Handbook of Blind Source Separation, Independent Component Analysis and Applications

P. Comon and C. Jutten Eds

October 6, 2009

Glossary

$\mathbf{x}$	vector of components x_p, $1 \leq p \leq P$
s, x, y	sources, observations, separator outputs
N	number of sources
P	number of sensors
T	number of observed samples
$\star$	convolution
$\mathbf{A}$	matrix with components A_{ij}
$\mathbf{A}$, $\mathbf{B}$	mixing and separation matrices
$\mathbf{G}$, $\mathbf{W}$, $\mathbf{Q}$	global, whitening, and separating unitary matrices
$\check{g}$	Fourier transform of g
$\widehat{\mathbf{s}}$	estimate of quantity $\mathbf{s}$
$p_{\mathbf{x}}$	probability density of $\mathbf{x}$
ψ	joint score function
φ_i	marginal score function of source s_i
Φ	first characteristic function
Ψ	second characteristic function
$\mathbb{E}\,\mathbf{x}$, $\mathbb{E}\{\mathbf{x}\}$	mathematical expectation of $\mathbf{x}$
$I\{\mathbf{y}\}$ or $I(p_{\mathbf{y}})$	mutual information of $\mathbf{y}$
$K\{\mathbf{x};\mathbf{y}\}$ or $K(p_{\mathbf{x}};p_{\mathbf{y}})$	Kullback divergence between $p_{\mathbf{x}}$ and $p_{\mathbf{y}}$
$H\{\mathbf{x}\}$ or $H(p_{\mathbf{x}})$	Shannon entropy $\mathbf{x}$
$\mathcal{L}$	likelihood
$\mathcal{A}$, $\mathcal{B}$	mixing, and separating (non linear) operators
$\mathrm{cum}\{x_1, \ldots, x_P\}$	joint cumulant of variables $\{x_1, \ldots, x_P\}$
$\mathrm{cum}_R\{y\}$	marginal cumulant of order R of variable y
$\mathbf{Q}^{\mathrm{T}}$	transposition
$\mathbf{Q}^{\mathrm{H}}$	conjugate transposition
$\mathbf{Q}^{*}$	complex conjugation
$\mathbf{Q}^{\dagger}$	pseudo-inverse

Υ	contrast function
$\mathbb{R}$	real field
$\mathbb{C}$	complex field
$\widehat{\mathbf{A}}$	estimator of mixing matrix
diag $\mathbf{A}$	vector whose composents are the diagonal of matrix $\mathbf{A}$
Diag $\mathbf{a}$	diagonal matrix whos entries are those of vector $\mathbf{a}$
trace $\mathbf{A}$	trace of matrix $\mathbf{A}$
det $\mathbf{A}$	determinant of matrix $\mathbf{A}$
mean $\mathbf{a}$	arithmetic average of component of vector $\mathbf{a}$
$\check{s}(\nu)$	Fourier transform of process $s(t)$
$\otimes$	Kronecker product between matrices
$\otimes$	tensor product
$\bullet_j$	contraction over index j
krank$\{\mathbf{A}\}$	Kruskal's k-rank of matrix $\mathbf{A}$

Contents

Chapter 1

Algebraic identification of under-determined mixtures

P. Comon and L. De Lathauwer

underdet-chap

As in the previous chapters, we consider the linear statistical model below

$$\mathbf{x} = \mathbf{A}\,\mathbf{s} + \mathbf{b} \tag{1.1}$$

obsmod-eq

where $\mathbf{x}$ denotes the P-dimensional vector of observations, $\mathbf{s}$ the N-dimensional source vector, $\mathbf{A}$ the $P \times N$ mixing matrix and $\mathbf{b}$ an additive noise, which stands for background noise as well as modeling errors. Matrix $\mathbf{A}$ is unknown deterministic, whereas $\mathbf{s}$ and $\mathbf{b}$ are random and also unobserved. All quantities involved are assumed to take their values in the real or complex field. It is assumed that components s_n of vector $\mathbf{s}$ are statistically mutually independent, and that random vectors $\mathbf{b}$ and $\mathbf{s}$ are statistically independent.

The particularity of this chapter is that the number of sources, N, is assumed to be strictly larger than the number of sensors, P. Even if the mixing matrix were known, it would in general be quite difficult to recover the sources. In fact the mixing matrix does not admit a left inverse, because the linear system is *under-determined*, which means that it has more unknowns than equations. The goal is to identify the mixing matrix $\mathbf{A}$ from the sole observation of realizations of vector $\mathbf{x}$. The recovery of sources themselves is not addressed in the present chapter.

Note that other approaches exist that do not assume statistical independence among sources s_n. One can mention non negativity of sources and mixture (see [50] and Chapter **??**), finite alphabet (see [19] and Chapters **??** and **??**) with possibly a sparsity assumption on source values (see [41, 35] and Chapter **??**).

This chapter is organized as follows. General assumptions are stated in Section 1.1. Necessary conditions under which the identification problem is

well posed are pointed out in Section 1.2. Various ways of posing the problem in mathematical terms are described in Section 1.3. Then tensor tools are introduced in Section 1.4 in order to describe numerical algorithms in Section 1.5.

