Separation Theorem for Independent Subspace Analysis and its Consequences[☆]

Zoltán Szabó[a,*], Barnabás Póczos[b], András Lőrincz[a]

[a]*Faculty of Informatics, Eötvös Loránd University,*
Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary
[b]*Carnegie Mellon University, Robotics Institute,*
5000 Forbes Ave, Pittsburgh, PA 15213, United States

Abstract

Independent component analysis (ICA)—the theory of mixed, independent, non-Gaussian sources—has a central role in signal processing, computer vision and pattern recognition. One of the most fundamental conjectures of this research field is that independent subspace analysis (ISA)—the extension of the ICA problem, where *groups* of sources are independent—can be solved by traditional ICA followed by grouping the ICA components. The conjecture, called *ISA separation principle*, (i) has been rigorously proven for some distribution types recently, (ii) forms the basis of the state-of-the-art ISA solvers, (iii) enables one to estimate the unknown number and the dimensions of the sources efficiently, and (iv) can be extended to generalizations of the ISA task, such as different linear-, controlled-, post nonlinear-, complex valued-, partially observed problems, as well as to problems dealing with nonparametric source dynamics. Here, we shall review the advances on this field.

Keywords: separation principles, independent subspace analysis, linear systems, controlled models, post nonlinear systems, complex valued models, partially observed systems, nonparametric source dynamics

[*]Corresponding author.
Email addresses: szzoli@cs.elte.hu (Zoltán Szabó), bapoczos@cs.cmu.edu (Barnabás Póczos), andras.lorincz@elte.hu (András Lőrincz)

1. Introduction

Independent component analysis (ICA) [1, 2, 3] has received considerable attention in signal processing, computer vision and pattern recognition, e.g., in face representation and recognition [4, 5], information theoretical image matching [6], fMRI analysis [7], feature extraction of natural images [8], texture segmentation [9], artifact separation in MEG recordings, and the exploration of hidden factors in financial data [10]. One may consider ICA as a cocktail party problem: we have some speakers (sources) and some microphones (sensors), which measure the mixed signals emitted by the sources. The task is to recover the original sources from the mixed observations. For a recent review about ICA, see [11, 12, 13].

Traditional ICA algorithms are *one-dimensional* in the sense that all sources are assumed to be independent *real* valued random variables. Nonetheless, applications in which only certain groups of the hidden sources are independent may be highly relevant in practice, because one cannot expect that all source components are statistically independent. In this case, the independent sources can be multidimensional. For instance, consider the generalization of the cocktail-party problem where *independent groups* of musicians are playing at the party. The separation task requires an extension of ICA, which is called multidimensional ICA [14], independent subspace analysis (ISA) [15], independent feature subspace analysis [16], subspace ICA [17], or group ICA [18] in the literature. We will use the ISA abbreviation throughout this paper. The several successful applications and the large number of different ISA algorithms—the authors are aware of more than 30 "different" ISA approaches (in terms of the applied cost function and the optimization technique)—show the importance of this field. Below, we list a few successful applications of ISA in signal processing, computer vision, and pattern recognition.

ECG analysis: An important task of ECG signal processing is to estimate fetal ECG signals from ECG recordings measured on the mother's skin (cutaneous recordings) [19, 14, 20, 18, 17]. The cardiac waveform of the fetal ECG can provide useful information for detecting certain diseases. Potential measurements on the mother's skin are the results of numerous bioelectric phenomena, such as maternal and fetal heart activity, respiration, stomach activity, and some other noise terms. The electric activity of the fetal and maternal hearts can be considered as independent multidimensional sources, and the ECG measurements on the mother's skin

are the mixture of these bioelectric and noise signals, where the transfer from the bioelectric sources to the electrodes on the body surface can be approximated by an unknown linear mixing. Our goal is to estimate this unknown linear mixing and the fetal heart activity.

fMRI and MEG data processing: In *fMRI* data processing, our goal is to detect and extract task-related signal components, artifacts, and noise terms from voxel activities. The main principles that allow us to perform these tasks are *localization* and *connectism*, which state that different brain parts are responsible for different cognitive tasks and these areas are spatially distributed. According to these principles, one may assume that the time series (a.k.a. time courses) of the voxels are linear mixtures of independent components [7], and it is of great importance to recover these independent signals (component maps) from the voxels. The component maps show the brain areas related to the independent components. Recently, [21] has shown that the assumption that all of these components are independent might be too strong in practice, and hence the application of ISA instead of ICA can give physiologically more meaningful results in certain cases.

Similarly, one might assume that there are hidden independent sources belonging to MEG measurements. [22] has also shown that the full independence assumption might be too restrictive in this case as well, and better results can be achieved if we allow dependent sources too by using ISA instead of ICA.

Natural image analysis, texture classification: It has been demonstrated several times that ICA on natural images leads to image filters that resemble to the simple cells in the V1 visual cortical area of the brain: They are localized, oriented, and selective only to certain frequencies (bandpass filters) [23]. If we use ISA instead of ICA on natural images, i.e, if we allow dependencies between some of the components, then ISA will provide independent subspaces that show phase- and shift-invariant properties [15, 24, 25]. ISA is naturally able to group similar components (with respect to frequency and orientation) into the same subspace. By exploiting this invariance property and selecting only one element from each subspace, [26] showed that one can get results similar to other state-of-the-art methods in the texture classification task using much smaller filter bank.

3

Action recognition: Another successful ISA application is *action recognition* in movies [27]. Here the key idea was the observation that ISA on spatiotemporal datasets provides subspaces that contain velocity selective invariant features. This ISA based approach outperformed several state-of-the-art methods using hand-crafted features (Harris3D, Cuboids, Hessian, HOG/HOF, HOG3D, extended SURF, etc).

Learning of face view-subspaces:
Multi-view face detection and recognition are very challenging problems. ICA, ISA and topographic ICA [28]—a generalization of ISA—can be used to learn view-specific feature representation [29]. In turn, ICA on multi-view face data sets leads to view-specific feature components, and ISA can organize these feature vectors into facial view-specific groups (subspaces). In addition to this, topographic ICA is able to arrange these subspaces in a topographically consistent way as well.

Single-channel source separation: An important problem in audio scene analysis is single-channel source separation. In this setting there are several independent sound sources and a microphone records the mixture of these sounds. The goal is to estimate the original signals from the microphone recording. [30] proposed an approach in which they applied ISA on the Fourier transformed windowed observations (spectogram). Using this approach on a Beethoven string quartet, they found that this method was able to separate the independent instruments into different subspaces.

Motion segmentation: Multibody motion segmentation is an important problem in computer vision. By observing the trajectories of certain points of a few objects, our goal is to decide which points belong to which objects. Assuming that we have a linear camera model, the multibody motion segmentation reduces to an ISA problem, where each subspace belongs to a single object [31].

Gene expression analysis: Gene clustering is a valuable tool for describing the characteristic patterns of cells and understanding properties of unknown genes. Linear latent variable models seem to outperform standard clustering approaches in this area. In this framework it is assumed that the gene expression profiles are the results of several biological processes, where each process affects few genes only. Since there are independent and highly inter-connected processes as well, [32] proposed to use ISA for processing gene expression data. Their approach lead to biologically

valuable and gene ontologically interpretable results.

One of the most exciting and fundamental hypotheses of the ICA research is due to Cardoso [14], who *conjectured* that the ISA task can be solved by ICA preprocessing and then clustering of the ICA elements into statistically independent groups. While the extent of this conjecture, the *ISA separation principle*, is still an open issue, it has recently been rigorously proven for some distribution types [33], and for this reason we call it ISA Separation Theorem. This principle (i) forms the basis of many state-of-the-art ISA algorithms, (ii) can be used to design algorithms that scale well and efficiently estimate the dimensions of the hidden sources and (iii) can be extended to different linear-, controlled-, post nonlinear-, complex valued-, partially observed systems, as well as to systems with nonparametric source dynamics. Here, we review such consequences of the theorem.

Beyond ISA, there exist numerous other exciting directions that relax the traditional assumptions of ICA (one-dimensional sources, i.i.d. sources in time, instantaneous mixture, complete observation). Below we list a few of these directions. We will see in the subsequent sections that the ISA separation technique can be extended to these models.

Linear systems: One may relax the ICA assumptions by assuming sources that have linear dynamics (e.g., autoregressive ones [34]), or echoes (moving average dynamics) may also be present leading to the blind source deconvolution (BSD) problem [35].

Post nonlinear models: The linear mixing restriction of ICA can be relaxed by assuming that there is an unknown component-wise nonlinear function superimposed on the linear mixture. This ICA generalization also has many successful applications, e.g., in sensor array processing, data processing in biological systems, and satellite communications. For an excellent review, see [36].

Complex valued sources and complex mixing: In the complex ICA problem, the sources and the mixing are both realized in the complex domain. Complex values naturally emerge in fMRI data processing, where in addition to the magnitude, the phase information can also be important. The complex-valued computations have been present from the "birth" of ICA [2, 3] and show nice potentials in the analysis of biomedical signals (EEG, fMRI), see e.g., [37, 38, 39].

Incomplete observations: In this setting certain parts (coordinates/time instants) of the mixture are not available for observation [40, 41].

Nonparametric dynamics: The general case of sources with unknown, nonparametric dynamics is quite challenging, and very few works focused on this direction [18, 42].

The paper is structured as follows: We define the ISA model in Section 2. We discuss the ISA Separation Theorem and its known sufficient conditions in Section 3. Section 4 is about the extensions of the ISA separation principle. Corollaries of the separation principles are summarized in Section 5. Numerical illustrations about these corollaries are presented in Section 6. Conclusions are drawn in Section 7. For the sake of convenience, we listed the abbreviations of the paper in the Appendix (Table A.1).

2. The Independent Subspace Analysis (ISA) Model

Here we review the basics of the ISA model and the related ISA cost function (Section 2.1). We elaborate on the ambiguities of the ISA task in Section 2.2 and present an ISA performance measure that can be applied to general sources of different dimensions.

