Blind Identification of Mixtures of Quasi-Stationary Sources using a Khatri-Rao Subspace Approach

Ka-Kit Lee, Wing-Kin Ma

Dept. Electronic Eng., Chinese Univ. Hong Kong
Shatin, N.T., Hong Kong
E-mail: kklee@ee.cuhk.edu.hk,wkma@ieee.org

Yi-Lin Chiou, Tsung-Han Chan, Chong-Yung Chi

Inst. Commun. Eng., National Tsing Hua Univ.
Hsinchu, Taiwan
E-mail: thchan@ieee.org,cychi@ee.nthu.edu.tw

***Abstract*—This paper addresses the problem of blind identification of a linear instantaneous overdetermined mixture of quasi-stationary sources, using a new formulation based on Khatri-Rao (KR) subspace. A salient feature of this formulation is that it decomposes the blind identification problem into a number of per-source, structurally less complex, blind identification problems. We tackle the per-source problems by developing a specialized alternating projections (AP) algorithm. Remarkably, we prove that AP almost surely converges to a true mixing matrix column in its first iteration, assuming an ideal model condition. Simulation results show that the proposed algorithm yields competitive complexity and performance.**

I. Introduction

In this paper, the problem of interest is blind identification of a linear instantaneous mixture of quasi-stationary sources (QSS). This problem has received much attention, motivated by its application to blind separation of speech and audio sources in microphone arrays. Its strong connection to fundamental frameworks, such as multi-way arrays and joint matrix diagonalization, also makes this topic interesting to pursue.

The idea behind QSS-based blind identification is to exploit statistically time-varying characteristics of QSS (generally of second order), thereby intending to retrieve the mixing system. Currently, there are two main classes of formulations for QSS-based blind identification. One is based on parallel factor analysis (PARAFAC) [1]–[3], where the blind identification problem is formulated as a three-way data array fitting problem. The other is joint diagonalization (JD), where the problem is formulated as a problem of diagonalizing multiple matrices [4]–[6]. Research on these two parallel formulations has revealed some fundamentally beautiful results, subsequently triggering much interest in the field. In PARAFAC, for example, there are elegant linear algebraic results that provide much insight into the identifiability conditions, especially under underdetermined mixture cases [1], [2]. As for JD, we now understand its connection to the optimal maximum-likelihood estimation formulation for the overdetermined Gaussian QSS case [4]. And, from these formulations, effective blind identification or separation algorithms have been developed. State-of-the-arts under PARAFAC are trilinear alternating least squares (TALS) [2] and alternating-columns diagonal-centers (AC-DC) [3], while those under JD include Pham's JD [4], fast Frobenius diagonalization (FFDiag) [5], and quadratic diagonalization (QDiag) [6].

This work was supported by a General Research Fund of Hong Kong Research Grant Council (Project No. CUHK415509) and by the National Science Council (R.O.C.) under Grant NSC 99-2221-E-007-003-MY3.

This paper aims at establishing an alternative formulation for QSS-based blind identification using Khatri-Rao (KR) subspace. KR subspace was previously proposed by the authors for DOA estimation of QSS [7]. In the context here, the KR subspace formulation decomposes the blind identification problem into a number of per-source, structurally less complex, blind identification problems; this is different from PARAFAC and JD which are inherently joint-source formulations. We focus on overdetermined mixtures, and devise a specialized alternating projections (AP) algorithm for dealing with the simpler per-source problems. A distinguishing result with this AP algorithm is its theoretically provable convergence—We will show by analysis that under an ideal model condition, AP almost surely converges to a true mixing matrix column in its first iteration. Using this important insight, we build a new blind identification algorithm which will be empirically demonstrated to have good runtime performance. Simulation results will also show that the proposed algorithm exhibits blind estimation performance on a par with the existing algorithms.

II. Problem Statement

Consider an N-by-K linear mixing model as follows:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t),\ t = 0, 1, 2, \ldots, \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ is the received signal vector, $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_K] \in \mathbb{C}^{N\times K}$ is the mixing matrix, $\mathbf{s}(t) = [s_1(t), \ldots, s_K(t)]^T$ is the source signal vector, and $\mathbf{v}(t)$ represents noise. Our model assumptions are as follows:

- (A1) The source signals $s_k(t), k = 1, \ldots, K$, are statistically independent of each other, with zero mean.
- (A2) $\mathbf{A} \in \mathbb{C}^{N\times K}$ is of full column rank.
- (A3) The noise vector $\mathbf{v}(t)$ is wide-sense stationary with zero mean and covariance $\sigma^2\mathbf{I}$, and it is statistically independent of the source signals.
- (A4) Each source signal $s_k(t)$ is wide-sense quasi-stationary with frame length L; specifically, $\mathrm{E}\{|s_k(t)|^2\} = d_{mk},\ \forall t \in [(m-1)L+1, mL]$.

