Identifiability for Blind Source Separation of Multiple Finite Alphabet Linear Mixtures

Merle Behr, Axel Munk

Abstract—We give under weak assumptions a complete combinatorial characterization of identifiability for linear mixtures of finite alphabet sources, with unknown mixing weights and unknown source signals, but known alphabet. This is based on a detailed treatment of the case of a single linear mixture. Notably, our identifiability analysis applies also to the case of unknown number of sources.

We provide sufficient and necessary conditions for identifiability and give a simple sufficient criterion together with an explicit construction to determine the weights and the source signals for deterministic data by taking advantage of the hierarchical structure within the possible mixture values. We show that the probability of identifiability is related to the distribution of a hitting time and converges exponentially fast to one when the underlying sources come from a discrete Markov process. Finally, we explore our theoretical results in a simulation study.

Our work extends and clarifies the scope of scenarios for which blind source separation becomes meaningful.

Index Terms—Blind source separation; BSS; finite alphabet signals; single mixture; instantaneous mixtures; Markov processes, stopping time

I. INTRODUCTION

A. Problem description

In this work we are concerned with identifiability in a particular kind of blind source separation (BSS) motivated by different applications in digital communication (see e.g., [1]–[3]), but also in cancer genetics (see e.g., [4], [5], [9], [10]). A prominent example is the separation of a mixture of audio or speech signals, which has been picked up by several microphones, simultaneously (see e.g., [6]). In this case the different speech signals correspond to the sources and the recordings of the microphones to the mixture of signals with unknown mixture weights. From this mixture the individual signals have to be separated.

More generally, in BSS problems one observes $P$ mixtures of $N$ sources and aims to recover the original sources from the available observations. In this paper we focus on the linear case (for the non-linear case see e.g., [7], [8]), where the blindness refers to the fact that neither the sources nor the mixing weights are known. We also treat the case of unknown number of sources.

A minimal requirement underlying any recovery algorithm for sources and mixture weights (in a deterministic or noisy setting) to be valid is identifiability, i.e., the unique decomposition of the mixture into sources and mixing weights. Without any additional information on the source signals $S$ identifiability cannot hold, of course, as from a linear system $X = AS$ the matrices $A$ and $S$ are not uniquely determined, in general. In particular this applies to single linear mixtures ($P = 1$). However, if we assume that the values of the sources are attained in a known finite set of real numbers (finite alphabet), then BSS identifiability holds under certain conditions, which is necessary for any recovery algorithm to be valid.

Therefore, the aim of this work is to give a comprehensive discussion of such conditions. We are able to give (under weak assumptions) a complete combinatorial characterization of identifiability. From this we derive sufficient conditions, which are easy to verify. Moreover, these conditions yield an explicit construction to recover sources and weights in the noiseless case from the deterministic mixture. More specifically, using the notation

$$v = (v_{i})_{1 \leq i \leq n} = (v_{1}, \ldots, v_{n})^{\top}, \quad V = (V_{ij})_{1 \leq i \leq n}^{1 \leq j \leq m}$$

for a vector $v \in \mathbb{R}^{n}$ and a matrix $V \in \mathbb{R}^{n \times m}$ with $n$ rows and $m$ columns and $V_{i}, V_{j}$ for the corresponding row and column vectors, respectively, we assume from now on that the observed signal is linked to the sources via

$$X = AS$$

where $X$ is the deterministic mixture, $A = (A_{pm})_{1 \leq p \leq P, 1 \leq n \leq N}$ are the unknown mixing weights, $S = (S_{1}, \ldots, S_{N})^{\top} \in \Omega^{N \times T}$ are the unknown source signals, and $\Omega := \{\omega_{1}, \ldots, \omega_{k}\} \subset \mathbb{R}$ is a known alphabet, i.e., the set of possible values the sources can attain. The row vectors $S_{n}$, with $1 \leq n \leq N$ denote the single source signals (of length $T$) and the row vectors $A_{p}$, with $1 \leq p \leq P$ denote the mixing vectors (of length $N$) of the different mixtures.

In an observational model we often have $Y = X + \epsilon$ where $\epsilon$ is some random noise term with zero mean. All identifiability results for (1) transfer to this situation as well, of course. However, we stress that the primarily aim of this paper is not to provide a method which reconstructs the mixing weights $A$ and the sources $S$ from the (possibly) noisy observations $Y$, but rather to clarify the scope of scenarios under which this is possible. To this end, we will analyze necessary and sufficient conditions under which the decomposition in (1) is unique.

For the moment, we assume that the number of sources $N$ is known. However, we point out that most of our results remain
true even when $N$ is unknown (see Section VII), a case which has never been treated before to best of our knowledge.

B. Related Work

As far as we are aware of, identifiability of model (1) has not yet been considered in the literature in this general form, although various special cases and variations of this BSS problem have been addressed. The particular case of a binary alphabet, i.e., when $k = 2$ and $\{\omega_1, \omega_2\} = \{-1, 1\}$, has been considered in [11]–[15], Diamantaras and Papadimitriou [16] and Rostami et al. [17] assume the alphabet to be equally spaced, i.e., $\{\omega_1, \ldots, \omega_k\} = \{\omega_0, \omega_0 + T, \omega_0 + 2T, \ldots, \omega_0 + kT\}$. We consider arbitrary finite alphabets $\{\omega_1, \ldots, \omega_k\} \subset \mathbb{R}$. Moreover, several authors (e.g., [12], [16], [17]) assume a specific distribution on the alphabet, e.g., uniform. Our results show that it is already sufficient to observe some specific combinations of alphabet values (which are minimal conditions in a sense), hence we do not need to assume such a specific distribution. Diamantaras [18] works with a general finite alphabet as well but considers a mixture of two sources, $N = 2$. Li et al. [19] gave necessary and sufficient identifiability criteria for sparse signals, i.e., signals having many zero entries, in contrast to our work. Although they consider underdetermined mixtures, their results require at least two sensors ($P > 1$) and do not hold for single linear mixture ($P = 1$). Bofill and Zibulevsky [20] suggest for $P = 2$ a method for estimating the mixing weights for sparse signals as well, although without giving any explicit identifiability criteria. Diamantaras [14] considers a general finite alphabet but assumes the mixing weights to be complex. Thus, he works with a 2-dimensional signal. Combing the results from [14] and [21] yields a sufficient identifiability criterion for finite alphabet sources with complex mixing weights. However, this result does not hold when the mixing weights are real as considered here.

There are further variations of the BSS problem. Some of them are associated with Independent Component Analysis (ICA) (see e.g., [22]), which is based on the stochastic independence of the different sources (assumed to be random). ICA can be a powerful tool for (over)determined models ($P \geq N$) [22] and there are approaches for underdetermined multiple linear mixture models ($1 < N < P$) as well [23]. However, ICA is not applicable for single linear mixtures ($P = 1$), as the error terms of the single sources sum up to a single error term such that stochastic independence of the sources becomes irrelevant.

Also conceptually related is blind deconvolution (see e.g., [24], [25]), however, the convolution model makes analysis and identifiability severely different [26].

Another related problem is non-negative matrix factorization (see e.g., [27]–[29]), where one assumes (1), but instead of $S \in \mathbb{Q}^{N \times T}$, both $S$ and $A$ are non-negative. Indeed, the identifiability conditions derived in [28], [29] are quite related in nature to ours in Section V where their simpliciality condition on $A$ corresponds to our condition (6) and their separability condition on $S$ corresponds to our assumption $\mathbb{A}$. However, whereas their assumptions necessarily imply $P > 1$,

ours yield identifiability for single linear mixtures ($P = 1$), explicitly exploring the finite alphabet.

For the non-blind scenario, i.e., when $A$ in model (1) is known, [30] considers identifiability in a probabilistic framework.

To the best of our knowledge, a comprehensive characterization and unifying treatment of identifiability of the mixing weights and the sources for model (1) has been elusive. This issue is, however, fundamental for identifying the scope of possible scenarios where recovery algorithms for $S$ and $A$ in (1) are applicable (see, e.g., [12], [14], [15], [17]). In this sense our work provides an almost “minimal” set of conditions under which any recovery algorithm for the BSS problem only can be expected to be valid; in the noiseless case as well as for the case with random error.

