Title	A linear source recovery method for underdetermined mixtures of uncorrelated AR-model signals without sparseness
Author(s)	Liu, Benxu; Reju, V.; Khong, Andy Wai Hoong
Citation	Liu, B., Reju, V. G., & Khong, A. W. H. (2014). A linear source recovery method for underdetermined mixtures of uncorrelated AR-model signals without sparseness. IEEE transactions on signal processing, 62(19), 4947-4958.
Date	2013
URL	http://hdl.handle.net/10220/20352
Rights	© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [http://dx.doi.org/10.1109/TSP.2014.2329646].

A Linear Source Recovery Method for Underdetermined Mixtures of Uncorrelated AR-Model Signals Without Sparseness

Benxu Liu, Vaninirappuputhenpurayil Gopalan Reju, *Member, IEEE,* Andy W. H. Khong, *Member, IEEE*

***Abstract*—Conventional sparseness-based approaches for instantaneous underdetermined blind source separation (UBSS) do not take into account the temporal structure of the source signals. In this work, we exploit the source temporal structure and propose a linear source recovery solution for the UBSS problem which does not require the source signals to be sparse. Assuming the source signals are uncorrelated and can be modeled by an autoregressive (AR) model, the proposed algorithm is able to estimate the source AR coefficients from the mixtures given the mixing matrix. We prove that the UBSS problem can be converted into a determined problem by combining the source AR model together with the original mixing equation to form a state-space model. The Kalman filter is then applied to obtain a linear source estimate in the minimum mean-squared error sense. Simulation results using both synthetic AR signals and speech utterances show that the proposed algorithm achieves better separation performance compared with conventional sparseness-based UBSS algorithms.**

***Index Terms*—Underdetermined blind source separation, source recovery, autoregressive model, Kalman filter, matrix rank**

I. Introduction

Blind source separation (BSS) refers to the recovery of the source signals from their mixtures without any knowledge of the mixing process or the signals [1]–[8]. Based on whether the mixing process is memoryless or not, BSS can be classified as instantaneous or convolutive mixing. The instantaneous mixing process can mathematically be expressed as

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

where $\mathbf{x}[n] = [x_1[n], ..., x_M[n]]^T$ is a vector containing the M mixtures, $\mathbf{s}[n] = [s_1[n], ..., s_Q[n]]^T$ is a $Q \times 1$ vector of source signals, $\mathbf{A}$ is the $M \times Q$ mixing matrix, $\mathbf{v}[n]$ denotes the additive noise and n is the discrete time index. In BSS, the objective is to estimate $\mathbf{s}[n]$ given $\mathbf{x}[n]$. When $M \geq Q$, the problem is well-determined and can be solved using independent component analysis (ICA) [2]. However, when $M < Q$, $\mathbf{A}$ is non-invertible and this renders the underdetermined BSS (UBSS) a challenging problem.

The UBSS problem is often addressed via a two-stage process: blind system identification (BSI) and blind source recovery (BSR) [4]. The BSI process estimates the mixing matrix whereas the BSR stage recovers the original source signals using the estimated mixing matrix and the mixed signals. Although many algorithms have been proposed to estimate $\mathbf{A}$ [9]–[17], recovery of the source signals is challenging. In this work, we assume that $\mathbf{A}$ is known a priori or can be estimated using methods such as those presented in [12]–[17] and the main objective here is to recover $\mathbf{s}[n]$.

Many of the existing UBSS algorithms estimate $\mathbf{s}[n]$ based on the assumption that the source signals are sparse [5], [18], i.e., only a limited number of sources are active at any time instant and hence the UBSS problem can be reformulated as a solvable determined problem. Since signals such as speech are more sparse in the time-frequency domain compared to that in the time domain, several algorithms have been proposed. These include, for example, the use of clustering and masking in the time-frequency domain. Assuming that the sources are W-disjoint, the degenerate unmixing estimation technique (DUET) [7] compares the elemental ratio of the mixed samples with the elemental ratio of each column in $\mathbf{A}$ to identify the active source. The DUET algorithm employs two sensors and is subsequently generalized to multiple sensors (MENUET) with arbitrary geometric arrangement using K-means clustering [10]. The W-disjoint assumption in DUET and MENUET requires at most one active source at any time-frequency point. A subspace-based nondisjoint algorithm relaxes this requirement such that the number of active sources can be at most equal to $M - 1$ [11] or M [19]. By assuming the sparse signal as one having a Laplacian probability distribution, the UBSS problem can be transformed into a minimum L_1-norm optimization problem [18], [20]. This optimization can then be solved efficiently using linear programming such as the simplex method, which implicitly requires the number of active sources to be less than or at most equal to M [20]. In addition, the above minimum L_1-norm algorithm assumes that $\mathbf{s}[n]$ is independent and identically distributed (i.i.d).

The above sparseness-based algorithms estimate the source signals by decomposing the overall underdetermined problem nonlinearly into locally well-determined problems at each time instant or time-frequency point. However, these algorithms ignore any temporal correlation of each source signal and hence they tend to generate artifacts [7], [10], [21]. It is worthwhile to note that many signals such as speech exhibit a temporal structure or correlation between neighboring samples, which can be described using the autoregressive (AR) model [22]. This temporal structure has already been exploited in well-determined BSS [3], [23]–[28] and speech enhancement [29], [30] problems to improve the performance of the algorithms.

In well-determined BSS, algorithms presented in [23], [24]

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore - 639798, e-mail: liub0008@e.ntu.edu.sg, reju@ntu.edu.sg, andykhong@ntu.edu.sg.

seek to achieve temporally white input signals by removing the temporal structure and as a result, adaptive algorithms used for estimating the mixing model can achieve better performance. Decorrelation has also been employed to identify the mixing matrix [27]. Assuming that the mixtures have lower predictability than the clean source signals, an algorithm maximizing the predictability of the extracted signals has been proposed in [25]. The expectation-maximization (EM) algorithm has also been used to estimate the AR coefficients of the source signals and the mixing matrix [26]. It is important to note that the above methods require $\mathbf{A}$ to be invertible, and hence cannot be directly extended to the UBSS problem, as will be explained in Section II.

For speech enhancement where only a single source signal is present, the AR coefficients of the speech signal, as well as the speech signal itself, can be estimated using an iterative EM method [29], [30]. The procedure starts with an initial guess of the AR coefficients and the source AR model is then combined with the observation to form a state-space model. This model is subsequently solved by the Kalman filter to obtain a minimum mean-square error (MMSE) estimate of the source signal. The estimated source signal is then used to further refine the AR coefficients using the Levinson-Durbin algorithm [31]. However, the EM algorithm only guarantees local optimality and therefore good initialization of the AR coefficients is essential to achieve a reliable estimate of the original speech signal [30]. This algorithm requires either a reasonably high signal-to-noise ratio (SNR) or a non-Gaussian source signal in the presence of Gaussian noise, i.e., the higher-order statistics of the received signal is solely determined by the source signal. However, the UBSS problem does not satisfy either condition 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.

In this work, we utilize the source temporal structure and develop a linear source recovery algorithm for UBSS which does not require the source signals to be sparse. Given $\mathbf{x}[n]$ and $\mathbf{A}$, the proposed algorithm first seeks to estimate the source AR coefficients based on the assumption that the source signals are uncorrelated with each other. The source AR models are then combined with (1) to form a state-space model which can subsequently be solved by the Kalman filter to estimate the source signals. We note that the proposed algorithm does not require source sparsity but it requires the temporal structure of the source signals. While this implies that the conditions of the proposed algorithm are not milder than the sparseness-based algorithms, the main contribution of this work is to recover the source signals from a new perspective: using AR structure instead of sparsity.

The remainder of the paper is organized as follows: Section II describes the source AR models and explains why existing methods used by well-determined BSS to estimate source AR coefficients cannot be directly extended to UBSS. Section III-A discusses the estimation of source AR coefficients for UBSS, followed by formulation of the state-space model in Section III-B. In Section III-C, we prove that the UBSS problem can be converted into a determined problem by including the source AR model. We further establish the relationship between the proposed method and conventional sparseness-based methods in Section III-D. Comparison of the proposed algorithm with existing techniques are presented in Section IV while concluding remarks are discussed in Section V.

II. The temporal structure of source signals

The temporal structure of speech source signals can be described by the AR model. If the source signal is given, the well-known linear prediction theory can be used to estimate the corresponding AR coefficients. In BSS problem, however, only mixtures of the source signals are available. In the case of well-determined BSS, linear prediction theory can be extended directly to solve for the source AR coefficients. This, however, does not hold true for UBSS, as will be explained next.

