UNDERDETERMINED INSTANTANEOUS BLIND SOURCE SEPARATION OF SPARSE SIGNALS WITH TEMPORAL STRUCTURE USING THE STATE-SPACE MODEL

Benxu Liu, V. G. Reju, Andy W. H. Khong

School of Electrical & Electronic Engineering,
Nanyang Technological University, Singapore.
Email: liub0008@e.ntu.edu.sg, {reju, andykhong}@ntu.edu.sg.

ABSTRACT

In this work, we exploit, in addition to sparseness, the temporal structure of the source signals to address the problem of underdetermined blind source separation. To achieve good separation performance and reduction of artifacts, a two-stage algorithm is proposed. In the first stage, the auto-regressive (AR) coefficients of the source signals are estimated using partially separated sources that have been derived from conventional sparseness-based algorithm. In the second stage, the AR model is combined with the mixing equation to form a state-space model. This model is subsequently solved using the Kalman filter in order to obtain the refined source estimate. Simulation results show the effectiveness of proposed sparseness-based AR-Kalman (SPARK) algorithm compared to the conventional sparseness-based algorithms.

Index Terms— Underdetermined blind source separation, autoregressive model, instantaneous mixing, state-space model

1. INTRODUCTION

Blind source separation (BSS) is the process of separating multiple sources from their mixtures without any prior knowledge about the sources or the mixing process [1]. In convolutive mixing, the mixed signals are the filtered sum of the source signals whereas for the case of instantaneous mixing they correspond to the sum of the scaled versions of the source signals. The instantaneous mixing problem can be expressed mathematically as

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n], \tag{1}$$

where $\mathbf{x}[n] = [x_1[n], ..., x_M[n]]^T$ is a vector containing the M mixtures, $\mathbf{A}$ is the $M \times Q$ mixing matrix, $\mathbf{s}[n] = [s_1[n], ..., s_Q[n]]^T$ contains the Q sources and n is the discrete time index.

The BSS problem is well-determined when $Q \leq M$ and it can be addressed using independent component analysis (ICA) [1]. However, when $Q > M$, the mixing matrix $\mathbf{A}$ is non-invertible and the problem is classified as underdetermined BSS (UBSS). The UBSS problem can be addressed by two stages: estimation of $\mathbf{A}$ followed by source recovery that employs the estimated $\mathbf{A}$ and $\mathbf{x}[n]$ [2]. Although many algorithms have already been proposed for the accurate estimation of $\mathbf{A}$ [3–7], recovery of the source signals still remains a challenging task. The main aim of this work is to recover the source signals using $\mathbf{x}[n]$ and $\mathbf{A}$.

Many existing algorithms estimate $\mathbf{s}[n]$ based on the assumption that the source signals are sparse [3, 8–12]. These methods assume that at any time instant the number of active sources are at most equal to the number of sensors such that the underdetermined problem can be reformulated as a determined problem. Assuming that the sources are W-disjoint in the time-frequency (TF) domain, the degenerate unmixing estimation technique (DUET) [8] compares the elemental ratio of the mixed samples with the elemental ratio of each column in $\mathbf{A}$ to identify the active source. DUET requires at most one active source at any single TF point whereas the subspace-based nondisjoint algorithm proposed in [6] relaxes this requirement; the number of active sources can be at most equal to $M - 1$. In [12, 13], under the assumption that the sparse source signals are Laplacian distributed, the UBSS problem is transformed into a minimum L_1-norm (MLN) optimization problem which is then solved using linear programming.

In all the sparseness-based algorithms discussed above, the inter-sample (temporal) correlation within the source signal has not been taken into account during separation. This results in artifacts in the separated signals [4, 8, 14]. It is well-known that in signals such as speech, there exists temporal correlation which can be described using auto-regressive (AR) coefficients or other alternative models as in [15–17]. The AR model is chosen in this work for its simplicity and tractable complexity. The AR model of the source signals has already been employed in well-determined BSS [18–20] as well as for speech enhancement [21, 22]. However, these techniques cannot be directly extended to UBSS as will be explained in Section 3.1. Our proposed algorithm takes advantage of both sparseness and temporal structure of the source signals such that signal sparsity is employed to obtain the partially separated sources from which source AR coefficients are estimated. The Kalman filter is then used to refine the source signal estimate to achieve good separation performance.