2.1. The ISA Equations and Cost Function

We define the independent subspace analysis (ISA) model. Assume that we have an observation ($\mathbf{x} \in \mathbb{R}^{D_x}$), which is instantaneous linear mixture ($\mathbf{A}$) of the hidden source ($\mathbf{e}$), that is,

$$\mathbf{x}_t = \mathbf{A}\mathbf{e}_t, \tag{1}$$

where (i) the unknown mixing matrix $\mathbf{A} \in \mathbb{R}^{D_x \times D_e}$ has full column rank, (ii) source $\mathbf{e}_t = \left[\mathbf{e}_t^1; \ldots; \mathbf{e}_t^M\right] \in \mathbb{R}^{D_e}$ is a vector concatenated (using Matlab notation ";") of components $\mathbf{e}_t^m \in \mathbb{R}^{d_m}$ ($D_e = \sum_{m=1}^M d_m$), subject to the following conditions:

1. $\mathbf{e}_t$ is assumed to be i.i.d. (independent and identically distributed) in time t,

2. there is at most one Gaussian variable among $\mathbf{e}^m$s; this assumption will be referred to as the "non-Gaussian" assumption,

6

3. $\mathbf{e}^m$s are independent, that is $I(\mathbf{e}^1, \dots, \mathbf{e}^M) = 0$, where I stands for mutual information [43]. Mutual information of $\{\mathbf{e}^m\}_{m=1}^M$ is non-negative and is zero, if and only if the $\{\mathbf{e}^m\}_{m=1}^M$ random variables are (jointly) independent.

The goal of the ISA problem is to eliminate the effect of the mixing $(\mathbf{A})$ with a suitable $\mathbf{W} \in \mathbb{R}^{D_e \times D_x}$ *demixing matrix* and estimate the original source components $\mathbf{e}^m$s by using observations $\{\mathbf{x}_t\}_{t=1}^T$ only $(\hat{\mathbf{e}} = \mathbf{W}\mathbf{x})$. If all the $\mathbf{e}^m$ source components are one-dimensional $(d_m = 1, \forall m)$, then the ICA task is recovered. For $D_x > D_e$ the problem is called *undercomplete*, while the case of $D_x = D_e$ is regarded as *complete*.[1]

In ISA, it can be assumed without any loss of generality—applying zero mean normalization and principal component analysis [46]—that (i) $\mathbf{x}$ and $\mathbf{e}$ are *white*, i.e., their expectation value is zero, and their covariance matrix is the identity matrix $(\mathbf{I})$, (ii) mixing matrix $\mathbf{A}$ is orthogonal, that is $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, where superscript T stands for transposition, and (iii) the task is complete $(D = D_x = D_e)$. In what follows, this assumption will be referred to as "whiteness".

The estimation of the demixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ is equivalent to the minimization of the mutual information between the estimated components, or equivalently to the minimization of the sum of the entropies of the estimated source components [47]:

$$ J_I(\mathbf{W}) := I\left(\mathbf{y}^1, \dots, \mathbf{y}^M\right), \qquad J_H(\mathbf{W}) := \sum_{m=1}^{M} H\left(\mathbf{y}^m\right), \qquad (2) $$

where $\mathbf{y} = \mathbf{W}\mathbf{x}$, $\mathbf{y} = \left[\mathbf{y}^1; \dots; \mathbf{y}^M\right]$, $\mathbf{y}^m \in \mathbb{R}^{d_m}$ and H denotes the Shannon's multidimensional differential entropy [43]. One can easily prove that due to the whiteness assumption, the optimization of the cost functions can be restricted to the *orthogonal* group $(\mathbf{W} \in \mathcal{O}^D)$. In the rest of the paper we will consider the J_H ISA cost function. In the special case when every hidden source component $\mathbf{e}^m$ is of one-dimensional $(d_m = 1, \forall m)$, the ICA problem/cost function is recovered [48]. Other equivalent entropy and mutual information based forms of the ISA cost function are given in [33].

---

[1]We shall not treat the *overcomplete* case $(D_x < D_e)$, but some efforts have been devoted to this area [44] extending the topographic ICA model [28] to subspaces with the quasi-orthogonal prior construction of [45].

2.2. The ISA Ambiguities and an ISA Performance Measure

Below, we list the ISA ambiguities that one can use to define a *general* performance measure for the ISA task.

Identification of the ISA model is ambiguous. However, the ambiguities of the model are simple: hidden components can be determined up to permutation of the subspaces and up to invertible linear transformations[2] within the subspaces [49, 50]. Therefore, in the ideal case, the product of the estimated ISA demixing matrix $\hat{\mathbf{W}}_{\mathrm{ISA}}$ and the ISA mixing matrix $\mathbf{A}$, i.e., matrix

$$\mathbf{G} = \hat{\mathbf{W}}_{\mathrm{ISA}} \mathbf{A} \tag{3}$$

is a block-permutation matrix (also called block-scaling matrix [18]). This property can be measured for source components with different dimensions by a simple extension of the Amari-index [51]. Namely, assume that we have a weight matrix $\mathbf{V} \in \mathbb{R}^{M \times M}$ made of positive matrix elements. Loosely speaking, we shrink the $d_i \times d_j$ sized blocks of matrix $\mathbf{G}$ according to the weights of matrix $\mathbf{V}$ and apply the traditional Amari-index for the resulting matrix. Formally, (i) assume without loss of generality that the component dimensions and their estimations are ordered in increasing order ($d_1 \leq \ldots \leq d_M$, $\hat{d}_1 \leq \ldots \leq \hat{d}_M$), (ii) decompose $\mathbf{G}$ into $d_i \times d_j$ sized blocks ($\mathbf{G} = [\mathbf{G}^{ij}]_{i,j=1,\ldots,M}$) and define g^{ij} as the sum of the absolute values of the elements of the matrix $\mathbf{G}^{ij} \in \mathbb{R}^{d_i \times d_j}$, weighted with V_{ij}:

$$g^{ij} = V_{ij} \sum_{k=1}^{d_i} \sum_{l=1}^{d_j} |\left(G^{ij}\right)_{k,l}|. \tag{4}$$

Then the Amari-index with parameters $\mathbf{V}$ can be adapted to the ISA task of possibly different component dimensions as follows

$$r_{\mathbf{V}}(\mathbf{G}) := \frac{1}{2M(M-1)} \left[\sum_{i=1}^{M} \left(\frac{\sum_{j=1}^{M} g^{ij}}{\max_j g^{ij}} - 1 \right) + \sum_{j=1}^{M} \left(\frac{\sum_{i=1}^{M} g^{ij}}{\max_i g^{ij}} - 1 \right) \right]. \tag{5}$$

One can see that $0 \leq r_{\mathbf{V}}(\mathbf{G}) \leq 1$ for any matrix $\mathbf{G}$, and $r_{\mathbf{V}}(\mathbf{G}) = 0$ if and only if $\mathbf{G}$ is block-permutation matrix with $d_i \times d_j$ sized blocks. $r_{\mathbf{V}}(\mathbf{G}) = 1$ is in the

---

[2]The condition of invertible linear transformations simplifies to orthogonal transformations for the "white" case.

worst case, i.e, when all the g^{ij} elements are equal. Note that this measure (5) is invariant to multiplication with a positive constant: $r_{c\mathbf{V}} = r_{\mathbf{V}} (\forall c > 0)$. Weight matrix $\mathbf{V}$ can be uniform ($V_{ij} = 1$) [52], or one can use weighing according to the size of the subspaces: $V_{ij} = 1/(d_i d_j)$.

3. The ISA Separation Theorem

This section is about the ISA Separation Theorem that targets one of the most relevant open conjectures of ICA research that dates back to 1998 [14]. The conjecture is the cornerstone of many state-of-the-art ISA solvers and has a number of implications. We show in Section 4 how to extend the conjecture to more general models, such as non-i.i.d. linear-, controlled-, post nonlinear-, complex valued-, partially observed systems, as well as for problems with nonparametric source dynamics. Corollaries of these extensions are discussed in Section 5.

According to the ISA Separation Theorem, the solution of the ISA task, i.e., the *global* optimum of the ISA cost function can be found by properly *grouping the ICA elements*, that is, "ISA=ICA followed by permutation search". In other words, one may think of the ISA problem (see J_H, Eq. (2)) as an ICA task with $d_m = 1 (\forall m)$. In this case, cost function J_H is the sum of entropies of *one-dimensional* variables, which—given their one-dimensional nature—can be estimated efficiently [48]. Then, if the ISA Separation Theorem holds, it is sufficient to permute the ICA elements (i.e., cluster them into statistically independent groups) to find the global solution of the ISA problem. One can provide sufficient conditions for the ISA Separation Theorem by using cost function J_H of the ISA task:

Theorem 1 (ISA Separation Theorem [33]) *Let* $\mathbf{y} = [y_1; \ldots; y_D] = \mathbf{Wx} \in \mathbb{R}^D$, *where* $\mathbf{W} \in \mathcal{O}^D$, $\mathbf{x} \in \mathbb{R}^D$ *is the whitened observation of the ISA model, and* $D = \sum_{m=1}^M d_m$. *Let* $\mathbb{S}_{\mathbb{R}}^{d_m}$ *denote the surface of the d_m-dimensional unit sphere, that is* $\mathbb{S}_{\mathbb{R}}^{d_m} := \{\mathbf{w} \in \mathbb{R}^{d_m} : \sum_{i=1}^{d_m} w_i^2 = 1\}$. *Presume that the* $\mathbf{v} := \mathbf{e}^m \in \mathbb{R}^{d_m}$ *sources* ($m = 1, \ldots, M$) *of the ISA model satisfy condition*

$$H\left(\sum_{i=1}^{d_m} w_i v_i\right) \geq \sum_{i=1}^{d_m} w_i^2 H(v_i), \forall \mathbf{w} \in \mathbb{S}_{\mathbb{R}}^{d_m}, \tag{6}$$

and that the ICA cost function $J_{ICA}(\mathbf{W}) = \sum_{i=1}^D H(y_i)$ *has minimum over the orthogonal matrices in* $\mathbf{W}_{\text{ICA}}$. *Then it is sufficient to search for the*

solution to the ISA task as a permutation of the solution to the ICA task. Using the concept of demixing matrices, it is sufficient to explore forms

$$\mathbf{W}_{\text{ISA}} = \mathbf{P}\mathbf{W}_{\text{ICA}}, \tag{7}$$

where $\mathbf{P} \in \mathbb{R}^{D \times D}$ is a permutation matrix to be determined and $\mathbf{W}_{\text{ISA}}$ is the ISA demixing matrix.

The general question whether a certain source satisfies the ISA Separation Theorem is now partially answered, since Eq. (6) provides a *sufficient* condition. Equation (6) holds, e.g., for variables ($\mathbf{v} := \mathbf{e}^m$) satisfying the so-called w-EPI condition

$$e^{2H\left(\sum_{i=1}^{d_m} w_i v_i\right)} \geq \sum_{i=1}^{d_m} e^{2H(w_i v_i)}, \forall \mathbf{w} \in \mathbb{S}_{\mathbb{R}}^{d_m}, \tag{8}$$

where EPI is a three letter acronym for the *entropy power inequality* [43].