C. Organisation of the paper

We will start our analysis by making two simplifications, which lead to a better interpretation of the corresponding identifiability conditions.

First, we will assume that $P = 1$, i.e., that we observe a single linear mixture in (1). Clearly, when $P$ increases the identification problem becomes easier, as more mixtures of the same sources are observed. Thus, the case $P = 1$ corresponds to the most difficult scenario and therefore we treat this case in detail. Generalizations to arbitrary $P$ will then follow easily from this case and are given in Section VI.

Second, we start with considering probability mixing weights, i.e., $A_{pn} > 0$ and $\sum_{n=1}^{N} A_{pn} = 1$ for all $p = 1, \ldots, P$. This is because the corresponding identifiability conditions have an easier interpretation when the mixing weights are positive. When we allow for negative mixing weights, the identifiability issue becomes slightly more difficult and the corresponding conditions become more complicated. Generalizations to negative mixing weights are given in Section VII.

The paper is organized as follows. After introducing a rigorous formulation of the problem and model (Section II), we will give a necessary and sufficient identifiability criterion for the mixing weights and the sources (Section III). To this end, we will first characterize the identifiability issue as a purely combinatorial problem. In Sections IV we will then generalize a result from Diamantaras and Chassioti [13] in order to derive a simple sufficient identifiability criterion. This characterizes those source signal combinations which make the variation of the mixture rich enough in order to become identifiable. This condition also provides an explicit construction (see the algorithm in Figure 2) for recovery of the weights $A$ and the sources $S$ from the mixture $X$ in (1).

In Section V we will shortly discuss how likely it is for the identification criterion of Section IV to be satisfied when the underlying sources $S_1, \ldots, S_N$ are discrete Markov processes. To this end, we will bound the probability of identification from below by a phase-type distribution. Using a stopping time argument, we will show that the mixture becomes identifiable exponentially fast, which reveals identifiability as less a critical issue in many practical situations as one might expect.
Although, we assume the number of source components $N$ to be known, in Section VII we show that most of our results remain true even when $N$ is unknown.

In Section IX we simulate source signals from a two-state Markov chain to illustrate how far our derived sufficient identifiability conditions (Section IV) are from being necessary in this setting. To this end, our results from Section III are fundamental as they give an explicit way to decide whether a given mixture is identifiable or not. Our simulation results reveal the simple sufficient condition in Section IV as quite sharp as we find that the number of observations needed for this to hold is quite close to the actual number of observations required for identifiability. This establishes the exponential bound in Section V as a useful tool to estimate the required number of observations guaranteeing identifiability with high probability.

We conclude in Section X.

II. PROBLEM STATEMENT

As mentioned above, for ease of presentation, we start with analyzing identifiability in (1) for single linear mixtures, i.e., $P = 1$ (for arbitrary $P$ see Section VI) and probability mixing weights (for arbitrary mixing weights see Section VII). To this end, let $A := \{\alpha \in \mathbb{R}^N : 0 < a_1 < \ldots < a_N < 1 \}$ and $\sum_{n=1}^{N} a_n = 1$ denote the set of positive (probability) mixing vectors. Note that $a_n \neq a_m$ for $n \neq m$ is always necessary to ensure identifiability of different sources $S_n, S_m$ (if $a_n = a_m$ interchanging of $S_n, S_m$ results in the same mixture $x$). For a given finite alphabet $\Omega = \{\omega_1, \ldots, \omega_k\} \subset \mathbb{R}$, with $\omega_1 < \ldots < \omega_k$ and given $N, T \in \mathbb{N}$ let $S = (S_1, \ldots, S_N)^\top \in \Omega^{N \times T}$ and $a = (a_1, \ldots, a_N) \in A$. Then the observed values are given by $x = aS$, i.e.,

$$x_t = \sum_{n=1}^{N} a_n S_{nt}, \quad t = 1, \ldots, T. \quad (2)$$

Definition II.1.

Let $x = aS$; $(a, S) \in A \times \Omega^{N \times T}$ as in (2). Then we denote the vector $a$ and the matrix $S$ as (jointly) identifiable from the observational vector $x$ when there exists exactly one $(\hat{a}, \hat{S}) \in A \times \Omega^{N \times T}$ such that $x = \hat{a}\hat{S}$.

In other words, identifiability means that $x = \hat{x}$ in (3) implies that $a = \hat{a}$ and $S = \hat{S}$. For simplicity, we refer to $a$ and $S$ being identifiable from $x$ just by saying that $(a, S)$ is identifiable. The aim of this paper is to study under which conditions $(a, S)$ is identifiable from $x$.

Even though we assume $N$ to be known, most of our results remain true when $N$ is unknown, i.e., $x = aS$ with $(a, S) \in \bigcup_{N \geq 2}(A \times \Omega^{N \times T})$ (see Section VIII).

Example II.2.

To illustrate the problem and notation, let us start with a simple example of model (3), where $N = 2$ and the alphabet is binary with $\Omega = \{0, 1\}$. This means that we consider mixing vectors of the form $a = (a_1, a_2)$ with $a_1, a_2 > 0$, $a_1 + a_2 = 1$ and two different sources $S_1 = (S_{11}, \ldots, S_{1T})$, $S_2 = (S_{21}, \ldots, S_{2T})$ with $S_{nt} \in \{0, 1\}$ for $n = 1, 2$ and $t = 1, \ldots, T$. The question we would like to answer is, under which conditions on $a$ and $S$ is $(a, S)$ uniquely determined via $x = aS$.

For a given observation $x_t$ the underlying source vector $S_t = (S_{1t}, S_{2t})^\top$ equals one of the four different values

$$\begin{align*}
(0, 0)^\top, (1, 0)^\top, (0, 1)^\top, (1, 1)^\top
\end{align*} \quad (3)$$

and hence,

$$x_t \in \{0, a_1, a_2, 1\}. \quad (4)$$

Clearly, if any two of the four values in the set on the r.h.s. of (3) coincide, then two different source values in (3) lead to the same mixture value for $x_t$ and hence the sources are not identifiable, i.e., they cannot be distinguished. Consequently, a necessary condition for identifiability is that all values in the r.h.s. of (4) are different, which is equivalent to

$$a_1 \neq a_2. \quad (5)$$

In other words, it is necessary that the alphabet values in $\Omega^N$ are well separated via the mixing weights $a \in A$. A generalization of this argument to arbitrary alphabets and number of sources is done later in (6). Further, we may assume w.l.o.g. $a_1 < a_2$, i.e., we denote that source as $S_1$, which comes with the smaller weight.

(5) alone, however, is necessary but not sufficient for identifiability. For instance, if $S_{1t} = S_{2t}$ for all $t = 1, \ldots, T$ then $x_t \in \{0, 1\}$ and hence, $a$ is not identifiable from $x$. Thus, a certain variability of the two sources $S_1$ and $S_2$ is necessary to guarantee identifiability of $a$. In this simple example, it is easy to check that a necessary and sufficient variability of $S_1$ and $S_2$ is that $S$ either takes the value $(1, 0)$ (i.e., $x_t = a_1$ for some $t$) or $(0, 1)$ (i.e., $x_t = a_2$ for some $t$) as by (3) and $a_1 + a_2 = 1$ it follows that $0 < a_1 < 1/2 < a_2 < 1$.

In other words, it is necessary that the mixing weight $a_1$ (or $a_2$ respectively) is seen somewhere in the mixture $x$ on its own, without the influence of the other mixing weight. A generalization of these assumptions and this argument to general systems (7) is done later in Theorem IV.7 and Theorem VII.7 respectively.