2. THE AR MODEL AND LINEAR PREDICTION (LP)

The temporal structure of a signal can be described using the AR model where the nth sample of the qth source signal $s_q[n]$ can be expressed as a function of its previous samples given by

$$s_q[n] = \sum_{k=1}^{p} a_{q,k} s_q[n-k] + u_q[n], \tag{2}$$

where $u_q[n]$ is the AR model input or innovation process, $a_{q,k}$, $1 \leq k \leq p$ is the kth AR coefficient of the qth source and p is the prediction order. Given $s_q[n]$, its AR coefficients can be estimated by minimizing the average forward LP error defined by [23]

$$\varepsilon_q = E\left\{\left(s_q[n] - \sum_{k=1}^{p} a_{q,k} s_q[n-k]\right)^2\right\}, \tag{3}$$

where $E\{\cdot\}$ is the expectation operator. The optimal $a_{q,k}$ which minimize ε_q can be obtained by employing the well-known linear prediction coefficient (LPC) approach [24]

$$\sum_{k=1}^{p} a_{q,k} r_{qq}^{s}[k-l] = r_{qq}^{s}[l],\ l = 1,\ 2, \cdots, p, \tag{4a}$$

$$r_{qq}^{s}[k-l] = E\left\{s_q[n-l]s_q[n-k]\right\}, \tag{4b}$$

where $r_{qq}^{s}[k-l] = r_{qq}^{s}[l-k]$ and the superscript s in r_{qq}^{s} denotes for the source signal. It is useful to note that the autocorrelation of $u_q[n]$, which can be estimated as [24]

$$r_{qq}^{u}[0] = E\left\{u_q^2[n]\right\} = r_{qq}^{s}[0] - \sum_{k=1}^{p} a_{q,k} r_{qq}^{s}[k], \tag{5}$$

quantifies the degree of predictability of $s_q[n]$ such that a higher value of $r_{qq}^{u}[0]$ implies larger difference between $s_q[n]$ and its predicted value $\sum_{k=1}^{p} a_{q,k} s_q[n-k]$.

3. THE PROPOSED SPARK ALGORITHM

For a UBSS problem, given $\mathbf{x}[n]$ and $\mathbf{A}$, there are an infinite number of $\mathbf{s}[n]$ satisfying (1). However, not all of them will have the expected temporal structure described by (2). Hence we propose to exploit the temporal structure of the source signal so as to improve the determinacy of the UBSS problem. The first stage of the proposed algorithm is to estimate the source AR coefficients from $\mathbf{x}[n]$ and the source signals will then be estimated using the Kalman filter in the second stage.

3.1. Estimation of source AR coefficients from $\mathbf{x}[n]$

In a well-determined BSS problem, the LPC method can directly be applied on $\mathbf{x}[n]$ to estimate the source AR coefficients. However, such technique cannot directly be extended to the UBSS problem. More specifically, we assume that the Q uncorrelated sources are generated by AR models of order p and if the order of the sources are different, the highest order will be taken as p. Considering all the Q sources, the vector AR (VAR) model for $\mathbf{s}[n]$ can be described by

$$\mathbf{s}[n] = \sum_{k=1}^{p} \mathbf{D}_k \mathbf{s}[n-k] + \mathbf{u}[n], \tag{6}$$

where $\mathbf{u}[n] = [u_1[n], u_2[n], \cdots, u_Q[n]]^T$ is the VAR model input for $\mathbf{s}[n]$ and $\mathbf{D}_k = \mathrm{diag}(a_{1,k}, a_{2,k}, \cdots, a_{Q,k})$ is a $Q \times Q$ diagonal matrix. Pre-multiplying both sides of (6) by $\mathbf{A}$, we obtain

$$\mathbf{x}[n] = \sum_{k=1}^{p} \mathbf{A}\mathbf{D}_k \mathbf{s}[n-k] + \mathbf{A}\mathbf{u}[n]. \tag{7}$$