Note that assumption (A4) is key to QSS-based blind identification techniques [2]–[6]. Physically, it means that the source

second-order statistics (SOSs) are static within a short, local period of time, but are time-variant over a global time scale.

Let us examine the local SOSs of the received signal. Define

$$\mathbf{R}_m = \mathrm{E}\{\mathbf{x}(t)\mathbf{x}(t)^H\}, \text{ for any } t \in [(m-1)L+1, mL], \tag{2}$$

to be the local covariance matrices of $\mathbf{x}(t)$, where $m = 1, 2, \ldots$ is the frame index. These $\mathbf{R}_m$ can be estimated by local time averaging; i.e., $\mathbf{R}_m \simeq \frac{1}{L}\sum_{t=(m-1)L+1}^{mL} \mathbf{x}(t)\mathbf{x}(t)^H$. Under (A1), (A3), and (A4), we obtain the following model for $\mathbf{R}_m$:

$$\mathbf{R}_m = \mathbf{A}\mathbf{D}_m\mathbf{A}^H + \sigma^2\mathbf{I}, \tag{3}$$

where $\mathbf{D}_m = \mathrm{Diag}(d_{m1}, d_{m2}, \ldots, d_{mK}) \in \mathbb{R}^{K\times K}$ is the source local covariance matrix of the mth frame. Suppose that we have $\mathbf{R}_1, \ldots, \mathbf{R}_M$ available, where M is the number of available frames. Our goal is to estimate the mixing matrix $\mathbf{A}$ from $\mathbf{R}_1, \ldots, \mathbf{R}_M$ without prior knowledge of $\mathbf{D}_1, \ldots, \mathbf{D}_M$ and σ^2.

For convenience, we will assume that the noise covariance matrix is absent from the local covariance model (3); i.e.,

$$\mathbf{R}_m = \mathbf{A}\mathbf{D}_m\mathbf{A}^H. \tag{4}$$

In practice, it has been known that the noise covariance matrix $\sigma^2\mathbf{I}$ can be easily removed for the case of $N > K$. In essence, one can make use of the fact that the minimum eigenvalue of $\mathbf{R}_m$, denoted here by $\lambda_{\min}(\mathbf{R}_m)$, is σ^2 for $N > K$. In this work, we will specifically use the following method to estimate σ^2

$$\hat{\sigma}^2 = \min_{m=1,\ldots,M} \lambda_{\min}(\mathbf{R}_m), \tag{5}$$

followed by subtracting the noise covariance from (3).

III. Khatri-Rao Subspace Criterion

We will first briefly review the KR product and some of its properties. Then, our KR subspace blind identification criterion will be described.

A. Some Basic Results for KR Product

Given two matrices $\mathbf{A} \in \mathbb{C}^{n\times k}$ and $\mathbf{B} \in \mathbb{C}^{m\times k}$ of identical number of columns, the KR product of $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} \odot \mathbf{B} = [\ \mathbf{a}_1 \otimes \mathbf{b}_1, \ldots, \mathbf{a}_k \otimes \mathbf{b}_k\ ] \in \mathbb{C}^{nm\times k}, \tag{6}$$

where $\otimes$ denotes the Kronecker product. KR product has many interesting algebraic properties; see, e.g., [1], [2] and the references therein. Here we list several results relevant to this work:

i) $\mathbf{a}\otimes\mathbf{b} = [\ a_1\mathbf{b}^T, \ldots, a_n\mathbf{b}^T\ ]^T = \mathrm{vec}(\mathbf{b}\mathbf{a}^T)$, where $\mathrm{vec}(\cdot)$ is the vectorization operation.

ii) Let $\mathbf{A} \in \mathbb{C}^{n\times k}$, $\mathbf{B} \in \mathbb{C}^{m\times k}$, and $\mathbf{d} \in \mathbb{C}^k$. Also denote $\mathbf{D} = \mathrm{Diag}(\mathbf{d})$, where $\mathrm{Diag}(\cdot)$ is the diagonal operator. Then we have [1]

$$\mathrm{vec}(\mathbf{A}\mathbf{D}\mathbf{B}^H) = (\mathbf{B}^* \odot \mathbf{A})\mathbf{d}. \tag{7}$$

iii) For any $\mathbf{a} \in \mathbb{C}^n$, $\mathbf{b} \in \mathbb{C}^m$, and $\mathbf{C} \in \mathbb{C}^{n\times m}$, we have

$$(\mathbf{b}^* \otimes \mathbf{a})^H \mathrm{vec}(\mathbf{C}) = \mathbf{a}^H\mathbf{C}\mathbf{b}. \tag{8}$$