III. A COMBINATORIAL CHARACTERISATION OF IDENTIFIABILITY: GENERAL THEORY

In model (3) every observation $x_t$, for $t = 1, \ldots, T$, is given by a linear combination of $a \in A$ (unknown) with one of the finitely many vectors in $\Omega^N$. So in order to identify $S_t$, we have to determine the corresponding vector in $\Omega^N$. Note that multiple observed values leave this identification problem invariant, i.e., do not contribute further to identifiability. Hence, w.l.o.g. we assume all observations $x_1, \ldots, x_T$ to be pairwise different. Note, that this implies $T \leq k^N = |\Omega^N|$.

Of course, when for a given mixing vector $a \in A$ there exist $w' \neq w'' \in \Omega^N$ with $aw' = aw''$ and one observes this value, it is not possible to identify the underlying sources $S_1, S_2, \ldots, S_N$, uniquely. Consequently, a necessary condition for identifiability is that those values are not observed, i.e., $x_t \notin \{aw : \exists w' \neq w \in \Omega^N \text{ s.t. } aw = aw'\}$ for all $t = 1, \ldots, T$. 

For arbitrary sources this is comprised in the condition of a positive alphabet separation boundary, i.e.,
\[ ASB(a) \coloneqq \min_{w \neq w' \in \Omega^N} |aw - aw'| > 0. \tag{6} \]

Let \( S_N^T \) be the collection of injective maps from \( \{1, \ldots, N\} \) to \( \{1, \ldots, T\} \), i.e., for \( \rho \in S_N^T \) the vector \((x_{\sigma(1)}, \ldots, x_{\sigma(N)})\) corresponds to a selection of elements from \( \{x_1, \ldots, x_T\} \).

**Theorem III.1.**

Assume model (2) with \( x = (x_1, \ldots, x_T) = aS; (a, S) \in \mathcal{A} \times \Omega^N \). Let \( E \in \Omega^{N \times N} \) be an arbitrary invertible \( N \times N \) matrix with elements in \( \Omega \). Assume that \( ASB(a) > 0 \) and \( A I \).

there exists \( \rho \in S_N^{T} \) such that \((S_{1\rho(r)}, \ldots, S_{N\rho(r)})\) for \( 1 \leq \rho \leq N = E \). Then \( (a, S) \) is identifiable if and only if

\[ \text{A 2.} \]

there exists exactly one \( \sigma \in S_N^T \) such that for \( \tilde{a} := (x_{\sigma(1)}, \ldots, x_{\sigma(N)})E^{-1} \)
\[ \tilde{a} \in \mathcal{A} \quad \text{and} \quad \{x_1, \ldots, x_T\} \in \{\tilde{a}w : w \in \Omega^N\}, \]
i.e., \( \tilde{a} \) is a valid mixing weight and can reproduce all observations.

For a proof see Appendix A-A.

**Theorem III.1** is fundamental for the following as \( A2 \) provides a necessary and sufficient condition for identifiability of \( (a, S) \), given \( A I \) holds. In Section IV it will serve to derive a simple sufficient identifiability condition which is easy to check. In the following, we will discuss condition \( A I \) and \( A2 \) more detailed.

First, we consider assumption \( A I \). Without further restrictions on the matrix \( E \), \( A I \) simplifies to \( \text{rank}(S) = N \) (in particular implying \( T \geq N \)), which is, indeed, an almost minimal condition. By simple linear algebra, it is easy to check that \( \text{rank}(S) < \text{dim}(\text{span}(\mathcal{A})) = N - 1 \) implies that for any \( a \in \mathcal{A} \) exists an \( \tilde{a} \neq a \in \mathcal{A} \) such that \( aS = \tilde{a}S \), i.e., \((a, S)\) is not identifiable. When we allow for arbitrary mixing weights (see Section VII), i.e., not necessarily summing up to one, then by the same argument \( \text{rank}(S) = N \) becomes even a necessary condition. Intuitively, \( A I \) ensures that the sources \( S_1, \ldots, S_N \) differ sufficiently such that one can identify \( a \) from their mixture. For instance, if \( S_1 = \ldots = S_N \), it follows that \( x = S_1 \), irrespective of \( a \). Note that if \( k \) different values \( x_1, \ldots, x_k \) are observed, i.e., \( T = kN \) (recall that w.l.o.g. in this section \( T \) equals the number of pairwise different \( x_i \)'s), then it must hold true that \( \{x_1, \ldots, x_k\} = \{aw : w \in \Omega^N\} \).

Thus, \( A I \) follows trivially for any invertible \( N \times N \) matrix \( E \) with elements in \( \Omega \). However, \( A I \) is a much weaker assumption than \( T = kN \).

Second, we comment on assumption \( A2 \). Given \( A I \) assumption \( A2 \) reveals \( a \) as identifiable as soon as we can assign a collection of observations to rows in \( E \) in a unique way. When \( a \) is determined, identifiability of \( S \) follows from \( ASB(a) > 0 \). For \( A2 \) to check, we need to assume a fixed, given \( E \) which depends on \( \rho \) in \( A I \). However, \( \rho \) does not need to be known, in order to be able to verify condition \( A2 \), i.e., one does not need to know which observations correspond to rows in \( E \). In practice, the choice of \( E \) will depend on the specific application and different choices of \( E \) lead to different identifiability conditions \( A2 \). In Section IV we will show that for some specific choices of \( E \), \( A2 \) always holds, i.e., \( A I \) already implies identifiability. The following example shows that this is not true in general, i.e., not for any choice of \( E \).

**Example III.2.**

With the notation of (2) and Theorem III.1 let \( T = N = 3 \), \( \Omega = \{0, 1, \frac{21}{6+\sqrt{15}}, \frac{6-\sqrt{15}}{30}, \frac{6+\sqrt{15}}{30}, 0.6\} \), and
\[ S = E = \begin{pmatrix} 0 & 0 & 0 \\ \frac{6}{6+\sqrt{15}} & 0 & 0 \\ 0 & \frac{21}{6+\sqrt{15}} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \]
i.e., \( \rho \in A I \) is the identity map and \( x = aS = \begin{pmatrix} 6 - \frac{21}{6+\sqrt{15}} \\ \frac{6-\sqrt{15}}{30} \\ 0.7 \end{pmatrix} \).

For \( \sigma : (1, 2, 3) \mapsto (3, 1, 2) \) we find that \( (x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)})E^{-1} = (0.1, 0.2, 0.7) =: \tilde{a} \), which is a valid mixing weight. Hence, \((a, S)\) is not identifiable.

As mentioned before, in Section IV we will show that some specific choices of \( E \) already lead to uniqueness of the selection \( \sigma \) in \( A2 \) and thus ensure identifiability. The following remark illustrates how specific choices of rows in \( E \) fix some subdomain of \( \sigma \) in \( A2 \).

**Remark III.3.**

If \( T = kN \) with \( x_1 < \ldots < x_{kN} \), then \( x_1 \), the smallest observed value, corresponds to the situation when all sources \( S_{1t}, \ldots, S_{Nt} \) take the smallest value of the alphabet (denoted with \( \omega_1 \)), i.e., if \( E_1 = (\omega_1, \ldots, \omega_1) \), then for all \( \sigma \) satisfying \( A2 \) \( \sigma(1) = 1 \). The second smallest observed value, \( x_2 \), corresponds to the situation when all sources \( S_{2t}, \ldots, S_{Nt} \) take the smallest value \( \omega_1 \), but the source \( S_{1t} \) with the smallest weight \( a_1 \) takes the second smallest value \( \omega_2 \), i.e., if \( E_2 = (\omega_2, \omega_1, \ldots, \omega_1) \), then for all \( \sigma \) satisfying \( A2 \) \( \sigma(2) = 2 \).

Analogous holds for the largest observed value and the second largest observed value.

