Uniqueness of linear factorizations into independent subspaces

Harold W. Gutch [a,c,*], Fabian J. Theis [b,c]

[a] *Max Planck Institute for Dynamics and Self-Organization (MPIDS), 37077 Goettingen, Germany*

[b] *CMB, Institute of Bioinformatics and Systems Biology, German Research Center for Environmental Health, 85764 Neuherberg, Germany*

[c] *Technical University of Munich, Arcisstrasse 21, 80333 München, Germany*

ARTICLE INFO

ABSTRACT

Given a random vector **X**, we address the question of linear separability of **X**, that is, the task of finding a linear operator **W** such that we have $(\mathbf{S}_1, \ldots, \mathbf{S}_M) = (\mathbf{WX})$ with statistically independent random vectors $\mathbf{S}_i$. As this requirement alone is already fulfilled trivially by **X** being independent of the empty rest, we require that the components be not further decomposable. We show that if **X** has finite covariance, such a representation is unique up to trivial indeterminacies. We propose an algorithm based on this proof and demonstrate its applicability. Related algorithms, however with fixed dimensionality of the subspaces, have already been successfully employed in biomedical applications, such as separation of fMRI recorded data. Based on the presented uniqueness result, it is now clear that also subspace dimensions can be determined in a unique and therefore meaningful fashion, which shows the advantages of independent subspace analysis in contrast to methods like principal component analysis.

1. Introduction

Assume a random vector **S** consisting of statistically independent components S_i, none of which is Gaussian (normally distributed). If the components of a linear mixing **AS** then again are statistically independent, one can show that **A** is at most the product of a permutation and scaling within the components, which originally was shown using the Darmois–Skitovitch theorem [11,21,30]. Under the additional assumption of finite covariance of **S**, one may additionally allow at most one of the S_i to be Gaussian [10]. The assumption of finite covariance actually is not required for this to hold [13], but if one assumes it, a simpler proof is possible, based on the idea that the characteristic function of **S** factorizes, so its logarithm has a diagonal Hessian almost everywhere [33].

This property of random variables has driven the development of algorithms performing the so-called Independent Component Analysis (ICA) under some approximations of statistical independence [4,10,20,37]. Exact cost functions, the so-called contrasts such as mutual information, are difficult to estimate in practice, where the random vector in question is only known up to some finite precision, so many approximations have been extensively studied. Such algorithms have been successfully used in various fields, e.g. signal processing, biomedical imaging and analysis of financial data, where it was argued that the data sets to be analyzed can be approximated well enough by modeling them as random variables mixed in a linear fashion; see [9,18] and references therein.

Apart from the question of validity when transferring the mathematical theory to real life data sets, another problem is apparent: What if a given random vector **X** has no such representation? This motivates the question if the original claim has a straight-forward extension to higher dimensions: if we write $(\mathbf{S}_1, \ldots, \mathbf{S}_M) := \mathbf{WX}$ with independent random vectors

---

* Corresponding author at: Max Planck Institute for Dynamics and Self-Organization (MPIDS), 37077 Goettingen, Germany.
 E-mail addresses: harold.gutch@ds.mpg.de (H.W. Gutch), fabian.theis@helmholtz-muenchen.de (F.J. Theis).

$\mathbf{S}_i$ of which at most one is Gaussian, again, are these unique up to permutation and invertible linear transformations (the multidimensional translation of scaling) within the $\mathbf{S}_i$? Loosened assumptions to the ICA model where specifically such subspaces in data were sought for have gained a lot of interest in recent years; see e.g. [3,8,14,24,28,40]. The task of finding a basis in which a random vector $\mathbf{X}$ has this property is usually denoted Independent Subspace Analysis (ISA), as one typically reads the independent random vectors gained here as data subsets or data subspaces [7]. Obviously this task requires some minimality constraint, as, given such an independent representation, we might arbitrarily group together some of the components and thus of course get two representations differing in more than just permutation and linear transformations. Our minimality constraint is the inability to decompose any of the components even further, a property we call *irreducibility* of the components.

Our main result is the following uniqueness theorem.

