ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10446804

COMPLEX BOUNDED COMPONENT ANALYSIS: IDENTIFIABILITY AND ALGORITHM

Jingzhou Hu and Kejun Huang

Department of CISE, University of Florida, Gainesville, FL 32611
(jingzhouhu,kejun.huang)@ufl.edu

ABSTRACT

In this paper, we study the complex bounded component analysis (BCA) problem, under the novel scenario that the magnitudes of the latent sources are bounded. This is opposed to the existing models that assume separate bounds on the real and imaginary parts, in which case the problem can be transformed into a real BCA problem, but loses the point of introducing complex numbers into the model in the first place. Unlike in the real case, it is hard to visualize a geometric interpretation for the complex BCA model. Nevertheless, we draw algebraic insights from prior work and propose a new formulation that uses the determinant of the mixing matrix as the identification criterion, and show that complex BCA is identifiable if the disked hull of the sources is sufficiently scattered in the complex hypercube. This result significantly extends the prior knowledge on BCA, and in a broader sense is perhaps the first identifiable unmixing model with parts of the sources being quadratically dependent (since the magnitude of the complex sources are bounded). We also present a new learning algorithm to solve the proposed complex BCA formulation based on linearized ADMM, and show numerically that the performance is surprisingly effective.

1. INTRODUCTION

Bounded component analysis (BCA) is an important type of method for blind source separation. As an unsupervised learning method, it makes the minimal assumption that the support of the latent components is element-wise bounded [1, 2]. In a lot of applications, it is used as an alternative to the well-knwon independent component analysis (ICA) [3] since the identifiability of ICA is based on the assumption that the latent sources are statistically independent, which may be hard to verify in practice. BCA, on the other hand, has recently been proven to be identifiable under the sufficiently scattered condition with a finite amount of data [4, 5], making it potentially more applicable than what most people realize.

In this paper we study BCA in the complex domain, i.e., when both the latent sources and the mixing matrix can take complex values. It is a problem that has not been well-studied in the literature. Among the few papers that considered the complex BCA problem, they all essentially transform it into a real BCA problem by assuming that the real and imaginary parts of the sources are independently bounded [6, 7]. However, when it is necessary to introduce complex numbers into the model, it is generally more suitable that the *magnitude* of the complex sources are bounded. This is a model that has never been studied before, yet we petition that this should be the correct problem formulation for complex BCA, which we elaborate in detail next.

Notations. Vectors and matrices are denoted with boldface italic lowercase and uppercase letters, e.g., $\boldsymbol{x}$ and $\boldsymbol{X}$. Superscripts $^\top$, $^\mathsf{H}$, $^{-1}$, and † denote transpose, Hermitian (conjugate transpose), inverse, and (Moore-Penrose) pseudo-inverse, respectively. We overload the absolute value $|\cdot| \leq 1$ with a vector or matrix argument to denote that the magnitude of its complex values are element-wise ≤ 1. Since we study the identifiability of an unsupervised learning model, we use superscript $^\natural$ to denote the groundtruth latent factors that generates the data, and superscript * to denote the output of a learning algorithm to see if they are (essentially) the same. We use ∂ in front of a set to denote its boundary. Finally, disk$(\cdot)$ denotes the disked hull of a set, which is officially defined in Definition 3.

Supported in part by NSF ECCS-2237640 and NIH R01LM014027.

1.1. The complex BCA model

Consider the classical blind source separation (BSS) model of a set of n data points generated as:

$$\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{s}_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{C}^d$ are the observations, $\boldsymbol{A} \in \mathbb{C}^{d\times k}$ is the *unknown* mixing matrix, and $\boldsymbol{s}_i \in \mathbb{C}^k$ are the latent sources that one is interested in recovering. Stacking all $\boldsymbol{x}_i$ as columns of the $d \times n$ matrix $\boldsymbol{X}$ and $\boldsymbol{s}_i$ as columns of the $k \times n$ matrix $\boldsymbol{S}$ gives the matrix factorization model $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$. Without additional assumptions on the latent factors, it is impossible to uniquely identify the mixing matrix $\boldsymbol{A}$ and the latent components $\boldsymbol{S}$, since we can always "insert" an invertible matrix $\boldsymbol{Q}$ and $\boldsymbol{Q}^{-1}$ as $\boldsymbol{X} = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{S}}$ where $\widetilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{Q}$ and $\widetilde{\boldsymbol{S}} = \boldsymbol{Q}^{-1}\boldsymbol{S}$, and one cannot distinguish whether $\boldsymbol{S}$ or $\widetilde{\boldsymbol{S}}$ are the groundtruth sources. Such rotation ambiguity cannot be resolved by the well-known principal component analysis (PCA) [8]. However, by imposing various structures onto the latent sources $\boldsymbol{S}$, identifiability can be established, such as independent component analysis (ICA) [3], nonnegative matrix factorization [9, 10], simplex structures [11–13], and dictionary learning [14].