**IV. A SIMPLE SUFFICIENT IDENTIFIABILITY CRITERION**

We have seen in Theorem III.1 that the problem of identifying \( a \) and \( S \) from the observations \( x_1, \ldots, x_T \) reduces to identify the corresponding vectors \((S_{1t}, \ldots, S_{Nt})^T \in \Omega^N \) for sufficiently many observations \( x_t \). Namely, one has to find the corresponding observations \( x_t \) for \( N \) linear independent rows of an invertible \( N \times N \) matrix \( E \) with elements in \( \Omega \). Remark III.3 points out that some observations \( x_t \) can always be uniquely assigned to source vectors \((S_{1t}, \ldots, S_{Nt})^T \in \Omega^N \) and thus, limit the possible maps \( \sigma \) in \( A2 \). The next theorem shows that there is even more structure in the observations \( x_1, \ldots, x_T \) and that certain variations in the sources, i.e., certain choices of \( E \) in \( A I \) already ensure identifiability. Moreover, the proof of the following theorem gives an explicit construction of the unique \((a, S)\) from \( x \).
Theorem IV.1.
Assume model (2) with $x = (x_1, \ldots, x_T) = aS; (a, S) \in A \times \Omega^{N \times T}$. Furthermore, assume that $ASB(a) > 0$ and $A3$, there exists $\rho \in S^T_N$ such that
\[
(S_{1\rho(r)}, \ldots, S_{N\rho(r)})^\top = s^r, \quad r = 1, \ldots, N,
\]
with $(s^r_n)_{1 \leq n \leq N} := (\omega_1 1_{n \neq r} + \omega_2 1_{n = r})_{1 \leq n \leq N}$. Then $(a, S)$ is identifiable.

Before we give a proof (which is based on an explicit algorithm to compute $a$ from $x_1, \ldots, x_T$) we will discuss relationships and differences between the previous results and their assumptions $A1 - A3$.

Assumption $A3$ of Theorem IV.1 has a simple interpretation. It means that each of the mixing weights $a_n$ appears somewhere in the mixture on its own (without the influence of any other mixing weight $a_m \neq a_n$) via the mixture value $x_{\rho(n)} = a_n \omega_n + \omega_1$. For instance, if the alphabet is of the form $\Omega = \{0, 1, \omega_3, \ldots, \omega_k\}$, $A3$ simplifies to the condition that the mixing weights appear somewhere in the mixture, i.e., $a_n \in \{x_1, x_2\}$ for all $n = 1, \ldots, N$.

Intuitively, $A3$ means that for each $n = 1, \ldots, N$ there exists one mixture observation $x_n$ that only $S_{ni}$ is active (taking the value $\omega_2$) and all other sources $S_{mt}$ with $m \neq n$ are silent (taking the value $\omega_1$). It is easy to check that the choice of alphabet values $\omega_1$ and $\omega_2$ (the smallest and second smallest alphabet value) in Theorem IV.1 can be replaced by $\omega_k$ and $\omega_{k-1}$ (the largest and second largest alphabet value).

Obviously, assumption $A3$ arises from assumption $A1$ for a specific choice of $E$ in Theorem III.1, namely with
\[
E = \begin{pmatrix}
\omega^1 & \cdots & \omega^1 \\
\vdots & \ddots & \vdots \\
\omega^m & \cdots & \omega^m
\end{pmatrix} \in \mathbb{R}^{m \times m}.
\]
Consequently, if the matrix in (7) is invertible, $A3$ implies $A1$.

The following Lemma shows that this holds under mild conditions on $\Omega$ and $N$.

Lemma IV.2.
For model (2) let $A3$ be as in Theorem IV.1 and $A1$ as in Theorem III.1. If $\omega_k \neq \omega_1(1 - N)$, then $A3$ implies $A1$.

For a proof see Appendix A-B.

Figure 1 summarizes all relations between $A1 - A3$ in a diagram.

Now we turn to the proof of Theorem IV.1 which is proven by explicit recovery of $(a, S)$. This generalizes an algorithm of Diamantaras and Chassiot [13] for the binary alphabet $\Omega = \{-1, 1\}$ to a general finite alphabet.

Proof of Theorem IV.1
Assumption $A3$ implies that
\[
x_{\rho(r)} = a \omega^r, \quad r = 1, \ldots, N,
\]
and hence,
\[
a_r = \frac{x_{\rho(r)} - \omega_1}{\omega_2 - \omega_1}.
\]

Thus, it suffices to determine the map $\rho$ and the values $x_{\rho(1)}, \ldots, x_{\rho(N)}$, respectively, in order to determine $a$. When $a$ is determined, identifiability of $S$ follows from $ASB(a) > 0$.

Recall Remark III.3 and note that
\[
x_{\rho(1)} = \min\{\{x_1, \ldots, x_T\} \setminus \{\omega_1\}\},
\]
which determines $a_1$ as in (9). The following two lemmas show that successively all the other $\rho(r)$ (and hence $a_r$) for $r = 2, \ldots, N$ can be determined as well, which finishes the proof.

Let $B = B(N, \Omega)$ be the $N \times k^N$-matrix, where the $j$th column of $B$ is the number $j - 1$ written in the positional notation based on the number $k$, identifying 0 with $\omega_1$, 1 with $\omega_2$, and so on, i.e.,
\[
B_{ij} = \sum_{r=1}^k \omega_r 1_{(j - 1) \mod k^i + 1} = r - 1,
\]
and $d := aB$ the $k^N$ dimensional vector of all possible values that $x$ can take.

Lemma IV.3.
From $a_1, \ldots, a_r$ one can determine $d_1, \ldots, d_{k^r}$ uniquely.

Lemma IV.4.
It holds that $x_{\rho(r)} = \min\{\{x_1, \ldots, x_T\} \setminus \{d_1, \ldots, d_{k^{r-1}}\}\}$.

For proofs see Appendix A-C and A-D respectively.

The proof of Theorem IV.1 gives an explicit recovery construction for $(a, S)$ which is summarized in Figure 2. For noisy data one may use algorithm 2 to proceed similar as Diamantaras and Chassiot [13] who suggest a clustering approach for estimating $a$ and $S$ from noisy observations of $x$. However, as the purpose of this paper is not to propose a practical method for recovery from noisy data, but rather to analyze the scope of scenarios under which this is possible in principle, we are not going to follow this approach here further.

V. IDENTIFIABILITY FOR MIXTURES OF STOCHASTIC PROCESSES

In this subsection we will shortly discuss how likely it is for the identifiability condition of Theorem IV.1 to be satisfied when $(S_{1t}, \ldots, S_{Nt})^\top$ is a stochastic process. Therefore, let $s^r$ be as in Theorem IV.1 and define the hitting times
\[
\mathcal{T}^r := \min\{t \in \mathbb{N} : (S_{1t}, \ldots, S_{Nt}) = s^r\},
\]
and
\[
\mathcal{E} := \{a, \omega \in \mathbb{R} : (1 - N) \omega + \omega_1 = \omega_2, \omega \neq \omega_1\}.
\]

As in Theorem IV.1 further.

Figure 1. Relation between $A1$, $A2$, and $A3$.
Then let identify the first \( T \) including iid sequences (see e.g., [10], [16], [17]).

Figure 2. Algorithm for weight identification in (2), with Example V.2. Assume that the source signals \( \{x_1, \ldots, x_T\} \) are generated under assumption A3 of Theorem IV.1. Let us consider (2) for the simple case where we have two sources \( S_1 \) and \( S_2 \) with transition matrix

\[
\begin{pmatrix}
pq & pq & p^2 & q^2 \\
pq & pq & p^2 & q^2 \\
pq & pq & p^2 & q^2 \\
pq & pq & p^2 & q^2 \\
\end{pmatrix} > 0.
\]

Hence, \( M = 1 \).