1.1 Observation model

obs-sec

In (1.1), mainly three cases can be envisaged concerning the noise:

10h1 **H1.1.** First, noise $\mathbf{b}$ may be assumed to be Gaussian. If one admits that noise is made of a large number of independent contributions, invoking the central limit theorem justifies this assumption.

10h2 **H1.2.** Second, it may be assumed to have independent components b_p. Or, more generally, it may be assumed to derive linearly from such a noise, so that it can be written as $\mathbf{b} = \mathbf{A}_2\,\mathbf{v}$, for some unknown matrix $\mathbf{A}_2$ and some random vector $\mathbf{v}$ with independent components.

10h3 **H1.3.** Third, the noise may not satisfy the assumptions above, in which case it is assumed to be of small variance.

Under hypotheses H1.1 or H1.2, (1.1) can be rewritten as a *noiseless* model. In fact, we have the following:

$$\mathbf{x} = [\mathbf{A},\ \mathbf{A}_2] \begin{pmatrix} \mathbf{s} \\ \mathbf{v} \end{pmatrix} \tag{1.2}$$

obsmod2-eq

where the random vector in the right hand side may be viewed as another source vector with statistically independent components. The price to pay is an increase in the number of sources.

On the other hand, models (1.1) or (1.2) will be approximations under hypothesis H1.3. We shall subsequently see that this leads to two different problems: the noiseless case corresponds to an exact fit of statistics such as cumulants, whereas the latter leads to an approximate fit.

In the algorithms developed in this chapter, we shall be mainly concerned by hypothesis H1.3, which is more realistic. However, identifiability results are known under hypothesis H1.2.

1.2 Intrinsic identifiability

identif-sec

Before we look at the identification problem, it is useful to examine the identifiability conditions that are inherent in the problem. It may happen that the actual necessary conditions that need to be satisfied in order to identify the mixing matrix are algorithm dependent, and eventually significantly stronger.

Linear mixtures of independent random variables have been studied for years in Statistics [38] [44], and the oldest result is probably due to Dugué (1951),

Darmois (1953) and Skitovich (1954). However, the latter results concern mainly identifiability and uniqueness, and were not constructive, in the sense that no numerical algorithm could be built from their proofs.

1.2.1 Equivalent representations

Before addressing the general case, it is convenient to have a look at the case where mixing matrix $\mathbf{A}$ has two (or more) collinear columns. Without restricting the generality, assume the Q first columns of $\mathbf{A}$, denoted $\mathbf{a}^{(q)}$, are collinear to the first one, that is:

$$\mathbf{a}^{(q)} = \alpha_q \, \mathbf{a}^{(1)}, \; 1 \leqslant q \leqslant Q, \; \alpha_1 \stackrel{\text{def}}{=} 1.$$

Then equation (1.1) can obviously be rewritten as

$$\mathbf{x} = \left[\mathbf{a}^{(1)}, \, \mathbf{a}^{(Q+1)}, \ldots \mathbf{a}^{(N)}\right] \cdot \begin{pmatrix} \sum_q \alpha_q \, s_q \\ s_{Q+1} \\ \vdots \\ s_N \end{pmatrix} + \mathbf{b}.$$

We end up with a linear statistical model similar to (1.1), but of smaller size, $N - Q + 1$, which satisfies the same independence assumption. It is clear that identifying the α_q is not possible without resorting to additional assumptions. Even if techniques do exist to solve this problem, they are out of the scope of this chapter. With the hypotheses we have assumed, only the direction of vector $\mathbf{a}^{(1)}$ can be estimated. Hence from now on, we shall assume that

10h4 **H1.4.** no columns of matrix $\mathbf{A}$ are collinear.

Now assume hypotheses H1.1 and H1.4, and let $\mathbf{y}$ admit two noiseless representations

$$\mathbf{y} = \mathbf{A}\,\mathbf{s} \;\text{ and }\; \mathbf{y} = \mathbf{B}\,\mathbf{z}$$

where components of $\mathbf{s}$ (resp. $\mathbf{z}$) have statistically independent components, and $\mathbf{A}$ (resp. $\mathbf{B}$) have pairwise noncollinear columns. Then we introduce the definition below [44]:

DEFINITION 1.1 *Two representations* $(\mathbf{A}, \mathbf{s})$ *and* $(\mathbf{B}, \mathbf{z})$ *are* equivalent *if every column of* $\mathbf{A}$ *is proportional to some column of* $\mathbf{B}$*, and vice versa.*

If all representations of $\mathbf{y}$ *are equivalent, they are said to be* essentially unique*, that is, they are equal up to permutation and scaling.*

1.2.2 Main theorem

Then, we have the following identifiability theorem [44]:

identif-th

THEOREM 1.1 (IDENTIFIABILITY) *Let* $\mathbf{y}$ *be a random vector of the form* $\mathbf{y} = \mathbf{A}\,\mathbf{s}$*, where* s_p *are independent, and* $\mathbf{A}$ *does not have any collinear columns. Then* $\mathbf{y}$ *can be represented as* $\mathbf{y} = \mathbf{A}_1\,\mathbf{s}_1 + \mathbf{A}_2\,\mathbf{s}_2$*, where* $\mathbf{s}_1$ *is non Gaussian,* $\mathbf{s}_2$ *is Gaussian independent of* $\mathbf{s}_1$*, and* $\mathbf{A}_1$ *is essentially unique.*

This theorem is quite difficult to prove, and we refer the readers to [44] for further readings. However, we shall give a proof in the case of dimension $P = 2$ in the next section page 12.