Bounded component analysis (BCA) is another such model that assumes each component of $\boldsymbol{s}$ is bounded [1, 2]. In the real case, it is recently proven that BCA is identifiable if the source matrix satisfies a so-called "sufficiently scattered" condition in the hypercube [4], and by allowing an additional *shift ambiguity* the groundtruth bound can even be unknown and assymetric around zero [5]. However, when the sources are complex, the only existing methods assume that the real and imaginary parts are independently bounded, thus any real BCA approach can be applied [7]. We argue that this is not the proper extension to the complex domain as it goes against the need to introduce complex sources in the first place. It intuitively makes more sense to assume that the *magnitude* of the sources are bounded, i.e., there exists a nonnegative vector $\boldsymbol{u}$ such that $|\boldsymbol{s}_i| \leq \boldsymbol{u}$ for all $i = 1, \ldots, n$, where the absolute value (magnitude) $|\cdot|$ and inequality $\leq$ are both taken element-wise.

Our proposed complex BCA model, while being more intuitive and much more applicable in practice, introduces additional hurdles in terms of both identifiability and algorithm. Unlike real BCA, our assumption on the bounded magnitude is inherently *quadratic*, making the identifiability of complex BCA seemingly impossible as all other identifiable unmixing models all assume some variant of a *linear* structure on the latent sources, such as a simplex (NMF), a hypercube (real BCA), or an orthoplex (dictionary learning), while a quadratic structure of uncorrelatedness is not identifiable (PCA). To much of our surprise, in this case with bounded magnitude, the model is still identifiable under the *complex sufficiently scattered* condition.

1.2. Proposed formulation

Most BSS models allow scaling and permutation ambiguity, i.e., a permutation matrix $\boldsymbol{\Pi}$ and a diagonal matrix $\boldsymbol{D}$ such that the recovered mixing matrix is $\boldsymbol{AD\Pi}$ and the recovered sources are $\boldsymbol{\Pi}^\top\boldsymbol{D}^{-1}\boldsymbol{S}$. For complex BCA, one can always scale the latent sources to lie in the unit circle, so without loss of generality, we assume that $|\boldsymbol{S}| \leq 1$. The identifiability of BCA is formally defined as follows:

Definition 1. Consider the generative model $\boldsymbol{X} = \boldsymbol{A}^\natural\boldsymbol{S}^\natural$, where $\boldsymbol{A}^\natural$ is the groundtruth mixing matrix and $|\boldsymbol{s}_i| \leq \boldsymbol{u}$ are the groundtruth latent components with element-wise bounded magnitude, where $\boldsymbol{s}_1^\natural, \ldots, \boldsymbol{s}_n^\natural$ are columns of $\boldsymbol{S}^\natural$ and the $\boldsymbol{u}$ is the element-wise unknown bounds. Let $(\boldsymbol{A}^\star, \boldsymbol{S}^\star)$ be optimal for an identification criterion q

$$(\boldsymbol{A}^\star, \boldsymbol{S}^\star) = \underset{\boldsymbol{X}=\boldsymbol{AS}, |\boldsymbol{S}|\leq 1}{\arg\min} \; q(\boldsymbol{A}, \boldsymbol{S}).$$

If $\boldsymbol{A}^\natural$ and/or $\boldsymbol{S}^\natural$ satisfy some condition such that for any $(\boldsymbol{A}^\star, \boldsymbol{S}^\star)$, there exist a permutation matrix $\boldsymbol{\Pi}$ and a complex diagonal matrix $\boldsymbol{D}$ with

$$\boldsymbol{A}^\natural = \boldsymbol{A}^\star\boldsymbol{D\Pi} \qquad \text{and} \qquad \boldsymbol{S}^\natural = \boldsymbol{\Pi}^\top\boldsymbol{D}^{-1}\boldsymbol{S}^\star,$$

then we say that the complex BCA model is essentially identifiable, up to permutation and complex scaling, under that condition.