If we assume that \( S_{1t} \) and \( S_{2t} \) are independent and identically distributed \( (i.i.d.) \) for all \( t = 1, \ldots, T \) with \( P(S_{nt} = \omega_1) = p \in (0,1) \) and \( P(S_{nt} = \omega_2) = 1 - p =: q \) for \( n = 1,2 \) and \( t = 1, \ldots, T \), then (V.1) constitutes an irreducible Markov process on the state space \( \{(\omega_2, \omega_1)^T, (\omega_1, \omega_2)^T, (\omega_1, \omega_1)^T, (\omega_2, \omega_2)^T\} \) with transition matrix

\[
P = \begin{pmatrix}
pq & pq & p^2 & q^2 \\
pq & pq & p^2 & q^2 \\
pq & pq & p^2 & q^2 \\
pq & pq & p^2 & q^2 \\
\end{pmatrix}.
\]

and \( c = qp + p^2 + q^2 = 1 - pq \). Thus, Theorem VI yields

\[
1 - P((a, S) \text{ is identifiable}) \leq 2(1 - pq)^T.
\]

In this simple setting we can even calculate the probability of identifiability exactly. Note that \( (a, S) \) is identifiable if and only if \( c_1 = a_1 \omega_2 + a_2 \omega_1 \) or \( c_2 = a_1 \omega_1 + a_2 \omega_2 \) is observed as \( a_1 < a_2 < a_2 < a_1 < \omega_1 + (\omega_2 - \omega_1)/2 < c_2 < c_2 < \omega_2 \) and

\[
\begin{align*}
(a_1, a_2) &= \left( \frac{c_1 - \omega_1}{\omega_2 - \omega_1}, \frac{\omega_2 - c_1}{\omega_2 - \omega_1} \right) \\
&= \left( \frac{\omega_2 - c_2}{\omega_2 - \omega_1}, \frac{c_2 - \omega_1}{\omega_2 - \omega_1} \right).
\end{align*}
\]

Therefore,

\[
1 - P((a, S) \text{ is identifiable}) = P(S_{11} = S_{21})^T = (1 - 2pq)^T.
\]

Example VII shows that the bound in Theorem V.1 does not need to be sharp in general but captures the exponential decay (in \( T \)) well. This is mainly because in Theorem VI the probability of \((a, S)\) being identifiable is bounded using the sufficient (and not necessary) identifiability condition A3 from Theorem VI. In Section VIII the gap between this bound and the true probability \( P((a, S) \text{ is identifiable}) \) in (14) is further explored in a simulation study.

**VI. Multiple Linear Mixtures**

After analyzing the most difficult scenario of a single linear mixture with \( P = 1 \) in (1), generalizations to arbitrary number of mixtures \( P \) now follow easily. To this end, for a vector \( v \) let \( \|v\|_1 := \sum_{i=1}^n |v_i| \) denote the \( l_1 \)-norm and define the set of \( P \)-mixtures as

\[
A_P := \{ A \in \mathbb{R}^{P \times N} : \|A_1\|_1 < \ldots < \|A_p\|_1, \|A_1\|_1 = \ldots = \|A_P\|_1 = 1 \}.
\]

Again, note that \( \|A_n\|_1 \neq \|A_m\|_1 \) for \( n \neq m \) is necessary to ensure identifiability of different sources. Then for \( S \in \Omega \) and \( A \in A_P \) the observed values are given by \( X = AS \), i.e.,

\[
X_{pt} = \sum_{n=1}^N A_{pn} S_{nt}, \quad t = 1, \ldots, T, \ p = 1, \ldots, P.
\]
Identifiability means to decompose the matrix \( X \in \mathbb{R}^{P \times T} \) uniquely into matrices \( A \in A_P \) and \( S \in \Omega^{N \times T} \) for given finite alphabet \( \Omega = \{ \omega_1, \ldots, \omega_k \} \) with \( \omega_1 < \ldots < \omega_k \) and given \( T, N, P \in \mathbb{N} \). Analog to before we define the alphabet separation boundary of a mixture matrix \( A \in A_P \) as
\[
\text{ASB}(A) := \min_{w \neq w' \in \Omega^N} \| Aw - Aw' \|_1.
\]
Clearly, \( \text{ASB}(A) > 0 \) is a necessary condition on \( A \) for \( (S, A) \) to be identifiable. Theorem III.1 Theorem IV.1 and Theorem VI.1 assume that \( P = 1 \). It is straightforward to check that Theorem III.1 holds unchanged when \( P > 1 \) with \( A \) replaced by \( A_P \). The same is true for Theorem IV.1 where in the proof the minimum of a set of observations \( x_i \) must be replaced by the minimum defined in terms of the ordering of \( \sum_{p=1}^P x_{ip} \).

Thus, clearly Theorem V.1 also holds unchanged when \( P > 1 \).

VII. ARBITRARY MIXING WEIGHTS

So far, we assumed the mixing weights to be positive and to sum up to one. However, in some applications this assumption is not satisfied (e.g., in digital communications) and in the following we discuss such generalizations. Let \( A \subset \mathbb{R} \) be an arbitrary mixing weights \( a_n \in \mathbb{R} \) an w.l.o.g. \( a_1 < \ldots < a_N \) in order to assign the mixing weight to a source.

It is easy to check that Theorem III.1 holds unchanged with \( A \) replaced by \( \hat{A} \). Note, however, that if \( \hat{A} \supseteq A \) condition A2 becomes more restrictive, i.e., a mixture \( (a, S) \) which is identifiable with respect to \( A \) might not be identifiable with respect to \( \hat{A} \).

Analogously, Theorem IV.1 can be generalized for \( \hat{A} \supseteq A \), where now the corresponding identifiability assumption A3 becomes more restrictive.

The following theorem considers the most general case of arbitrary mixing weights in \( A_0 \).

**Theorem VII.1.** Assume model (2) \( x = aS \) with \( (a, S) \in A_0 \times \Omega^{N \times T} \). Furthermore, assume that \( \text{ASB}(a) > 0 \) and there exists \( \rho, \mu \in S^T_{N+1} \) such that
\[
(S_{j_1(i)}, \ldots, S_{j_N(i)} \mathbb{T}) = \begin{cases} 
\rho^{-1}, & r = 1, \ldots, N + 1, \\
\rho^{-1}, & r = 1, \ldots, N + 1,
\end{cases}
\]
with \( \rho^{-1}, \mu^{-1} \in \Omega^N \) for \( r = 0, \ldots, N \) defined as
\[
(\rho^{-1})_i := \omega_2 \mathbb{1}_{i=r} + \omega_k \mathbb{1}_{a_i < a_{i+1}}, \quad \mu^{-1} := \omega_1 \mathbb{1}_{a_i < a_{i+1}} > 0,
\]
for \( i = 1, \ldots, N \) and \( a_0 := 1 \). Then \( (a, S) \) is identifiable.

The proof of Theorem VII.1 is given in Appendix A.1. Recall that for positive mixing weights the identifiability condition A3 in Theorem IV.1 had a very simple interpretation, namely that each of the single mixing weights \( a_n \) appears somewhere in the mixture \( x \) on its own, without the influence of any of the other mixing weights \( a_m \) for \( m \neq n \). The interpretation of (13) is somewhat more difficult, but similar. In the case of probability mixing weights \( a \in A \) as in

Theorem VII.1 both, the sum and the absolute sum of the mixing weights were fixed via \( \sum_{n=1}^N a_n = \sum_{n=1}^N |a_n| = 1 \) and this determined the scaling in which the mixing weights appear in the mixture \( x \). Now for general mixing weights \( a \in A_0 \) as in Theorem VII.1 both, the sum and the absolute sum (or equivalently the sum of the negative mixing weights and the sum of the positive mixing weights) are unknown and thus, additional conditions to determined these unknown scaling parameters are needed. These correspond to \( s^0 \) and \( v^0 \).

They ensure that the smallest possible mixture value (which corresponds to \( s^0 \)) and the largest possible mixture value (which corresponds to \( v^0 \)) are observed and thus determine the scaling parameters. Now analog to \( s^0 \) in A3 of Theorem IV.1 \( s^0 \) and \( v^0 \) in (15) of Theorem VII.1 ensure that \( a_n \) appears somewhere in the mixture \( x \) on its own and can thus be determined. However, as the sign of \( a_n \) is now unknown, too, we get the additional unambiguity that a mixture value can be increased either by increasing a source which corresponds to a positive mixing weight or by decreasing a source which corresponds to a negative mixing weight.

From Theorem VII.1 it follows directly that Theorem VII.1 holds with \( N \) replaced by \( 2N + 2 \), when we allow for arbitrary mixing weights in \( A_0 \).