The temporal structure can be described using the AR model where samples of the qth source signal are related by

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

where $u_q[n]$ is the AR model input, and $\alpha_{q,k}$ is the kth AR coefficient of the qth source and K is the AR order. Given $s_q[n]$, the AR coefficients can be estimated by minimizing the average forward linear prediction squared error defined by [31]

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

where $E\{\cdot\}$ is the expectation operator. Assuming the sources are real and wide-sense stationary, the optimal $\alpha_{q,k}$ which minimizes ε_q can be obtained by solving the following set of Yule-Walker equations

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

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

using the well-known linear prediction coefficient (LPC) approach [31] 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 $u_q[n]$, estimated using [31]

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

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

When using the above LPC method for BSS, we assume that the Q uncorrelated source signals are generated by AR models of order K and if the order of the source signals are different, the highest order will be taken as K. Considering all Q sources, the vector AR (VAR) model for $\mathbf{s}[n]$ can be described by

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

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 $\mathbf{D}_k = \mathrm{diag}(\alpha_{1,k}, \alpha_{2,k}, \cdots, \alpha_{Q,k})$ is a $Q \times$

Q diagonal matrix. Considering the noiseless case and pre-multiplying both sides of (6) by $\mathbf{A}$, we obtain

$$\mathbf{x}[n] = \sum_{k=1}^{K} \mathbf{A}\mathbf{D}_k\mathbf{s}[n-k] + \mathbf{A}\mathbf{u}[n]. \quad (7)$$

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

$$\mathbf{B}_k\mathbf{A} \quad = \quad \mathbf{A}\mathbf{D}_k, \quad 1 \leq k \leq K, \quad (8)$$

(7) can be written as

$$\mathbf{x}[n] = \sum_{k=1}^{K} \mathbf{B}_k\mathbf{A}\mathbf{s}[n-k] + \mathbf{A}\mathbf{u}[n] = \sum_{k=1}^{K} \mathbf{B}_k\mathbf{x}[n-k] + \mathbf{w}[n] \quad (9)$$

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 $\mathbf{B}_k$. 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, the source AR coefficients in $\mathbf{D}_k$ cannot be solved even if $\mathbf{B}_k$ is available.

III. The Proposed Algorithm

In the proposed algorithm, we first estimate the source AR coefficients which are then used to obtain a state-space model. The Kalman filter is then applied to estimate the source signals. We assume, in this work:

1) The source signals are zero mean and uncorrelated to each other.
2) The mixing matrix $\mathbf{A}$ has full row rank.
3) The Khatri-Rao product of $\mathbf{A}$ to itself given by $\mathbf{A} \odot \mathbf{A}$ is of full rank and in good condition.
4) Different source signals have different AR coefficients.
5) The source AR coefficients and the elements in $\mathbf{A}$ are independent to each other.

The first assumption is necessary for the estimation of source AR coefficients given only their mixtures, as will be shown in Section III-A. The second assumption is fundamental in BSS to ensure spatial diversity and separability [2], [5], [7]. The third assumption is to ensure that the source AR coefficients can be estimated from the mixtures. The conditions required for the Khatri-Rao product to be full-column rank have been investigated in [32]–[35]. As explained in [12], [35], this is a mild assumption for any random mixing matrix $\mathbf{A}$ whose columns tend to be maximally independent. The fourth assumption is, in general, valid for independent sources and is common in BSS algorithms exploiting temporal structure of the source signals [3], [12], [28]. The fifth assumption is generally valid since $\mathbf{A}$ is determined by the mixing channel which, in turn, is independent to the source signals and hence to the source AR coefficients. As will be shown in Section III-C, the last two assumptions are used to convert the UBSS problem into a determined problem.

A. Estimation of source AR coefficients in UBSS

We estimate the source AR coefficients $\alpha_{q,k}$ by exploiting the fact that the Khatri-Rao product of two rank-deficient matrices can result in a full-column rank matrix [32]. Since $\alpha_{q,k}$ depends on $r^s_{qq}[k], 0 \leq k \leq K$ as can be seen from (4), it is therefore desirable to estimate $r^s_{qq}[k]$ first. For ease of representation, we first consider the noiseless case where $\mathbf{v}[n] = \mathbf{0}$ and we will consider the noisy case at the end of the derivation.

We start by computing the mixture autocorrelation

$$\begin{aligned}\mathbf{R}^x[k] &= E\left\{\mathbf{x}[n]\mathbf{x}^T[n-k]\right\} \\ &= \mathbf{A}\mathbf{R}^s[k]\mathbf{A}^T, \end{aligned} \quad (10)$$

where $\mathbf{R}^s[k] = E\left\{\mathbf{s}[n]\mathbf{s}^T[n-k]\right\}$. Vectorizing (10) and utilizing the assumption of uncorrelated source signals,

$$\mathbf{r}^x[k] = (\mathbf{A} \odot \mathbf{A})\mathbf{r}^s[k], \quad (11)$$

where the operator $\odot$ denotes Khatri-Rao product, the $M^2 \times 1$ vector $\mathbf{r}^x[k]$ is obtained by concatenating columns of $\mathbf{R}^x[k]$ and the $Q \times 1$ vector

$$\begin{aligned}\mathbf{r}^s[k] &= \mathrm{diag}(\mathbf{R}^s[k]) \\ &= [r^s_{11}[k], r^s_{22}[k], \cdots, r^s_{QQ}[k]]^T. \end{aligned} \quad (12)$$

It is important to note that our aim of solving for $\mathbf{r}^s[k]$ can be achieved if $\mathbf{A} \odot \mathbf{A}$ is of full-column rank, i.e.,

$$\mathrm{rank}(\mathbf{A} \odot \mathbf{A}) = Q. \quad (13)$$

If the above property holds, $\mathbf{r}^s[k]$ can then be solved by

$$\mathbf{r}^s[k] = (\mathbf{A} \odot \mathbf{A})^\dagger \mathbf{r}^x[k], \quad (14)$$

where the pseudo-inverse

$$(\mathbf{A} \odot \mathbf{A})^\dagger = \left((\mathbf{A} \odot \mathbf{A})^T(\mathbf{A} \odot \mathbf{A})\right)^{-1}(\mathbf{A} \odot \mathbf{A})^T. \quad (15)$$

The AR coefficients $\alpha_{q,k}$ can subsequently be computed by solving the Yule-Walker equation described by (4).

Note that a full-rank matrix may still be ill-conditioned. This ill-conditioning may reduce the robustness of estimating $\mathbf{r}^s[k]$ from noisy $\mathbf{x}[n]$. More specifically, in the presence of noise, (14) can now be expressed as

$$\mathbf{r}^s[k] = \boldsymbol{\mathcal{A}}^{-1}(\mathbf{A} \odot \mathbf{A})^T\mathbf{r}^x[k] - \boldsymbol{\mathcal{A}}^{-1}(\mathbf{A} \odot \mathbf{A})^T\mathbf{r}^v[k], \quad (16)$$

where

$$\boldsymbol{\mathcal{A}} \triangleq (\mathbf{A} \odot \mathbf{A})^T(\mathbf{A} \odot \mathbf{A}) \quad (17)$$

and $\mathbf{r}^v[k]$ is obtained by vectorizing the noise autocovariance matrix $\mathbf{R}^{\mathbf{v}}[k] = E\left\{\mathbf{v}[n]\mathbf{v}^T[n-k]\right\}$. According to (16), if $\boldsymbol{\mathcal{A}}$ is well-conditioned and that the elements in $\mathbf{r}^v[k]$ are small, a good estimate of $\mathbf{r}^s[k]$ can be achieved. However, the estimation performance will be reduced if $\boldsymbol{\mathcal{A}}$ is ill-conditioned. According to (17), the condition number of $\boldsymbol{\mathcal{A}}$ is determined by the condition number of $\mathbf{A} \odot \mathbf{A}$. As a consequence, for the noisy case, we also assume that $\mathbf{A} \odot \mathbf{A}$ is in good condition to achieve robustness in the estimation of $\mathbf{r}^s[k]$.

B. State-space model and the Kalman filter

After $\alpha_{q,k}$ 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 be reformulated linearly as the following state-space model

$$\vec{\mathbf{s}}[n] = \boldsymbol{\Phi}\vec{\mathbf{s}}[n-1] + \vec{\mathbf{u}}[n], \quad (18a)$$

$$\mathbf{x}[n] = \mathbf{H}\vec{\mathbf{s}}[n] + \mathbf{v}[n]. \quad (18b)$$

Equation (18a) can be obtained by expressing (2) in its companion form where the $KQ \times 1$ state vector

$$\vec{\mathbf{s}}[n] = \left[\vec{\mathbf{s}}_1^T[n], \vec{\mathbf{s}}_2^T[n], \cdots, \vec{\mathbf{s}}_Q^T[n]\right]^T, \tag{19}$$

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

$$\mathbf{\Phi} = \text{diag}(\mathbf{\Phi}_1, \mathbf{\Phi}_2, \cdots, \mathbf{\Phi}_Q), \tag{20a}$$

$$\mathbf{\Phi}_q = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ \alpha_{q,K} & \alpha_{q,K-1} & \cdots & \cdots & \alpha_{q,2} & \alpha_{q,1} \end{bmatrix}, \tag{20b}$$

where $\text{diag}(\cdot)$ in (20a) denotes block diagonal operation. In addition, the $KQ \times 1$ AR model input $\vec{\mathbf{u}}[n]$ is given by

$$\vec{\mathbf{u}}[n] = \left[\mathbf{u}_1^T[n], \mathbf{u}_2^T[n], \cdots, \mathbf{u}_Q^T[n]\right]^T, \tag{21}$$

where $\mathbf{u}_q[n] = [\mathbf{0}_{1\times K-1}, u_q[n]]^T$, and $\mathbf{0}_{1\times K-1}$ is a $1 \times K-1$ null vector. Equation (18b) can be obtained by replacing $\mathbf{A}$ in (1) with the $M \times KQ$ measurement sensitivity matrix

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_Q], \tag{22}$$

where the $M \times K$ matrix $\mathbf{H}_q = [\mathbf{0}_{M\times K-1}, \mathbf{a}_q]$ and $\mathbf{a}_q$ is the qth column of $\mathbf{A}$.