Note that if there exists a set of matrices $\mathbf{B}_k$ such that

$$\mathbf{B}_k \mathbf{A} \quad = \quad \mathbf{A}\mathbf{D}_k,\ \ 1 \le k \le p, \tag{8}$$

equation (7) can be written as

$$\mathbf{x}[n] = \sum_{k=1}^{p} \mathbf{B}_k \mathbf{A}\mathbf{s}[n-k] + \mathbf{A}\mathbf{u}[n] = \sum_{k=1}^{p} \mathbf{B}_k \mathbf{x}[n-k] + \mathbf{w}[n] \tag{9}$$

which describes the VAR model for $\mathbf{x}[n]$ where $\mathbf{w}[n] = \mathbf{A}\mathbf{u}[n]$ is the VAR model input. We therefore see that the LPC approach can be applied on $\mathbf{x}[n]$ to estimate $\mathbf{B}_k$. Note that for the case of well-determined BSS where $\mathbf{A}$ is invertible, the source AR coefficients in $\mathbf{D}_k$ can then be estimated using $\mathbf{A}^{-1}\mathbf{B}_k\mathbf{A}$. However, since $\mathbf{A}$ is non-invertible in UBSS, $\mathbf{D}_k$ cannot be solved even if $\mathbf{B}_k$ is available.

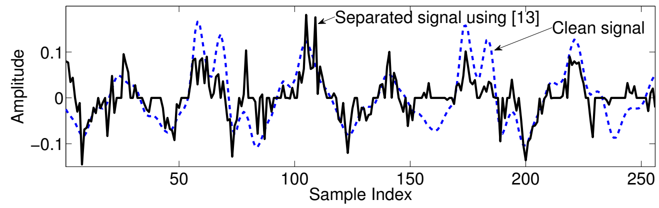

Fig. 1. Temporal structure of the clean signal and the separated signal obtained by conventional sparseness-based method [13] are similar.

In [22], an algorithm to estimate the source AR coefficients has been proposed for speech enhancement. This algorithm either requires the noise to be Gaussian and source signal to be non-Gaussian or a reasonably high signal-to-noise ratio (SNR). That is, if the SNR is high, the LPC method can be applied on $\mathbf{x}[n]$ to estimate $a_{q,k}$. However, this algorithm cannot be directly extended to the UBSS problem due to the similarity in probability distribution across the source signals and the low source-to-interference ratio (SIR) of $\mathbf{x}[n]$ where the interfering sources correspond to noise.

To this end, if the sources can partially be separated to achieve a higher SIR, these partially-separated signals can then be used to estimate $a_{q,k}$. It is important to note that even though the conventional sparseness-based algorithms for UBSS generate significant amount of of artifacts, the amount of interference is significantly reduced [8,14]. This implies that conventional sparseness-based UBSS methods can be used to obtain the partially-separated signals. It is also useful to note that conventional sparseness-based UBSS methods are able to preserve the source temporal structure, characterized by the similarity in long-term variations of the clean and separated speech signal as depicted in Fig. 1. We therefore propose to employ conventional sparseness-based algorithms such as those presented in [6, 8, 13] to obtain the partially-separated source estimate $\tilde{s}_q[n]$ based on which $a_{q,k}$ and $r_{qq}^{u}[0]$ can be estimated using (4) and (5), respectively.

3.2. State-space model and the Kalman filter

After $a_{q,k}$ and $r_{qq}^{u}[0]$ have been estimated for each source, the UBSS problem can be translated into a problem of finding $\mathbf{s}[n]$ which satisfies both (1) and (2). This problem can, in turn, be reformulated as the following state-space model [25]:

$$\vec{\mathbf{s}}[n] = \mathbf{\Phi}\vec{\mathbf{s}}[n-1] + \vec{\mathbf{u}}[n], \tag{10a}$$

$$\mathbf{x}[n] = \mathbf{H}\vec{\mathbf{s}}[n]. \tag{10b}$$

Equation (10a) can be obtained by expressing (2) in its companion form where the $pQ \times 1$ state vector $\vec{\mathbf{s}}[n]$ is defined by

$$\vec{\mathbf{s}}[n] = \left[\vec{\mathbf{s}}_1^T[n], \vec{\mathbf{s}}_2^T[n], \cdots, \vec{\mathbf{s}}_Q^T[n]\right]^T, \tag{11}$$