Remark 1.1 If $\mathbf{s}_2$ is 1-dimensional, then $\mathbf{A}_2$ is also essentially unique, because it has a single column.

Remark 1.2 Note that Theorem 1.1 does not tell anything about the uniqueness of the source vector itself. It turns out that if, in addition, the columns of $\mathbf{A}_1$ are linearly independent, then the distribution of $\mathbf{s}_1$ is unique up to scale and location indeterminacies [44]. But in our framework, the number of sources exceeds the number of sensors so that this condition cannot be fulfilled. We just give two examples below in order to make this issue clearer.

uniqueness-ex

Example 1.1 (Uniqueness) Let s_i be independent with no Gaussian component, and b_i be independent Gaussian. Then the linear model below is identifiable, but $\mathbf{A}_2$ is not essentially unique whereas $\mathbf{A}_1$ is:

$$\begin{pmatrix} s_1 + s_2 + 2\,b_1 \\ s_1 + 2\,b_2 \end{pmatrix} = \mathbf{A}_1\,\mathbf{s} + \mathbf{A}_2 \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{A}_1\,\mathbf{s} + \mathbf{A}_3 \begin{pmatrix} b_1 + b_2 \\ b_1 - b_2 \end{pmatrix}$$

with

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},\ \mathbf{A}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } \mathbf{A}_3 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Hence the distribution of $\mathbf{s}$ is essentially unique. But $(\mathbf{A}_1,\ \mathbf{A}_2)$ is not equivalent to $(\mathbf{A}_1,\ \mathbf{A}_3)$.

nonuniqueness-ex

Example 1.2 (Non uniqueness) Let s_i be independent with no Gaussian component, and b_i be independent Gaussian. Then the linear model below is identifiable, but the distribution of $\mathbf{s}$ is not unique [36]:

$$\begin{pmatrix} s_1 + s_3 + s_4 + 2\,b_1 \\ s_2 + s_3 - s_4 + 2\,b_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1 \\ s_2 \\ s_3 + b_1 + b_2 \\ s_4 + b_1 - b_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1 + 2\,b_1 \\ s_2 + 2\,b_2 \\ s_3 \\ s_4 \end{pmatrix}$$

with

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}.$$

Further details may be found in [36], [8] and [44, ch.10]. In particular, it is pointed out in [36] that the source distributions can be obtained in a unique fashion, even when sources cannot be extracted, provided various sufficient conditions are satisfied.

1.2.3 Core equation

defCAF-sec

The first characteristic function of a real random variable is defined as the conjugated Fourier transform of its probability distribution, that is, for a real random variable x with distribution dF_x, it takes the form: $\Phi_x(t) = \int_u e^{\jmath\, tu}\, dF_x(u)$, which is nothing else but $\mathbb{E}\, e^{\jmath\, tx}$. If the random variable is multi-dimensional, the Fourier transform is taken on all variables, leading to:

$$\Phi_{\mathbf{x}}(\mathbf{t}) \stackrel{\text{def}}{=} \mathbb{E}\, e^{\jmath\, \mathbf{t}^{\mathrm{T}}\mathbf{x}} = \int_{\mathbf{u}} e^{\jmath\, \mathbf{t}^{\mathrm{T}}\mathbf{u}}\, dF_{\mathbf{x}}(\mathbf{u}). \tag{1.3}$$

defcf-eq

The second characteristic function is defined as the logarithm of the first one: $\Psi_x(\mathbf{t}) \stackrel{\text{def}}{=} \log \Phi_x(\mathbf{t})$. It always exists in the neighborhood of the origin, and is hence uniquely defined as long as $\Phi_x(\mathbf{t}) \neq 0$. An important property of the second characteristic function is given by the *Marcinkiewicz* theorem that we recall below without proof:

Marcinkiewicz-th

THEOREM 1.2 (MARCINKIEWICZ, 1938) *If a second characteristic function $\Psi_x(t)$ is a polynomial, then its degree is at most 2 and x is Gaussian.*

Another basic property will be useful for our further developments. If x and y are two statistically independent random variables, then the joint second c.f. splits into the sum of the marginals:

$$\Psi_{x,y}(u, v) = \Psi_x(u) + \Psi_y(v). \tag{1.4}$$

sumcf-eq

This property can be easily proved by direct use of the definition $\Psi_{x,y}(u, v) = \log[\mathbb{E} \exp \jmath(ux + vy)]$, which yields $\Psi_{x,y}(u, v) = \log[\mathbb{E} \exp(\jmath\, ux)\ \mathbb{E} \exp(\jmath\, vy)]$ as soon as x and y are independent.

Denote by $\Psi_x(\mathbf{u})$ the joint second characteristic function of the observed random vector $\mathbf{x}$, and $\psi_p(v_p)$ the marginal second characteristic function of source s_p. Then, the following core equation is established:

PROPOSITION 1.3 *If s_p are mutually statistically independent random variables, and if $\mathbf{x} = \mathbf{A}\,\mathbf{s}$, then we have the* core equation*:*

$$\Psi_x(\mathbf{u}) = \sum_p \Psi_{s_p}\left(\sum_q u_q\, A_{qp}\right) \tag{1.5}$$

core-eq

for any $\mathbf{u}$ in a neighborhood of the origin.