VIII. UNKNOWN NUMBER OF SOURCE COMPONENTS

So far, we assumed that the number of sources \( N \) is fixed and known. Now we consider the case where \( N \) is unknown, i.e. \( x = aS \) with
\[
(a, S) \in \bigcup_{N \geq 2} (A \times \Omega^{N \times T}).
\]
While it is not clear how to generalize Theorem III.1 for \( (a, S) \) as in (16), condition A3 in Theorem IV.1 is still sufficient for identifiability when \( N \) is unknown.

To see this, note that the proof of Theorem IV.1 (in particular Lemma V.3) does not require the number of sources \( N \) to be known, where \( N \) is determined via
\[
N = \min(r \in \mathbb{N} \geq 2 \text{ s.t. } \{x_1, \ldots, x_T\} \subset \{d_1, \ldots, d_k\}),
\]
and thus, we obtain the following theorem.

**Theorem VIII.1.** Assume model (2) with \( x = (x_1, \ldots, x_T) = aS; (a, S) \in \bigcup_{N \geq 2} (A \times \Omega^{N \times T}) \). Furthermore, assume that \( \text{ASB}(a) > 0 \) and A3 holds. Then \( (a, S) \) (and thus \( N \)) is identifiable.

Analogously, Theorem VI.1 and Theorem VII.1 do not require \( N \) to be known.

IX. SIMULATIONS

Finally, we explore in a simulation study how far assumption A3 from Theorem IV.1 is from being necessary when the sources come from an irreducible Markov process; which corresponds to exploring the tightness of the bound in Theorem VII.1. To this end, Theorem III.1 is fundamental as it enables us to explicitly examine identifiability of \( (a, S) \). We consider an example with a binary alphabet \( \Omega = \{-2, 1\} \) and a mixture of \( N = 3 \) sources. The matrix \( E \) in Theorem III.1
was chosen randomly over the set of invertible $N \times N$ matrices with elements in $\Omega$. Simulation runs were always 1,000.

First, we assume the mixing weights $a = (0.2, 0.35, 0.45)$. Note that $a\omega \neq a\omega'$ for all $\omega \neq \omega' \in \{-2, 1\}^3$, i.e., $\text{ASB}(a) > 0$.

A. Bernoulli Model

Assuming the sources to be i.i.d. for all $n = 1, 2, 3$ and $t = 1, \ldots, T$, with $P(S_{nt} = -2) = P(S_{nt} = 1) = 0.5$ we find that $(a, S)$ is already identifiable on average for $t \geq 4.39$ observations. For the sufficient identifiability condition from Theorem IV.1 to hold we find an average value of $t \geq 14.79$ observations. Figure 3 shows the corresponding histograms and cumulative distribution functions. The results indicate that the number of observations needed for the sufficient identifiability condition from Theorem IV.1 to be satisfied is not considerably higher than the actual number of observations needed for the sufficient identifiability condition. This is confirmed in Figure 5, which shows the corresponding histograms and cumulative distribution functions. The histograms and cumulative distribution functions for $P = 1$ and $P = 4$ differ only slightly and for $P = 1$ they look almost the same as in Figure 3, although in Figure 3 $a$ is fixed, whereas in Figure 5 it is random. For the sufficient identifiability condition from Theorem IV.1 to hold we find an average value of $t \geq 15.13$ observations. Note that this condition depends on the sources only and, thus, is the same for all $P$.

B. Markov Model

We consider a more general Markov model for generating the sources, i.e., we assume the sources to be independent Markov processes on the state space $\Omega = \{-2, 1\}$ with transition matrix

$$P = \begin{pmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{pmatrix}.$$  \hspace{1cm} (17)

In Figure 4 we display the average numbers of observations until $(a, S)$ is identifiable and until the sufficient identifiability condition from Theorem IV.1 is fulfilled, respectively, for each $(p_1, p_2) \in \{0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9\}^2$. Note that $p_1 = 1-p_2$ corresponds to i.i.d. observations, with $1-P(S_{nt} = 1) = P(S_{nt} = -2) = p_1$.

From Figure 4 we draw that identifiability is achieved faster when $p_1$ and $p_2$ are close to 0.5, which corresponds to the i.i.d. Bernoulli model from Section IX-A. This is explained by condition $\text{A}2^\prime$ in Theorem III.1, where a richer variation in the sources, i.e., many different observations $x_1, \ldots, x_T$, reduces the set of possible valid mixing weights and thus favors identifiability. The sufficient identifiability condition $\text{A}3^\prime$ from Theorem IV.1 however, requires repeated occurrence of the smallest alphabet values $\omega_1$. Consequently, small $p_1$ and large $p_2$ discriminate against those variations.

C. Multiple Linear Mixtures

Now, we consider multiple linear mixtures, i.e., $P > 1$. Therefore, for each run, we draw $P$ mixing weights, each of length $N = 3$, independently from the uniform distribution on $A$ (implying $\text{ASB}(a) > 0$). For the sources, we consider a Bernoulli model as in Section IX-A.

We find that $(a, S)$ is identifiable on average after $t \geq 4.17, 4.07, 4.01, 3.99$ for $P = 1, 2, 3, 4$ observations, revealing that identifiability (condition $\text{A}2^\prime$ in Theorem III.1) depends much more on the variability of the sources than on the specific mixing weights. This is confirmed in Figure 5, which shows the corresponding histograms and cumulative distribution functions. The histograms and cumulative distribution functions for $P = 1$ and $P = 4$ differ only slightly and for $P = 1$ they look almost the same as in Figure 3, although in Figure 3 $a$ is fixed, whereas in Figure 5 it is random. For the sufficient identifiability condition from Theorem IV.1 to hold we find an average value of $t \geq 15.13$ observations. Note that this condition depends on the sources only and, thus, is the same for all $P$.

D. Arbitrary Mixing Weights

Finally, we consider arbitrary mixing weights in $A_0$. Therefore, for each run, we draw mixing weights ($P = 1, N = 3$) independently from the uniform distribution on $\{a \in [-10, 10]^N : a_1 < \ldots < a_N\}$ and for the sources we consider a Bernoulli model as in Section IX-A.
The derived sufficient identifiability conditions were briefly investigated in a simulation study and the required sample size for their validity was found to be quite close to the minimal sample size for identifiability.

This work is intended to give a solid theoretical background for a model that is used in a variety of applications in digital communications, but also in bioinformatics.

APPENDIX A

Proofs

A. Proof of Theorem VI.1

Proof:

For \( \sigma \in S_N^T \) we define \( x^\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(N)}) \).

By assumption \( A[1] \) \( x^\rho = aE \), i.e., \( x^\rho E^{-1} = a \) and, consequently,

\[
x^\rho E^{-1} \in A \text{ and } \{x_1, \ldots, x_T\} \in \{(x^\rho E^{-1})\omega : \omega \in \Omega^N\},
\]

which, by assumption \( A[2] \) is not fulfilled for any other \( \sigma \in S_N^T \). Thus, \( a \) is uniquely determined. Moreover, as \( ASB(a) > 0 \), \( S \) is uniquely determined as well.

Assume \( A[2] \) does not hold, i.e., there exists \( \sigma \neq \rho \in S_N^T \) such that \( \tilde{a} := x^\rho E^{-1} \) fulfills

\[
\tilde{a} \in A \text{ and } \{x_1, \ldots, x_T\} \in \{\tilde{a}\omega : \omega \in \Omega^N\}.
\]

As we assume all observations to be pairwise different, \( \tilde{a} \neq a \) and \( \tilde{a} \) with the corresponding \( S \) lead to the same observations \( x_1, \ldots, x_T \). Therefore, \( a, S \) is not identifiable.