So far the complex BCA model with bounded magnitude has not been studied before, to the best of our knowledge, let alone its identifiability issue. When all the elements are restricted to be real, then the bounded constraint is essentially equivalent to an arbitrary but symmetric bound $-\boldsymbol{u} \leq \boldsymbol{s}_i \leq \boldsymbol{u}$, and recently it has been shown that such a model is identifiable if the convex hull of $\boldsymbol{S}^\natural$, after rescaling it to be between $[-1, 1]^{k\times n}$, satisfies a so-called sufficiently scattered condition [4]. It is later extended to allow the sources to be bounded in an assymetric box $\boldsymbol{l} \leq \boldsymbol{s}_i \leq \boldsymbol{u}$, and the model is still identifiable by allowing an additional shift-ambiguity [5]. Neither results can be applied to the complex BCA model unless the bounds to the latent sources are on the real and imaginary parts separately, which makes little sense for a model that involves complex numbers in the first place.

Nevertheless, inspired by novel formulations of real BCA proposed in [5], we propose the following problem formulation for complex BCA:

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{S}}{\text{minimize}} \quad & \det \boldsymbol{A}^{\mathsf{H}}\boldsymbol{A} \\ \text{subject to} \quad & \boldsymbol{X} = \boldsymbol{AS}, |\boldsymbol{S}| \leq 1. \end{aligned} \tag{2}$$

Formulation (2) falls into the framework introduced in Definition 1, with the identifiability criterion being specifically $\det \boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}$. In the rest of this paper, we will show that complex BCA is identifiable via solving (2) under some conditions, and we propose a highly efficient algorithm to solve it approximately, with surprising effectiveness.

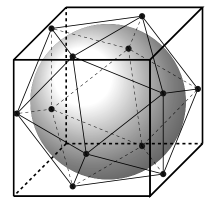

Fig. 1: An example of sufficiently scattered in the hypercube of $\mathbb{R}^3$.

2. IDENTIFIABILITY

Problem (2) provides an intuitive identification criterion for BCA. In the real case, the same identification criterion has been applied in various models, such as NMF [10], dictionary learning [14], and, above all else, real BCA [5], each having an intuitive geometric interpretation of finding the minimum-volume enclosing polytope for a set of points. In the complex domain, it does not seem immediately obvious how those geometric interpretation could provide insight. Nevertheless, we will continue with the algebraic representations of the concepts that could be otherwise visualized in the real domain, and show mathematically that similar results still hold even though they may not be geometrically interpreted with ease.

Definition 2 (Disked set)**.** A set $\mathcal{S} \in \mathbb{C}^k$ is said to be disked if for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}$ and $\theta_1, \theta_2 \in \mathbb{C}$ such that $|\theta_1| + |\theta_2| = 1$, we have $\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2 \in \mathcal{S}$.

In the real Euclidean space, a disked set is also said to be *absolutely convex* [15], as the definition differs from a convex set only by allowing the coefficients of the linear combination to be absolutely sum to one.

Definition 3 (Disked hull)**.** The disked hull of a set $\mathcal{S}$, denoted as $\operatorname{disk}\mathcal{S}$, is the smallest disked set that contains $\mathcal{S}$. When $\mathcal{S}$ is a finite set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, then its disked hull is

$$\operatorname{disk}\mathcal{S} = \left\{ \sum_{i=1}^n \theta_i\boldsymbol{s}_i \;\middle|\; \sum_{i=1}^n |\theta_i| \leq 1 \right\}.$$

Assumption 1 (Sufficiently scattered in the complex hypercube)**.** Let $\mathcal{B}$ denote the complex Euclidean ball $\mathcal{B} = \{\boldsymbol{x} \in \mathbb{C}^k \mid \|\boldsymbol{x}\| \leq 1\}$ and $\mathcal{C}$ denote the complex hypercube $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{C}^k \mid \|\boldsymbol{x}\|_\infty \leq 1\}$. A set $\mathcal{S}$ is sufficiently scattered in the complex hypercube if:

1. $\mathcal{B} \subseteq \mathcal{S} \subseteq \mathcal{C}$;
2. $\partial\mathcal{B} \cap \partial\mathcal{S} = \{\alpha\boldsymbol{e}_i \mid |\alpha| = 1, i = 1, \ldots, k\}$, where ∂ denotes the boundary of the set, and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ are the k unit vectors in $\mathbb{R}^k$.

If we restrict ourselves in the real domain, then a geometric illustration of a polytope that satisfies the sufficiently scattered condition is shown in Figure 1. The term "sufficiently scattered" first appeared in [16] to characterize the identifiability condition for nonnegative matrix factorization that has already appeared in [9]. The difference is that in [9, 10, 16], the condition is defined over the conic hull of a set of points in the nonnegative orthant containing a specific hyperbolic cone. It has also been defined over the convex hull of a set of points in the probability simplex [11–13]. The most related case is the sufficiently scattered condition defined for the standard hypercube (L_∞-norm ball) [4], as well as the standard orthoplex (L_1-norm ball). In this paper, we further extend this important notion into the complex domain.