where the target source signal $s_q[n]$ can be extracted from $\vec{\mathbf{s}}[n]$ using the relationship $\vec{\mathbf{s}}_q[n] = [s_q[n-p+1], s_q[n-p+2], \cdots, s_q[n]]^T$. The $pQ \times pQ$ block-diagonal state transition matrix $\mathbf{\Phi}$ is given by

$$\mathbf{\Phi} = \mathrm{diag}(\mathbf{\Phi}_1, \mathbf{\Phi}_2, \cdots, \mathbf{\Phi}_Q), \tag{12a}$$

$$\mathbf{\Phi}_q = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ a_{q,p} & a_{q,p-1} & \cdots & \cdots & a_{q,2} & a_{q,1} \end{bmatrix}, \tag{12b}$$

where $\mathrm{diag}(\cdot)$ in (12a) represents the block diagonal operation. In addition, the $pQ \times 1$ AR model input $\vec{\mathbf{u}}[n]$ is given by

$$\vec{\mathbf{u}}[n] = \left[\mathbf{u}_1^T[n], \mathbf{u}_2^T[n], \cdots, \mathbf{u}_Q^T[n]\right]^T, \tag{13}$$

where $\mathbf{u}_q[n] = [0, \cdots, 0, u_q[n]]^T$. Equation (10b) can be obtained by replacing $\mathbf{s}[n]$ in (1) with $\vec{\mathbf{s}}[n]$ and $\mathbf{A}$ with the $M \times pQ$ measurement sensitivity matrix $\mathbf{H}$, defined as

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_Q], \tag{14}$$

where $\mathbf{H}_q = [\mathbf{0}, \cdots, \mathbf{0}, \mathbf{a}_q]$ and $\mathbf{a}_q$ is the qth column of $\mathbf{A}$.

Since $\mathbf{x}[n]$ contains contributions from all the sources, it can be proved that the UBSS problem in (1) has now been converted into a determined problem described by (10), i.e., the state-space model is observable. The Kalman filter can then be used to estimate $\vec{\mathbf{s}}[n]$ [25] by first predicting the state value from the previous estimate using

$$\vec{\mathbf{s}}[n|n-1] = \mathbf{\Phi}\vec{\mathbf{s}}[n-1]. \tag{15}$$

The above ensures that the temporal correlation between neighboring samples is preserved. The error covariance of the above prediction is then evaluated as

$$\mathbf{P}[n|n-1] = \mathbf{\Phi}\mathbf{P}[n-1]\mathbf{\Phi}^T + \mathbf{R}^{\vec{\mathbf{u}}}, \tag{16}$$

where the $pQ \times pQ$ covariance matrix $\mathbf{R}^{\vec{\mathbf{u}}}$ is given by

$$\mathbf{R}^{\vec{\mathbf{u}}} = \mathrm{diag}(\mathbf{R}^{\mathbf{u}_1}, \mathbf{R}^{\mathbf{u}_2}, \cdots, \mathbf{R}^{\mathbf{u}_Q}), \tag{17}$$

with $\mathbf{R}^{\mathbf{u}_q} = E\left\{\mathbf{u}_q[n]\mathbf{u}_q^T[n]\right\} = \mathrm{diag}(0, \cdots, 0, r_{qq}^u[0])$. We note from (16) and (17) that $\mathbf{P}[n|n-1]$ increases with $r_{qq}^u[0]$. Since a higher value of $r_{qq}^u[0]$ results in lower predictability of $s_q[n]$, this implies that a larger $\mathbf{P}[n|n-1]$ as expected. The Kalman filter gain is then computed as

$$\mathbf{K}[n] = \mathbf{P}[n|n-1]\mathbf{H}^T\left[\mathbf{H}\mathbf{P}[n|n-1]\mathbf{H}^T + \mathbf{R}^{\mathbf{v}}\right]^{-1} \tag{18}$$