By including the source AR models, the UBSS problem in (1) has now been converted into a determined problem described by (18), as will be explained in Section III-C. The Kalman filter can then be used to obtain, in the MMSE sense, the estimate of $\vec{\mathbf{s}}[n]$ [36] by first predicting the state value from the previous estimate using

$$\vec{\mathbf{s}}[n|n-1] = \mathbf{\Phi}\vec{\mathbf{s}}[n-1]. \tag{23}$$

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

$$\mathbf{P}[n|n-1] = \mathbf{\Phi}\mathbf{P}[n-1]\mathbf{\Phi}^T + \mathbf{R}^{\vec{\mathbf{u}}}, \tag{24}$$

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

$$\mathbf{R}^{\vec{\mathbf{u}}} = \text{diag}(\mathbf{R}^{\mathbf{u}_1}, \mathbf{R}^{\mathbf{u}_2}, \cdots, \mathbf{R}^{\mathbf{u}_Q}), \tag{25}$$

with $\mathbf{R}^{\mathbf{u}_q} = E\left\{\mathbf{u}_q[n]\mathbf{u}_q^T[n]\right\} = \text{diag}(0, \cdots, 0, r_{qq}^u[0])$. The variable $r_{qq}^u[0]$ can be estimated using (5) after obtaining $\alpha_{q,k}$ and $r_{qq}^s[k]$ as described in Section III-A. We note from (24) and (25) 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 $s_q[n]$, a larger $\mathbf{P}[n|n-1]$ is expected. The Kalman filter gain is then computed as

$$\mathbf{K}[n] = \mathbf{P}[n|n-1]\mathbf{H}^T\left[\mathbf{H}\mathbf{P}[n|n-1]\mathbf{H}^T + \mathbf{R}^{\mathbf{v}}\right]^{-1} \tag{26}$$

where the observation noise covariance $\mathbf{R}^{\mathbf{v}} = E\left\{\mathbf{v}[n]\mathbf{v}^T[n]\right\}$. For the noiseless case where $\mathbf{v}[n] = \mathbf{0}$, in order to regularize the inverse operation, we have used $\mathbf{R}^{\mathbf{v}} = \delta\mathbf{I}$ where $\delta = 0.01\overline{r}_{qq}^u[0]$ with $\overline{r}_{qq}^u[0]$ 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 computed using

$$\vec{\mathbf{s}}[n] = \vec{\mathbf{s}}[n|n-1] + \mathbf{K}[n][\mathbf{x}[n] - \mathbf{H}\vec{\mathbf{s}}[n|n-1]] \tag{27}$$

TABLE I
SUMMARY OF THE PROPOSED AR-KALMAN ALGORITHM.

Input: $\mathbf{A}$, $\mathbf{x}[n]$
Output: $\widehat{s}_q[n]$

1) Compute $\mathbf{R}^x[k]$ using time average as
$$\mathbf{R}^x[k] = \frac{1}{T}\sum_{n=1}^{T}\mathbf{x}[n]\mathbf{x}[n-k],$$
where T is the sample length or frame length of the mixed signal.
2) Vectorize $\mathbf{R}^x[k]$ to get $\mathbf{r}^x[k]$.
3) Solve for $\mathbf{r}^s[k]$ using (14).
4) For each source q, solve for $\alpha_{q,k}$ using (4).
5) Form $\mathbf{\Phi}$ using (20).
6) Solve for $\mathbf{R}^{\vec{\mathbf{u}}}$ using (25) and (5).
7) Apply Kalman filter on the state-space model (18) to estimate $\vec{\mathbf{s}}[n]$ by iteratively implementing the following steps [36]:
   7.1) extrapolate state estimate using (23),
   7.2) extrapolate error covariance using (24),
   7.3) compute Kalman gain using (26),
   7.4) update state estimate by new observation using (27),
   7.5) update error covariance using (28).
8) Extract $\widehat{s}_q[n]$ from state estimate $\vec{\mathbf{s}}[n]$ according to (19).

with the new state estimate error covariance being

$$\mathbf{P}[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{H})\,\mathbf{P}[n|n-1]. \tag{28}$$

Finally, after convergence, the estimate of $s_q[n]$, denoted as $\widehat{s}_q[n]$, can be extracted from $\vec{\mathbf{s}}[n]$ using (19). The above algorithm uses the same K for all sources and can directly be extended to source signals having different AR order [30], [37]. In addition, the conversion from AR model in (6) to the state-space model in (18) may be ill-conditioned for a large value of K [36]. Hence, a relatively small value of K (such as $K \leq 4$) is preferred as will be shown in Section IV.

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 be applied on each frame. For ease of reference, the proposed AR-Kalman algorithm is summarized in Table I.

C. Conversion of UBSS into a determined problem

The UBSS is underdetermined because given $\mathbf{x}[n]$ and $\mathbf{A}$, infinite number of $\mathbf{s}[n]$ satisfies (1). However, not all of them will possess the expected source AR model. This implies that combining the source AR model with (1) to form the state-space model defined by (18) is expected to improve the degree of determinacy of the UBSS problem. Conversion of the UBSS into a determined problem requires an improvement of the determinacy to a degree such that a unique correspondence between the model observation $\mathbf{x}[n]$ and model state $\vec{\mathbf{s}}[n]$ can be established. The unique correspondence between $\mathbf{x}[n]$ and $\vec{\mathbf{s}}[n]$ implies that (18) is observable [36] and we now prove the observability of (18).

The observability of (18) is determined by the rank of the

$KMQ \times KQ$ observability matrix [36]

$$\boldsymbol{\mathcal{O}} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}\boldsymbol{\Phi} \\ \mathbf{H}\boldsymbol{\Phi}^2 \\ \vdots \\ \mathbf{H}\boldsymbol{\Phi}^{KQ-1} \end{bmatrix}. \tag{29}$$

The model will be observable if $\text{rank}(\boldsymbol{\mathcal{O}}) = KQ$, i.e., the dimension of $\vec{\mathbf{s}}[n]$. For clarity of presentation, we consider the case where $K = 1$ although this proof can be extended for higher values of K. For $K = 1$, we have

$$\mathbf{H} = \mathbf{A}, \tag{30}$$

$$\boldsymbol{\Phi} = \text{diag}(\alpha_{1,1}, \alpha_{2,1}, \cdots, \alpha_{Q,1}), \tag{31}$$

and the model will be observable if the $MQ \times Q$ matrix $\boldsymbol{\mathcal{O}}$ satisfies

$$\text{rank}(\boldsymbol{\mathcal{O}}) = Q. \tag{32}$$

Proof: Denoting $[\cdot]_{1:i}$ as the first i rows of a matrix, (32) is proved based on the observation that $\text{rank}([\boldsymbol{\mathcal{O}}]_{1:i})$ increases with i until $i = Q$. We first note that

$$\text{rank}([\boldsymbol{\mathcal{O}}]_{1:M}) = M, \tag{33}$$

since $[\boldsymbol{\mathcal{O}}]_{1:M} = \mathbf{H}$ and $\text{rank}(\mathbf{H}) = M$. We next show that

$$\text{rank}([\boldsymbol{\mathcal{O}}]_{1:M+1}) = M + 1, \tag{34}$$

which implies that the $(M+1)$th row $\mathbf{h}_1^T\boldsymbol{\Phi}$ (where $\mathbf{h}_i^T$ is the ith row of $\mathbf{H}$) cannot be expressed as a linear combination of rows in $[\boldsymbol{\mathcal{O}}]_{1:M}$. To show this, suppose $\mathbf{h}_1^T\boldsymbol{\Phi}$ is a linear combination of rows in $[\boldsymbol{\mathcal{O}}]_{1:M}$. Since $[\boldsymbol{\mathcal{O}}]_{1:M} = \mathbf{H}$, there must exist a vector $\mathbf{c} = [c_1, c_2, \cdots, c_M]^T$ such that

$$\mathbf{c}^T\mathbf{H} = \mathbf{h}_1^T\boldsymbol{\Phi}. \tag{35}$$

Solving for $\mathbf{c}$ satisfying the above constraint would result in an over-determined problem as there are Q equations but only M unknown variables. More specifically, transposing (35), we obtain

$$\mathbf{H}^T\mathbf{c} = \boldsymbol{\Phi}^T\mathbf{h}_1, \tag{36}$$

where $\mathbf{H}^T$ is an over-determined matrix. We show, in Appendix A, that $\mathbf{c}$ satisfying (36) does not exist if the source AR coefficients are independent of the elements in $\mathbf{A}$ and that the AR coefficients are independent and different across all the source signals. Failure of expressing the $(M+1)$th row using $[\boldsymbol{\mathcal{O}}]_{1:M}$ implies that

$$\text{rank}([\boldsymbol{\mathcal{O}}]_{1:M+1}) = \text{rank}([\boldsymbol{\mathcal{O}}]_{1:M}) + 1 = M + 1. \tag{37}$$

If $M + 1 < Q$, expressing the $(M+2)$th row of $\boldsymbol{\mathcal{O}}$ given by $\mathbf{h}_2^T\boldsymbol{\Phi}$, as a linear combination of rows in $[\boldsymbol{\mathcal{O}}]_{1:M+1}$, will again fail similarly. We therefore have

$$\text{rank}([\boldsymbol{\mathcal{O}}]_{1:M+2}) = \text{rank}([\boldsymbol{\mathcal{O}}]_{1:M+1}) + 1 = M + 2. \tag{38}$$