Our main result on the identifiability of complex BCA is presented as follows:

Theorem 1. *Consider the complex BCA model* $\boldsymbol{X} = \boldsymbol{A}^\natural \boldsymbol{S}^\natural$*, where* $\boldsymbol{A}^\natural \in \mathbb{C}^{d\times k}$ *is the groundtruth mixing matrix and* $|\boldsymbol{S}^\natural| \leq 1$ *are the groundtruth latent complex components with element-wise bounded magnitude. If* $\operatorname{rank}(\boldsymbol{A}^\natural) = k$ *and* $\operatorname{disk}(\boldsymbol{S}^\natural)$ *is sufficiently scattered in the complex hypercube as in Assumption 1, then for any solution of* (2)*, denoted as* $(\boldsymbol{A}^\star, \boldsymbol{S}^\star)$*, there exist a permutation matrix* $\boldsymbol{\Pi}$ *and a complex diagonal matrix* $\boldsymbol{D}$ *with unimodular values on the diagonal such that*

$$\boldsymbol{A}^\natural = \boldsymbol{A}^\star \boldsymbol{D}\boldsymbol{\Pi} \qquad \text{and} \qquad \boldsymbol{S}^\natural = \boldsymbol{\Pi}^\top \boldsymbol{D}^{-1}\boldsymbol{S}^\star.$$

In other words, complex BCA is identifiable if the groundtruth $\boldsymbol{A}^\natural$ *has full column rank and the disked hull of* $\boldsymbol{S}^\natural$ *is sufficiently scattered.*

Due to space limitation, we provide a sketch of the proof. The full proof is relegated to the journal version.

Proof sketch. Since both $(\boldsymbol{A}^\natural, \boldsymbol{S}^\natural)$ and $(\boldsymbol{A}^\star, \boldsymbol{S}^\star)$ are feasible for (2), we immediately have that $\det \boldsymbol{A}^{\star\mathsf{H}}\boldsymbol{A}^\star \leq \det \boldsymbol{A}^{\natural\mathsf{H}}\boldsymbol{A}^\natural$. Define $\boldsymbol{M} = (\boldsymbol{A}^\star)^\dagger \boldsymbol{A}^\natural$, then

$$|\det \boldsymbol{M}| = \sqrt{\det(\boldsymbol{A}^{\star\mathsf{H}}\boldsymbol{A}^\star)^{-1}\boldsymbol{A}^{\natural\mathsf{H}}\boldsymbol{A}^\natural} \geq 1. \tag{3}$$

On the other hand, since $\boldsymbol{M}\boldsymbol{S}^\natural = (\boldsymbol{A}^\star)^\dagger \boldsymbol{A}^\natural \boldsymbol{S}^\natural = \boldsymbol{S}^\star$, we also have that $|\boldsymbol{M}\boldsymbol{S}^\natural| \leq 1$. Let $\boldsymbol{m}^\mathsf{H}$ be any row of $\boldsymbol{M}$, then $|\boldsymbol{m}^\mathsf{H}\boldsymbol{s}_i^\natural| \leq 1$ for all $i = 1, \ldots, n$. Now we invoke the assumption that $\operatorname{disk}(\boldsymbol{S}^\natural)$ is sufficiently scattered, then for any $\boldsymbol{v}$ with unit norm $\|\boldsymbol{v}\| = 1$, there must exist $\theta_1, \ldots, \theta_n$ with $|\theta_1| + \cdots + |\theta_n| = 1$ such that $\boldsymbol{v} = \theta_1 \boldsymbol{s}_1^\natural + \cdots + \theta_n \boldsymbol{s}_n^\natural$. Therefore

$$\left|\boldsymbol{m}^\mathsf{H}\boldsymbol{v}\right| = \left|\sum_{i=1}^n \theta_i \boldsymbol{m}^\mathsf{H}\boldsymbol{s}_i^\natural\right| \leq \sum_{i=1}^n |\theta_i| \left|\boldsymbol{m}^\mathsf{H}\boldsymbol{s}_i^\natural\right| \leq 1. \tag{4}$$