The w-EPI condition is fulfilled, e.g., by spherical variables [53], whose distributions are invariant to orthogonal transformations. One can show that in the 2-dimensional case ($d_m = 2$) invariance to 90° rotation, a condition weaker than invariance to spherical transformation, is sufficient. A special case of this requirement is invariance to permutation and sign changes, which also includes distributions having constant density over the spheres of the L^p-space, the so-called L^p-spherical variables [54]. The case $p = 2$ corresponds to spherical variables. For an illustration of distributions with 90° rotation and sign change invariance, see Fig. A.1.

Takano has shown [55] that the w-EPI condition is satisfied by certain weakly dependent variables subject to the dimensionality constraint that $d_m = 2$.

These sufficient conditions of the ISA Separation Theorem, i.e. conditions ensuring that the global optimum can be found by ICA followed by clustering of the ICA elements, are summarized schematically in Fig. A.2.

It is intriguing that if (6) is satisfied, then this simple separation principle provides the *global* minimum of the ISA cost function. About joint block diagonalization (JBD), Meraim and Belouchrani [56] has put forth a similar *conjecture* recently: the JBD of a finite matrix set can be obtained by the joint diagonalization of the set up to permutation. JBD based ISA solvers [18, 57, 50, 58] make efficient use of this conjecture in practice. We also note that this principle was justified for *local* minimum points in [50].

10

4. Extensions of the ISA Separation Principle

Below we review the extensions of the ISA separation principle. The principle is extended to different linear (Section 4.1), post nonlinear- (Section 4.2), complex valued- (Section 4.3), controlled-, partially observed models, as well as for nonparametric source dynamics (Section 4.4). These different methods, however, can be used in combinations, too. It is important to note that the separation principle is valid for *all* of these models, and thus (i) the dimension of the source components (d_m) can be different and unknown in all of these models (see also the clustering algorithms in Section 5), and (ii) the Amari-index detailed in Section 2.2 can be applied as performance measure to all of them. Traditionally, the ISA problem considers the *instantaneous linear* mixture of *independent and identically distributed* (i.i.d.) hidden sources (see Section 2.1). These constraints will be alleviated below. The corresponding general problem family will be referred to as independent process analysis (IPA). The relationships of the different generalizations and separation principles are illustrated in Fig. A.4(a) and (b).

4.1. Linear Systems

In this section we focus on linear models: Section 4.1.1 is about autoregressive models, and Section 4.1.2 treats moving average (convolutive) based models.

4.1.1. The AR-IPA Model

In the AR-IPA (autoregressive-IPA) task [59], the traditional *i.i.d.* assumption for the sources is generalized to AR time series: the hidden sources ($\mathbf{s}^m \in \mathbb{R}^{d_m}$) are not necessarily independent, only their driving noises ($\mathbf{e}^m \in \mathbb{R}^{d_m}$) are. The observation ($\mathbf{x} \in \mathbb{R}^D$, $D = \sum_{m=1}^{M} d_m$) is an instantaneous linear mixture ($\mathbf{A}$) of the source $\mathbf{s}$:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \qquad \mathbf{s}_t = \sum_{i=1}^{L_s} \mathbf{F}_i \mathbf{s}_{t-i} + \mathbf{e}_t, \qquad (9)$$

where L_s is the order of the AR process, $\mathbf{s}_t = [\mathbf{s}_t^1; \ldots; \mathbf{s}_t^M]$ and $\mathbf{e}_t = [\mathbf{e}_t^1; \ldots; \mathbf{e}_t^M] \in \mathbb{R}^D$ denote the hidden sources and the hidden driving noises, respectively. (9) can be rewritten in the following concise form: $\mathbf{x} = \mathbf{A}\mathbf{s}$, and $\mathbf{F}[z]\mathbf{s} = \mathbf{e}$ using the polynomial of the time-shift operator

$\mathbf{F}[z] := \mathbf{I} - \sum_{i=1}^{L_s} \mathbf{F}_i z^i \in \mathbb{R}[z]^{D \times D}$ [60]. We assume that (i) polynomial matrix $\mathbf{F}[z]$ is *stable*, that is $\det(\mathbf{F}[z]) \neq 0$, for all $z \in \mathbb{C}, |z| \leq 1$, (ii) mixing matrix $\mathbf{A} \in \mathbb{R}^{D \times D}$ is invertible, and (iii) $\mathbf{e}$ satisfies the ISA assumptions (see Section 2.1). The aim of the AR-IPA task is to estimate hidden sources $\mathbf{s}^m$, dynamics $\mathbf{F}[z]$, driving noises $\mathbf{e}^m$ and mixing matrix $\mathbf{A}$ or its $\mathbf{W}$ inverse given observations $\{\mathbf{x}_t\}_{t=1}^T$. For the special case of $L_s = 0$, the ISA task is obtained.

Making use of the basis transformation rule of AR processes, it can be shown that the observation process $\mathbf{x}$ is also AR

$$\mathbf{x}_t = \sum_{i=1}^{L_s} \left(\mathbf{A} \mathbf{F}_i \mathbf{A}^{-1} \right) \mathbf{x}_{t-i} + \mathbf{n}_t, \tag{10}$$

with innovation $\mathbf{n}_t = \mathbf{A}\mathbf{e}_t$, whose marginals are approximately Gaussian according to the d-dependent central limit theorem [61]. Using this form and that sources $\mathbf{e}^m$ are independent according to our assumptions, the AR-IPA estimation can be carried out by (i) applying AR fit to observation $\mathbf{x}$, (ii) followed by ISA on $\hat{\mathbf{n}}_t$, the estimated innovation of $\mathbf{x}$. AR identification can be performed, e.g., by the methods detailed in [62, 63]. The pseudocode of this AR-IPA solution can be found in Table A.2. The presented approach extends [34] to multidimensional ($d_m \geq 1$) sources. We note that in the one-dimensional case ($d_m = 1$), simple temporal differentiating might be sufficient for the reduction step [64].

4.1.2. The MA-IPA Model and its Extensions

In this section the assumption on *instantaneous* linear mixture of the ISA model is weakened to convolutions. This problem is called moving average independent process analysis (MA-IPA, also known as blind subspace deconvolution) [33]. We describe this task for the undercomplete case. Assume that the convolutive mixture of hidden sources $\mathbf{e}^m \in \mathbb{R}^{d_m}$ is available for observation ($\mathbf{x} \in \mathbb{R}^{D_x}$)

$$\mathbf{x}_t = \sum_{l=0}^{L_e} \mathbf{H}_l \mathbf{e}_{t-l}, \tag{11}$$

where (i) $D_x > D_e$ (undercomplete, $D_e = \sum_{m=1}^{M} d_m$), (ii) the polynomial matrix $\mathbf{H}[z] = \sum_{l=0}^{L_e} \mathbf{H}_l z^l \in \mathbb{R}[z]^{D_x \times D_e}$ has a (polynomial matrix) left inverse[3] and (iii) source $\mathbf{e} = [\mathbf{e}^1; \dots; \mathbf{e}^M] \in \mathbb{R}^{D_e}$ satisfies the conditions of ISA. The goal of this undercomplete MA-IPA problem (uMA-IPA problem, where "u" stands for undercomplete) is to estimate the original $\mathbf{e}^m$ sources by using observations $\{\mathbf{x}_t\}_{t=1}^{T}$ only. The case $L_e = 0$ corresponds to the ISA task, and in the blind source deconvolution problem [35] $d_m = 1$ $(\forall m)$, and L_e is a non-negative integer. We note that in the ISA task the full column rank of matrix $\mathbf{H}_0$ was presumed, which is equivalent to the assumption that matrix $\mathbf{H}_0$ has left inverse. This left inverse assumption is extended in the uMA-IPA model for the polynomial matrix $\mathbf{H}[z]$.

The separation principle below claims that by applying temporal concatenation (TCC) on the observation, one can reduce the uMA-IPA estimation problem to ISA.

Theorem 2 (uMA-IPA via TCC [33]) *Let L' be such that $D_x L' \geq D_e(L_e + L')$ is fulfilled. Then we end up with an $\mathbf{X}_t = \mathbf{A E}_t$ ISA task with an $\mathbf{A} \in \mathbb{R}^{D_x L' \times D_e(L_e + L')}$ ($\mathbf{H}[z]$ dependent) Toeplitz matrix upon applying temporal concatenation of depth $L_e + L'$ and L' on the sources and the observations, respectively.*

Choosing the minimal value for L', the dimension of the obtained ISA task is $D_{\min} = D_e(L_e + L') = D_e \left(L_e + \left\lceil \frac{D_e L_e}{D_x - D_e} \right\rceil \right)$. Unfortunately D_{min} can easily become too large. This dimensionality problem can be alleviated by the linear prediction approximation (LPA) approach, which is formulated in the following theorem.

Theorem 3 (uMA-IPA via LPA [66]) *In the uMA-IPA task, observation process $\mathbf{x}_t$ is autoregressive (finite but of unknown order), and its innovation $\tilde{\mathbf{x}}_t := \mathbf{x}_t - E[\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots]$ is $\mathbf{H}_0 \mathbf{e}_t$, where $E[\cdot | \cdot]$ denotes the conditional expectation value. Consequently, there is a polynomial matrix $\mathbf{W}_{\mathrm{AR}}^{\mathrm{LPA}}[z] \in \mathbb{R}[z]^{D_x \times D_x}$ such that $\mathbf{W}_{\mathrm{AR}}^{\mathrm{LPA}}[z]\mathbf{x} = \mathbf{H}_0 \mathbf{e}$, and thus the solution becomes an AR-IPA task.*

In the undercomplete case, one can extend the LPA approach to the solution of the more general ARIMA-IPA (integrated autoregressive moving

---

[3]One can show for $D_x > D_e$ that under mild conditions $\mathbf{H}[z]$ has a left inverse with probability 1 [65]; e.g., when the matrix $[\mathbf{H}_0, \dots, \mathbf{H}_{L_e}]$ is drawn from a continuous distribution.

average IPA) [67] as well as to the complete MA-IPA [68] problems in an asymptotically consistent way. The ARIMA-IPA problem allows both AR and MA terms in the evolution of $\mathbf{s}_t$ hidden sources, and it also defines a non-stationary process by means of rth order temporal difference.