It can be verified that the increase in $\text{rank}([\boldsymbol{\mathcal{O}}]_{1:i})$ with i will continue until $i = Q$ such that

$$\text{rank}([\boldsymbol{\mathcal{O}}]_{1:Q}) = Q. \tag{39}$$

Beyond that, using $[\boldsymbol{\mathcal{O}}]_{1:Q}$ to express the subsequent rows in $\boldsymbol{\mathcal{O}}$ will result in a well-determined problem, of which the linear combination weight vector $\mathbf{c}$ exist. This implies that $\text{rank}(\boldsymbol{\mathcal{O}}) = \text{rank}([\boldsymbol{\mathcal{O}}]_{1:Q})$. As a result, (32) and hence the observability of the state-space model is proved. ■

Note that for the case of $K = 1$, $\boldsymbol{\Phi}^i$, $0 \leq i \leq Q-1$ are diagonal. The observability matrix can then be formulated as the Khatri-Rao product of $\mathbf{H}$ and a Vandermonde matrix containing the diagonal elements of $\boldsymbol{\Phi}^i$. Theorems presented in [38], [39] can be exploited to prove the observability of the state-space model. However, extension of the above algorithms [38], [39] for cases of $K > 1$ requires separate treatment since for such a case, $\boldsymbol{\Phi}^i$ is not diagonal. On the other hand, the proof contained in this work follows standard algebra which can be extended for higher values of K.

D. Relationship with existing sparseness-based algorithms

The proposed algorithm is significantly different from existing sparseness-based UBSS algorithms. Existing solutions for UBSS based on sparsity are well-determined at each time instant or time-frequency point apart from the fact that the locally active sources are unknown. In order to identify the active sources, techniques such as clustering [7], [10], maximum likelihood [5], [18], or subspace projection [11] have been used. The proposed method, on the other hand, does not require sparseness but assumes that successive samples of each source signal $s_q[n]$ are correlated. This inter-sample correlation, described by the AR model, allows one to utilize neighboring samples for the estimation of the current source sample, as has been achieved by the Kalman filter. Therefore, the proposed algorithm exploits the temporal structure of the source signals while existing sparseness-based algorithms exploit the source signal sparsity. Note that this does not imply that the working conditions of the proposed algorithm are milder than that used in existing sparseness-based algorithms since existing algorithms do not impose any restrictions on the temporal structure of the source signal. The main contribution of the proposed algorithm is to recover the source signals using AR structure instead of sparsity.

Although the methods differ, the assumption of source signals being uncorrelated with each other in the proposed method can be related to the sparseness assumption in conventional UBSS algorithms. More specifically, vectorizing (10) without the assumption of uncorrelated sources, we obtain

$$\mathbf{r}^x[k] = \boldsymbol{\Psi}\mathbf{r}_{\text{v}}^s[k] \tag{40}$$

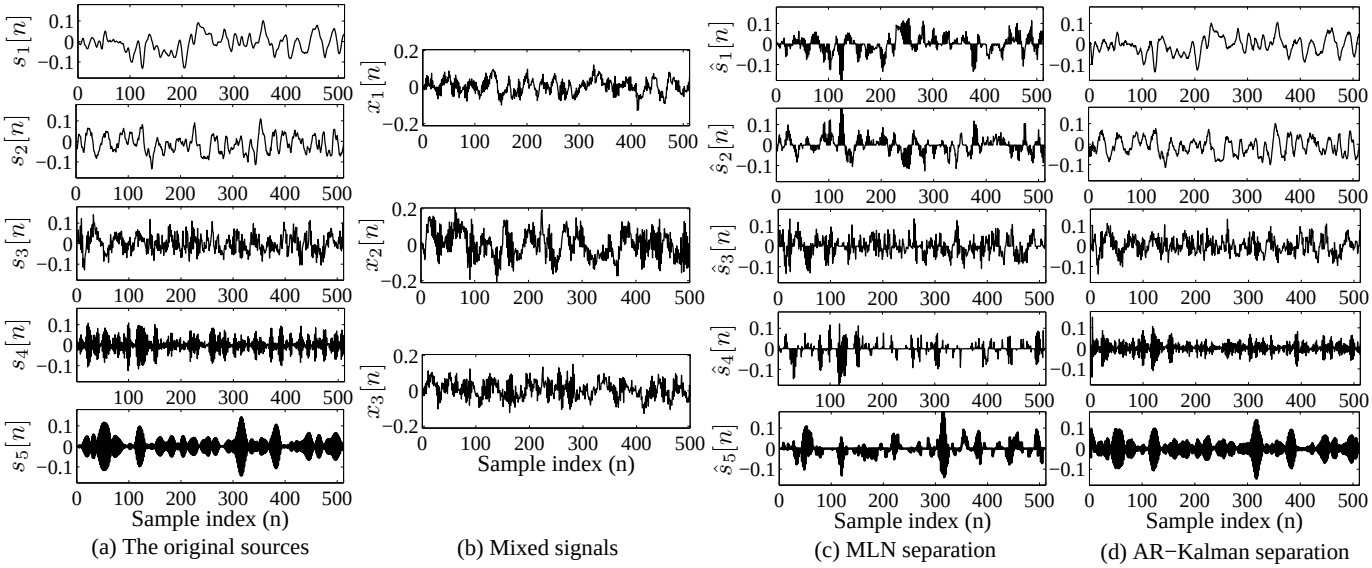
where

$$\boldsymbol{\Psi} = \mathbf{A} \otimes \mathbf{A} \tag{41}$$

and $\otimes$ denotes the Kronecker product. The $M^2 \times 1$ vector $\mathbf{r}^x[k]$ and the $Q^2 \times 1$ vector $\mathbf{r}_{\text{v}}^s[k]$ are obtained by concatenating columns of $\mathbf{R}^x[k]$ and $\mathbf{R}^s[k]$, respectively. Since $\mathbf{R}^x[k]$ and $\mathbf{R}^s[k]$ are symmetrical, the dimension of (40) can be reduced by vectorizing only the lower triangular part of the matrices on both sides of (10). Such reduction results in

$$\mathbf{r}_{\text{h}}^x[k] = \boldsymbol{\Psi}_{\text{h}}\mathbf{r}_{\text{h}}^s[k] \tag{42}$$

where the $M(M+1)/2 \times 1$ vector $\mathbf{r}_{\text{h}}^x[k]$ is formed by vectorizing the lower triangular part of $\mathbf{R}^x[k]$ and the $Q(Q+1)/2 \times 1$ vector $\mathbf{r}_{\text{h}}^s[k]$ is formed by vectorizing the lower triangular part



(a) The original sources (b) Mixed signals (c) MLN separation (d) AR−Kalman separation

Fig. 1. Waveforms of the original signals, their mixtures and the separated signals using MLN and AR-Kalman.

of $\mathbf{R}^s[k]$. The $M(M+1)/2 \times Q(Q+1)/2$ matrix $\mathbf{\Psi}_{\rm h}$ can be formed using $\mathbf{\Psi}$ as described in Appendix B.

Since $M(M+1)/2 < Q(Q+1)/2$, given $\mathbf{r}^x_{\rm h}[k]$ and $\mathbf{\Psi}_{\rm h}$, solving for $\mathbf{r}^s_{\rm h}[k]$ using (42) will result in an underdetermined problem, which is similar to (1). While conventional approaches solve (1) by assuming that the source signals are sparse, i.e., some of the source signals are zero [5], [7], [10], here we solve (42) by assuming that the sources are uncorrelated such that elements corresponding to the correlation between different sources in $\mathbf{r}^s_{\rm h}[k]$ are zero. This will reduce the number of unknown variables in $\mathbf{r}^s_{\rm h}[k]$ from $Q(Q+1)/2$ to Q, which corresponds to the number of autocorrelation elements $r^s_{qq}[k]$. As a result, if $M(M+1)/2 \geq Q$, it is possible to solve for $\mathbf{r}^s_{\rm h}[k]$ using (42). Thus in this work, the conventional requirement of sparseness has now been converted into a more ensemble 'sparseness' sense such that the correlation between different sources are zero.

IV. Simulation results

To evaluate the performance of the proposed AR-Kalman algorithm, simulations are conducted using both synthetic AR signals and speech utterances. The performance of the proposed algorithm is compared with the minimum L_1-norm method (MLN) [18], disjoint algorithm [7], [10] and subspace-based nondisjoint algorithm [11]. In this work, the separation performance is quantified in terms of source-to-interference ratio (SIR), source-to-distortion ratio (SDR), source-to-artifacts ratio (SAR) and normalized mean-squared error (NMSE). The SIR, SDR and SAR are performance measures proposed in [21] to evaluate separation of audio signals. These measures can be computed using the *BSS-EVAL* Toolbox [40]. Since SDR considers both interference and artifacts, it is expected to be more comprehensive than SIR and SAR [21]. The NMSE for the qth source is defined as

$$\mathrm{NMSE}_q = 10\log_{10}\left(\frac{\sum_n(\widehat{s}_q[n]-s_q[n])^2}{\sum_n s_q^2[n]}\right). \tag{43}$$

In addition, columns of $\mathbf{A}$ are normalized to unit norm so as to remove the scaling ambiguity.