B. KR Subspace Criterion

Let us consider the local covariance model in (4). By (7), we have

$$\mathbf{y}_m \triangleq \mathrm{vec}(\mathbf{R}_m) = (\mathbf{A}^* \odot \mathbf{A})\mathbf{d}_m \in \mathbb{C}^{N^2}, \tag{9}$$

where $\mathbf{d}_m = [d_{m1}, \ldots, d_{mK}]^T$. Then, stacking all $\mathbf{y}_m$ yields

$$\mathbf{Y} \triangleq [\mathbf{y}_1, \ldots, \mathbf{y}_M] = (\mathbf{A}^* \odot \mathbf{A})\mathbf{\Psi}^T \in \mathbb{C}^{N^2\times M}, \tag{10}$$

where $\mathbf{\Psi} = [\mathbf{d}_1, \ldots, \mathbf{d}_M]^T$. We are interested in the subspace characteristics of $\mathbf{A}$. To do this, we first show that

Lemma 1. *Under (A2), $\mathbf{A}^* \odot \mathbf{A}$ is of full column rank.*

Proof: We aim to show that $(\mathbf{A}^* \odot \mathbf{A})\mathbf{x} \neq \mathbf{0}$ for all $\mathbf{x} \neq \mathbf{0}$. By (7), we get

$$\mathrm{vec}^{-1}((\mathbf{A}^* \odot \mathbf{A})\mathbf{x}) = \mathbf{A}\mathrm{Diag}(\mathbf{x})\mathbf{A}^H. \tag{11}$$

Since $\mathbf{A}$ is of full column rank, $\mathbf{A}\mathrm{Diag}(\mathbf{x})\mathbf{A}^H = \mathbf{0}$ only when $\mathrm{Diag}(\mathbf{x}) = \mathbf{0}$. Hence, we complete the proof. ■

Moreover, we assume

(A5) $\mathbf{\Psi}$ is of full column rank.

Assumption (A5) is justified as follows. Physically, each column of $\mathbf{\Psi}$ describes the power distribution of a source over a long-term time scale. Hence, for sources yielding significantly different time-variant power distributions and for $M \geq K$, we would expect that the columns of $\mathbf{\Psi}$ be linearly independent.

By Lemma 1 and (A5), the rank of $\mathbf{Y}$ is K. Hence, $\mathbf{Y}$ admits a compact singular value decomposition (SVD)

$$\mathbf{Y} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^H, \tag{12}$$

where $\mathbf{\Sigma}_s \in \mathbb{R}^{K\times K}$ is the nonzero singular value matrix, and $\mathbf{U}_s \in \mathbb{C}^{N^2\times K}$ and $\mathbf{V}_s \in \mathbb{C}^{M\times K}$ are the associated left and right singular matrices, resp. As a basic SVD result, we have

$$\mathcal{R}(\mathbf{U}_s) = \mathcal{R}(\mathbf{Y}) = \mathcal{R}(\mathbf{A}^* \odot \mathbf{A}) \tag{13}$$

where $\mathcal{R}(\mathbf{X})$ denotes the range space of $\mathbf{X}$. The implication with (13) is that every mixing matrix column $\mathbf{a}_k$ satisfies the condition $\mathbf{a}_k^* \otimes \mathbf{a}_k \in \mathcal{R}(\mathbf{U}_s)$. This leads us to the following blind identification criterion [7]:

$$\boxed{\begin{array}{l} \text{find } \ \mathbf{a} \\ \text{such that } \ \mathbf{a}^* \otimes \mathbf{a} \in \mathcal{R}(\mathbf{U}_s),\ \mathbf{a} \in \mathbb{C}^N \end{array}} \tag{14}$$

Eq. (14) will be called the *KR subspace criterion* in the sequel.

We are concerned with the identifiability of criterion (14).

Lemma 2. (Identifiability) *Under (A2), a vector $\mathbf{a} \in \mathbb{C}^N$ satisfies the KR subspace criterion $\mathbf{a}^* \otimes \mathbf{a} \in \mathcal{R}(\mathbf{U}_s)$ if and only if $\mathbf{a} = c\mathbf{a}_k$ for any $k = 1, \ldots, K$ and for any non-zero constant $c \in \mathbb{C}$.*

Lemma 2 implies that the columns of the true mixing matrix $\mathbf{A}$ can be unambiguously identified by the KR subspace criterion (up to a scale factor). The proof of Lemma 2 is similar to that of Lemma 1, and is skipped due to page limit.

We should note that the KR subspace criterion (14) suggests a per-source decomposition approach to identifying the mixing matrix $\mathbf{A}$. This is in contrast to the PARAFAC and JD criteria, where the columns of $\mathbf{A}$ are jointly identified.