Proof: First notice that, by definition (1.3) of the characteristic function, we have $\Psi_x(\mathbf{u}) = \Psi_{As}(\mathbf{u}) = \Psi_s(\mathbf{A}^{\mathrm{T}}\mathbf{u})$. Then the remainder of the proof is an immediate consequence of property (1.4). To see this, just notice that because s_p are independent, $\Psi_s(\mathbf{v}) = \sum_p \Psi_{s_p}(v_p)$. Replacing v_p by its value, *i.e.* the pth row of $\mathbf{A}^{\mathrm{T}}\mathbf{u}$, yields (1.5). □

In (1.5), we see that identifying the mixing matrix $\mathbf{A}$ amounts to decomposing the multivariate joint characteristic function of $\mathbf{x}$ into a sum of univariate functions, or in other words, to finding the linear combinations entering latter univariate functions. Various approach to this problem will be surveyed in Section 1.3.

1.2.4 Identifiability in the 2-dimensional case

We are now in a position to state the proof of the identifiability theorem 1.1 in the case where $P = 2$.

Darmois-th

THEOREM 1.4 (DARMOIS-SKITOVIC) *Let s_i be statistically* independent *random variables, and two linear statistics:*

$$y_1 = \sum_i a_i\, s_i \quad \textit{and} \quad y_2 = \sum_i b_i\, s_i.$$

If y_1 and y_2 are statistically independent, then random variables s_k for which $a_k\, b_k \neq 0$ are Gaussian.

It does not restrict the generality to assume that column vectors $[a_k,\ b_k]^{\mathrm{T}}$ are not collinear, as pointed out earlier in this section. This is equivalent to saying that (1) one can group variables that are mixed with collinear columns, and (2) one makes a change of variable by summing the corresponding sources together. Note however that if variables are grouped together, say $s_p + s_q$, they need to be both non Gaussian, or both Gaussian. In the latter case we may apply a theorem on the infinite decomposability of the normal law to conclude that s_p and s_q are both Gaussian.

Proof: To simplify the proof and make it more readable, we also assume that ψ_p are differentiable, which is not a necessary assumption. Assume the notations below for the characteristic functions involved:

$$\begin{aligned}
\Psi_{1,2}(u,v) &= \log \mathbb{E} \exp(\jmath\, y_1\, u + \jmath\, y_2\, v) \\
\Psi_k(w) &= \log \mathbb{E} \exp(\jmath\, y_k\, w),\ k \in \{1,1\} \\
\psi_p(w) &= \log \mathbb{E} \exp(\jmath\, s_p\, w),\ p \in \{1,\ldots,N\}.
\end{aligned}$$

From (1.4), the statistical independence between s_p's implies:

$$\begin{aligned}
\Psi_{1,2}(u,v) &= \textstyle\sum_{k=1}^N \psi_k(u\, a_k + v\, b_k) \\
\Psi_1(u) &= \textstyle\sum_{k=1}^N \psi_k(u\, a_k) \\
\Psi_2(v) &= \textstyle\sum_{k=1}^N \psi_k(v\, b_k)
\end{aligned}$$

which are in fact core equations similar to (1.5), whereas statistical independence between y_1 and y_2 implies

$$\Psi_{1,2}(u,v) = \Psi_1(u) + \Psi_2(v).$$

Hence we have $\sum_{k=1}^{N} \psi_p(u\,a_k + v\,b_k) = \sum_{k=1}^{N} \psi_k(u\,a_k) + \psi_k(v\,b_k)$. The equality is trivial for terms for which $a_k b_k = 0$. So from now on, one can restrict the sum to terms corresponding to $a_k b_k \neq 0$. Since the equations hold true for any (u, v), write this equation at $u + \alpha/a_N$ and $v - \alpha/b_N$ for an arbitrary α:

$$\sum_{k=1}^{N} \psi_k \left(u\,a_k + v\,b_k + \alpha(\frac{a_k}{a_N} - \frac{b_k}{b_N}) \right) = f(u) + g(v).$$

Now perform subtraction so as to cancel the Nth term, divide the result by α, and let $\alpha \to 0$; we obtain:

$$\sum_{k=1}^{N-1} (\frac{a_k}{a_N} - \frac{b_k}{b_N})\, \psi_k^{(1)}(u\,a_k + v\,b_k) = f^{(1)}(u) + g^{(1)}(v)$$

for some *univariate functions* $f^{(1)}(u)$ and $g^{(1)}(u)$.

Hence, we have now a similar expression, but with one term less than before in the sum. The idea is then to repeat the procedure $(N-1)$ times in order to eventually get:

$$\prod_{j=2}^{N} (\frac{a_1}{a_j} - \frac{b_1}{b_j})\, \psi_1^{(N-1)}(u\,a_1 + v\,b_1) = f^{(N-1)}(u) + g^{(N-1)}(v).$$

As a consequence, $\psi_1^{(N-1)}(u\,a_1 + v\,b_1)$ is linear, as a sum of two univariate functions (in fact $\psi_1^{(N)}$ is a constant because $a_1 b_1 \neq 0$). By succesive integrations, we eventually see that ψ_1 is a polynomial. Lastly invoke Theorem 1.2 to conclude that s_1 is Gaussian.