Both the noiseless and noisy mixtures are considered. A white Gaussian noise is added into the noiseless mixtures to form the noisy mixtures. The amount of noise added to the mth mixture is quantified by

$$\mathrm{SNR}_m = 10\log_{10}\left(\frac{\sum_n x_m^2[n]}{\sum_n v_m^2[n]}\right). \tag{44}$$

The mixing matrix $\mathbf{A}$ is assumed to be known in this work while in practice only an estimate is available. As a consequence, simulations have been conducted to investigate the sensitivity of the proposed algorithm to any error in the estimation of $\mathbf{A}$. Note that the proposed algorithm focuses on source recovery and in order to systematically evaluate the sensitivity, the estimate $\widehat{\mathbf{A}}$ in this work is generated by adding white Gaussian noise to $\mathbf{A}$ according to a predefined normalized mean-squared error

$$\mathrm{NMSE}_{\mathbf{A}} = 10\log_{10}\left(\frac{\sum_{i,j}(\widehat{a}_{ij}-a_{ij})^2}{\sum_{i,j}a_{ij}^2}\right), \tag{45}$$

where $\widehat{a}_{ij}$ is the (i,j)th element of the estimated mixing matrix, $\widehat{\mathbf{A}}$.

A. Separation of synthetic signals

The methods to generate source AR coefficients, synthetic source signals and $\mathbf{A}$ used in each simulation are described in Table II. Unless otherwise stated, we have used $M = 3$, $Q = 5$ and $K = 3$ throughout the simulations. Figure 1 shows separation results for synthetic signals using MLN and AR-Kalman. The source signals, shown in Fig. 1 (a), are used to generate the three mixtures as depicted in Fig. 1 (b). It can be seen that the source signals are not sparse and hence conventional sparseness-based methods (such as MLN) achieve poor separation performance as shown in Fig. 1 (c). By forcefully assuming that the number of non-zero sources are at most equivalent to the number of mixtures (three in this example,) significant portions of the sources are separated as zero samples which, as a consequence, lead to significant amount of artifacts and poor separation performance. On the other hand, the proposed AR-Kalman method, which does not

TABLE II
DEFAULT PARAMETERS USED FOR SYNTHETIC SIGNALS IN EACH SIMULATION TRIAL.

Source signals	Generated by filtering a white Gaussian data sequence via a randomly generated source AR filter of order 10.
Source AR coefficients	Generated by selecting poles, of the corresponding AR filter, that are randomly distributed within $[-0.8, 0.8]$ for stability.
$\mathbf{A}$	Generated randomly with each element having a uniform distribution within the range $[-1\ 1]$ and column-wise normalized and as explained in Section III-A, only those with $\min_{i,j\in[1,Q],i\neq j} \angle(\mathbf{a}_i, \mathbf{a}_j) \geq 5^\circ$ are used.
$\widehat{\mathbf{A}}$	Generated by adding white Gaussian noise, according to a predefined $\text{NMSE}_{\mathbf{A}}$, to $\mathbf{A}$ obtained using the above method
Q	5
M	3
K	3

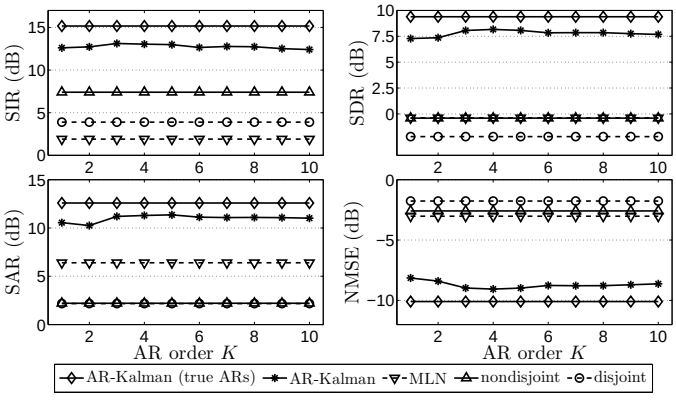


Fig. 2. Separation results on synthetic AR signals for different K.

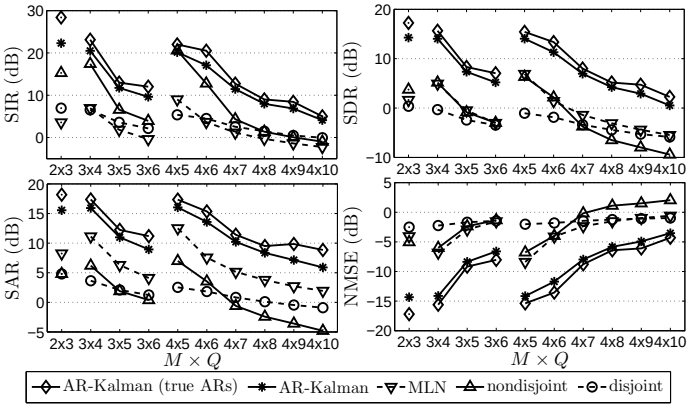


Fig. 3. Separation results on synthetic AR signals for different $M \times Q$.

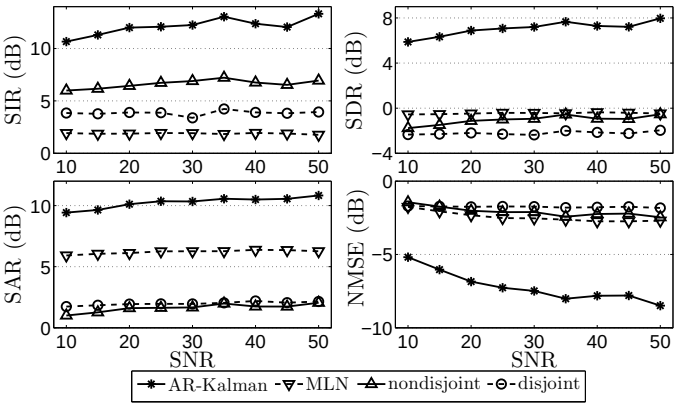


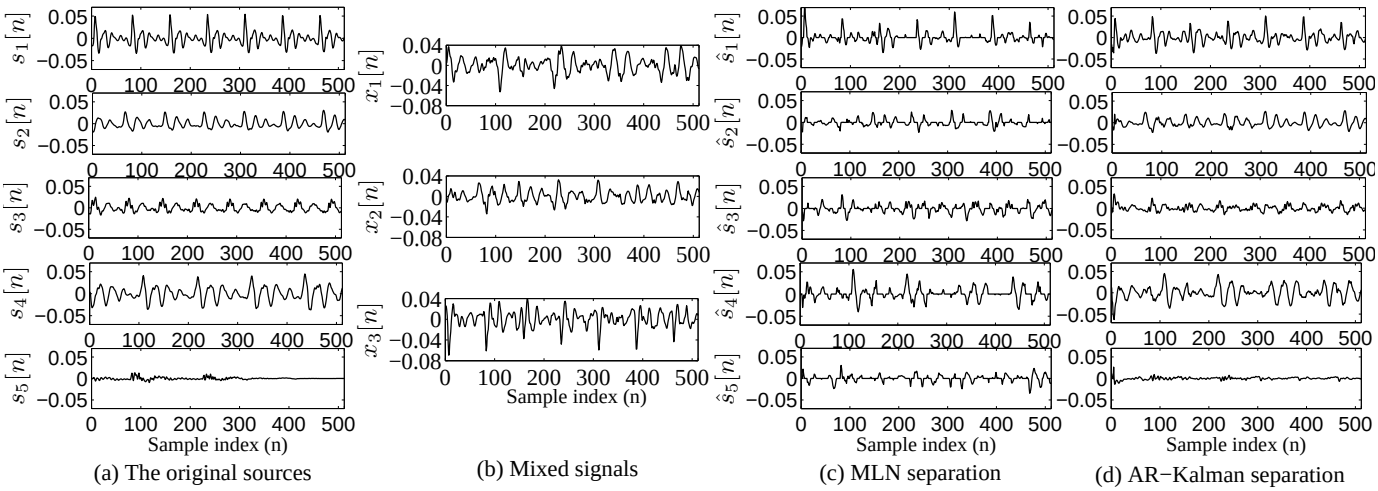
Fig. 4. Separation results on synthetic AR signals for different SNR.

assume sparseness, achieves better separation performance as shown in Fig. 1 (d). It can also be seen from the first few samples of Fig. 1 (d) that the Kalman filter converges within 20 samples of the simulation.