where, similar to [3, 6] and according to (1) for the noiseless condition, the observation noise covariance $\mathbf{R}^{\mathbf{v}} = \mathbf{0}$. However, in order to regularize the inverse operation, we have used $\mathbf{R}^{\mathbf{v}} = \delta\mathbf{I}$ where $\delta = 0.01\bar{r}_{qq}^u[0]$ with $\bar{r}_{qq}^u[0]$ being the average value of $r_{qq}^u[0]$ over q and $\mathbf{I}$ is the identity matrix. The updated state estimate $\vec{\mathbf{s}}[n]$ is then calculated as

$$\vec{\mathbf{s}}[n] = \vec{\mathbf{s}}[n|n-1] + \mathbf{K}[n][\mathbf{x}[n] - \mathbf{H}\vec{\mathbf{s}}[n|n-1]] \tag{19}$$

with the new state estimate error covariance being

$$\mathbf{P}[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{H})\,\mathbf{P}[n|n-1]. \tag{20}$$

Finally, after convergence, $s_q[n]$ can be extracted from $\vec{\mathbf{s}}[n]$ using (11).

The Kalman filter requires initialization and in this work we have used $\vec{\mathbf{s}}[0] = \mathbf{0}$ and $\mathbf{P}[0] = \mathbf{I}$. When the source signals are short-term stationary, $\mathbf{x}[n]$ can be divided into frames, and the proposed algorithm can then operate on each of the frames in a similar way. Figure 2 shows the schematic flow of the proposed two-stage sparseness-based AR-Kalman (SPARK) algorithm. Note that the proposed approach of solving the UBSS problem is significantly different from conventional sparseness-based algorithms. The UBSS problem based on sparsity is indeed a set of determined and independent problems at each time instant or TF point. On the other hand, the proposed method assumes that the successive samples of $\mathbf{s}[n]$ are correlated and exploits this inter-sample correlation to convert the UBSS problem into a determined problem.

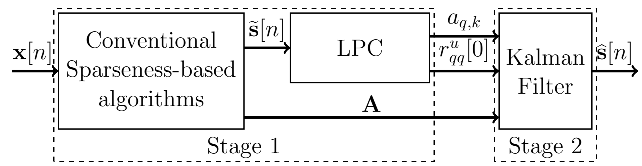

Fig. 2. Schematic flow of the SPARK algorithm.

4. SIMULATION RESULTS

To evaluate the performance of the proposed SPARK algorithm, simulations are conducted using speech utterances from the TIMIT database with a sampling frequency of 16 kHz. The performance of the proposed algorithm is compared with the sparseness-based algorithms including the disjoint algorithm [8], MLN [13] and the subspace-based nondisjoint algorithm [6]. The proposed algorithm employing disjoint (DSJ), MLN and nondisjoint (NDSJ) in the first stage are named as SPARK-DSJ, SPARK-MLN and SPARK-NDSJ, respectively. In this work, the separation performance is quantified in terms of SIR, source-to-distortion ratio (SDR), source-to-artifacts ratio (SAR) and normalized mean-square error (NMSE). The first three performance measures are computed according to [14] which first decomposes the qth source estimate $\widehat{s}_q[n]$ as

$$\widehat{s}_q[n] = s_{\mathrm{target}_q}[n] + e_{\mathrm{interf}_q}[n] + e_{\mathrm{artif}_q}[n], \tag{21}$$

where $s_{\mathrm{target}_q}[n]$ is the portion attributed to $s_q[n]$, $e_{\mathrm{interf}_q}[n]$ is the interference caused by other sources and $e_{\mathrm{artif}_q}[n]$ is the artifacts, such as musical noise, introduced by the separation algorithm. The SIR, SDR and SAR for source q are then defined as

$$\mathrm{SIR}_q = 10\log_{10}\frac{\sum_n s^2_{\mathrm{target}_q}[n]}{\sum_n e^2_{\mathrm{interf}_q}[n]}, \tag{22}$$

$$\mathrm{SDR}_q = 10\log_{10}\frac{\sum_n s^2_{\mathrm{target}_q}[n]}{\sum_n\left(e_{\mathrm{interf}_q}[n] + e_{\mathrm{artif}_q}[n]\right)^2}, \tag{23}$$

$$\mathrm{SAR}_q = 10\log_{10}\frac{\sum_n\left(s_{\mathrm{target}_q}[n] + e_{\mathrm{interf}_q}[n]\right)^2}{\sum_n e^2_{\mathrm{artif}_q}[n]}. \tag{24}$$