Now, the reasoning we have made for s_1 holds valid for any s_p. By repeating the proof for any ψ_p such that $a_p b_p \neq 0$, we would also prove that s_p is Gaussian. This concludes the proof. □

Remark 1.3 The proof found in the literature is given when ψ_p are not all differentiable, but is more complicated [44, 38]. The proof can also be extended to infinitely many variables.

Remark 1.4 The proof was derived above for real variables, but it also holds true in the complex field. Some other interesting remarks concerning specificities of the complex field may be found in [37]. Some issues related to complex variables will be stated in the next sections.

1.3 Problem formulation

formul-sec

In equation (1.5), $\Psi_x(\mathbf{u})$ is written as a sum of contributions of the individual sources, i.e., the sources have been separated. This remains true when we apply a linear transformation to both sides of the equation. One could for instance compute derivatives at the origin. This leads to cumulant-based methods, see Section 1.3.2. Specific algorithms are discussed in Sections 1.5.2.2, 1.5.4 and 1.5.5. One could also compute derivatives in different points than the origin, see Section 1.3.1. Specific algorithms are discussed in Sections 1.5.2.3 and 1.5.6.

1.3.1 Approach based on derivatives of the joint characteristic function

sect:charderfit

The goal is hence to produce simple equations from the core equation (1.5), whose coefficients can be estimated from realizations of observation $\mathbf{x}$.

The idea discussed in this section has its roots in the proof of the theorem of Darmois-Skitovic [44, 38]. It has been first proposed in [71] and further developed in [25]. A variant has been proposed in [76], see also Section 1.5.3.

Consider the core equation (1.5) for values $\mathbf{u}$ belonging to some finite set $\mathcal{G}$ of cardinality L. Assume source characteristic functions ψ_p all admit finite derivatives up to order r in a neighborhood of the origin containing $\mathcal{G}$. Then, taking $r = 3$ as a working example [25]:

$$\frac{\partial^3 \Psi_x}{\partial u_i \partial u_j \partial u_k}(\mathbf{u}) = \sum_{p=1}^{N} A_{ip}\, A_{jp}\, A_{kp}\;\; \psi_p^{(3)}(\sum_{q=1}^{P} u_q\, A_{qp}). \tag{1.6}$$

CAF-eq

Now denote $\mathbf{u}(\ell)$ the L points of the grid $\mathcal{G}$, $1 \leqslant \ell \leqslant L$, and define $B_{\ell p} \stackrel{\text{def}}{=} \psi_p^{(3)}(\sum_{q=1}^{P} u_q(\ell)\, A_{qp})$. The array of 3rd order derivatives may be estimated from the observations of $\mathbf{x}$. Hence the problem has now been translated into the following: given an array $T_{ijk\ell}$, find two matrices $\mathbf{A}$ and $\mathbf{B}$ such that

$$T_{ijk\ell} \approx \sum_{p} A_{ip}\, A_{jp}\, A_{kp}\, B_{\ell p}. \tag{1.7}$$

CAF2-eq

In Sections 1.5.2.3 and 1.5.6, algorithms will be proposed to perform this matching, ignoring the dependence of $\mathbf{B}$ on $\mathbf{A}$.

Up to now, random variables have been assumed to take their values in the real field. However, in a number of applications, *e.g.*, telecommunications, identification problems are posed in the complex field. For complex random variables, one can work with proper generalizations of the characteristic functions.

1.3.2 Approach based on cumulants

sect:cumfit

In this section, we shall show how the blind identification problem can be seen, in a first approximation, as a cumulant matching problem, which will allow us to solve it in different ways in Sections 1.5.2.2, 1.5.4 and 1.5.5.

1.3.2.1 Definitions

If Φ can be expanded in Taylor series about the origin, then its coefficients are related to *moments*:

$$\mu'^{x}_{(r)} \stackrel{\text{def}}{=} \mathbb{E}\, X^r = (-\jmath)^r \left.\frac{\partial^r \Phi(t)}{\partial t^r}\right|_{t=0}. \tag{1.8}$$

PhiExpand-eq

It is usual to introduce *central moments* $\mu^x_{(r)}$ as the moments of the centered variable $x - \mu'^{x}_{(1)}$.

Similarly, if Ψ may be expanded in Taylor series about the origin, then its coefficients are the *cumulants*:

$$\kappa^x_{(r)} \stackrel{\text{def}}{=} \text{cum}\{\underbrace{X, X, \ldots, X}_{r \text{ times}}\} = (-\jmath)^r \left.\frac{\partial^r \Psi(t)}{\partial t^r}\right|_{t=0}. \tag{1.9}$$

PsiExpand-eq

The relation between moments and cumulants can be obtained by expanding the logarithm and grouping terms of same order together.

Example 1.3 Cumulants of order 2, 3 and 4 The cumulant of 2nd order, κ_2, is nothing else but the variance: $\mu'_{(2)} - {\mu'_{(1)}}^2 = \kappa_{(2)}$. And for zero-mean random variables, cumulants of order 3 and 4 are related to moments by: $\kappa_3 = \mu_3$ and $\kappa_4 = \mu_4 - 3\,{\mu_2}^2$.