Since (4) holds for every $\boldsymbol{v}$ with unit norm, we must have $\|\boldsymbol{m}\| \leq 1$ as well (otherwise we would let $\boldsymbol{v} = \boldsymbol{m}/\|\boldsymbol{m}\|$ and have that $|\boldsymbol{m}^\mathsf{H}\boldsymbol{v}| = \|\boldsymbol{m}\| > 1$, contradicting (4)), which means every row of $\boldsymbol{M}$ has norm no greater than 1. This gives us

$$|\det \boldsymbol{M}| \leq \prod_{j=1}^k \|\boldsymbol{m}_j\| \leq 1, \tag{5}$$

where the first inequality is due to the Hadamard inequality. Combining (3) and (5) shows that $(\boldsymbol{A}^\natural, \boldsymbol{S}^\natural)$ or any of their column permutation and unimodular scaling is in the set of optimal solutions of (2). In the journal version of this paper, we will complete the proof by showing that if the second requirement of the sufficiently scattered condition (c.f. Assumption 1) is satisfied, then column permutation and unimodular scaling of $(\boldsymbol{A}^\natural, \boldsymbol{S}^\natural)$ are the only optimal solutions of (2), which leads to the identifiability of complex BCA. □

Remark. Although the overall steps of showing identifiability of complex BCA seems somewhat similar to that of the real case in [4,5], we would like to reemphasize that the implication is highly nontrivial and very significant, as it is the first identifiable unmixing model with parts of the sources being *quadratically dependent* with each other, in this case the real and imaginary parts of each individual sources. What is more, the introduction of the concept of the *disked hull* provides new insights to the real BCA model, as it effortlessly shows that real BCA is identifiable if $\operatorname{disk}(\boldsymbol{S}^\natural)$ is sufficiently scattered, which is a more relaxed condition than that of [4] that $\operatorname{conv}(\boldsymbol{S}^\natural)$ is sufficiently scattered, since we always have $\operatorname{conv}(\boldsymbol{S}^\natural) \subset \operatorname{disk}(\boldsymbol{S}^\natural)$.

Similar to the real case, one may show that a bounded source matrix is more likely to be sufficiently scattered if many of its entries lie on the boundary of the complex hypercube, i.e., with magnitudes exactly equal to 1. We will use this intuition to synthetically generate bounded complex sources that are identifiable as per Theorem 1.

3. ALGORITHM

In this section, we propose an algorithm based on the linearized alternating direction method of multipliers (L-ADMM) with well-defined and low-complexity iterations.

Note that in problem (2), since we assume that the columns of $\boldsymbol{A}$ are linearly independent, we can define $\boldsymbol{P} = \boldsymbol{A}^\dagger$ and apply a change of variable to problem (2). Then, it becomes an equivalent object of minimizing $1/\det \boldsymbol{P}\boldsymbol{P}^\mathsf{H}$, after we apply the log function and make it $-\log\det \boldsymbol{P}\boldsymbol{P}^\mathsf{H}$, and in the constraints we can now eliminate the $\boldsymbol{S}$ variables by simply requiring $|\boldsymbol{P}\boldsymbol{X}| \leq 1$. This leads to the following reformulation

$$\begin{aligned} \underset{\boldsymbol{P}}{\text{minimize}} \quad & -\log\det \boldsymbol{P}\boldsymbol{P}^\mathsf{H} \\ \text{subject to} \quad & |\boldsymbol{P}\boldsymbol{X}| \leq 1. \end{aligned} \tag{6}$$

Problem (6) now has a convex, or more specifically linear, constraint set, although the objective is still not convex.

We further modify the formulation (6) by introducing an auxiliary variable $\boldsymbol{S}$:

$$\begin{aligned} \underset{\boldsymbol{P},\boldsymbol{S}}{\text{minimize}} \quad & -\log\det \boldsymbol{P}\boldsymbol{P}^\mathsf{H} \\ \text{subject to} \quad & \boldsymbol{P}\boldsymbol{X} = \boldsymbol{S}, \;\; |\boldsymbol{S}| \leq 1. \end{aligned} \tag{7}$$