In the one-dimensional ($d_m = 1$) case, it has been shown that the uMA-IPA problem can be solved by means of spatio-temporal decorrelation [69], and by the TCC technique [70]. Furthermore, the uMA-IPA and uARMA-IPA problems can be reduced to ICA by LPA [71, 72].

4.2. Post Nonlinear Models

Below the *linear* mixing assumption of the ISA model is alleviated by presenting the post nonlinear ISA (PNL-ISA) problem [73]. Assume that the observations ($\mathbf{x} \in \mathbb{R}^D$) are post nonlinear mixtures ($\mathbf{f}(\mathbf{A}\cdot)$) of multidimensional independent sources ($\mathbf{e} \in \mathbb{R}^D$):

$$\mathbf{x}_t = \mathbf{f}(\mathbf{A}\mathbf{e}_t), \tag{12}$$

where the (i) unknown function $\mathbf{f} : \mathbb{R}^D \to \mathbb{R}^D$ is a component-wise transformation, i.e, $\mathbf{f}(\mathbf{v}) = [f_1(v_1); \ldots ; f_D(v_D)]$ and $\mathbf{f}$ is invertible, and (ii) mixing matrix $\mathbf{A} \in \mathbb{R}^{D \times D}$ and hidden source $\mathbf{e}$ satisfy the ISA assumptions of Section 2.1. The PNL-ISA problem is to estimate the hidden source components $\mathbf{e}^m$ knowing only the observations $\{\mathbf{x}_t\}_{t=1}^T$. For $d_m = 1$, we get back the PNL-ICA problem [74] (for a review see [36]), whereas "$\mathbf{f}$=identity" leads to the ISA task.

Under certain technical conditions, one can carry out the estimation of the hidden source $\mathbf{e}$ on the basis of the *mirror structure* of the PNL-ISA system (12). Formally, we use the $\hat{\mathbf{e}} = [\hat{\mathbf{e}}^1; \ldots ; \hat{\mathbf{e}}^M] = \mathbf{W}\mathbf{g}(\mathbf{x}) = \mathbf{W}\mathbf{g}(\mathbf{f}(\mathbf{A}\mathbf{e}))$ equation, where we need to estimate $\mathbf{W}$ and $\mathbf{g}$. In the ideal case the component-wise acting $\mathbf{g}$ transformation and matrix $\mathbf{W}$ inverts function $\mathbf{f}$ and matrix $\mathbf{A}$, respectively. Independence of the estimated $\hat{\mathbf{e}}^m$s implies the recovery of sources $\mathbf{e}^m$ up to permutation and invertible affine transformations within each subspace. According to the d-dependent central limit theorem [61], marginals of $\mathbf{A}\mathbf{e}$ can be considered as approximately Gaussian variables. Therefore, the nonlinearity $\mathbf{g}$ can be optimized to make the distribution of observation $\mathbf{x}$ the most similar to a Gaussian distribution. This is called "gaussianization (transformation)" [75, 76]. In these works the one-dimensional ($d_m = 1$) PNL-ICA special case was treated; however, the ideas can be generalized to PNL-ISA as well in the following way [73]: After gaussianization,

the next step is to estimate $\mathbf{W}$ by means of linear ISA. This second step transforms the result of the gaussianization transformation $[\mathbf{g}(\mathbf{x})]$ into the opposite direction, i.e., to the most non-Gaussian direction.

4.3. Complex Valued Models

We summarize a few basic concepts for complex variables. An excellent review on this topic can be found in [77]. Define the $\varphi_v : \mathbb{C}^L \mapsto \mathbb{R}^{2L}$, $\varphi_M : \mathbb{C}^{L_1 \times L_2} \mapsto \mathbb{R}^{2L_1 \times 2L_2}$ mappings as

$$\varphi_v(\mathbf{v}) = \mathbf{v} \otimes \left[\begin{array}{c} \Re(\cdot) \\ \Im(\cdot) \end{array} \right], \qquad \varphi_M(\mathbf{M}) = \mathbf{M} \otimes \left[\begin{array}{cc} \Re(\cdot) & -\Im(\cdot) \\ \Im(\cdot) & \Re(\cdot) \end{array} \right], \qquad (13)$$

where $\otimes$ is the Kronecker product, $\Re$ stands for the real part, $\Im$ for the imaginary part. Subscript v and M denote vector and matrix, respectively. Independence of complex random variables $\mathbf{v}_m \in \mathbb{C}^{d_m}$ $(m = 1, \ldots, M)$ is defined as the independence of variables $\varphi_v(\mathbf{v}_m)$. The entropy of a complex independent variable $\mathbf{v} \in \mathbb{C}^d$ is $H(\mathbf{v}) = H(\varphi_v(\mathbf{v}))$.

By the definition of independence for complex random variables, the complex valued ISA ($\mathbb{C}$-ISA) task [78, 79] can be defined similarly to the real case (Section 2.1) as $\mathbf{x} = \mathbf{Ae}$.

We review two approaches to solve the complex ISA problem. First, suppose that the "non-Gaussian" assumption is made in the $\mathbb{C}$-ISA model for variables $\varphi_v(\mathbf{e}^m) \in \mathbb{R}^{2d_m}$. Now, applying φ_v to the complex ISA equation (Eq. (1)), one gets

$$\varphi_v(\mathbf{x}) = \varphi_M(\mathbf{A})\varphi_v(\mathbf{e}). \qquad (14)$$

Given that (i) the independence of $\mathbf{e}^m \in \mathbb{C}^{d_m}$ is equivalent to that of $\varphi_v(\mathbf{e}^m) \in \mathbb{R}^{2d}$, and (ii) the existence of the inverse of $\varphi_M(\mathbf{A})$ is inherited from $\mathbf{A}$, we end up with a real valued ISA task with observation $\varphi_v(\mathbf{x})$ and M pieces of $2d_m$-dimensional hidden components $\varphi_v(\mathbf{e}^m)$. The consideration can be extended to the complex variants of the linear models of Section 4.1 including the ARIMA-IPA model, too.

Another possible solution is to use the ISA Separation Theorem, which remains valid even for complex variables [33] when the following condition holds for the $\mathbf{v} = \mathbf{e}^m \in \mathbb{C}^{d_m}$ hidden sources:

$$H\left(\sum_{i=1}^{d_m} w_i v_i\right) \geq \sum_{i=1}^{d_m} |w_i|^2 H(v_i) \quad \forall \mathbf{w} \in S_{\mathbb{C}}^{d_m}. \qquad (15)$$

Sources $\mathbf{v} = \mathbf{e}^m \in \mathbb{C}^{d_m}$ that satisfy the complex w-EPI property—which is similar to (8), but "$2H$" needs to be replaced by "H" and "$\mathbb{S}_{\mathbb{R}}^{d_m}$" by "$\mathbb{S}_{\mathbb{C}}^{d_m}$"—also satisfy the sufficient condition (15). Complex spherical variables [80], whose distribution are invariant to unitary transformations, make one of the examples. The relation of these sufficient conditions is illustrated in Fig. A.3.

4.4. Controlled, Partially Observed Models, and Nonparametric Source Dynamics

In what follows we briefly review the generalization of the IPA problem to controlled (ARX-IPA, "X" stands for exogenous input) and partially observed problems (mAR-IPA, "m" denotes missing observations), as well as to problems with nonparametric source dynamics (fAR-IPA, "f" means functional). All of these three problems can be solved with the tricks used for solving the AR-IPA problem (Section 4.1.1). Formally, the ARX-IPA [81], mAR-IPA [82], and fAR-IPA [52] problems are defined as follows:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t \qquad \mathbf{s}_t = \sum_{i=1}^{L_s} \mathbf{F}_i \mathbf{s}_{t-i} + \sum_{j=1}^{L_u} \mathbf{B}_j \mathbf{u}_{t+1-j} + \mathbf{e}_t, \qquad (16)$$

$$\mathbf{y}_t = \mathcal{M}_t(\mathbf{x}_t) \qquad \mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \quad \mathbf{s}_t = \sum_{i=1}^{L_s} \mathbf{F}_i \mathbf{s}_{t-i} + \mathbf{e}_t, \qquad (17)$$

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \qquad \mathbf{s}_t = \mathbf{f}(\mathbf{s}_{t-1}, \dots, \mathbf{s}_{t-L_s}) + \mathbf{e}_t, \qquad (18)$$

where the notation will be explained below.

In the *ARX-IPA* problem (Eq. (16)) the AR-IPA assumption holds (Eq. (9)), but the time evolution of the hidden source $\mathbf{s}$ can be influenced via *control* variable $\mathbf{u}_t \in \mathbb{R}^{D_u}$ through matrices $\mathbf{B}_j \in \mathbb{R}^{D \times D_u}$. The goal is to estimate the hidden source $\mathbf{s}$, the driving noise $\mathbf{e}$, the parameters of the dynamics and control matrices ($\mathbf{F}_i$ and $\mathbf{B}_j$), as well as the mixing matrix $\mathbf{A}$ or its inverse $\mathbf{W}$ by using observations $\{\mathbf{x}_t\}_{t=1}^T$. In the special case of $L_u = 0$, the ARX-IPA task reduces to AR-IPA.

In the *mAR-IPA* problem (Eq. (17)), the AR-IPA assumptions (Eq. (9)) are relaxed by allowing a few coordinates of the mixed AR sources $\mathbf{x}_t \in \mathbb{R}^D$ to be *missing* at certain time instants. Formally, we observe $\mathbf{y}_t \in \mathbb{R}^D$ instead of $\mathbf{x}_t$, where "mask mappings" $\mathcal{M}_t : \mathbb{R}^D \mapsto \mathbb{R}^D$ represent the coordinates and the time indices of the non-missing observations. Our task is the estimation of the hidden source $\mathbf{s}$, its driving noise $\mathbf{e}$, parameters of the dynamics $\mathbf{F}[z]$,

mixing matrix $\mathbf{A}$ (or its inverse $\mathbf{W}$) from observation $\{\mathbf{y}_t\}_{t=1}^T$. The special case of "$\boldsymbol{\mathcal{M}}_t = $ identity" corresponds to the AR-IPA task.