IV. KR Subspace Alternating Projections

In this section we focus on per-source identification in accordance with the KR subspace criterion (14). The results obtained here will be used to build a systematic all-sources blind identification algorithm in the next section.

We employ alternating projections (AP) to deal with the KR subspace criterion. AP is a simple technique for finding a vector in the intersection of some given sets [8]. To apply AP, we formulate criterion (14) as an optimization problem:

$$\min_{\alpha\in\mathbb{C},\mathbf{a}\in\mathbb{C}^N,\mathbf{h}\in\mathbb{C}^{N^2}} \|\alpha\mathbf{a}^*\otimes\mathbf{a}-\mathbf{h}\|^2 \quad \text{s.t. } |\alpha|=1,\ \ \|\mathbf{a}\|^2=1,\ \ \mathbf{h}\in\mathcal{R}(\mathbf{U}_s), \tag{15}$$

where $\|\cdot\|$ is the Euclidean norm. The operation of AP is as follows: at one time, minimize (15) with respect to (w.r.t.) $\mathbf{h}$ with $(\alpha,\mathbf{a})$ fixed, and, at another time, minimize (15) w.r.t. $(\alpha,\mathbf{a})$ with $\mathbf{h}$ fixed. Let us first consider the partial minimization w.r.t. $\mathbf{h}$:

$$\min_{\mathbf{h}\in\mathcal{R}(\mathbf{U}_s)} \|\alpha\mathbf{a}^*\otimes\mathbf{a}-\mathbf{h}\|^2. \tag{16}$$

Problem (16) is a linear projection problem, whose solution is

$$\mathbf{h}=\mathbf{U}_s\mathbf{U}_s^H(\alpha\mathbf{a}^*\otimes\mathbf{a}). \tag{17}$$

Moreover, the partial minimization w.r.t. $(\alpha,\mathbf{a})$

$$\min_{|\alpha|=1,\|\mathbf{a}\|^2=1} \|\alpha\mathbf{a}^*\otimes\mathbf{a}-\mathbf{h}\|^2 \tag{18}$$

also has a closed form. To show this, consider the objective function of (18). We have, for any $|\alpha|=1,\|\mathbf{a}\|^2=1$,

$$\begin{aligned}
&\|\alpha\mathbf{a}^*\otimes\mathbf{a}-\mathbf{h}\|^2 && (19)\\
&\quad=|\alpha|^2\|\mathbf{a}^*\otimes\mathbf{a}\|^2-2\mathrm{Re}\{\alpha^*(\mathbf{a}^*\otimes\mathbf{a})^H\mathbf{h}\}+\|\mathbf{h}\|^2\\
&\quad=1-2\mathrm{Re}\{\alpha^*\mathbf{a}^H\mathrm{vec}^{-1}(\mathbf{h})\mathbf{a}\}+\|\mathbf{h}\|^2 && (20)\\
&\quad\geq 1-2|\mathbf{a}^H\mathrm{vec}^{-1}(\mathbf{h})\mathbf{a}|+\|\mathbf{h}\|^2, && (21)
\end{aligned}$$

where (20) is due to (8). Note that equality in (21) holds when $\alpha=\exp(j\phi(\mathbf{a}^H\mathrm{vec}^{-1}(\mathbf{h})\mathbf{a}))$ where $\phi(x)$ is the phase of x. Moreover, (21) is minimized when $\mathbf{a}$ equals a unit-norm eigenvector of $\mathrm{vec}^{-1}(\mathbf{h})$ associated with an eigenvalue of the largest absolute value, which we will denote by $\mathbf{q}_{\max}(\mathrm{vec}^{-1}(\mathbf{h}))$. In summary, a solution to (18) is

$$\mathbf{a}=\mathbf{q}_{\max}(\mathrm{vec}^{-1}(\mathbf{h})),\alpha=\exp(j\phi(\mathbf{a}^H\mathrm{vec}^{-1}(\mathbf{h})\mathbf{a})). \tag{22}$$

We summarize the AP method as follows:

Alternating Projection Algorithm for (14)	
given	an orthogonal basis matrix $\mathbf{U}_s\in\mathbb{C}^{N^2\times K}$, and an initial point $\mathbf{h}\in\mathcal{R}(\mathbf{U}_s)$.
repeat	
	$\mathbf{a}:=\mathbf{q}_{\max}(\mathrm{vec}^{-1}(\mathbf{h}))$, $\alpha:=\mathrm{e}^{j\phi(\mathbf{a}^H\mathrm{vec}^{-1}(\mathbf{h})\mathbf{a})}$, $\mathbf{h}:=\mathbf{U}_s\mathbf{U}_s^H(\alpha\mathbf{a}^*\otimes\mathbf{a})$,
until	a stopping criterion is satisfied.
output	the vector $\mathbf{a}$.