Figure 2 shows the variation of average separation performance for $1 \leq K \leq 10$ over 100 simulation trials using synthetic signals of length 5000 samples. In this simulation, we compare the separation performance of the AR-Kalman, MLN, nondisjoint and disjoint methods. For the proposed AR-Kalman method, results are generated using both the AR coefficients estimated from the method proposed in Section III-A, and the ten true AR coefficients used in the synthetic source signal generation process. This allows us to evaluate the best separation performance as well as the amount of performance degradation when AR coefficients or AR order are unknown. Although AR-Kalman with true AR coefficients, MLN, nondisjoint and disjoint methods are independent of K, they are plotted against K for ease of comparison. Compared with the sparseness-based MLN, nondisjoint and disjoint algorithms, the AR-Kalman algorithm achieved at least 7.5, 9.5, 6 and 7 dB improvement, respectively, in terms of SIR, SDR, SAR and NMSE when the true AR coefficients are used. The significant improvement in the separation performance validates our theory that by including the source AR models, it is possible to convert the UBSS problem into a determined problem. The proposed AR-Kalman algorithm achieved performance close to that using the true AR coefficients and more than 5.5, 8, 5 and 6 dB improvement respectively in terms of SIR, SDR, SAR and NMSE over the three sparseness-based algorithms. In Fig. 2, for all the four performance measures, the proposed AR-Kalman algorithm achieves the best performance for $K = 3$ or 4 and the performance reduces modestly beyond that. Although it is expected that $K = 10$ should achieve the best performance (since $K = 10$ has been used to generate the source signals,) additional simulations revealed that this degradation is due to an increase in the condition number of $\mathbf{\Phi}$ with increasing K [36].

Figure 3 shows the average separation performance over 100 simulation trials for different $M \times Q$ configurations. We note that, for each M, the separation performance reduces with increasing Q as expected. Again we can see that for all the $M \times Q$ configurations the proposed algorithm achieves better performance than the conventional sparseness-based algorithms. In addition, the proposed AR-Kalman algorithm can achieve performance very close to that using the true AR coefficients.

Figure 4 shows the average separation performance for the case of noisy observations (over 100 simulation trials) for $\text{SNR} = 10$ to 50 dB. In this simulation, we have assumed that $\mathbf{A}$ is known. It can be seen that the proposed AR-Kalman algorithm is robust to noise since $\mathbf{A} \odot \mathbf{A}$ is in good condition for randomly generated $\mathbf{A}$. Note that the sparseness-based algorithms are also robust against noise because of the availability of $\mathbf{A}$. For $\text{SNR} = 20$ dB, the proposed AR-Kalman algorithm achieves more than 5.5, 7, 4 and 4.5 dB



(a) The original sources (b) Mixed signals (c) MLN separation (d) AR−Kalman separation

Fig. 6. Speech utterances frame of length 512 samples, their mixtures and the separated signals using MLN and AR-Kalman methods.

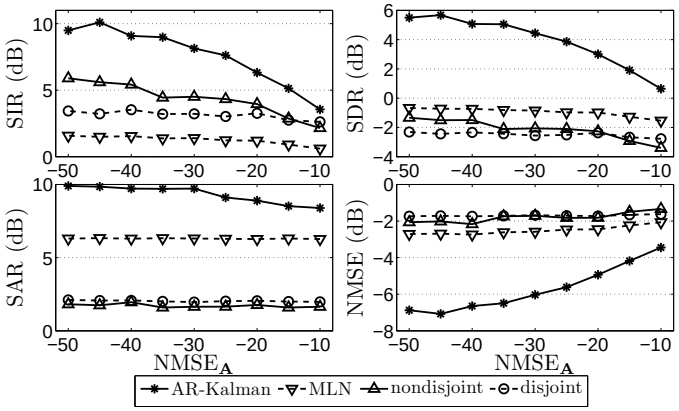


Fig. 5. Separation results on synthetic AR signals for different $\mathrm{NMSE}_{\mathbf{A}}$.

improvement, respectively, in terms of SIR, SDR, SAR and NMSE over the three sparseness-based algorithms.

We next proceed to investigate the sensitivity of the separation performance against error mismatch in $\mathbf{A}$. Figure 5 shows the average separation performance over 100 simulation trials for $\mathrm{NMSE}_{\mathbf{A}} = -50$ to -10 dB. It can be seen that the performance of AR-Kalman reduces with $\mathrm{NMSE}_{\mathbf{A}}$ as expected. Note that the source signals are not sparse and hence existing sparseness-based algorithms cannot achieve good performance even when $\mathrm{NMSE}_{\mathbf{A}}$ is low. For $\mathrm{NMSE}_{\mathbf{A}} = -30$ dB, the proposed AR-Kalman algorithm achieves more than 3.6, 5.4, 3.4 and 3.5 dB improvement, respectively, in terms of SIR, SDR, SAR and NMSE over the three sparseness-based algorithms.

B. Separation of speech signals

Unlike synthetic signals, speech signals are only short-term stationary and hence segmentation of the signals into frames is necessary. Information pertaining to the choice of speech utterances and parameters used in this section are summarized in Table III. Since the proposed algorithm operates on each frame separately, separation results for a single frame are presented first, followed by the average performance using the whole 2 s speech utterances over 100 trials.

TABLE III
DEFAULT PARAMETERS USED FOR SPEECH SIGNALS IN EACH SIMULATION TRIAL.

Source signals	Randomly selected from a set of 20 speech utterances from the TIMIT database.
Sampling Frequency	16 kHz
Signal length	2 s
Frame length	512
Frame overlap	50%
$\mathbf{A}$	Same as in Table II
$\widehat{\mathbf{A}}$	Same as in Table II
Q	5, same as in Table II
M	3, same as in Table II
K	3, same as in Table II

Figure 6 shows the source signals and separation results of a typical frame of five speech utterances using MLN and AR-Kalman. The source signals are illustrated in Fig. 6 (a) and we have used a randomly generated matrix

$$\mathbf{A} = \begin{bmatrix} -0.6534 & -0.2945 & 0.2623 & 0.7666 & -0.8504 \\ -0.5240 & 0.9492 & 0.0378 & -0.6202 & -0.2768 \\ -0.5464 & -0.1105 & -0.9643 & -0.1661 & 0.4475 \end{bmatrix}$$

to obtain the three mixtures shown in Fig. 6 (b). The five separated speech signals using MLN and AR-Kalman are illustrated in Fig. 6 (c) and (d), respectively. It can be seen that compared with MLN, the AR-Kalman algorithm can achieve better separation performance with less artifacts.

Figure 7 shows the average separation performance for $1 \leq K \leq 10$ over 100 simulation trials. We note that since speech signals are, in general, sparse the nondisjoint algorithm can achieve the best SIR performance at the expense of higher distortion and artifacts, as illustrated by its lower SDR, SAR and higher NMSE compared with the proposed AR-Kalman method. This agrees with the finding that sparseness-based UBSS algorithms tend to introduce significant amount of artifacts [7], [21]. The modestly lower SIR but higher SDR of AR-Kalman compared with the nondisjoint algorithm implies that the proposed AR-Kalman algorithm can achieve interference suppression while introducing modest amount of artifacts. Compared with MLN, AR-Kalman (with $K = 3$) achieves an improvement of approximately 7, 3, 0.5 and 2 dB in terms of SIR, SDR, SAR and NMSE, respectively.

Compared with the disjoint algorithm, AR-Kalman achieved an improvement of approximately 7, 6, 3 and 5 dB in terms of SIR, SDR, SAR and NMSE while AR-Kalman achieved 1, 1.5 and 1.5 dB improvement over the nondisjoint algorithm in terms of SDR, SAR and NMSE, respectively.

Figure 8 shows the average separation performance over 100 simulation trials for $\mathrm{SNR} = 10$ to 50 dB. As before, it can be seen that AR-Kalman achieves the best performance in terms of SDR and NMSE across the SNRs being considered. For $\mathrm{SNR} < 20$ dB, AR-Kalman achieves lower SAR compared with MLN. This is attributed to the estimation mismatch in source AR coefficients caused by noise. Compared with the disjoint algorithm for $\mathrm{SNR} = 20$ dB, the proposed algorithm achieved 7, 5.7, 2.7 and 5 dB improvement in terms of SIR, SDR, SAR and NMSE, respectively. Compared with MLN for the same SNR, the proposed algorithm achieved 7, 2.2 and 2 dB improvement in performance in terms of SIR, SDR and NMSE, respectively. For the same SNR, the proposed algorithm achieved 1.2, 1.6 and 1.3 dB improvement in terms of SDR, SAR and NMSE, respectively, compared to the nondisjoint algorithm.

Figure 9 shows the average separation performance over 100 simulation trials for $\mathrm{NMSE}_{\mathbf{A}} = -50$ to -10 dB. It can be seen that AR-Kalman and nondisjoint are more sensitive to error mismatch in $\mathbf{A}$ compared to the MLN and disjoint algorithms. This can be attributed to the fact that AR-Kalman and nondisjoint involve more processing steps compared with the MLN and disjoint algorithms and each of the steps requires an accurate $\mathbf{A}$. Hence any mismatch in $\mathbf{A}$ may introduce more error. Compared with the subspace-based nondisjoint algorithm, the AR-Kalman achieves similar robustness in terms of SDR, SAR and NMSE, but is more robust in terms of SIR. This can be attributed to the sensitivity of the subspace projection on the error mismatch in $\mathbf{A}$. For $\mathrm{NMSE}_{\mathbf{A}} < -20$ dB, the AR-Kalman achieves the best performance in terms of SDR and NMSE. For $\mathrm{NMSE}_{\mathbf{A}} = -30$ dB, compared with MLN, the proposed AR-Kalman algorithm achieves 4.5, 0.5 and 0.5 dB improvement in terms of SIR, SDR and NMSE respectively, but is inferior in performance by 1.4 dB in terms of SAR. This is attributed to the error in estimated source temporal structure caused by the mismatch in $\mathbf{A}$. For $\mathrm{NMSE}_{\mathbf{A}} = -30$ dB, compared with the nondisjoint algorithm, AR-Kalman achieves similar SIR with 1.5, 1, and 1 dB improvement in terms of SDR, SAR and NMSE.