In the *fAR-IPA* problem, the *parametric* assumption for the dynamics of the hidden sources is circumvented by fAR sources ((18)). The goal is the same as before; we are to estimate hidden sources $\mathbf{s}^m \in \mathbb{R}^{d_m}$ including their dynamics $\mathbf{f}$ and their driving innovations $\mathbf{e}^m \in \mathbb{R}^{d_m}$ as well as mixing matrix $\mathbf{A}$ (or its inverse $\mathbf{W}$) given observations $\{\mathbf{x}_t\}_{t=1}^T$. If we knew the parametric form of $\mathbf{f}$ and if it were linear, then the problem would be AR-IPA.

By exploiting the facts that the linear invertible transformations ($\mathbf{A}$) of fAR, ARX, and AR (see Eq. (10)) processes also belong to the family of fAR, ARX, and AR processes with $\mathbf{n}_t = \mathbf{A}\mathbf{e}_t$ innovation,[4] we can see that a reasonable approach is to fit ARX, mAR, or fAR processes to the observations and then use ISA on the estimated $\hat{\mathbf{n}}_t$ innovations.

The parameter estimation of ARX processes can be done by either recent active learning methods [84], or by means of more traditional approaches [85, 86]. The mAR fit can be accomplished, e.g., by the maximum likelihood principle [85], the subspace technique [86], or in a Bayesian framework [87]. For the identification of fAR processes, one can use nonparametric regression [88, 89].

5. Consequences of the Separation Principles

The ISA Separation Theorem (Section 3) and its extensions (Section 4) have a number of important consequences that we discuss here.

According to the ISA Separation Theorem, the ISA task can be solved by finding the optimal permutation of the ICA elements by rendering the elements into statistically dependent subspaces. State-of-the-art solvers use this approach since it scales well and enables the estimation for unknown component dimensions ($\{d_m\}_{m=1}^M$, M) too. These properties are detailed below.

First, assume that the dimensions (d_m) of the hidden sources are given. Then according to the ISA Separation Theorem, cost function (2) can be minimized by considering all demixing matrices $\mathbf{W} = \mathbf{P}\mathbf{W}_{\text{ICA}}$, where $\mathbf{P}$ denotes a permutation. Below we list a few possibilities for finding $\mathbf{P}$.

---

[4] We note that these tricks can be used for other processes as well which have this property, such as Markov-switching AR-IPA processes. For Markov-switching AR processes, see [83].

Exhaustive way: the possible number of all permutations, i.e., the number of $\mathbf{P}$ matrices is $D!$, where "!" denotes the factorial function. Considering that the ISA cost function is invariant to the exchange of elements *within* the subspaces (see, e.g., (2)), the number of relevant permutations decreases to $\frac{D!}{\prod_{m=1}^{M} d_m!}$. This number can still be enormous, and the related computations could be formidable justifying searches for efficient approximations that we detail below.

Greedy way [78]: We exchange two estimated ICA components belonging to different subspaces, if the exchange decreases the value of the ISA cost J as long as such pairs exist.

"Global" way: Our experiences show that greedy permutation search is often sufficient for the estimation of the ISA subspaces. However, if the greedy approach cannot find the true ISA subspaces, then global permutation search method of higher computational burden may become necessary [90]: the cross-entropy solution suggested for the traveling salesman problem [91] can be adapted to this case.

Now, let us assume that source dimensions (d_m) are not known in advance. The lack of such knowledge causes *combinatorial difficulty* in such a sense that one should try all possible $D = d_1 + \ldots + d_M \quad (d_m > 0, M \le D)$ dimension allocations to the subspace $(\mathbf{e}^m)$ dimensions, where D is the dimension of the hidden source $\mathbf{e}$. The number of these $f(D)$ possibilities grows quickly with the argument, its asymptotic behaviour is known [92, 93]: $f(D) \sim \exp(\pi\sqrt{2D/3})/(4D\sqrt{3})$ as $D \to \infty$. An efficient method with good scaling properties has been put forth in [67] for searching the permutation group for the ISA Separation Theorem (see Table A.3). This approach builds upon the fact that the mutual information between different ISA subspaces $\mathbf{e}^m$ is zero due the assumption of independence. The method assumes that coordinates of $\mathbf{e}^m$ that fall into the same subspace can be paired by using the *pairwise* dependence of the coordinates. For the clustering of the coordinates, one may apply different approaches:

Greedy solutions: The AR-IPA task (Section 4.1.1) can be solved by greedy solutions as in [94] for $L_s = 1$ and [95] for $L_s \ge 1$ assuming that $\mathbf{F}[z]$ is block-diagonal, i.e, the sources $\mathbf{s}^m$ are not coupled through the dynamics. Using the basis transformation rule of AR processes, we can see that after ICA preprocessing of the estimated innovation of the observations, it is sufficient to jointly block diagonalize the coefficient matrices of

18

the polynomial matrix $\hat{\mathbf{F}}^s[z] := \mathbf{W}_{\text{ICA}}\hat{\mathbf{F}}[z]\mathbf{W}_{\text{ICA}}^{-1}$. The optimal permutation to get block diagonal matrices can be estimated using, e.g., greedy clustering of the coordinates. Similarly, greedy considerations can be applied in the ISA problem by replacing $\{|\hat{F}^s_{ij}| \in \mathbb{R}^{D \times D}\}^D_{i,j=1}$, for example, with generalized variance [96], or cumulant based [50] matrices.

Robust approaches: It has been reported that the previous greedy approach is not robust enough in certain applications, and more robust clustering methods have been proposed to overcome this difficulty. These robust approaches include hierarchical clustering [20, 97], tree-structured clustering [98], deterministic annealing [30], and spectral clustering methods [67]. We note that spectral clustering methods scale well; for those ISA problems that satisfy the conditions detailed in Table A.3, a single general desktop computer can handle about a million observations (in our case estimated ICA elements) within several minutes [99]. Spectral clustering thus fits large-scale applications, and for large ISA tasks *ICA is the main bottleneck.*

It is worth noting that one can construct examples where algorithms that only use pairwise dependencies cannot work well [100, 90].

For all of the problems defined in Section 3-4, the dimensions of the source components may differ or may even be unknown. The quality of the solution can be measured by the Amari-index detailed in Section 2.2. Thanks to the separation principle, these problems can be reduced to well-known subtasks, such as ICA, clustering, estimation of mutual information between one-dimensional random variables, gaussianization, estimation of linear models, principal component analysis, nonparametric regression, etc, enabling the exploitation of well-studied solution techniques of these subproblems.

6. Numerical Illustrations

In this section we provide three numerical experiments to demonstrate some of the algorithms presented above.

In our first experiment we compare ISA and AR-IPA methods on a facial dataset [59]. Separating mixed facial images is a common application of ISA. We chose 6 different facial images ($M = 6$) with 50×50 pixels (Fig. A.5(a)). The pixel values were linearly scaled and truncated to integers such that their sum was $100,000$. Then we scanned the images from left-to-right and from top-to-bottom and took the 2D coordinate samples of the

pixels as many times as the value of each pixel. When we mixed these two-dimensional sources, ISA was not able to find the proper subspaces because the sampling is very far from being temporally independent (Fig. A.5(c), (e)). Nevertheless, the AR-IPA method was able to estimate the subspaces of the faces (Fig. A.5(d), (f)).

In our second experiment we compare the TCC and LPA methods on the separation task of the convolutive mixture of stereo Beatles songs (A Hard Day's Night, Can't Buy Me Love) [66]. The sources are not i.i.d., and their dimension is $d_m = 2$ (stereo songs). We studied the $D_x = 2D_e$ case in the uBSSD problem, using sample number $1,000 \leq T \leq 75,000$ and convolution length $1 \leq L_e \leq 30$. The performances measured by the Amari-index are shown in Figs. A.6 (here we averaged 50 independent experiments). Fig A.6(a) demonstrates that the TCC performed well when the sample size was $T \geq 50,000$. The LPA method provided good results for $T \geq 30,000$ (Fig. A.6(b)), and for this sample size it worked better than TCC. For larger convolution parameter L_e, the LPA method is even more superior. For $T = 75,000$ and $L_e = 1, 2, 5, 10, 20, 30$, Fig. A.6(c) shows that on average LPA performs $1.50, 2.24, 4.33, 4.42, 9.03, 11.13$ times better than TCC, respectively.

In our third experiment we compared the AR-IPA and fAR-IPA metohds on the *ikeda* dataset [52]. Here, the hidden $\mathbf{s}_t^m = [s_{t,1}^m, s_{t,2}^m] \in \mathbb{R}^2$ sources ($M = 2$) are realized by the ikeda map $s_{t+1,1}^m = 1 + \lambda_m[s_{t,1}^m \cos(w_t^m) - s_{t,2}^m \sin(w_t^m)]$, $s_{t+1,2}^m = \lambda_m[s_{t,1}^m \sin(w_t^m) + s_{t,2}^m \cos(w_t^m)]$, where λ_m is a parameter of the dynamical system and $w_t^m = 0.4 - \frac{6}{1+(s_{t,1}^m)^2+(s_{t,2}^m)^2}$, see Fig. A.7(a). We mixed these sources by a random $\mathbf{A}$ mixing matrix; they formed our observations $\mathbf{x}_t$. We compared the AR-IPA and fAR-IPA methods on this datset. The results of 10 independent experiments can be seen in Fig. A.7(b). As the results show, the standard AR-IPA method could not find the proper subspaces, but the fAR-IPA method was able to estimate the subspaces for $T \geq 10,000$ sample size.

7. Conclusions

We have reviewed known results on several different generalizations of independent subspace analysis and the ISA separation principle. According to this principle, the ISA task can be solved by applying ICA, and then clustering the ICA components into statistically dependent subspaces. The theorem has recently been rigorously proven for some distribution types. Joint block

diagonalization based methods have an analogous approach with a similar separation principle. The Separation Theorem enables one to construct ISA methods that scale well with the dimensions even if the dimensions of the subspaces differ or are unknown. It also makes possible to extend the ISA problem to different linear-, controlled-, post nonlinear-, complex valued-, and partially observed systems, as well as to systems with nonparametric source dynamics.

Acknowledgments

References

[1] C. Jutten, J. Hérault, Blind separation of sources: An adaptive algorithm based on neuromimetic architecture, Signal Process. 24 (1991) 1–10.

[2] J. Cardoso, A. Souloumiac, Blind beamforming for non-gaussian signals, IEE-Proceedings-F 140 (6) (1993) 362–370.

[3] P. Comon, Independent component analysis, a new concept?, Signal Process. 36 (3) (1994) 287–314.

[4] M. S. Bartlett, J. R. Movellan, T. J. Sejnowski, Face recognition by independent component analysis, IEEE Trans. Neural Networks 13 (6) (2002) 1450–1464.