It is known that for general applications, AP may exhibit slow convergence. Rather unexpectedly, we find that for semi-unitary $\mathbf{A}$ (i.e., $\mathbf{A}^H\mathbf{A}=\mathbf{I}$), the convergence of AP becomes dramatically different:

Theorem 1. (Convergence of AP for semi-unitary A) *Suppose that $\mathbf{A}$ is semi-unitary, that the initialization is generated by $\mathbf{h}=\mathbf{U}_s\boldsymbol{\zeta}$ where $\boldsymbol{\zeta}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$ (i.e., i.i.d. zero-mean unit variance complex Gaussian), and that there is no modeling error in* (4)*. Then, with probability one and in one iteration, the AP algorithm converges to any one of $\mathbf{a}_1,\ldots,\mathbf{a}_K$ up to a scale factor.*

Proof: It can be verified that for a semi-unitary $\mathbf{A}$, $(\mathbf{A}^*\odot\mathbf{A})$ is also semi-unitary. With this property, we can show that

$$\mathbf{U}_s=(\mathbf{A}^*\odot\mathbf{A})\boldsymbol{\Gamma} \tag{23}$$

for some unitary $\boldsymbol{\Gamma}\in\mathbb{C}^{K\times K}$. Let $\boldsymbol{\eta}=\boldsymbol{\Gamma}\boldsymbol{\zeta}$. Then, we have $\boldsymbol{\eta}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$. Subsequently, the initialization can be expressed as

$$\mathbf{h}=\mathbf{U}_s\boldsymbol{\zeta}=(\mathbf{A}^*\odot\mathbf{A})\boldsymbol{\eta}. \tag{24}$$

Let us consider the devectorization of $\mathbf{h}$. Applying (7) to (24) yields

$$\mathrm{vec}^{-1}(\mathbf{h})=\mathbf{A}\mathrm{Diag}(\boldsymbol{\eta})\mathbf{A}^H. \tag{25}$$

Since $\mathbf{A}$ is semi-unitary, the right hand side of (25) is in fact an eigenvalue decomposition (EVD) of $\mathrm{vec}^{-1}(\mathbf{h})$. The remaining question is whether (25) is the unique EVD. If it does, then the AP step $\mathbf{a}:=\mathbf{q}_{\max}(\mathrm{vec}^{-1}(\mathbf{h}))$ will pick up a principal eigenvector of $\mathrm{vec}^{-1}(\mathbf{h})$, which is one of the $\mathbf{a}_k$ up to a scale factor. It is known that if the eigenvalues $\eta_1,\ldots,\eta_K$ are distinct, then the respective EVD is unique. As $\eta_i=\eta_j$ holds with probability zero for any $i\neq j$, we assert that the AP algorithm almost surely converges to a true mixing matrix column in its first iteration. ■

V. KR Subspace AP-based Algorithm

We now establish a blind identification algorithm that utilizes AP to identify the whole mixing matrix $\mathbf{A}$. In particular, we intend to make use of the rapid convergence characteristic of AP shown in Theorem 1. While Theorem 1 applies to semi-unitary $\mathbf{A}$ only, we can enforce mixing matrix semi-unitarity through prewhitening. To illustrate this, let

$$\bar{\mathbf{R}}\triangleq\frac{1}{M}\sum_{m=1}^{M}\mathbf{R}_m=\mathbf{A}\bar{\mathbf{D}}\mathbf{A}^H \tag{26}$$

be the time-average covariance matrix, where $\bar{\mathbf{D}}=\frac{1}{M}\sum_{m=1}^{M}\mathbf{D}_m$. We perform a square root factorization on $\bar{\mathbf{R}}$ (e.g., EVD) to obtain $\bar{\mathbf{R}}=\mathbf{B}\mathbf{B}^H$, where $\mathbf{B}\in\mathbb{C}^{N\times K}$. Then, we carry out the following prewhitening procedure:

$$\tilde{\mathbf{R}}_m=\mathbf{B}^\dagger\mathbf{R}_m(\mathbf{B}^\dagger)^H,\quad m=1,\ldots,M. \tag{27}$$

where $\mathbf{B}^\dagger$ is the pseudo-inverse of $\mathbf{B}$. It can be shown that

$$\tilde{\mathbf{R}}_m=\mathbf{G}\tilde{\mathbf{D}}_m\mathbf{G}^H,\quad m=1,\ldots,M, \tag{28}$$

where $\tilde{\mathbf{D}}_m = \bar{\mathbf{D}}^{-1}\mathbf{D}_m$, and $\mathbf{G} \in \mathbb{C}^{K\times K}$ is *unitary* and satisfies

$$\mathbf{G} = \mathbf{B}^\dagger \mathbf{A}\bar{\mathbf{D}}^{1/2}. \tag{29}$$

Hence, we can apply AP on $\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_M$ to obtain the columns of $\mathbf{G}$. Once the whole $\mathbf{G}$ is identified, we estimate the original mixing matrix by $\mathbf{BG}$; cf., (29).

