34.3641	-34.7653
	Time (sec.)	0.0275	0.0404	0.0537	0.0666	0.0802

Table 2: The MSEs and running times under various numbers of users K. $N = K + 1$; signal length= 6sec.; SNR= 16dB.

Method \ K		3	4	5	6	7	8
Proposed with projection	MSE (dB)	**-36.0243**	**-36.3974**	**-36.1934**	**-35.1247**	**-35.0058**	**-33.9706**
	Time (sec.)	**0.0014**	**0.0026**	**0.0035**	**0.0053**	**0.0092**	**0.0155**
Clustering-based method	MSE (dB)	-23.3953	-21.1632	-19.0816	-16.8285	-14.7481	-12.8761
	Time (sec.)	0.0754	0.1000	0.1212	0.1354	0.1465	0.1677
FFDiag	MSE (dB)	-34.1806	-34.3838	-34.0448	-34.3268	-33.6833	-33.5266
	Time (sec.)	0.0529	0.0553	0.0732	0.0790	0.1183	0.1318

Fig. 1 also shows the performance of the proposed algorithm without using the projection process in Section 4, where we can see performance loss relative to that with projection. This verifies that modeling errors caused by source cross-correlations do exist in practice, and the projection proposed in Section 4 is useful in mitigating the modeling error effects.

In Tables 1-2, we illustrate the MSEs and running times of the algorithms under various settings. In general, the proposed algorithm and FFDIAG are quite on a par in terms of MSE performance. However, the running times of the proposed algorithm are lower than FFDIAG, as well as the clustering-based algorithm, by at least 8 times.

In conclusion, we have developed a simple blind identification algorithm for BSS of locally dominant quasi-stationary sources. The simplicity of the proposed algorithm is made possible by utilizing the local dominance and BSS-QSS problem structures. Simulation results have illustrated that the proposed algorithm is computationally much more efficient than some existing algorithms.

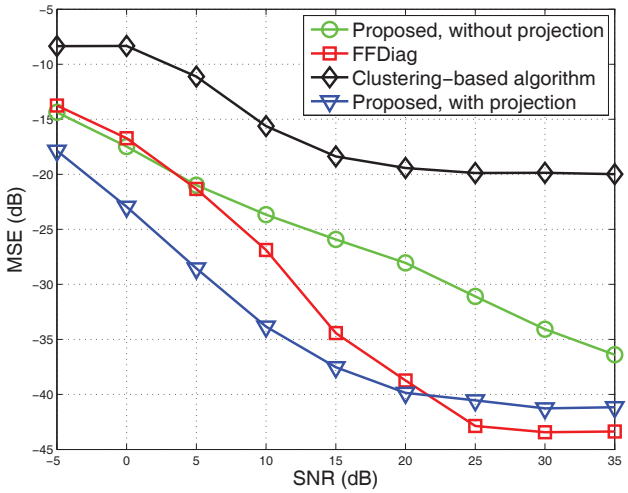

Fig. 1: MSE comparison of the various algorithms. $K = 5$; $N = 6$; signal length= 6sec.

6. REFERENCES

[1] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep 2001.

[2] K. Rahbar, J. P. Reilly, and J. H. Manton, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 832–844, Sep 2005.

[3] D. Nion, K. N. Mokios, N. D. Sidiropoulos, and A. Potamianos, "Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures," *IEEE Trans. Speech Audio Process.*, vol. 18, no. 6, pp. 1193–1207, 2010.

[4] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, Aug 2000.

[5] A. Ziehe, P. Laskov, and G. Nolte, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 801–818, 2004.

[6] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[7] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul 2004.

[8] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time-frequency distributions," *EURASIP J. Adv. Sig. Proc.*, vol. 2005, no. 17, pp. 2828–2847, 2005.

[9] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, "Underdetermined blind separation of nondisjoint sources in the time-frequency domain," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 897–907, 2007.

[10] W.-K. Ma, T.-H. Chan, C.-Y. Chi, and Y. Wang, "Convex analysis for non-negative blind source separation with application in imaging," in *Convex Optimization in Signal Processing and Communications*, D. P. Palomar and Y. Eldar, Eds. Cambridge Univ. Press, 2010.

[11] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov 2011.

[12] G. H. Golub and C. F. V. Loan, *Matrix Computations. 3rd ed.* Baltimore: Johns Hopkins University Press, 1996.

[13] P. Common and C. Jutten, *Handbook of Blind Source Separation*. Elsevier, 2010.