Example 1.4 Skewness and Kurtosis The *skewness* is a 3rd order normalized cumulant: $\mathcal{K}_3 \stackrel{\text{def}}{=} \kappa_3/{\kappa_2}^{3/2}$. The *kurtosis* is a normalized 4th order cumulant $\mathcal{K}_4 \stackrel{\text{def}}{=} \kappa_4/{\kappa_2}^2$.

Skewness and kurtosis are null for any Gaussian random variable. These quantities can serve as measures of deviation from Gaussianity. In fact, random variables having a negative (resp. positive) kurtosis can be called *platykurtic* (resp. *leptokurtic*) [45]. Conversely, random variables having zero kurtosis (referred to as *mesokurtic*) are not necessarily Gaussian.

For multivariate random variables, denote the cumulants

$$\kappa_{ij..\ell} = \text{cum}\{X_i,\, X_j, ...X_\ell\}.$$

As explained above, expressions of moments as a function of cumulants can be obtained by expanding the logarithm in the definition of the second characteristic function and grouping terms of same order together. This yields for instance:

$$\begin{aligned} \mu'_i &= \kappa_i, \\ \mu'_{ij} &= \kappa_{ij} + \kappa_i\,\kappa_j, \\ \mu'_{ijk} &= \kappa_{ijk} + [3]\,\kappa_i\,\kappa_{jk} + \kappa_i\,\kappa_j\,\kappa_k. \end{aligned} \tag{1.10}$$

In the relation above, we have used McCullagh's *bracket notation* [54] defined below.

Bracket notation A sum of k terms that can be deduced from each other by permutation of indices is denoted by the number k between brackets followed by a single monomial describing the generic term. This is McCullagh's *bracket notation* [54].

Example 1.5 Simple examples will do it better than a long explanation.

$$[3]\,\delta_{ij}\,\delta_{kl} = \delta_{ij}\,\delta_{kl} + \delta_{ik}\,\delta_{jl} + \delta_{il}\,\delta_{jk}, \tag{1.11}$$

$$[3]\,a_{ij}\,b_k\,c_{ijk} = a_{ij}\,b_k\,c_{ijk} + a_{ik}\,b_j\,c_{ijk} + a_{jk}\,b_i\,c_{ijk}. \tag{1.12}$$

The presence of the bracket yields an implicit summation; all terms with r indices are completely symmetric order-r tensors. The number of distinct monomials that may be obtained by permutation is equal to the integer appearing between brackets. As additional examples, the following expressions are consistent:

$$[3]\, a_i \delta_{jk}, \quad [6]\, a_i\, a_j\, \delta_{kl}, \quad [10]\, b_i\, b_j\, b_k\, \delta_{lm}, \quad [35]\, A_{ijk}\, B_{abcd}\, C_{ijkabcd}.$$

1.3.2.2 Relations between moments and cumulants

Relations (1.10) can be inverted in order to obtain cumulants as a function of moments [45] [54]. In the case of non central random variables, *multivariate cumulants* of order 3 and 4 can be given in a compact form as a function of *multivariate moments* as:

$$\begin{aligned}
\kappa_{ij} &= \mu'_{ij} - \mu'_i\, \mu'_j, && (1.13)\\
\kappa_{ijk} &= \mu'_{ijk} - [3]\, \mu'_i\, \mu'_{jk} + 2\mu'_i\, \mu'_j\, \mu'_k, && (1.14)\\
\kappa_{ijkl} &= \mu'_{ijkl} - [4]\, \mu'_i\, \mu'_{jkl} - [3]\mu'_{ij}\, \mu'_{kl} \\
&+ 2\,[6]\, \mu'_i\, \mu'_j\, \mu'_{kl} - 6\mu'_i\, \mu'_j\, \mu'_k\, \mu'_l. && (1.15)
\end{aligned}$$

On the other hand, if variables are all zero-mean, then the expressions simplify to:

$$\begin{aligned}
\kappa_{ij} &= \mu_{ij}, && (1.16)\\
\kappa_{ijk} &= \mu_{ijk}, && (1.17)\\
\kappa_{ijkl} &= \mu_{ijkl} - [3]\, \mu_{ij}\, \mu_{kl}. && (1.18)
\end{aligned}$$

At order 5 and 6 we have:

$$\begin{aligned}
\kappa_{ijklm} &= \mu_{ijklm} - [10]\, \mu_{ij}\, \mu_{klm}, \\
\kappa_{ijklmn} &= \mu_{ijklmn} - [15]\, \mu_{ij}\, \mu_{klmn} \\
&- [10]\, \mu_{ijk}\, \mu_{lmn} + 2\,[15]\, \mu_{ij}\, \mu_{kl}\, \mu_{mn}.
\end{aligned}$$

1.3.2.3 Properties of cumulants

Cumulants enjoy several useful properties [55]. Some of them are shared by moments, but others are not. See also Chapter **??**.

First of all, moments and cumulants enjoy the so-called *multi-linearity property*:

multilinearity-prty

PROPERTY 1.5 *If two random variables are related by a linear transform,* $\mathbf{y} = \mathbf{A}\,\mathbf{x}$*, then their cumulants are related multi-linearly:*

$$\mathrm{cum}\{y_i, y_j, \ldots y_k\} = \sum_{p,q,\ldots r} A_{ip}\, A_{jq}\, \ldots A_{kr}\, \mathrm{cum}\{x_p, x_q, \ldots x_r\}. \qquad (1.19)$$

multilinearity-eq

A similar property holds for moments.