Theorem 1.1. *The decomposition of a random vector* $\mathbf{X}$ *with existing covariance into independent, irreducible components is unique up to order and invertible transformations within the components and an invertible transformation in the possibly higher-dimensional Gaussian component.*

Transferring this mathematical statement to the real world, the conclusion now is that *every* kind of data that one can model as a random vector inherently has a unique factorization into subspaces. This statement has far wider applicability than ICA, which states uniqueness of a factorization only *if* one exists at all within the limits of the model—i.e. if there is a decomposition into independent one-dimensional components. In contrast to this, ISA is applicable to almost any kind of high-dimensional data that can be modeled as a random vector—the only requirement is finite covariance.

This manuscript is structured as follows. In Section 2 we define the notation used and the framework we work in and state a few simple lemmata used. The main part of this manuscript, Section 3, consists of the proof of Theorem 1.1. In Section 4, we propose an algorithm whose main ideas are based on the theoretical proof and demonstrate its applicability in a simulation study. Finally, in Section 5 we shed some light on the practical usefulness of this result, compare it with the literature and address some open questions.

Some parts of this work were presented at the ICA 2007 conference [15], stating Theorem 1.1, however lacking the proof and the algorithmic approach.

2. Definition of independent subspace analysis

We will define the analyzed model and review a few properties of characteristic functions. We restrict this analysis to the real case, that is, real valued random vectors and real linear mixings thereof, although extensions to the complex case are possible.

2.1. Notation

In order to be able to quickly differentiate between scalars and vectors, scalars are depicted in regular font, e.g. $x \in \mathbb{R}$ while vectors and matrices are depicted in bold font, e.g. $\mathbf{x} \in \mathbb{R}^n$. Random values and vectors are always depicted in uppercase letters, e.g. S and $\mathbf{S}$, and we will only need the two letters $\mathbf{S}$, and $\mathbf{X}$ (and regular typeface versions thereof) for these; all other uppercase letters used represent real valued matrices. In order to keep the notation as simple as possible, vectors will often be written in rows, e.g. $\mathbf{x} = \mathbf{A}(\mathbf{v}_1, \mathbf{v}_2)$ instead of $\mathbf{x} = \mathbf{A}(\mathbf{v}_1^\top, \mathbf{v}_2^\top)^\top$. The symbol ∂_i denotes the i-th partial derivative operator, so for a function depending on $\mathbf{x} = (x_1, \ldots, x_n)$, we have $\partial_i = \frac{\partial}{\partial x_i}$. We write $\mathbf{d}f$ for the differential of an $f \in C^1$, and we depict the Hessian of an $f \in C^2$ with $\mathbf{H}_f$, that is $\mathbf{e}_i^\top \mathbf{H}_f \mathbf{e}_j = \partial_i \partial_j f$. We write $\mathbf{d}f|_\mathbf{x}$ instead of $(\mathbf{d}f)(\mathbf{x})$, the differential of f evaluated at $\mathbf{x}$, and similarly $\mathbf{H}_f|_\mathbf{x}$ instead of $\mathbf{H}_f(\mathbf{x})$, the Hessian of f evaluated at $\mathbf{x}$.

2.2. Irreducibility

Let us now introduce the key notion of irreducibility and point out the special role of Gaussian random vectors.

Definition 2.1. An n-dimensional random vector $\mathbf{X}$ is said to be *reducible* if it can be written as $\mathbf{X} = \mathbf{A}(\mathbf{S}_1, \mathbf{S}_2)$ with some invertible $n \times n$-matrix $\mathbf{A}$, a k-dimensional random vector $\mathbf{S}_1$ and an $(n-k)$-dimensional random vector $\mathbf{S}_2$, where $\mathbf{S}_1$ is independent of $\mathbf{S}_2$. A random vector that is not reducible is called *irreducible*.

Remark 2.1. For example, any n-dimensional Gaussian random vector is reducible if $n > 1$: Gaussians are fully defined by their first and second order moments, so here independence is equivalent to decorrelation, and for every random vector $\mathbf{X}$ with finite covariance there is some invertible matrix $\mathbf{A}$ such that $\mathbf{AX}$ is decorrelated (see Lemma 3.1). Therefore, an n-dimensional Gaussian can always be fully reduced to one-dimensional components.

Obviously both properties, irreducibility and reducibility, are preserved under any invertible linear transformation.

A decomposition $(\mathbf{X}_1, \ldots, \mathbf{X}_L) = \mathbf{X}$ of a random vector $\mathbf{X}$ is said to be *independent*, if the random vectors $\mathbf{X}_j$ $(j = 1, \ldots, L)$ are mutually statistically independent. It is said to be *irreducible*, if the vectors $\mathbf{X}_j$ additionally are irreducible.