B. Proof of Lemma IV.2

Proof: Obviously, \( A[3] \) arises from \( A[1] \) when we choose the matrix \( E \) in Theorem III.1 as in (7). Hence, \( A[3] \) implies \( A[1] \) if the matrix in (7) is invertible. (7) can be written as

\[
\begin{pmatrix}
\omega_1 & \cdots & \omega_1 \\
\vdots & \ddots & \vdots \\
\omega_1 & \cdots & \omega_1
\end{pmatrix} - \begin{pmatrix}
\omega_1 - \omega_2 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1
\end{pmatrix} \begin{pmatrix}
1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1
\end{pmatrix}
\]

and consequently, the matrix in (7) has zero determinant if and only if \( \omega_1 - \omega_2 \) is an eigenvalue of

\[
\begin{pmatrix}
\omega_1 & \cdots & \omega_1 \\
\vdots & \ddots & \vdots \\
\omega_1 & \cdots & \omega_1
\end{pmatrix},
\]

i.e., \( \omega_1 - \omega_2 = 0 \) or \( \omega_1 - \omega_2 = N \cdot \omega_1 \). As \( \omega_1 < \omega_2 \) the assertion follows.

C. Proof of Lemma IV.3

Proof: For \( r = N \) the assertion is obvious. So let \( r < N \). Then for \( i \in \{1 \ldots k^r\} \) we have that

\[
d_i = \sum_{j=1}^r a_j B_{ji} + \sum_{j=r+1}^N a_j \omega_1 = \sum_{j=1}^r a_j B_{ji} + \omega_1 (1 - \sum_{j=1}^r a_j),
\]

where \( B_{ji} \) are the entries of the matrix \( B \) in (11).
D. Proof of Lemma \[\text{IV.4}\]

Proof: By definition $x_{\rho(r)} = d_{kr-1}$. Therefore, $x_{\rho(r)} \geq \min\{x_1, \ldots, x_T\} \setminus \{d_1, \ldots, d_{kr-1}\}$ and for $i \in \{kr+1, \ldots, kN\}$

$$d_i - x_{\rho(r)} = \sum_{l=1}^{r-1} a_l(B_l - \omega_1) + a_{r}(B_{ri} - \omega_2) + \sum_{l=r+1}^{N} a_l(B_l - \omega_1).$$

If $B_{ri} \in \{\omega_2, \ldots, \omega_k\}$, then obviously $d_i - x_{\rho(r)} \geq 0$. If $B_{ri} = \omega_1$, then by definition of $B$ there exists an $s \in \{r+1, \ldots, N\}$ such that $B_{si} \in \{\omega_2, \ldots, \omega_k\}$, and therefore,

$$d_i - x_{\rho(r)} = \sum_{l=1}^{r-1} a_l(B_l - \omega_1) + a_{r}(\omega_1 - \omega_2)
+ a_s(B_s - \omega_1) + \sum_{l=r+1, l \neq s}^{N} a_l(B_l - \omega_1)
\geq \sum_{l=1}^{r-1} a_l(B_l - \omega_1) + (a_s - a_r)(\omega_2 - \omega_1)
+ \sum_{l=r+1, l \neq s}^{N} a_l(B_l - \omega_1) \geq 0.$$

Consequently, $x_{\rho(r)} \leq \min\{x_1, \ldots, x_T\} \setminus \{d_1, \ldots, d_{kr-1}\}$. \hfill $\blacksquare$

E. Proof of Theorem \[\text{VI.1}\]

Proof: Let $\mathcal{T}^T$ be as in \[\text{12}\] and let $p_0$ be the initial distribution of $(S_1, \ldots, S_N)^T$. Define the stopped process

$$\tilde{S}_t^r := \begin{cases} (S_1, \ldots, S_N)^T & \text{if } t < \mathcal{T}^T \\ \sigma^r & \text{otherwise,} \end{cases}$$

for $r = 1, \ldots, N$, which is a Markov process as well (see e.g., \[\text{32}\] Proposition 4.11.1.1.). It is obvious that for the Markov process $(\tilde{S}_t^r)$, the state $\sigma^r$ is absorbing and all other states are transient. Moreover, when we reorder the states in $\Omega^N$ such that $\sigma^r$ is the first state, the transition matrix of $(\tilde{S}_t^r)$ is given by

$$\tilde{P}_r = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ p_{2r} & Q_r \\ \vdots & & \ddots & \vdots \\ p_{Nr} & & & Q_r \end{pmatrix}. $$

The distribution of $\mathcal{T}^T$ is a discrete phase type distribution (see e.g., \[\text{33}\] Section 2.2.2.), i.e.,

$$P(\mathcal{T}^T > T) = p_0 Q_r^T 1 \leq \|Q_r^T 1\|_\infty. \tag{18}$$

As $P^M > 0$

$$\tilde{P}_r^M = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ s_{2r} & \vdots & & Q^M \\ \vdots & & \ddots & \vdots \\ s_{Nr} & & & Q^M \end{pmatrix}$$

with $s_2, \ldots, s_{N_r} > 0$ for $r = 1, \ldots, N$. Consequently, all row sums of $Q_r^M$ are smaller than 1, i.e.,

$$c_r := \|Q_r^M 1\|_\infty < 1 \tag{19}$$

and hence $c = \max_{1 \leq r \leq N} c_r < 1$.

Next, we show by induction that $\|Q_r^T 1\|_\infty \leq c_r^{\lfloor T/M \rfloor}$ for all $T \geq M$. For $T = M$ this holds by definition. So assume that $\|Q_r^T 1\|_\infty \leq c_r^{\lfloor t/M \rfloor}$ for all $M \leq l \leq T$ and define $A = (a_{ij}) := Q_r^T$, i.e.,

$$\max_i \sum_j a_{ij} \leq c_r^{\lfloor t/M \rfloor}. $$

If $|M_T| = |T+1/M_T|$, then

$$\|Q_r^{T+1} 1\|_\infty = \|AQ_r 1\|_\infty = \max_i \sum_j \sum_k a_{ik} q_{kj} = \max_i a_{ik} \sum_j q_{kj} \leq c_r^T c_r^{\lfloor T+1/M \rfloor},$$

as $\max_i \sum_k a_{ik} \leq c_r^T$ and $\sum_j q_{kj} \leq 1$.

If $|M_T| \neq |T+1/M_T|$, then $|M_T| + 1 = \lfloor T+1/M \rfloor$ and $T + 1 = M |M_T| + M$, with $M |M_T| := l \in \{M, \ldots, T\}$. Therefore,

$$\|Q_r^{T+1} 1\|_\infty = \|Q_r^T Q_r^1 1\|_\infty \leq c_r^T c_r^{\lfloor T/M \rfloor} + 1 = c_r^{T+1}. $$

With \[\text{18}\] and \[\text{15}\] it follows that

$$1 - P((a, S) \text{ is identifiable}) \leq N c^{\lfloor T/M \rfloor}$$

and as

$$\frac{NC^{\lfloor T/M \rfloor}}{c^{\lfloor T/M \rfloor}} \leq N < \infty$$

the assertion follows. \hfill $\blacksquare$
If $\sigma_- = 0$, all weights are positive and, as $\sigma_+$ is identified and thus w.l.o.g equal to one, Theorem [VI] applies. Thus, assume that $\sigma_- < 0$ and define $\mathcal{G}_0 := \{ \min \mathcal{G}, \max \mathcal{G} \}$ and

$$N_0 := \max\{n = 1, \ldots, N \text{ s.t. } a_n < 0\},$$

$$N_0^+ := N_0 + 1,$$

i.e., $a_1 < \ldots < a_{N_0} < 0 < a_{N_0^+} < \ldots < a_N$. Second, note that analog to [10]

$$\min \mathcal{G}_0 = \min(x_i \mathcal{S}_0^+, x_j \mathcal{S}_0^+), \quad \max \mathcal{G}_0 = \max(x_i \mathcal{S}_0^+, x_j \mathcal{S}_0^+),$$

and thus

$$\min \mathcal{G}_0 - \omega_k \sigma_+ - \omega_1 \sigma_- = \min \left( \frac{\omega_2 - \omega_1}{\omega_k - \omega_{k-1}} a_{N_0^+}^+, - a_{N_0} \right)$$

and

$$\omega_k \sigma_+ + \omega_1 \sigma_- - \max \mathcal{G}_0 = \min \left( -a_{N_0}^+, \frac{\omega_2 - \omega_1}{\omega_k - \omega_{k-1}} a_{N_0^+}^+ \right).$$