Formulation (7) consists of two sets of variables ($\boldsymbol{P}$ and $\boldsymbol{S}$) over two separable functions (one of them being an indicator function of $\boldsymbol{S}$ that the ℓ_∞ norms of its rows are no bigger than 1) and linear equality constraints. It is easy to derive the alternating direction method of multipliers (ADMM) [17] for (7):

$$\begin{cases} \boldsymbol{P}_{(t+1)} \leftarrow \arg\min\limits_{\boldsymbol{P}} \; -\log\det \boldsymbol{P}\boldsymbol{P}^\mathsf{H} + \rho\|\boldsymbol{P}\boldsymbol{X} - \boldsymbol{S}_{(t)} + \boldsymbol{U}_{(t)}\|^2, & \text{(8a)} \\ \boldsymbol{S}_{(t+1)} \leftarrow \arg\min\limits_{\boldsymbol{S}} \; \mathbb{1}_{|\cdot|\leq 1}(\boldsymbol{S}) + \rho\|\boldsymbol{P}_{(t+1)}\boldsymbol{X} - \boldsymbol{S} + \boldsymbol{U}_{(t)}\|^2, & \text{(8b)} \\ \boldsymbol{U}_{(t+1)} \leftarrow \boldsymbol{U}_{(t)} + \boldsymbol{P}_{(t+1)}\boldsymbol{X} - \boldsymbol{S}_{(t+1)}. & \text{(8c)} \end{cases}$$

The second step (8b) is well-defined, as it projects each element of the complex matrix $\boldsymbol{P}_{(t+1)}\boldsymbol{X} + \boldsymbol{U}_{(t)}$ to the unit circle, which boils down to rescaling the elements with magnitude bigger than 1 to be unimodular. The first step (8a), however, is not clear how to compute. One popular method to mitigate this issue, which is prevalent in ADMM, is to take a linear approximation of the loss function of $\boldsymbol{P}$ at the previous iterate $\boldsymbol{P}_{(t)}$ [18]. The gradient of $\log\det \boldsymbol{P}\boldsymbol{P}^\mathsf{H}$ is $2(\boldsymbol{P}^\dagger)^\mathsf{H}$, therefore the linear approximation of $-\log\det \boldsymbol{P}\boldsymbol{P}^\mathsf{H}$ at $\boldsymbol{P}_{(t)}$ is $-\log\det \boldsymbol{P}_{(t)}\boldsymbol{P}_{(t)}^\mathsf{H} - 2\operatorname{Tr}\boldsymbol{P}_{(t)}^\dagger(\boldsymbol{P} - \boldsymbol{P}_{(t)})$. The overall update of $\boldsymbol{P}_{(t+1)}$ becomes minimizing a convex quadratic function, which can be done in closed-form. The derived linearized ADMM (L-ADMM) iterates are

$$\begin{cases} \boldsymbol{P}_{(t+1)} \leftarrow ((\boldsymbol{S}_{(t)} - \boldsymbol{U}_{(t)})\boldsymbol{X}^\mathsf{H} + (1/\rho)(\boldsymbol{P}_{(t)}^\dagger)^\mathsf{H})(\boldsymbol{X}\boldsymbol{X}^\mathsf{H})^{-1}, & \text{(9a)} \\ \boldsymbol{S}_{(t+1)} \leftarrow \operatorname{Proj}_{|\cdot|\leq 1}(\boldsymbol{P}_{(t+1)}\boldsymbol{X} + \boldsymbol{U}_{(t)}), & \text{(9b)} \\ \boldsymbol{U}_{(t+1)} \leftarrow \boldsymbol{U}_{(t)} + \boldsymbol{P}_{(t+1)}\boldsymbol{X} - \boldsymbol{S}_{(t+1)}. & \text{(9c)} \end{cases}$$

Algorithm 1 Solving (6) with L-ADMM

take the QR factorization of $\boldsymbol{X}^{\mathsf{H}} = \boldsymbol{Q}\boldsymbol{R}$
set $\rho = nk$ and initialize $\boldsymbol{P}_{(0)}$
for $t = 0, 1, 2, \dots$ until convergence **do**
  $\boldsymbol{P}_{(t+1)} \leftarrow (\boldsymbol{S}_{(t)} - \boldsymbol{U}_{(t)})\boldsymbol{Q} + (1/\rho)\boldsymbol{P}_{(t)}^{-\mathsf{H}}$
  $\boldsymbol{S}_{(t+1)} \leftarrow \mathrm{Proj}_{|\cdot|\leq 1}(\boldsymbol{P}_{(t+1)}\boldsymbol{Q}^{\mathsf{H}} + \boldsymbol{U}_{(t)})$
  $\boldsymbol{U}_{(t+1)} \leftarrow \boldsymbol{U}_{(t)} + \boldsymbol{P}_{(t+1)}\boldsymbol{Q}^{\mathsf{H}} - \boldsymbol{S}_{(t+1)}$
end for