This property actually makes that we can call moments and cumulants *tensors* [54]. As a corollary, we have the following:

$$\begin{aligned} \mathrm{cum}\{\alpha x_1,\, x_2, .., x_n\} &= \alpha\, \mathrm{cum}\{x_1,\, x_2, .., x_n\}, \\ \mathrm{cum}\{x_1 + y_1,\, x_2, .., x_n\} &= \mathrm{cum}\{x_1,\, x_2, .., x_n\} + \mathrm{cum}\{y_2,\, x_2, .., x_n\}. \end{aligned} \tag{1.20}$$

Another obvious property directly results from the definition, namely that of invariance by permutation of indices:

$$\mathrm{cum}\{X_1,\, X_2, ..X_r\} = \mathrm{cum}\{X_{\sigma(1)},\, X_{\sigma(2)}, ..X_{\sigma(r)}\}.$$

In other words, rth order cumulants (and moments) are fully *symmetric* rth order tensors.

Let's now turn to properties that are specific to cumulants. First, they are invariant with respect to translation; this means that $\forall r > 1$ and $\forall h$ constant:

$$\mathrm{cum}\{X_1 + h,\, X_2, .., X_r\} = \mathrm{cum}\{X_1,\, X_2, .., X_r\}.$$

This property is sometimes referred to as the *shift invariance of cumulants*. Next, cumulants of a set of random variables are null as soon as this set can be split into two statistically independent subsets:

$$\{X_1, \ldots,\, X_p\} \text{ independent of } \{Y_1, \ldots,\, Y_q\} \Rightarrow \mathrm{cum}\{X_1, \ldots,\, X_p,\, Y_1, \ldots,\, Y_q\} = 0.$$

A consequence of this property is the *additivity of cumulants*.

additivity-prty

PROPERTY 1.6 *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be (possibly multivariate) statistically independent random variables. Then*

$$\begin{aligned} \mathrm{cum}\{X_1 + Y_1,\, X_2 + Y_2, .., X_r + Y_r\} &= \mathrm{cum}\{X_1,\, X_2, .., X_r\} \\ &+ \mathrm{cum}\{Y_1,\, Y_2, .., Y_r\}. \end{aligned} \tag{1.21}$$

cumCmplx-sec

1.3.2.4 Cumulants of complex random variables

Up to now, random variables have been assumed to take their values in the real field. Moments and cumulants of complex random variables can be defined starting from proper generalizations of the characteristic functions. Contrary to real variables, there is not a unique way of defining a cumulant (or a moment) of order r of a complex random variable; in fact, it depends on the number of conjugated terms. It is thus necessary to be able to distinguish between complex random variables that are conjugated and those that are not. For this purpose, one introduces a specific notation, with superscripts:

$$\mathrm{cum}\{X_i, ..X_j, X_k^*, ..X_\ell^*\} \stackrel{\mathrm{def}}{=} \mathcal{C}_{i..j}^{k..\ell}. \tag{1.22}$$

cumCdef-eq

Example 1.6 The covariance matrix of a complex random variable is

$$\mathbb{E}\, z_i\, z_j^* - \mathbb{E}\, z_i\, \mathbb{E}\, {z_j}^* = \kappa_i^{\mathbf{z}\, j}.$$

Among all the possible definitions, only one is called *circular cumulant*, namely the one having exactly half of its arguments conjugated. All other cumulants may be called *non circular* cumulants. Note that there exist circular cumulants only at even orders. For instance, the cumulant below is circular

$$\kappa_{ij}^{\mathbf{z}\,k\ell} = \mathrm{cum}\{z_i,\, z_j,\, z_k^*,\, z_\ell^*\}.$$

Exact expressions of cumulants of multivariate complex random variables are given in the Appendix.

1.3.2.5 Blind identification via cumulant matching

cummat-sec

Combining the model (1.1) with the multi-linearity (1.19) and additivity (1.21) properties of cumulants leads to a system of equations, which we shall solve in Section 1.5. In order to fix the ideas, consider first cumulants of order 4 of the observed random vector $\mathbf{x}$:

$$\kappa_{i\ell}^{\mathbf{x}\,jk} = \sum_{n=1}^{N} A_{in}\, A_{jn}^*\, A_{kn}^*\, A_{\ell n}\, \kappa_{nn}^{\mathbf{s}\,nn}, \tag{1.23}$$

$$\kappa_{ijk}^{\mathbf{x}\,\ell} = \sum_{n=1}^{N} A_{in}\, A_{jn}\, A_{kn}\, A_{\ell n}^*\, \kappa_{nnn}^{\mathbf{s}\,n}, \tag{1.24}$$

$$\kappa_{ijk\ell}^{\mathbf{x}} = \sum_{n=1}^{N} A_{in}\, A_{jn}\, A_{kn}\, A_{\ell n}\, \kappa_{nnnn}^{\mathbf{s}}. \tag{1.25}$$

In the equations above, the notation introduced in (1.22) has been used.