All the above measures can be computed using the *BSS-EVAL* Toolbox [26]. The NMSE in dB for the qth source is defined as

$$\mathrm{NMSE}_q = 10\log_{10}\left(\frac{\sum_n(\widehat{s}_q[n] - s_q[n])^2}{\sum_n s_q^2[n]}\right). \tag{25}$$

In this work, $\mathbf{A}$ is assumed to be known and hence NMSE will not suffer from the scaling ambiguity. Figure 3 illustrates the separation result using the nondisjoint algorithm [6] and SPARK-NDSJ for $M = 3$ and $Q = 5$. The speech source signals of 2 s duration are from the TIMIT database and we have used a randomly generated

$$\mathbf{A} = \begin{bmatrix} -0.8087 & -0.2021 & 0.8590 & -0.3528 & 0.1647 \\ 0.3784 & 0.3857 & -0.5118 & 0.9355 & -0.9861 \\ -0.4504 & 0.9002 & 0.0104 & -0.0204 & -0.0232 \end{bmatrix}.$$

From Fig. 3 it can be seen that the SPARK-NDSJ algorithm enhances the harmonic structure of all the separated signals and the listening quality is expected to be improved. We next generate simulation results using five sources ($Q = 5$) and three sensors ($M = 3$) as before. For each trial, the five speech utterances, each of duration 2 s, are randomly selected from a set of 16 speech utterances from the TIMIT database. Matrix $\mathbf{A}$ is generated randomly with each

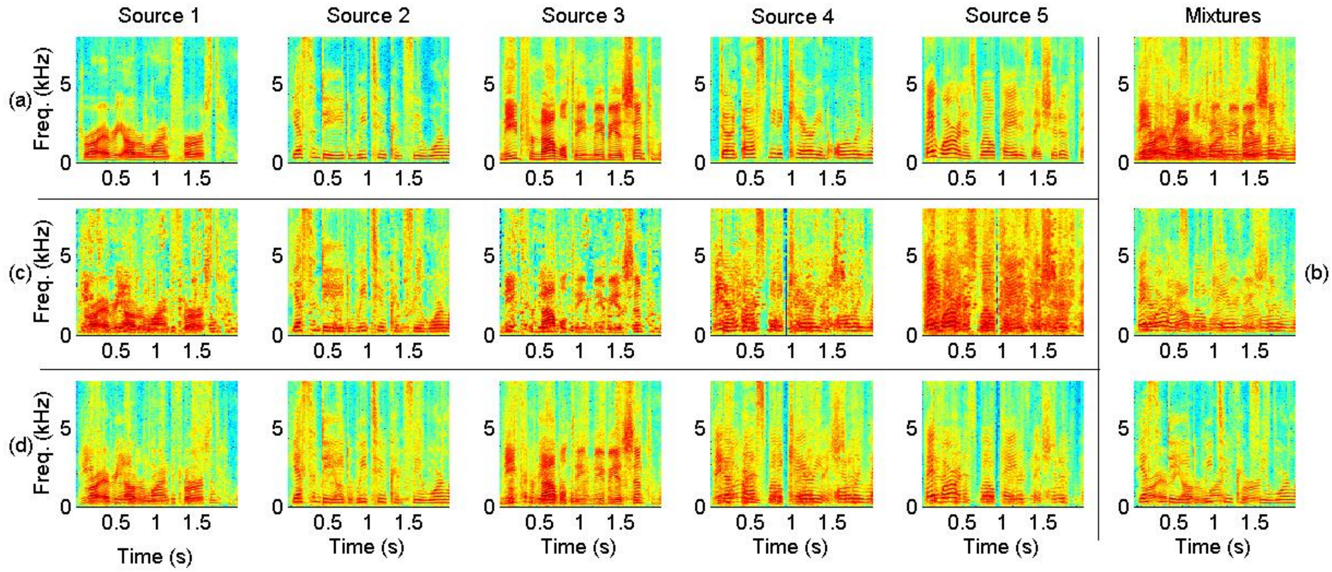

Fig. 3. Spectrogram of a separation example: (a) Clean speech signals, (b) mixtures (on the right-hand side), (c) separation results by nondisjoint algorithm [6], (d) Separation results by SPARK-NDSJ for $p = 4$.