Finally, we notice that when $k = d$, i.e., when the mixing matrix $\boldsymbol{A}$ is square, formulation (7) and the derived L-ADMM algorithm are invariant under linear transformation of columns of $\boldsymbol{X}$, meaning if we replace $\boldsymbol{X}$ with $\widetilde{\boldsymbol{X}} = \boldsymbol{G}\boldsymbol{X}$ in (7), where $\boldsymbol{G}$ is a $k \times k$ invertible matrix, then every iterate $\boldsymbol{P}_{(t)}$ is uniquely mapped to $\boldsymbol{P}_{(t)}\boldsymbol{G}^{-1}$ (while $\boldsymbol{S}_{(t)}$ and $\boldsymbol{U}_{(t)}$ are exactly the same), and the objective value has a constant difference $-\log|\det \boldsymbol{P}_{(t)}\boldsymbol{G}^{-1}| = -\log|\det \boldsymbol{P}_{(t)}| + \log|\det \boldsymbol{G}|$. This observation allows us to preprocess the data matrix $\boldsymbol{X}$ by orthogonalizing its rows, so that the update for $\boldsymbol{P}$ in (9a) can be further simplified. The overall algorithm is summarized in Algorithm 1. We empirically found that setting $\rho = 1$ works very well in practice. We would like to emphasize that the affine invariance does not hold in general when the mixing matrix $\boldsymbol{A}$ is tall, as $\log\det \boldsymbol{P}\boldsymbol{G}^{-1}\boldsymbol{G}^{-\top}\boldsymbol{P}^{\top}$ is not in general a constant difference from $\log\det \boldsymbol{P}\boldsymbol{P}^{\top}$; moreover, if $\boldsymbol{P}$ is wide, we generally do not have that $(\boldsymbol{P}\boldsymbol{G}^{-1})^{\dagger} = \boldsymbol{G}\boldsymbol{P}^{\dagger}$. Therefore, in the over-determined case one should still stick to the updates given in (9). Computations can still be made more efficient by caching the Cholesky factorization and perform only the forward/backward substitutions in every iteration, as suggested in [17].

4. NUMERICAL VALIDATION

We conclude the proposed theoretical result by providing some numerical validation. We fix $n = 1000$ and $d = k = 20$. To evaluate the identifiability result, we generate the mixing matrix $\boldsymbol{A}^{\natural}$ with elements (real and imaginary parts) drawn from i.i.d. normal distributions; with respect to the source matrix $\boldsymbol{S}^{\natural}$, we generate it elementwise with each entry's phase uniformly distributed in $[0, 2\pi]$, while its magnitude equals to 1 with probability p or uniformly distributed in $[0, 1]$ with probability $1 - p$. As we explained earlier, the higher the p, the more likely the disked hull of the generated source matrix satisfies the complex sufficiently scattered condition, thus can be uniquely identified via solving (2). We will refer to this factor p as the "scattering level" in the sequel.

4.1. Convergence of Algorithm 1

We start by evaluating the performance of Algorithm 1. Since Problem (6) is nonconvex, one would expect that the algorithm may sometimes stuck at a local optimum. Much to our surprise, Algorithm 1 seems to always find the optimal solution when the BCA model is identifiable, meaning it always recovers the groundtruth factors up to column permutation and sign ambiguities as we know they are the optimal solution as per our identifiability result given in Theorem 1.

The convergence of Algorithm 1 on 10 random instances are shown in Fig. 2. In order to guarantee that the model is identifiable, we fix the scattering level $p = 0.5$. Since we know $\boldsymbol{A}^{\natural}$ is optimal for (2), then the optimal value of (6) must be $-2\log|\det \boldsymbol{A}^{\natural}|$. On the other hand, since L-ADMM directly tackles formulation 7, it is not guaranteed that $\boldsymbol{P}_{(t)}$ is feasible in every iteration, which makes little sense to check the difference $-\log|\det \boldsymbol{P}_{(t)}| + \log|\det \boldsymbol{P}^{\star}|$. We

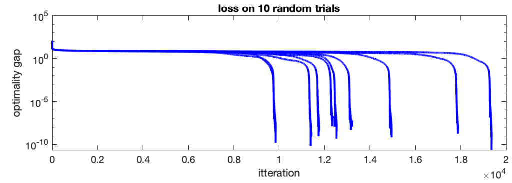

Fig. 2: The convergence of Algorithm 1 on 10 random instances.