As a side benefit of prewhitening, we can systematically identify all the columns of $\mathbf{G}$ by exploiting its column orthogonality. Suppose that we have identified $\mathbf{g}_r$. Let $\mathbf{G}_{-r} = [\mathbf{g}_1, \ldots, \mathbf{g}_{r-1}, \mathbf{g}_{r+1}, \ldots, \mathbf{g}_K]$. It can be easily shown that

$$\mathcal{R}(\mathbf{G}_{-r}^* \odot \mathbf{G}_{-r}) = \mathcal{R}\left(\mathbf{P}^\perp_{\mathbf{g}_r^*\otimes\mathbf{g}_r}(\mathbf{G}^* \odot \mathbf{G})\right) = \mathcal{R}\left(\mathbf{P}^\perp_{\mathbf{g}_r^*\otimes\mathbf{g}_r}\mathbf{U}_s\right)$$

where $\mathbf{P}^\perp_{\mathbf{g}_r^*\otimes\mathbf{g}_r} = \mathbf{I} - (\mathbf{g}_r^* \otimes \mathbf{g}_r)(\mathbf{g}_r^* \otimes \mathbf{g}_r)^H$ is the orthogonal complement projector of $\mathbf{g}_r^* \otimes \mathbf{g}_r$. By applying AP on $\mathbf{P}^\perp_{\mathbf{g}_r^*\otimes\mathbf{g}_r}\mathbf{U}_s$, we can identify another $\mathbf{g}_i$, $i \neq r$.

We finish this section by providing the complete pseudo-code of the proposed algorithm in Table I. For convenience, we will call this algorithm the *prewhitened AP algorithm* (PAPA).

TABLE I: The proposed prewhitened AP algorithm.

Given	local covariance matrices $\mathbf{R}_1, \ldots, \mathbf{R}_M$.
Step 1.	compute $\hat{\sigma}^2 = \min_{m=1,\ldots,M} \lambda_{\min}(\mathbf{R}_m)$, and obtain $\hat{\mathbf{R}}_m = \mathbf{R}_m - \hat{\sigma}^2\mathbf{I}$, $m = 1, \ldots, M$.
Step 2.	compute $\bar{\mathbf{R}} = \frac{1}{M}\sum_{m=1}^M \hat{\mathbf{R}}_m$, and perform a square-root factorization $\bar{\mathbf{R}} = \mathbf{B}\mathbf{B}^H$, where $\mathbf{B} \in \mathbb{C}^{N\times K}$.
Step 3.	compute $\tilde{\mathbf{R}}_m = \mathbf{B}^\dagger\hat{\mathbf{R}}_m(\mathbf{B}^\dagger)^H$, $m = 1, 2, \ldots, M$.
Step 4.	form $\mathbf{Y} = [\mathrm{vec}(\tilde{\mathbf{R}}_1), \ldots, \mathrm{vec}(\tilde{\mathbf{R}}_M)]$, and compute its compact SVD; i.e., $\mathbf{Y} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^H$, and set $i = 1$.
Step 5.	apply the AP algorithm with the basis matrix $\mathbf{U}_s$ and an initial point $\mathbf{h} = \mathbf{U}_s\boldsymbol{\zeta}$ for $\boldsymbol{\zeta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and record its output as $\mathbf{g}_i$.
Step 6.	compute $\mathbf{P}^\perp_{\mathbf{g}_i^*\otimes\mathbf{g}_i} = \mathbf{I} - (\mathbf{g}_i^*\otimes\mathbf{g}_i)(\mathbf{g}_i^*\otimes\mathbf{g}_i)^H$ and obtain the basis matrix of $\mathbf{P}^\perp_{\mathbf{g}_i^*\otimes\mathbf{g}_i}\mathbf{U}_s$, denoted by $\mathbf{Q}_s \in \mathbb{C}^{K^2\times(K-i)}$. Then, update $\mathbf{U}_s := \mathbf{Q}_s$.
Step 7.	set $i := i + 1$ and goto **Step 5.** until $i > K$.
Step 8.	output $\hat{\mathbf{A}} = \mathbf{BG}$, where $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_K]$.