C. Computational complexity

We proceed to investigate the computational complexity of the proposed algorithm in terms of the number of multiplications and divisions, which is more expensive than additions. The proposed AR-Kalman consists of two stages: The first stage of AR model estimation mainly involves the estimation of $\mathbf{R}^x[k]$ and inversion of $\boldsymbol{\mathcal{A}}$ while the second stage involves conventional Kalman filtering. Assuming T samples have been used for the estimation of $\mathbf{R}^x[k]$, the number of multiplications and divisions required by each stage in AR-Kalman are presented in Table IV. For clarity of presentation, the number of multiplications and divisions shown for AR model

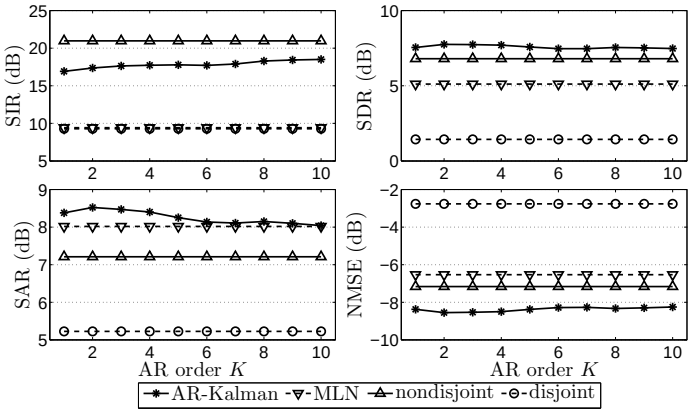


Fig. 7. Separation results of speech sources with $M = 3$ and $Q = 5$.

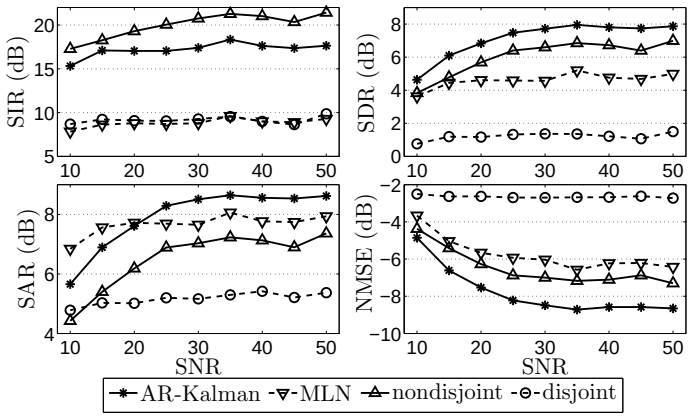


Fig. 8. Separation results of speech sources for different SNR.

estimation is the total number required for T samples whereas the number for the conventional Kalman filtering stage is for one sample of the data. It can be seen that both stages are of polynomial computational complexity and hence can be implemented efficiently.

Given $\mathbf{A}$, the disjoint algorithm is only required to compute the inner product of $\mathbf{x}[n]$ with each $\mathbf{a}_q$ in $\mathbf{A}$. Hence, the disjoint algorithm requires a computational complexity of $O(MQ)$. The MLN algorithm can be solved by the simplex [20] or the interior point method [41], which, in general, have polynomial complexity. The nondisjoint algorithm operates in the combinatorial manner and hence has an exponential computational complexity of $O(C^Q_{M-1}M^2)$. As a consequence, the disjoint algorithm is the most computationally efficient. The proposed algorithm requires similar computational complexity as MLN and both are of lower computational complexity compared to the nondisjoint algorithm.

V. Conclusion

A linear source recovery algorithm for UBSS has been developed for uncorrelated signals without requiring the source signals to be sparse. We described the source temporal structure using the AR model and subsequently combining it with the UBSS mixing equation to form a state-space model. Exploiting the rank change due to Khatri-Rao product, the proposed algorithm is able to estimate the source AR coefficients from $\mathbf{x}[n]$ given $\mathbf{A}$ if $\mathbf{A} \odot \mathbf{A}$ is of full-rank and in good

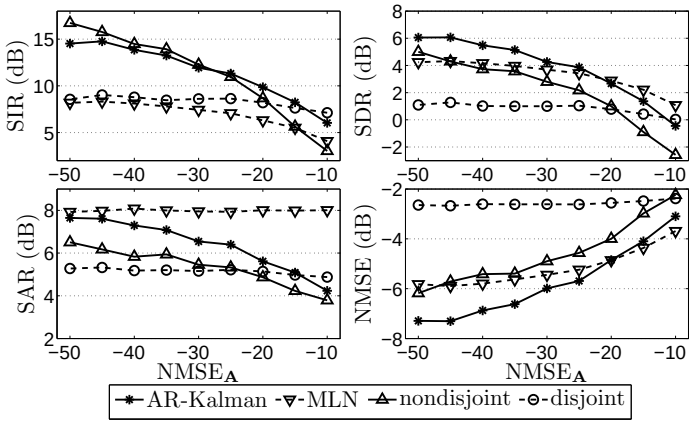


Fig. 9. Separation results of speech sources for different NMSE$_{\mathbf{A}}$.

TABLE IV
NUMBER OF MULTIPLICATIONS AND DIVISIONS REQUIRED BY AR-KALMAN.

Stage	Eq. No.	Number of multiplications and divisions
AR model Estimation	(14)	$QM^2 + 2Q^2M^2 + O(Q^3) + KM^2T + KQM^2$
	(4)	$O(K^3)$
Standard Kalman filtering	(23)	K^2Q^2
	(24)	K^3Q^3
	(26)	$2K^2Q^2M + 2KQM^2 + O(M^3)$
	(27)	$2KQM$
	(28)	$K^2Q^2M + K^3Q^3$

condition. By including the source AR models, the proposed algorithm is able to convert the UBSS into a determined problem when the AR coefficients differ across all the source signals and that they are independent of $\mathbf{A}$. The Kalman filter is then applied to separate the source signals linearly in the MMSE sense. We have also shown that the sparseness criterion in existing sparseness-based algorithms is transformed into a more ensemble sense such that sources are uncorrelated to each other. Simulations using both synthetic AR signals and speech signals show that the proposed AR-Kalman algorithm can achieve better performance than existing sparseness-based algorithms in terms of SDR, SIR, SAR and NMSE.

APPENDIX A

We prove that $\mathbf{c}$ satisfying (36) does not exist when the source signals have independent and different AR coefficients and that the source AR coefficients are independent of the elements in $\mathbf{A}$. For the $M \times Q$ matrix $\mathbf{H} = \mathbf{A}$ and $\mathrm{rank}(\mathbf{H}) = M$, at least one $M \times M$ submatrix of $\mathbf{H}$ is invertible. Without loss of generality, we assume that the submatrix consisting of the first M columns of $\mathbf{H}$ is invertible and denote it as $[\mathbf{H}]_{1:M,1:M}$, while the submatrix containing the remaining $Q-M$ columns is denoted as $[\mathbf{H}]_{1:M,M+1:Q}$. Following straightforward algebraic manipulation and noting that $\mathbf{\Phi}$ is a block diagonal matrix, we separate the constraints in (36) into two parts

$$[\mathbf{H}]^T_{1:M,1:M}\,\mathbf{c} = [\mathbf{\Phi}]^T_{1:M,1:M}[\mathbf{h}_1]_{1:M}, \tag{46a}$$

$$[\mathbf{H}]^T_{1:M,M+1:Q}\,\mathbf{c} = [\mathbf{\Phi}]^T_{M+1:Q,M+1:Q}[\mathbf{h}_1]_{M+1:Q}, \tag{46b}$$

where $[\mathbf{\Phi}]_{1:M,1:M}$ is the upper left $M \times M$ submatrix of $\mathbf{\Phi}$ and $[\mathbf{\Phi}]_{M+1:Q,M+1:Q}$ is the lower right $Q-M \times Q-M$ submatrix of $\mathbf{\Phi}$. The operator $[\cdot]^T_{i:j,k:l}$ refers to the transpose of $[\cdot]_{i:j,k:l}$. The $M \times 1$ vector $[\mathbf{h}_1]_{1:M}$ and the $Q-M \times 1$ vector $[\mathbf{h}_1]_{M+1:Q}$ contains the first M and last $Q-M$ elements of $\mathbf{h}_1$, respectively. Since $[\mathbf{H}]_{1:M,1:M}$ is invertible, (46a) uniquely determines $\mathbf{c}$ as

$$\mathbf{c} = \left([\mathbf{H}]^T_{1:M,1:M}\right)^{-1}[\mathbf{\Phi}]^T_{1:M,1:M}[\mathbf{h}_1]_{1:M}. \tag{47}$$