Hence, if $\min \mathcal{G}_0 - \omega_k \sigma_+ - \omega_1 \sigma_- < \omega_2 - \omega_1$, we find that

$$\sigma_{N_0^+} := \frac{\min \mathcal{G}_0 - \omega_k \sigma_+ - \omega_1 \sigma_-}{\omega_2 - \omega_1},$$

and if $\min \mathcal{G}_0 - \omega_k \sigma_+ - \omega_1 \sigma_- = \omega_2 - \omega_1$ then

$$\sigma_{N_0^+} := \frac{\min \mathcal{G}_0 - \omega_k \sigma_+ - \omega_1 \sigma_-}{\omega_2 - \omega_1}.$$

Thus, we have identified the $(l+1)$-th weight as

$$a_{l+1}^* := \begin{cases} a_{N_0^+}^+ & \text{if } \min \mathcal{G}_0 - \omega_k \sigma_+ - \omega_1 \sigma_- < \frac{\omega_2 - \omega_1}{\omega_k - \omega_{k-1}} a_{N_0^+}^+, \\ a_{N_0} & \text{otherwise}. \end{cases}$$

By induction, we can identify all weights and thus, by $ASB(a) > 0$ the assertion follows.

ACKNOWLEDGMENT

The authors acknowledge support of DFG CRC 803, 755, RTG 2088, and FOR 916. Helpful comments of an editor, two referees, C. Holmes, P. Rigollet, and H. Siebel are gratefully acknowledged.

REFERENCES

[1] S. Verdú, *Multilayer Detection*. Cambridge University Press, 1998.

[2] J. Proakis, *Digital Communications*, ser. Communications and signal processing. McGraw-Hill, 1995.

[3] Y. Zhang and S. A. Kassam, “Blind separation and equalization using fractional sampling of digital communications signals,” *Signal Processing*, vol. 81, no. 12, pp. 2591–2608, 2001.

[4] S. L. Carter, K. Cibulskis, E. Helman, A. McKenna, H. Shen, T. Zack, P. W. Laird, R. C. Onofrio, W. Winckler, B. A. Weir et al., “Absolute quantification of somatic dna alterations in human cancer,” *Nature biotechnology*, vol. 30, no. 5, pp. 413–421, 2012.

[5] L. Li, C. D. Morris, C. S. Johnson, D. L. Trump, M. Qin, J. C. Conroy, J. Wang, and S. Liu, “Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges,” *Oncoarget*, vol. 4, no. 11, p. 1868, 2013.

[6] A. Aissa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Undetermined blind audio source separation using modal decomposition,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, no. 1, pp. 14–14, 2007.

[7] B. Lachover and A. Yeredor, “Separation of polynomial post nonlinear mixtures of discrete sources,” *Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on on*, pp. 1126–1131.

[8] M. Castella, “Inversion of polynomial systems and separation of nonlinear mixtures of finite-alphabet sources,” *vol. 56, no. 8, pp. 3905–3917. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4567638*

[9] G. Ha, A. Roth, J. Khattra, J. Ho, D. Yap, L. M. Prentice, N. Melnyk, A. McPherson, A. Bashashati, E. Lakis et al., “Titan: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data,” *Genome research*, vol. 24, no. 11, pp. 1881–1893, 2014.

[10] C. Yao, O. Papaspiliopoulos, G. O. Roberts, and C. Holmes, “Bayesian non-parametric hidden markov models with applications in genomics,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 1, pp. 37–57, 2011.

[11] S. Talwar, M. Viberg, and A. Paulraj, “Blind separation of synchronous co-channel digital signals using an antenna array,” *I Algorithms*, *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, 1996.

[12] P. Pajunen, “Blind separation of binary sources with less sensors than sources,” *International Conference on Neural Networks*, vol. 3, pp. 1994–1997, 1997.

[13] K. Diamantaras and E. Chassioti, “Blind separation of a binary source from one observation: A deterministic approach,” *International Workshop on Independent Component Analysis and blind Signal Separation*, pp. 93–98, 2000.

[14] K. Diamantaras, “A clustering approach for the blind separation of multiple finite alphabet sequences from a single linear mixture,” *Signal Processing*, vol. 86, pp. 877–891, 2006.

[15] F. Gu, Z. Zhang, N. Li, and W. Lu, “Blind separation of multiple sequences from a single linear mixture using finite alphabet,” *International Conference on Wireless Communications and Signal Processing*, pp. 1–5, 2010.
[16] K. Diamantaras and T. Papadimitriou, “Blind miso deconvolution using the distribution of output differences,” IEEE International Workshop on Machine Learning for Signal Processing, pp. 1–6, 2009.

[17] M. Rostami, M. Babaie-Zadeh, S. Samadi, and C. Jutten, “Blind source separation of discrete finite alphabet sources using a single mixture,” IEEE Statistical Signal Processing Workshop (SSP), pp. 709–712, 2011.

[18] K. Diamantaras, “Blind separation of two multi-level sources from a single linear mixture,” IEEE Workshop on Machine Learning for Signal Processing, pp. 67–72, 2008.

[19] Y. Li, S.-I. Amari, A. Cichocki, D. W. Ho, and S. Xie, “Underdetermined blind source separation based on sparse representation,” IEEE Transactions on Signal Processing, vol. 54, no. 2, pp. 423–437, 2006.

[20] P. Bofill and M. Zibulevsky, “Underdetermined blind source separation using sparse representations,” Signal Processing, vol. 81, pp. 2353–2362, 2001.

[21] K. Diamantaras, “Blind channel identification based on the geometry of the received signal constellation,” Signal Processing, IEEE Transactions on, vol. 50, no. 5, pp. 1133–1143, 2002.

[22] P. Comon, “Independent component analysis. a new concept?” Signal Processing, vol. 36, pp. 287–314, 1994.

[23] T.-W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, “Blind source separation of more sources than mixtures using overcomplete representations,” IEEE Signal Processing Letters, vol. 6, no. 4, pp. 87–90, 1999.

[24] T.-H. Li, “Finite-alphabet information and multivariate blind deconvolution and identification of linear systems,” Information Theory, IEEE Transactions on, vol. 49, no. 1, pp. 330–337, 2003.

[25] K. Diamantaras and T. Papadimitriou, “Blind deconvolution of multi-input single-output systems using the distribution of point distances,” Journal of Signal Processing Systems, vol. 65, no. 3, pp. 525–534, 2011.

[26] D. Yellin and B. Porat, “Blind identification of fir systems excited by discrete-alphabet inputs,” Signal Processing, IEEE Transactions on, vol. 41, no. 3, pp. 1331–1339, 1993.

[27] D. Lee and S. Seung, “Learning the parts of objects by non-negative matrix factorization,” Nature, vol. 401, pp. 788–791, 1999.

[28] D. Donoho and V. Stodden, “When does non-negative matrix factorization give a correct decomposition into parts?” in Advances in neural information processing systems, 2003.

[29] S. Arora, R. Ge, R. Kannan, and A. Moitra, “Computing a nonnegative matrix factorization-provably,” in Proceedings of the forty-fourth annual ACM symposium on Theory of computing, 2012, pp. 145–162.

[30] A. Aissa-El-Bey, D. Pastor, S. M. A. Shai, and Y. Fadlallah, “Sparsity-based recovery of finite alphabet solutions to underdetermined linear systems,” Information Theory, IEEE Transactions on, vol. 61, no. 4, pp. 2008–2018, 2015.

[31] F. Gamboa and E. Gassiat, “Source separation when the input sources are discrete or have constant modulus,” Signal Processing, IEEE Transactions on, vol. 45, no. 12, pp. 3062–3072, 1997.

[32] V. Kolokol’cov, Markov Processes, Semigroups, and Generators, ser. De Gruyter studies in mathematics. De Gruyter, 2011.

[33] M. Neuts, Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach, ser. Algorithmic Approach. Dover Publications, 1981.