VI. Simulations and Conclusion

We use simulations to demonstrate the performance of PAPA and compare it to some benchmarked QSS-based blind algorithms. The simulation settings are described as follows. We consider real-valued mixtures and sources, with $N = 11$ and $K = 10$. The mixing matrix $\mathbf{A}$ is randomly generated at each trial. The sources are real speech. We have a database of 23 speech signals, and at each trial we randomly pick K of them as the source signals. The total signal length is $T = 25600$. In order to obtain more local covariances under limited signal lengths, we employ 50% overlapping frames to estimate $\mathbf{R}_m$; specifically,

$$\mathbf{R}_m = \frac{1}{L}\sum_{t=0.5(m-1)L+1}^{0.5(m-1)L+L} \mathbf{x}(t)\mathbf{x}^H(t). \tag{30}$$

We set $L = 256$, and the subsequent number of local covariances is $M = 199$. These $\mathbf{R}_m$'s are then processed by the noise covariance removal procedure mentioned in Section II.

The benchmarked algorithms are TALS, Pham's JD, and FFDIAG. All the algorithms under test were run on the same set of noise covariance removed $\{\mathbf{R}_m\}$. In PAPA, we adopt a standard stopping criterion for its per-source AP processes, specifically, $|f^{(n)} - f^{(n-1)}|/|f^{(n)}| < \epsilon = 10^{-4}$, where $f^{(n)}$ is the objective value of the algorithm at the nth iteration. The other algorithms also use the same type of criterion to stop. The performance measure employed is the average mean square error (MSE), defined as

$$\mathrm{MSE} = \min_{\substack{\boldsymbol{\pi}\in\Pi,\\ c_1,\ldots,c_K\in\{\pm1\}}} \frac{1}{K}\sum_{k=1}^K \left\| \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|} - c_k\frac{\hat{\mathbf{a}}_{\pi_k}}{\|\hat{\mathbf{a}}_{\pi_k}\|}\right\|^2,$$

where Π is the set of all permutations of $\{1, 2, \ldots, K\}$; $\mathbf{A}$ and $\hat{\mathbf{A}}$ are the true and estimated mixing matrices, respectively. The signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = \frac{1}{T}\sum_{t=0}^{T-1} E\{||\mathbf{As}(t)||^2\}/E\{||\mathbf{v}(t)||^2\}$.

Fig. 1 shows the MSEs of the various algorithms; the results were obtained by 1000 independent trials. We see that for $\mathrm{SNR} \leq 30$dB, PAPA generally provides the best MSE performance compared to the other algorithms. FFDIAG is also competitive, yielding MSEs almost identical to that of PAPA for $\mathrm{SNR} \geq 15$dB. PAPA nevertheless gives better MSEs than FFDIAG for $\mathrm{SNR} < 15$dB. Pham's JD yields better MSEs for $\mathrm{SNR} \geq 35$dB, which is very high[1]; otherwise its MSEs are higher than those of PAPA and FFDIAG. We tried two different implementations of TALS: In the legend of Fig. 1, "TALS" is a standard TALS, while "TALS with warm start" has its initialization done by a cheap version of PAPA, specifically, restricting PAPA's AP iterations to one only. As seen in Fig. 1, our warm start attempt improves the MSEs of TALS. This illustrates that PAPA can also be used to warm start other blind algorithms, especially those requiring reasonable initializations, to help improve their performance.

Fig. 2 plots the average runtimes of the various algorithms corresponding to the above simulation. All the algorithms were run on a 2.80GHz Desktop PC, written in MATLAB. PAPA is seen to have the lowest runtime. FFDIAG also shows good runtime, with its gap relative to PAPA being higher for low SNRs. Moreover, TALS yields an improved runtime performance when warm start by PAPA is employed.

To get a better idea of the computational efficiency of PAPA, we examine its number of iterations used. Recall that in Theorem 1, we prove a key result that AP should converge to the ground truth in one iteration, when there is no modeling error in $\{\mathbf{R}_m\}$. Table II shows the average number of AP iterations corresponding to the above simulation. From the second row of the table, we see that the number of AP iterations is higher than what we expect in theory; this is owing to noise corruption and signal covariance estimation errors, which constitute the modeling errors. To verify Theorem 1, we perform an additional numerical evaluation where $\{\mathbf{R}_m\}$ is generated following the model in (3) exactly; i.e., synthesizing an ideal model. The results are tabulated in the third row of

[1]This high-SNR performance advantage may be attributed to the fact that Pham's JD criterion is ML in the noiseless Gaussian QSS case [4].

TABLE II: Average number of iterations of AP in PAPA.

SNR(dB)	-10	0	10	20
real speech	9.45	5.28	4.70	4.98
ideal model (ref. only)	2	2	2	2

Table II, where we see that the number of AP iterations are all 2 (the 2nd iteration is required for AP to realize that the estimate is good enough), verifying our claim in Theorem 1. We also note that for real speech, the number of AP iterations is about 5 for moderate to high SNRs, which is quite modest.

In Figs. 3-4, we illustrate the MSEs and runtimes of the various algorithms w.r.t. the number of frames M. Note that $L = 256$, the SNR is fixed at 10dB, and the number of trials is 200. Again, PAPA is seen to yield competitive estimation accuracy and complexity in general.