Next we proceed to show that the above value of $\mathbf{c}$ is unable to satisfy (46b) and hence $\mathbf{c}$ satisfying (36) does not exist. Substituting (47) into (46b), we obtain

$$[\mathbf{H}]^T_{1:M,M+1:Q}\left([\mathbf{H}]^T_{1:M,1:M}\right)^{-1}[\mathbf{\Phi}]^T_{1:M,1:M}[\mathbf{h}_1]_{1:M} = [\mathbf{\Phi}]^T_{M+1:Q,M+1:Q}[\mathbf{h}_1]_{M+1:Q} \tag{48}$$

which establishes a relationship between $[\mathbf{\Phi}]_{1:M,1:M}$ and $[\mathbf{\Phi}]_{M+1:Q,M+1:Q}$ and hence violates our assumption that the AR coefficients of different sources are independent. This implies that (48) will fail with probability of one when elements in $\mathbf{\Phi}$ are drawn from a continuous probability distribution. In addition, similar to the method of separation of variables [42], if two sides of an equation are independent, in order for them to be equal, both sides must be invariant and equal to the same constant, i.e.,

$$[\mathbf{H}]^T_{1:M,M+1:Q}\left([\mathbf{H}]^T_{1:M,1:M}\right)^{-1}[\mathbf{\Phi}]^T_{1:M,1:M}[\mathbf{h}_1]_{1:M} = \mathbf{c}^o, \tag{49a}$$

$$[\mathbf{\Phi}]^T_{M+1:Q,M+1:Q}[\mathbf{h}_1]_{M+1:Q} = \mathbf{c}^o, \tag{49b}$$

where the $Q-M \times 1$ vector $\mathbf{c}^o$ denotes the separation constant. We again note that (49) imposes a relationship between $\mathbf{\Phi}$ and $\mathbf{H}$ which again violates our assumption that $\mathbf{\Phi}$ and $\mathbf{H}$ are independent with each other. As a result, (48) will fail with probability of one, achieving the desired aim of this proof to show that $\mathbf{c}$ satisfying (36) does not exist.

We next consider a case where all the source signals have the same AR coefficient, denoted as α_s, hence giving $[\mathbf{\Phi}]_{1:M,1:M} = \alpha_s\mathbf{I}_M$ and $[\mathbf{\Phi}]_{M+1:Q,M+1:Q} = \alpha_s\mathbf{I}_{Q-M}$ where $\mathbf{I}_M$ is the identity matrix of size $M \times M$. Defining the $M \times 1$ vector $\mathbf{e} = [1, 0, \ldots, 0]^T$, it can be verified that

$$\mathbf{c} = \alpha_s\left([\mathbf{H}]^T_{1:M,1:M}\right)^{-1}[\mathbf{h}_1]_{1:M} = \alpha_s\mathbf{e} \tag{50}$$

satisfies both (46a) and (46b) and hence is a valid solution for (36). In such a case, the state-space model will *not* be observable. This is as expected since if all the sources have exactly the same AR coefficients, taking into account the source AR model does not provide any diversity that is necessary to improve the degree of determinacy. Therefore it is necessary to assume that the source signals have different AR coefficients. Note that the source AR coefficients are independent and hence the probability for the above case to happen is close to zero which will not affect the generality of the proof.

APPENDIX B

Following standard algebra, $\mathbf{\Psi}_{\mathrm{h}}$ can be formed using elements in $\mathbf{\Psi}$. We start by defining the following index sets

$$\mathcal{I}_M = \{i + (j-1)M : 1 \le j \le i \le M\}, \tag{51a}$$

$$\mathcal{I}_Q = \{i + (j-1)Q : 1 \le j \le i \le Q\}. \tag{51b}$$

The formation of $\mathbf{\Psi}_{\mathrm{h}}$ has two steps. In the first step, the rows in $\mathbf{\Psi}$ of indices contained in $\mathcal{I}_M$ will be selected to form an intermediary matrix $\widetilde{\mathbf{\Psi}}$ of size $M(M+1)/2 \times Q^2$. In the second step, the columns in $\widetilde{\mathbf{\Psi}}$ of indices contained in $\mathcal{I}_Q$ will

be selected to form $\mathbf{\Psi}_{\rm h}$ using the following operation: Noting that each index in $\mathcal{I}_Q$ depends on both i and j, if the index corresponds to the case where $i = j$, the column of index $i + (j-1)Q$ will be selected from $\widetilde{\mathbf{\Psi}}$ and used directly as a column in $\mathbf{\Psi}_{\rm h}$. If the index corresponds to the case where $i \neq j$, column vectors in $\widetilde{\mathbf{\Psi}}$ of indices $i + (j-1)Q$ and $j + (i-1)Q$ will be selected and their element-wise sum will be used as the column in $\mathbf{\Psi}_{\rm h}$.

REFERENCES

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

[2] A. Hyvrinen, K. Juhu, and O. Erkki, *Independent Component Analysis*, Wiley-Interscience, 2001.

[3] S. Choi, A. Cichocki, H. Park, and S. Lee, "Blind source separation and independent component analysis: A review," *Neural Information Process.*, vol. 6, no. 1, pp. 1–57, 2005.

[4] S. Winter, H. Sawada, and S. Makino, "On real and complex valued ℓ_1-norm minimization for overcomplete blind source separation," in *Proc. IEEE Workshop App. Signal Process. Audio and Acoust. (WASPAA)*, Oct. 2005, pp. 86 – 89.

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

[6] B. Liu, V. G. Reju, and A. W. H. Khong, "Underdetermined instantaneous blind source separation of sparse signals with temporal structure using the state-space model," in *Proc. Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, May 2013, pp. 81–85.

[7] 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.

[8] B. Liu, V. G. Reju, A. W. H. Khong, and V. V. Reddy, "A GMM post-filter for residual crosstalk suppression in blind source separation," *IEEE Signal Process. Letters*, vol. 21, no. 8, pp. 942–946, Aug. 2014.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] L. D. Lathauwer, J. Castaing, and J. F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2965–2973, Jun. 2007.

[14] A. Ferréol, L. Albera, and P. Chevalier, "Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM)," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1640–1653, May 2005.

[15] L. Albera, A. Ferréol, P. Comon, and P. Chevalier, "Blind identification of overcomplete mixtures of sources (BIOME)," *Linear alg. appl.*, vol. 391, pp. 3–30, 2004.

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

[17] P. Comon and M. Rajih, "Blind identification of under-determined mixtures based on the characteristic function," *Signal Process.*, vol. 86, no. 9, pp. 2271–2281, 2006.

[18] 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.

[19] D. Peng and Y. Xiang, "Underdetermined blind source separation based on relaxed sparsity condition of sources," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 809 –814, Feb. 2009.

[20] 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.

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

[22] L. R. Rabiner and R. W. Schafer, *Theory and Application of Digital Speech Processing*, Pearson, 2011.

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

[24] 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.

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

[26] K. E. Hild, H. T. Attias, and S. S. Nagarajan, "An Expectation-Maximization method for spatio-temporal blind source separation using an AR-MOG source model," *IEEE Trans. Neural Networks*, vol. 19, no. 3, pp. 508 –519, Mar. 2008.

[27] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1 – 24, 2001.

[28] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[29] 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.

[30] 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.

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

[32] X. Guo, S. Miron, D. Brie, S. Zhu, and X. Liao, "A CANDECOMP/PARAFAC perspective on uniqueness of DOA estimation using a vector sensor array," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3475–3481, Jul. 2011.

[33] 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.

[34] N. D. Sidiropoulos and X. Liu, "Identifiability results for blind beamforming in incoherent multipath with small delay spread," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 228–236, Jan. 2001.

[35] L. D. Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.

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

[37] D. C. Popescu and I. Zeljkovic, "Kalman filtering of colored noise for speech enhancement," in *Proc. Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. IEEE, 1998, vol. 2, pp. 997–1000.

[38] T. Jiang, N. D. Sidiropoulos, and J. M. F. Ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1849–1859, 2001.

[39] M. Sorensen and L. D. Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507–5519, Nov. 2013.

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

[41] Y. Nesterov and A. S. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, Philadelphia, PA: SIAM, 1994.

[42] R. P. Agarwal and D. O'Regan, "The method of separation of variables," in *Ordinary and Partial Differential Equations*, pp. 227–233. Springer New York, 2009.

Benxu Liu received his B.Eng. degree in Automation from the department of Control Science and Engineering and the M.Eng. degree majoring in Systems Engineering in 2006 and 2008 from Huazhong University of Science and Technology, China. He is now pursuing the Ph.D. degree in the field of Information Engineering at Nanyang Technological University (NTU). His current research interests include statistical signal processing and blind source separation.

Vaninirappuputhenpurayil Gopalan Reju received the B.Tech. degree in electrical and electronics engineering from University of Kerala, India, in 1992, M.Tech. degree in electronics engineering from Cochin University of Science and Technology, India, in 1994 and Ph.D. from Nanyang Technological University, Singapore in 2010. He was with Cochin University of Science and Technology from 1995 to 2002 as a Lecturer. He is currently a postdoctoral research fellow at Infocomm Centre of Excellence, Nanyang Technological University. His research interests include array signal processing and blind source separation.

Andy W. H. Khong is currently an Assistant Professor in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Prior to that, he obtained his Ph.D. ('02-'05) from the Department of Electrical and Electronic Engineering, Imperial College London, after which he also served as a research associate ('05-'08) in the same department. He obtained his B.Eng. ('98-'02) from the Nanyang Technological University, Singapore. His postdoctoral research involved the development of signal processing algorithms for vehicle destination inference as well as the design and implementation of acoustic array and seismic fusion algorithms for perimeter security systems. He has also published works on adaptive filtering, tangible human-computer interfaces, acoustic blind channel identification and equalization for speech dereverberation as well as acoustic source localization and tracking. Dr. Khong is serving as an Associate Editor for the Multidimensional Systems and Signal Processing, Springer.