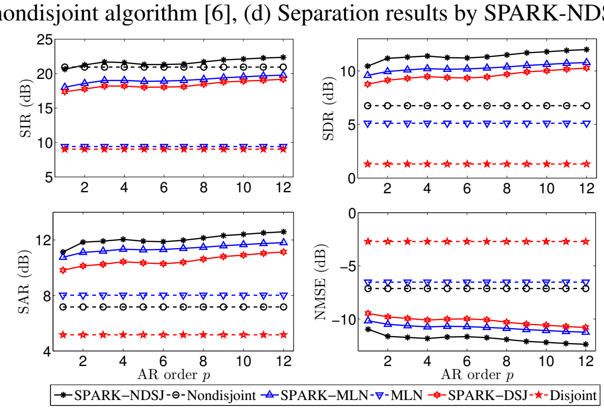

Fig. 4. Separation results on speech signals with changing p.

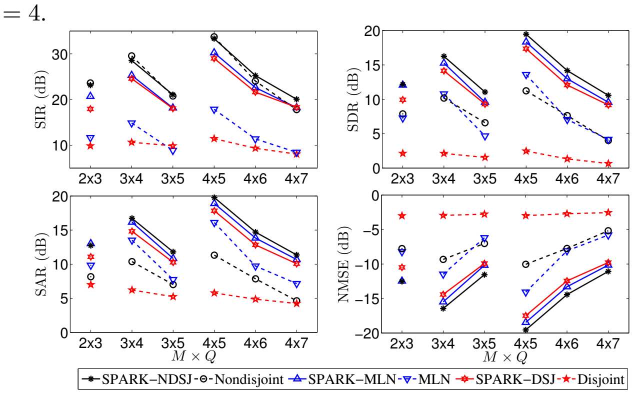

Fig. 5. Separation results on speech signals with different $M \times Q$.

element having a uniform distribution within $[-1, 1]$. Figure 4 shows the average separation performance over 100 simulation trials for $1 \leq p \leq 12$. It is clear from the figure that, by incorporating an additional stage to the conventional sparseness-based algorithms, the overall separation performance is significantly improved. Out of the three algorithms used for comparison, the disjoint algorithm showed a poor performance compared to other algorithms. This could be due to its stringent assumption that there is only one active source in each TF point. For $p = 10$, SPARK-MLN outperforms MLN by approximately 10, 5.5, 3.5 and 4.5 dB in terms of SIR, SDR, SAR and NMSE, respectively whereas SPARK-DSJ outperforms the disjoint algorithm by 9, 8.5, 5.5 and 7.5 dB in terms of SIR, SDR, SAR and NMSE, respectively. For the same AR order, SPARK-NDSJ outperforms the nondisjoint algorithm by approximately 1, 5, 5 and 5 dB in terms of SIR, SDR, SAR and NMSE, respectively.

Figure 5 shows the average separation performance of 100 trials against different $M \times Q$ configurations for $p = 4$. It can be seen that all the algorithms show the same expected trend, i.e., for a fixed value of M, the separation performance reduces with increasing Q. For $M = 4$ and $Q = 5$, SPARK-MLN outperforms MLN by approximately 12, 5, 2.5 and 4.5 dB in terms of SIR, SDR, SAR and NMSE, respectively whereas SPARK-DSJ outperforms the disjoint algorithm by approximately 18, 15, 12 and 15 dB in terms of SIR, SDR, SAR and NMSE, respectively. For the same configuration, SPARK-NDSJ outperforms the nondisjoint algorithm by approximately 8, 8 and 9 dB in terms of SDR, SAR and NMSE, respectively while the two methods have similar SIR performance.