To conclude, we have established an alternative approach to QSS-based blind identification using a KR subspace formulation. In particular, we have developed a blind identification algorithm that has a desirable theoretical convergence property. Simulation results have shown that the proposed algorithm is competitive in performance and complexity.

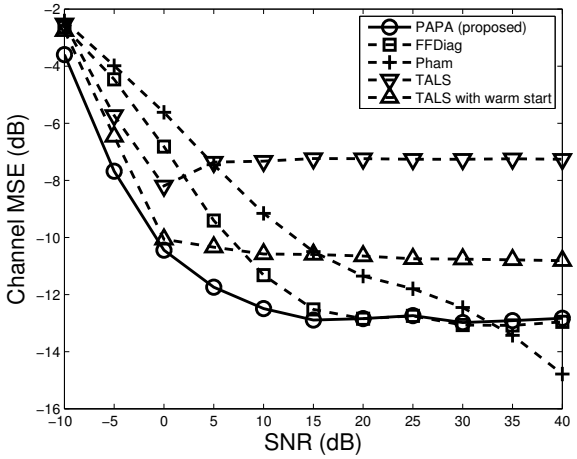

Fig. 1: The average MSE of the various algorithms.

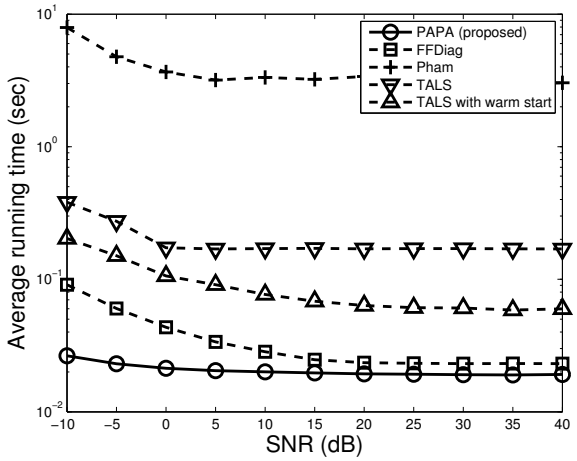

Fig. 2: The average running time of the various algorithms.

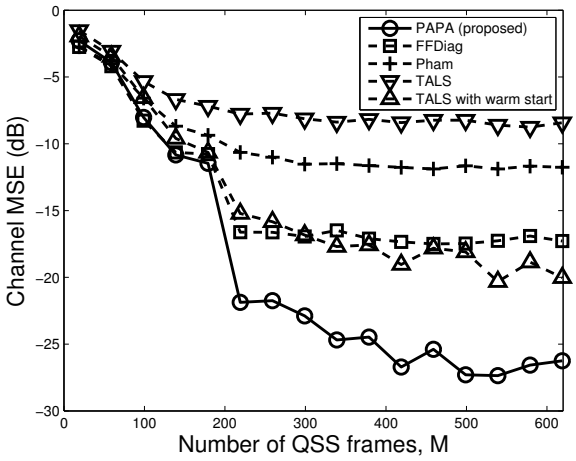

Fig. 3: The average MSE of the various algorithms.

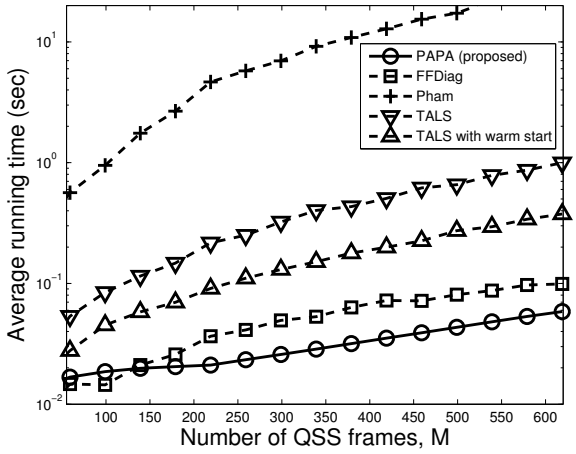

Fig. 4: The average running time of the various algorithms.

REFERENCES

[1] 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.

[2] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1697 – 1710, May. 2005.

[3] 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.

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

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

[6] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270 –3278, Sep. 2006.

[7] W.-K. Ma, T.-H. Hsieh, and C.-Y. Chi, "DOA estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: A Khatri-Rao subspace approach," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2168–2180, Apr. 2010.

[8] S. Boyd and J. Dattoro, *Alternating Projections*, EE392o: Optimization Projects, Stanford University, 2003, available online: http://www.stanford.edu/class/ee392o/alt_proj.pdf