[5] P. C. Yuen, J. H. Lai, Face representation using independent component analysis, Pattern Recognit. 35 (6) (2002) 1247–1257.

[6] H. Neemuchwalaa, A. Hero, P. Carson, Image matching using alpha-entropy measures and entropic graphs, Signal Process. 85 (2) (2005) 277–296.

[7] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, T. J. Sejnowski, Analysis of fMRI data by blind separation into independent spatial components, Human Brain Mapping 6 (1998) 160–188.

[8] A. J. Bell, T. J. Sejnowski, The 'independent components' of natural scenes are edge filters, Vision Res. 37 (1997) 3327–3338.

21

[9] R. Jenssen, T. Eltoft, Independent component analysis for texture segmentation, Pattern Recognit. 36 (13) (2003) 2301–2315.

[10] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (4-5) (2000) 411–430.

[11] A. Cichocki, S. Amari, Adaptive blind signal and image processing, Wiley, 2002.

[12] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, 2001.

[13] S. Choi, A. Cichocki, H. Park, S. Lee, Blind source separation and independent component analysis, Neural Inf. Proc. Lett. Rev. 6 (2005) 1–57.

[14] J. Cardoso, Multidimensional independent component analysis, in: ICASSP '98, pp. 1941–1944.

[15] A. Hyvärinen, P. O. Hoyer, Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces, Neural Comput. 12 (2000) 1705–1720.

[16] H. Kim, S. Choi, S. Bang, Membership scoring via independent feature subspace analysis for grouping co-expressed genes, in: IJCNN 2003, pp. 1690–1695.

[17] A. Sharma, K. K. Paliwal, Subspace independent component analysis using vector kurtosis, Pattern Recognit. 39 (2006) 2227–2232.

[18] F. J. Theis, Blind signal separation into groups of dependent signals using joint block diagonalization, in: ISCAS 2005, pp. 5878–5881.

[19] L. D. Lathauwer, B. D. Moor, J. Vandewalle, Fetal electrocardiogram extraction by blind source subspace separation, IEEE Trans. Biomed. Eng. 47 (5) (2000) 567–572.

[20] H. Stögbauer, A. Kraskov, S. A. Astakhov, P. Grassberger, Least dependent component analysis based on mutual information, Phys. Rev. E Stat. Nonlin. Soft. Matter Phys. 70 (6) (2004) 066123.

[21] S. Ma, X.-L. Li, N. Correa, T. Adali, V. Calhoun, Independent subspace analysis with prior information for fMRI data, in: ICASSP-2010, pp. 1922–1925.

[22] F. Kohl, G. Wübbeler, D. Kolossa, C. Elster, M. Bär, R. Orglmeister, Non-independent BSS: A model for evoked MEG signals with controllable dependencies, in: ICA 2009, pp. 443–450.

[23] B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.

[24] Y. Nishimori, S. Akaho, M. D. Plumbley, Riemannian optimization method on the flag manifold for independent subspace analysis, in: ICA 2006, pp. 295–302.

[25] H. Choi, S. Choi, Relative gradient learning for independent subspace analysis, in: IJCNN-2006, pp. 3919–3924.

[26] C. S. Santos, J. E. Kögler, E. D. M. Hernandez, Using independent subspace analysis for selecting filters used in texture processing, in: ICIP 2005, pp. 465–468.

[27] Q. Le, W. Zou, S. Yeung, A. Ng, Stacked convolutional independent subspace analysis for action recognition, in: CVPR 2011, pp. 3361–3368.

[28] A. Hyvärinen, P. O. Hoyer, M. Inki, Topographic independent component analysis, Neural Comput. 13 (7) (2001) 1527–1558.

[29] S. Z. Li, X. Lv, H. Zhang, View-subspace analysis of multi-view face patterns, in: RATFG-RTS-2001, pp. 125–132.

[30] M. A. Casey, A. Westner, Separation of mixed audio sources by independent subspace analysis, in: ICMC 2000, pp. 154–161.

[31] Z. Fan, J. Zhou, Y. Wu, Motion segmentation based on independent subspace analysis, in: ACCV 2004.

[32] J. K. Kim, S. Choi, Tree-dependent components of gene expression data for clustering, in: ICANN 2006, pp. 837–846.

[33] Z. Szabó, B. Póczos, A. Lőrincz, Undercomplete blind subspace deconvolution, J. Mach. Learn. Res. 8 (2007) 1063–1095.

[34] A. Hyvärinen, Independent component analysis for time-dependent stochastic processes, in: ICANN '98, pp. 541–546.

[35] M. S. Pedersen, J. Larsen, U. Kjems, L. C. Parra, A survey of convolutive blind source separation methods, in: Springer Handbook of Speech Processing, Springer, 2007.

[36] C. Jutten, J. Karhunen, Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear systems, Int. J. Neural Syst. 14 (5) (2004) 267–292.

[37] J. Anemüller, T. J. Sejnowski, S. Makeig, Complex independent component analysis of frequency-domain electroencephalographic data, Neural Networks 16 (2003) 1311–1323.

[38] V. Calhoun, T. Adali, G. Pearlson, P. van Zijl, J. Pekar, Independent component analysis of fMRI data in the complex domain, Magn. Reson. Med. 48 (2002) 180–192.

[39] J. Anemüller, J. Duann, T. J. Sejnowski, S. Makeig, Spatio-temporal dynamics in fMRI recordings revealed with complex independent component analysis, Neurocomputing 69 (13-15) (2006) 1502–1512.

[40] K. Chan, T.-W. Lee, T. J. Sejnowski, Variational Bayesian learning of ICA with missing data, Neural Comput. 15 (8) (2003) 1991–2011.

[41] A. T. Cemgil, C. Févotte, S. J. Godsill, Variational and stochastic inference for Bayesian source separation, Digit. Signal Process. 17 (2007) 891–913.

[42] J. Anemüller, Second-order separation of multidimensional sources with constrained mixing system, in: ICA 2006, pp. 16–23.

[43] T. M. Cover, J. A. Thomas, Elements of information theory, Wiley, New York, USA, 1991.

[44] L. Ma, L. Zhang, Bayesian estimation of overcomplete independent feature subspaces for natural images, in: ICA 2007, pp. 746–753.

[45] A. Hyvärinen, M. Inki, Estimating overcomplete independent component bases for image windows, J. Math. Imagng Vision 17 (2) (2002) 139–152.

[46] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psych. 24 (1933) 417–441.

[47] B. Póczos, A. Lőrincz, Independent subspace analysis using k-nearest neighborhood distances, ICANN 2005 163–168.

[48] E. G. Learned-Miller, J. W. Fisher III, ICA using spacings estimates of entropy, J. Mach. Learn. Res. 4 (2003) 1271–1295.

[49] F. J. Theis, Uniqueness of complex and multidimensional independent component analysis, Signal Process. 84 (5) (2004) 951–956.

[50] F. J. Theis, Towards a general independent subspace analysis, in: NIPS, 2007, pp. 1361–1368.

[51] S. Amari, A. Cichocki, H. H. Yang, A new learning algorithm for blind signal separation, NIPS '96 757–763.

[52] Z. Szabó, B. Póczos, Nonparametric independent process analysis, in: EUSIPCO-2011, pp. 1718–1722.

[53] K.-T. Fang, S. Kotz, K. W. Ng, Symmetric multivariate and related distributions, Chapman and Hall, 1990.

[54] A. K. Gupta, D. Song, L^p-norm spherical distribution, J. Stat. Plan. Inference 60 (2) (1997) 241–260.

[55] S. Takano, The inequalities of Fisher information and entropy power for dependent variables, in: Proc. of the 7th Japan-Russia Symposium on Probability Theory and Mathematical Statistics, 1995.

[56] K. Abed-Meraim, A. Belouchrani, Algorithms for joint block diagonalization, in: EUSIPCO 2004, pp. 209–212.

[57] F. J. Theis, Multidimensional independent component analysis using characteristic functions, in: EUSIPCO 2005.

[58] K. Nordhausen, H. Oja, Independent subspace analysis using three scatter matrices, Austrian J. Stat. 40 (1-2) (2011) 93–101.

[59] B. Póczos, B. Takács, A. Lőrincz, Independent subspace analysis on innovations, in: ECML 2005, pp. 698–706.

[60] R. H. Lambert, Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures, Ph.D. thesis, University of Southern California (1996).

[61] V. Petrov, Central limit theorem for m-dependent variables, in: Proc. of the All-Union Conf. on Prob. Theory and Math. Statistics, 1958, pp. 38–44.

[62] A. Neumaier, T. Schneider, Estimation of parameters and eigenmodes of multivariate autoregressive models, ACM Trans. Math. Soft. 27 (1) (2001) 27–57.

[63] T. Schneider, A. Neumaier, Algorithm 808: ARfit - a Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models, ACM Trans. Math. Soft. 27 (1) (2001) 58–65.

[64] S. Choi, Differential learning algorithms for decorrelation and independent component analysis, Neural Networks 19 (10) (2006) 1558–1567.

[65] R. Rajagopal, L. C. Potter, Multivariate MIMO FIR inverses, IEEE Trans. Image Proc. 12 (2003) 458–465.

[66] Z. Szabó, B. Póczos, A. Lőrincz, Undercomplete blind subspace deconvolution via linear prediction, in: ECML 2007, pp. 740–747.

[67] B. Póczos, Z. Szabó, M. Kiszlinger, A. Lőrincz, Independent process analysis without a priori dimensional information, in: ICA 2007, pp. 252–259.

[68] Z. Szabó, Complete blind subspace deconvolution, in: ICA 2009, pp. 138–145.

[69] S. Choi, A. Cichocki, Blind signal deconvolution by spatio-temporal decorrelation and demixing, Neural Networks for Signal Process. 7 (1997) 426–435.

[70] A. Mansour, C. Jutten, P. Loubaton, Subspace method for blind separation of sources in convolutive mixture, in: EUSIPCO '96, pp. 2081–2084.

[71] S. Icart, R. Gautier, Blind separation of convolutive mixtures using second and fourth order moments, in: ICASSP '96, pp. 3018–3021.

[72] A. Gorokhov, P. Loubaton, Multiple-input multiple-output ARMA systems: Second order blind identification for signal extractions, in: SSAP '96, pp. 348–351.

[73] Z. Szabó, B. Póczos, G. Szirtes, A. Lőrincz, Post nonlinear independent subspace analysis, in: ICANN 2007, pp. 677–686.