5. CONCLUSIONS

We proposed a two-stage approach for source separation by exploiting the sparseness and temporal structure of the source signals in UBSS. In the first stage, the sources are partially separated using the conventional sparseness-based algorithms available in literature and the AR coefficients of the source signals are estimated. We have shown that the conventional methods of AR coefficient estimation used for determined BSS or speech enhancement applications cannot be employed directly for UBSS. In the second stage of the algorithm, the AR model obtained from the first stage are combined with the mixing equation to form a state-space model which is then solved by Kalman filter to obtain the refined source signals estimate. Simulations on speech utterances showed that the proposed SPARK algorithm achieves significant separation performance improvement compared with conventional sparseness-based UBSS algorithms.

6. REFERENCES

[1] S. Haykin, Ed., *Unsupervised Adaptive Filtering, Volume 1, Blind Source Separation*, Wiley-Interscience, 2000.

[2] S. Winter, W. Kellermann, H. Sawada, and S. Makino, "Underdetermined blind source separation of convolutive mixtures by hierarchical clustering and l_1-norm minimization," *Blind Speech Separation*, pp. 271–304, 2007.

[3] V. G. Reju, S. N. Koh, and I. Y. Soon, "An algorithm for mixing matrix estimation in instantaneous blind source separation," *Signal Process.*, vol. 89, no. 9, pp. 1762 – 1773, 2009.

[4] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Process.*, vol. 87, no. 8, pp. 1833 – 1847, 2007.

[5] F. M. Naini, G. H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "Estimating the mixing matrix in sparse component analysis (SCA) based on partial k-dimensional subspace clustering," *Neurocomputing*, vol. 71, no. 10-12, pp. 2330 – 2343, 2008.

[6] A. Aissa-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, Mar. 2007.

[7] L. D. Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1096 –1105, Mar. 2008.

[8] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: demixing N sources from two mixtures," in *Proc. Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 2000, vol. 5, pp. 2985 –2988.

[9] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, 2010.

[10] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Process.*, vol. 81, no. 11, pp. 2353 – 2362, 2001.

[11] S. Winter, H. Sawada, and S. Makino, "On real and complex valued L1-norm minimization for overcomplete blind source separation," in *Proc. IEEE Workshop App. Signal Process. Audio and Acoust. (WASPAA)*, Oct. 2005, pp. 86 – 89.

[12] I. Takigawa, M. Kudo, and J. Toyama, "Performance analysis of minimum l_1-norm solutions for underdetermined source separation," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 582 – 591, Mar. 2004.

[13] 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.

[14] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 4, pp. 1462 –1469, Jul. 2006.

[15] D. Ellis and R. Weiss, "Model-based monaural source separation using a vector-quantized phase-vocoder representation," in *Proc. Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. IEEE, 2006, vol. 5, pp. V–V.

[16] M. Kowalski, E. Vincent, and R. Gribonval, "Underdetermined source separation via mixed-norm regularized minimization," in *Proc. European Signal Process. Conf. (EUSIPCO)*, 2008.

[17] R. J Weiss, *Underdetermined source separation using speaker subspace models*, Ph.D. thesis, Columbia University, 2009.

[18] A. Hyvarinen, "Independent component analysis for time-dependent stochastic processes," in *Proc. Int. Conf. Artificial Neural Networks (ICANN)*, 1998, pp. 541–546.

[19] K. Kokkinakis and A. K. Nandi, "Multichannel blind deconvolution for source separation in convolutive mixtures of speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 200 – 212, Jan. 2006.

[20] J. V. Stone, "Blind deconvolution using temporal predictability," *Neurocomputing*, vol. 49, no. 1-4, pp. 79 – 86, 2002.

[21] B. G. Lee, K. Y. Lee, and S. Ann, "An EM-based approach for parameter enhancement with an application to speech signals," *Signal Process.*, vol. 46, pp. 1–14, 1995.

[22] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 6, no. 4, pp. 373 –385, jul 1998.

[23] S. Haykin, *Adaptive filter theory*, Prentice Hall, c2002.

[24] P. P Vaidyanathan, *The theory of linear prediction*, Morgan & Claypool, 2008.

[25] M. S. Grewal and A. P. Andrews, *Kalman filtering : theory and practice using MATLAB*, Wiley, 2001.

[26] C. Fevotte, R. Gribonval, and E. Vincent, "BSS EVAL toolbox user guide," Tech. Rep., IRISA 1706, Rennes, France, Apr. 2005.