instead check the optimality gap of the Lagrangian function values, using the optimal dual variable $\boldsymbol{\Lambda}$, since we have

$$-\log|\det \boldsymbol{P}| + \mathrm{Tr}(\boldsymbol{P}\boldsymbol{X} - \boldsymbol{S})\boldsymbol{\Lambda}^{\star} \geq -\log|\det \boldsymbol{P}^{\star}|,$$

for any $\boldsymbol{P}$ and a feasible $\boldsymbol{S}$. Obviously, the gap equals to zero when $\boldsymbol{P} = \boldsymbol{P}^{\star}$ and $\boldsymbol{S} = \boldsymbol{S}^{\star}$, in which case $\boldsymbol{P}^{\star}\boldsymbol{X} - \boldsymbol{S}^{\star} = 0$. Furthermore, it is easy to show that an optimal $\boldsymbol{\Lambda}$ is $(\boldsymbol{S}^{\natural})^{\dagger}$. In this simulation with known groundtruth factors, we will use this to measure the optimality gap shown on the vertical axis of Fig. 2. As we can see, in all 10 instances a global optimum is attained. The surprising effectiveness is well-worth further investigation.

4.2. Identifiability performance

Finally, we showcase how the "scattered level" of the latent sources affect the identifiability performance. As we explained in the previous subsection, a source matrix is more likely to be sufficiently scattered if a lot of its entries are unimodular. Therefore, we define the "scattered level" of a complex matrix $\boldsymbol{S} \in \mathbb{C}^{k\times n}$ as the percentage of entries whose magnitude equals to either 1: the more unimodular entries in $\boldsymbol{S}$, the higher the "scattered level" and thus more likely to be identifiable. For various scattered levels p, we fix $n = 1000$ and $k = 20$, then use Algorithm 1 to try to exactly recover the mixing matrix, and equivalently the latent sources. It is easy to see that their optimal are dual to each other once we can exactly recover any of them. If after resolving the permutation and sign ambiguities, the estimation error is less than 10^{-5}, then we declare success. The results are shown in Fig. 3. The transition threshold seems to be around 40–50%.

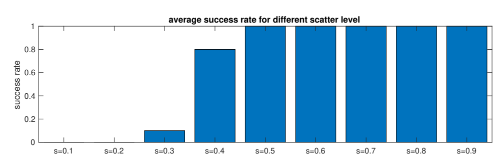

Fig. 3: Probability of exact recovery of the latent factors as we vary the "scattered level" of the latent sources.

5. CONCLUSION

We studied complex BCA problem with the novel scenario that the magnitudes of the complex sources are bounded, which has never in the literature possibly due to its challenging nature. We showed that it is still possible to uniquely identify the latent sources (up to permutation and complex scaling) if their disked hull is sufficiently scattered in the complex hypercube. We also present a new learning algorithm to based on linearized ADMM, and show numerically that the performance is surprisingly effective.

6. REFERENCES

[1] S. Cruces, “Bounded component analysis of linear mixtures: A criterion of minimum convex perimeter,” *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2141–2154, 2010.

[2] A. T. Erdogan, “A class of bounded component analysis algorithms for the separation of both independent and dependent sources,” *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5730–5743, 2013.

[3] P. Comon, “Independent component analysis, a new concept?” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[4] G. Tatli and A. T. Erdogan, “Polytopic matrix factorization: Determinant maximization based criterion and identifiability,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 5431–5447, 2021.

[5] J. Hu and K. Huang, “Identifiable bounded component analysis via minimum volume enclosing parallelotope,” in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023.

[6] A. T. Erdogan, “A blind separation approach for magnitude bounded sources,” in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5. IEEE, 2005, pp. v–177.

[7] ——, “A simple geometric blind source separation method for bounded magnitude sources,” *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 438–449, 2006.

[8] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.

[9] K. Huang, N. D. Sidiropoulos, and A. Swami, “Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition,” *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 211–224, 2013.

[10] X. Fu, K. Huang, and N. D. Sidiropoulos, “On identifiability of nonnegative matrix factorization,” *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 328–332, 2018.

[11] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, “Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain,” *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2306–2320, 2015.

[12] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, “Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 10, pp. 5530–5546, 2015.

[13] K. Huang and X. Fu, “Detecting overlapping and correlated communities without pure nodes: Identifiability and algorithm,” in *International Conference on Machine Learning*. PMLR, 2019, pp. 2859–2868.

[14] J. Hu and K. Huang, “Global identifiability of ℓ_1-based dictionary learning via matrix volume optimization,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.

[15] L. Narici and E. Beckenstein, *Topological Vector Spaces*. CRC Press, 2010.

[16] K. Huang, X. Fu, and N. D. Sidiropoulos, “Anchor-free correlated topic modeling: Identifiability and algorithm,” *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[18] S. Lu, J. D. Lee, M. Razaviyayn, and M. Hong, “Linearized admm converges to second-order stationary points for nonconvex problems,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 4859–4874, 2021.