[74] A. Taleb, C. Jutten, Source separation in post-nonlinear mixtures, IEEE Trans. Signal Process. 10 (47) (1999) 2807–2820.

[75] A. Ziehe, M. Kawanabe, S. Harmeling, K.-R. Müller, Blind separation of post-nonlinear mixtures using linearizing transformations and temporal decorrelation, J. Mach. Learn. Res. 4 (7-8) (2003) 1319–1338.

[76] J. Solé-Casals, C. Jutten, D. Pham, Fast approximation of nonlinearities for improving inversion algorithms of PNL mixtures and Wiener systems, Signal Process. 85 (2005) 1780–1786.

26

[77] J. Eriksson, Complex random vectors and ICA models: Identifiability, uniqueness and separability, IEEE Trans. Inf. Theor. 52 (3) (2006) 1017–1029.

[78] Z. Szabó, A. Lőrincz, Real and complex independent subspace analysis by generalized variance, in: ICARN 2006, pp. 85–88.

[79] Y. Nishimori, S. Akaho, M. D. Plumbley, Natural conjugate gradient on complex flag manifolds for complex independent subspace analysis, in: ICANN 2008, pp. 165–174.

[80] P. Krishnaiah, J. Lin, Complex elliptically symmetric distributions, Comm. Stat. 15 (12) (1986) 3693–3718.

[81] Z. Szabó, A. Lőrincz, Towards independent subspace analysis in controlled dynamical systems, in: ICARN 2008, pp. 9–12.

[82] Z. Szabó, Autoregressive independent process analysis with missing observations, in: ESANN 2010, pp. 159–164.

[83] H.-M. Krolzig, Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis., Springer, 1997.

[84] B. Póczos, A. Lőrincz, Identification of recurrent neural networks by Bayesian interrogation techniques, J. Mach. Learn. Res. 10 (2009) 515–554.

[85] J. T. Lomba, Estimation of Dynamic Econometric Models with Errors in Variables, Vol. 339 of Lecture notes in econ. and math. systems, Springer, 1990.

[86] A. García-Hiernaux, J. Casals, M. Jerez, Fast estimation methods for time series models in state-space form, J. Stat. Comput. Simul. 79 (2) (2009) 121–134.

[87] K. R. Kadiyala, S. Karlsson, Numerical methods for estimation and inference in Bayesian VAR-models, J. Appl. Econ. 12 (1997) 99–132.

[88] D. Bosq, Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, Lecture Notes in Statistics, Springer, 1998.

[89] N. Hilgert, B. Portier, Strong uniform consistency and asymptotic normality of a kernel based error density estimator in functional autoregressive models, Tech. rep., http://arxiv.org/abs/0905.2327v1 (2009).

[90] Z. Szabó, B. Póczos, A. Lőrincz, Cross-entropy optimization for independent process analysis, in: ICA 2006, pp. 909–916.

[91] R. Y. Rubinstein, D. P. Kroese, The Cross-Entropy Method, Springer, 2004.

[92] G. H. Hardy, S. I. Ramanujan, Asymptotic formulae in combinatory analysis, Proc. London Math. Soc. 17 (1) (1918) 75–115.

[93] J. V. Uspensky, Asymptotic formulae for numerical functions which occur in the theory of partitions, Bull. Russ. Acad. Sci. 14 (6) (1920) 199–218.

[94] B. Póczos, A. Lőrincz, Non-combinatorial estimation of independent autoregressive sources, Neurocomputing Lett. 69 (2006) 2416–2419.

[95] Z. Szabó, B. Póczos, A. Lőrincz, Auto-regressive independent process analysis without combinatorial efforts, Pattern Anal. Appl. 13 (2010) 1–13.

[96] Z. Szabó, A. Lőrincz, Independent subspace analysis can cope with the „curse of dimensionality", Acta Cybernet. (+Symp. Intell. Syst. 2006) 18 (2007) 213–221.

[97] P. Gruber, H. W. Gutch, F. J. Theis, Hierarchical extraction of independent subspaces of unknown dimensions, in: ICA 2009, pp. 259–266.

[98] F. R. Bach, M. I. Jordan, Beyond independent components: Trees and clusters, J. Mach. Learn. Res. 4 (2003) 1205–1233.

[99] D. Yan, L. Huang, M. I. Jordan, Fast approximate spectral clustering, in: KDD 2009, pp. 907–916.

[100] B. Póczos, A. Lőrincz, Independent subspace analysis using geodesic spanning trees, in: ICML 2005, pp. 673–680.

Zoltán Szabó received M.Sc. in Applied Mathematics, 2006 and Ph.D. in Computer Science, 2009. He is a Ph.D. candidate (Applied Mathematics) and a research fellow at the Eötvös Loránd University. His research interest includes independent subspace analysis and its extensions, information theory and kernel methods.

Barnabás Póczos received M.Sc. in Applied Mathematics, 2002 and Ph.D. in Computer Science, 2007. He is a postdoctoral fellow at the School of Computer Science, Carnegie Mellon University. His current research interest lies in the area of unsupervised learning, manifold learning, Bayesian methods, entropy and mutual information estimation.

András Lőrincz received M.Sc. in Physics, 1975; Ph.D. in Solid State Physics, 1978; C.Sc. in Molecular Physics, 1986; Laser Physics habilitation, 1998; Information Technology habilitation, 2009. He is working on intelligent systems and human computer collaboration at the Eötvös Loránd University.

Appendix A. Abbreviations

We summarize the notations in Table A.1. The first part of the list contains the acronyms. It is followed by a few mathematical notations.

Acronyms	
Abbreviation	Meaning
AR, MA, ARMA	AutoRegressive, Moving Average, AR MA
ARX, ARIMA	AR with eXogenous input, Integrated ARMA
fAR, mAR	Functional AR, AR with Missing values
PNL, "u-"	Post NonLinear, prefix for "undercomplete"
ECG, EEG,	Electro-Cardiography, Electro-Encephalography
EPI	Entropy Power Inequality
FIR	Finite Impulse Response
fMRI	Functional Magnetic Resonance Imaging
ICA/ISA/IPA	Independent Component/Subspace/Process Analysis
i.i.d.	Independent Identically Distributed
JBD	Joint Block Diagonalization
LPA	Linear Prediction Approximation
TCC	Temporal Concatenation
Mathematical notations	
$\mathbb{R}, \mathbb{R}^L, \mathbb{R}^{L_1 \times L_2}, \mathbb{R}[z]^{L_1 \times L_2}$ $\mathbb{C}, \mathbb{C}^L, \mathbb{C}^{L_1 \times L_2}$	real and complex numbers, L-dimensional vectors, $L_1 \times L_2$ sized matrices, -polynomial matrices
$\mathcal{O}^D, \mathcal{U}^D$	$D \times D$ sized orthogonal and unitary matrices
H, I	entropy, mutual information
$\mathbb{S}_{\mathbb{R}}^d, \mathbb{S}_{\mathbb{C}}^d$	d-dimensional unit sphere over $\mathbb{R}$ and $\mathbb{C}$
$\otimes, \Re(\cdot), \Im(\cdot)$	Kronecker product, real and imaginary part

Table A.1: Acronyms and mathematical notations.

Input of the algorithm:
observation ($\{\mathbf{x}_t\}_{t=1}^T$), AR order (L_s),
optional (depending on the applied ISA solver):
number of components (M) and source dimensions ($\{d_m\}_{m=1}^M$).
Optimization:
AR estimation of order L_s on $\mathbf{x} \Rightarrow \hat{\mathbf{F}}_{\mathrm{AR}}[z]$,
Estimation of the innovation of $\mathbf{x} \Rightarrow \hat{\mathbf{n}} = \hat{\mathbf{F}}_{\mathrm{AR}}[z]\mathbf{x}$,
ISA on the estimated innovation $\hat{\mathbf{n}} \Rightarrow$
ISA demixing matrix: $\hat{\mathbf{W}}_{\mathrm{ISA}}$; optional: $\hat{M}, \{\hat{d}_m\}_{m=1}^{\hat{M}}$.
Output of the algorithm:
estimated mixing matrix, hidden source: $\hat{\mathbf{A}} = \hat{\mathbf{W}}_{\mathrm{ISA}}^{-1}$, $\hat{\mathbf{s}} = \hat{\mathbf{W}}_{\mathrm{ISA}}\mathbf{x}$,
source dynamics, driving noise: $\hat{\mathbf{F}}[z] = \hat{\mathbf{W}}_{\mathrm{ISA}}\hat{\mathbf{F}}_{\mathrm{AR}}[z]\hat{\mathbf{W}}_{\mathrm{ISA}}^{-1}$, $\hat{\mathbf{e}} = \hat{\mathbf{W}}_{\mathrm{ISA}}\hat{\mathbf{n}}$,
optional:
number of components and dimensions of sources: $\hat{M}, \{\hat{d}_m\}_{m=1}^{\hat{M}}$.

Table A.2: AR-IPA Algorithm - pseudocode.

Construct an undirected graph with nodes corresponding to ICA coordinates and edge weights (similarities) defined by the *pairwise* statistical dependencies, i.e., the mutual information of the estimated ICA elements: $\mathbf{S} = [\hat{I}(\hat{e}_{\mathrm{ICA},i}, \hat{e}_{\mathrm{ICA},j})]_{i,j=1}^D$. Cluster the ICA elements, i.e., the nodes using similarity matrix $\mathbf{S}$.

Table A.3: Approximation that scales well for the permutation search task in the ISA Separation Theorem.
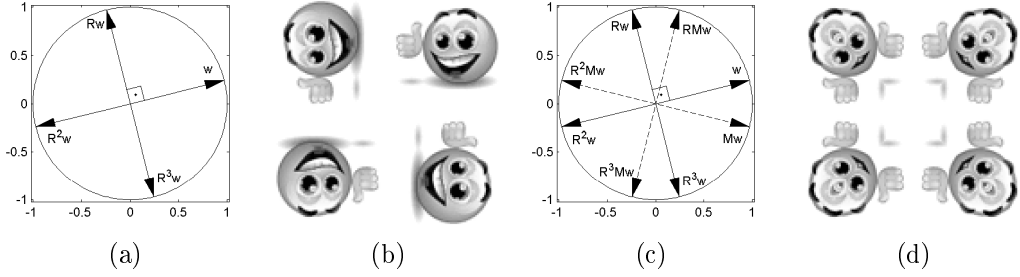
Figure A.1: Illustration: Density functions (for variables $\mathbf{e}^m$) invariant to 90° rotation *or* permutation and sign changes. (a) and (c): density function f takes identical values at the arrowheads. Matrix $\mathbf{R}$ and matrix $\mathbf{M}$ are 90° counter-clockwise rotation and reflection to axis x, respectively. (b) and (d): density functions (illustrated as 2D images) for (a) and (c), respectively.
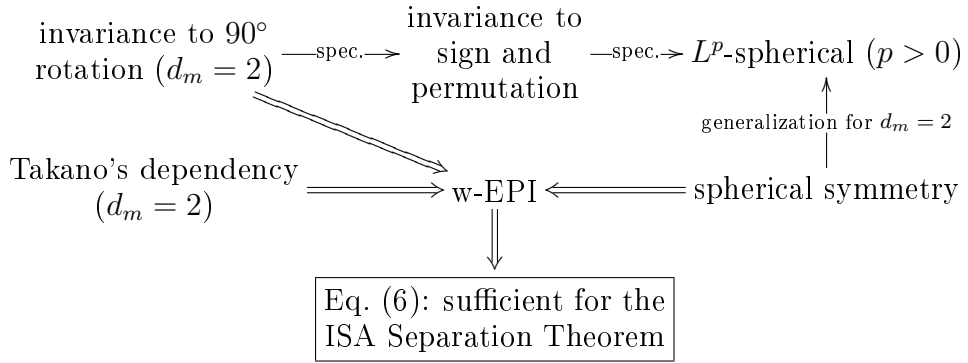

Figure A.2: Sufficient conditions for the ISA Separation Theorem. For details, see Theorem 1.
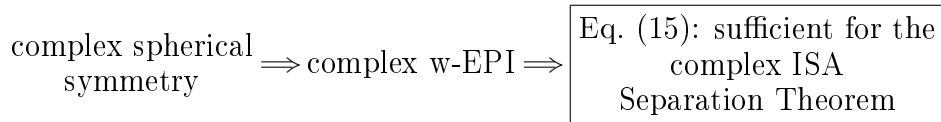

Figure A.3: Sufficient conditions for the complex ISA Separation Theorem. For details, see Section 4.3.

31

ARIMA-IPA   ARX-IPA   mAR-IPA

$\downarrow r=0$ $\qquad$ $\downarrow L_u=0$ $\qquad$ $\mathbf{M}_t=$identity

ARMA-IPA $\xrightarrow{L_e=0}$ AR-IPA $\xleftarrow{\mathbf{f}=\text{linear}}$ fAR-IPA

$\downarrow L_s=0$ $\qquad\qquad$ $\downarrow L_s=1$

MA-IPA (BSSD) $\xrightarrow{L_e=0}$ ISA $\xleftarrow{\mathbf{f}=\text{identity}}$ PNL-ISA

$\downarrow d_m=1$ $\qquad$ $\downarrow d_m=1$ $\qquad$ $\downarrow d_m=1$

BSD $\xrightarrow{L_e=0}$ ICA $\xleftarrow{\mathbf{f}=\text{identity}}$ PNL-ICA

(a)

$\mathbb{C}$ linear $\xrightarrow{\varphi_v,\varphi_M}$ $\mathbb{R}$ linear

PNL $\xrightarrow{\text{gaussianization}}$ linear

uMA-IPA $\xrightarrow{\text{TCC}}$ ISA

uARIMA-IPA, MA-IPA $\xrightarrow{\text{LPA}}$ ISA

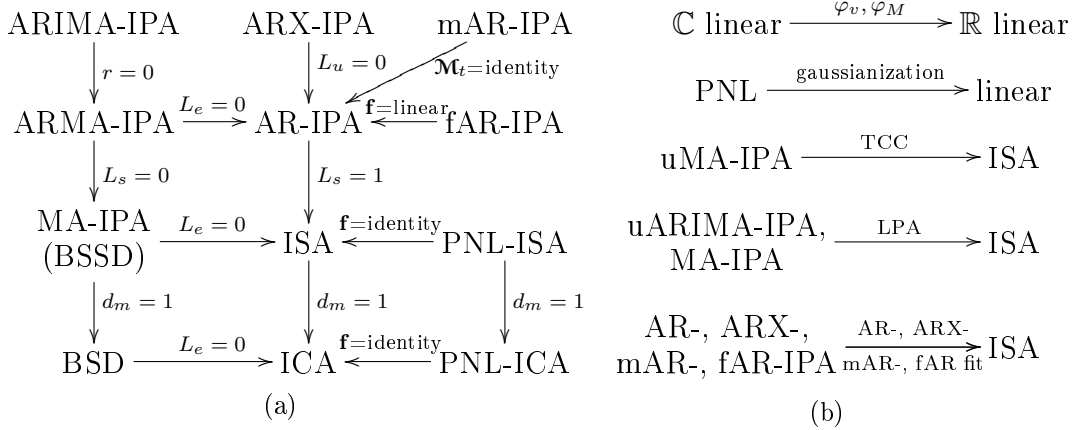AR-, ARX-, mAR-, fAR-IPA $\xrightarrow[\text{mAR-, fAR fit}]{\text{AR-, ARX-}}$ ISA

(b)

Figure A.4: Illustration of the IPA problem family. (a): Connections between problems; arrows point to special cases. (b): Separation principles. Prefix "u" denotes undercomplete case. According to the figure, complex linear models (Section4.3) can be reduced to linear models (Section 4.1) using the φ_v, φ_M transformations. Similarly, PNL problems (Section 4.2) can be reduced to linear problems (Section 4.1) using gaussianization, etc.
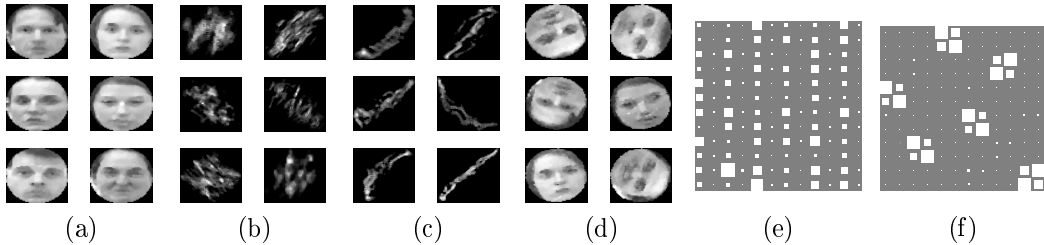
(a)　(b)　(c)　(d)　(e)　(f)

Figure A.5: AR-IPA vs ISA illustration on facial dataset. (a): Original facial images, the hidden $\mathbf{s}_t$ sources. (b): Mixed sources, the observation ($\mathbf{x}_t$). (c)-(d): Independent sources estimated ($\hat{\mathbf{s}}_t$) by ISA and AR-IPA, respectively. (e)-(f): Hinton-diagram of $\mathbf{G}$, the product of the estimated demixing matrix and the mixing matrix for ISA and AR-IPA, respectively. When the separation is perfect, it is a block-permutation matrix with 2×2 blocks.

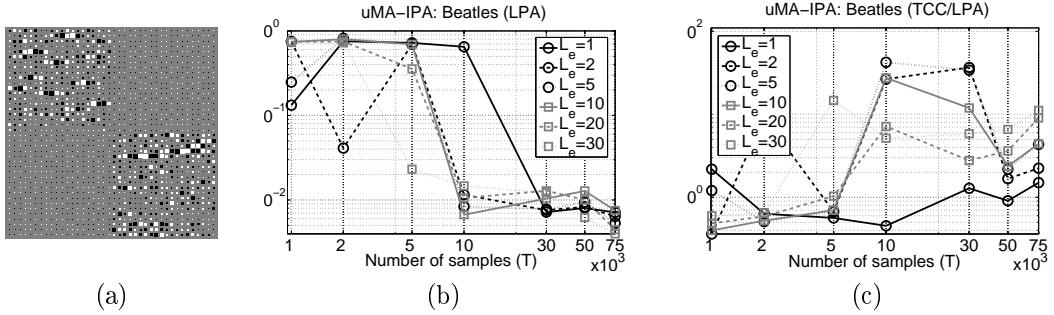<p align="center">(a)   (b)   (c)</p>

Figure A.6: LPA vs TCC illustration on convolved *Beatles* songs. (a): Hinton-diagram of $\mathbf{G}$ using the TCC method with $L_e = 5$. When the separation is perfect this is a block-permutation matrix of two blocks. (b): LPA performance as a function of the sample size (T) and the convolution length (L_e). (c): same as (b), but here we show the quotient of the TCC and LPA Amari-indices (showing how many times LPA is better than TCC).

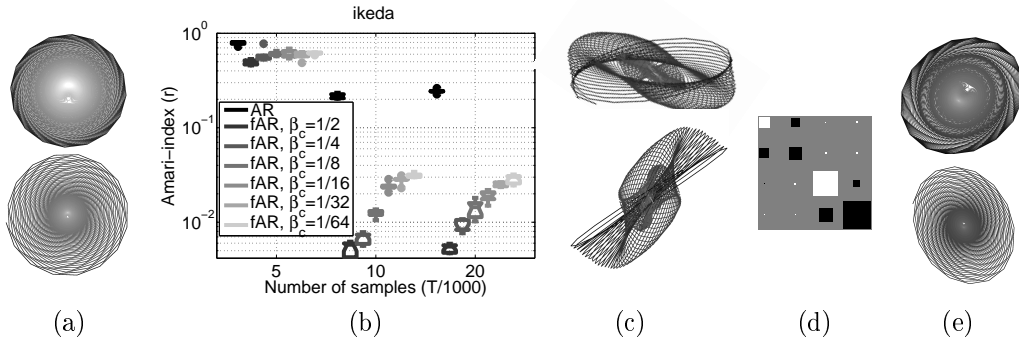<p align="center">(a)   (b)   (c)   (d)   (e)</p>

Figure A.7: AR-IPA vs fAR-IPA illustration on the *ikeda* dataset. (a): hidden $\mathbf{s}_t$ sources. (b): Amari-index as a function of the sample number for the AR-IPA and the fAR-IPA method; $\beta_c \in (0,1)$: kernel regression bandwith parameter. (c): Observation, $\mathbf{x}_t$. (d): Hinton-diagram of $\mathbf{G}$ with average Amari-index. When the separation is perfect it is block-permutation matrix of 2×2 sized blocks. (e): Estimated subspaces ($\hat{\mathbf{s}}_t$) using the fAR-IPA method ($\beta_c = \frac{1}{2}$, $T = 20,000$).