In general, the use of the circular cumulant $\kappa_{ij}^{\mathbf{x}\,k\ell}$ is preferred. But cumulant $\kappa_{ijk\ell}^{\mathbf{x}}$ has also been successfully used in the identification of communication channels [18] [30] [31] [40]. On the other hand, cumulant $\kappa_{ijl}^{\mathbf{x}\,\ell}$ is rarely used because it is generally close to zero. The choice of the cumulants to use depends on the *a priori* we have on the source cumulants.

In Section 1.5.5, the circular cumulant of order 6,

$$\kappa_{ijn}^{\mathbf{x}\,k\ell m} \stackrel{\mathrm{def}}{=} \mathrm{cum}\{x_i, x_j, x_k^*, x_\ell^*, x_m^*, x_n\},$$

will be also used. Because again of properties (1.19) and (1.21), the equation to solve for $\mathbf{A}$ is then the following, for all 6-uplet of indices:

$$\kappa_{ijp}^{\mathbf{x}\,k\ell m} = \sum_{n=1}^{N} A_{in}\, A_{jn}\, A_{kn}^*\, A_{\ell n}^*\, A_{mn}^*\, A_{pn}\, \kappa_{nnn}^{\mathbf{s}\,nnn}. \tag{1.26}$$

6thCum-eq

1.4 Higher-order tensors

tensors-sec

For our purposes, a tensor may be assimilated to its array of coordinates. Justifications may be found in [22, sec. 2.1]; see also [26, sec. 2.1]. Hence we shall

not make the distinction in the remainder. A tensor of order r defined on the tensor product of r vector spaces of dimension N_i, $1 \leqslant i \leqslant r$, will be represented by an array of numbers, of dimensions $N_1 \times N_2 \times \cdots \times N_r$.

1.4.1 Canonical tensor decomposition

1.4.1.1 Definition

rank1-def DEFINITION 1.2 *If a rth order tensor $[\![T_{ij\ldots\ell}]\!]$ can be written as an outer product $T_{ij\ldots\ell} = u_i\, v_j \ldots w_\ell$, then it is called a rank-1 tensor.*

In a compact form, one shall write rank-1 tensors as $\mathbf{T} = \mathbf{u} \otimes \mathbf{v} \otimes \ldots \otimes \mathbf{w}$, where $\otimes$ denotes the outer (tensor) product.

cand-def DEFINITION 1.3 (TENSOR RANK) *A tensor $[\![T_{ij\ldots\ell}]\!]$ can always be decomposed into a sum of rank-1 tensors as*

$$\mathbf{T} = \sum_{n=1}^{r} \mathbf{u}_n \otimes \mathbf{v}_n \otimes \ldots \otimes \mathbf{w}_n. \tag{1.27}$$ cand-eq

The minimal value of r for which the equality holds is called the rank of tensor $\mathbf{T}$.

For order-2 tensors, which are actually matrices, this definition coincides with the usual definition of matrix rank.

When r equals rank$\{\mathbf{T}\}$, the decomposition given in (1.27) is often referred to as the *Canonical Decomposition* (CANDECOMP) of tensor $\mathbf{T}$ [13]. Other names appear in the literature, such as PARAFAC in psychometrics and chemometrics [42] [47] [49] [66], or POLYADIC FORM [43] in mathematical physics. In linear algebra, the acronym CP is now often used, and stands for CANDECOMP-PARAFAC; see [26] and references therein.

An alternative representation of (1.27) is obtained by imposing each vector to be of unit modulus, and by inserting a scale factor λ_p in the decomposition, which yields:

$$\mathbf{T} = \sum_{n=1}^{r} \lambda_n\ \mathbf{u}_n \otimes \mathbf{v}_n \otimes \ldots \otimes \mathbf{w}_n. \tag{1.28}$$ normalizedCanD-eq

1.4.1.2 Symmetry

A tensor is *symmetric* if its entries do not change when permuting the indices. The terminology of "supersymmetry" should be avoided [21]. When decomposing a symmetric tensor, it may be relevant to impose all vectors in each outer product of (1.28) to be the same. This leads to a CANDECOMP of the form

$$\mathbf{T} = \sum_{n=1}^{r} \lambda_n\ \mathbf{u}_n \otimes \mathbf{u}_n \otimes \ldots \otimes \mathbf{u}_n. \tag{1.29}$$ symRank-eq

Up to now, it has not yet been proved that the rank defined above under the symmetry constraint is the same as that defined in (1.28). Hence, the rank defined above is called the *symmetric rank* of $\mathbf{T}$ [21]. In this chapter, by the rank of a structured tensor it will be always meant the structured rank of that tensor. This definition applies to both real or complex *symmetric tensors*.

In the complex field, Hermitian symmetry is also quite important, and often more useful than plain symmetry. Similar definitions can be introduced for various symmetry properties. For instance, for the 4th order cumulant tensor defined in (1.23), the rank is the minimal integer r such that the following CanD holds:

$$\mathbf{T} = \sum_{n=1}^{r} \lambda_n \, \mathbf{u}_n \otimes \mathbf{u}_n^* \otimes \mathbf{u}_n^* \otimes \mathbf{u}_n \tag{1.30}$$ CanD-4th-eq

and similarly for the 6th order cumulant tensor (1.26),

$$\mathbf{T} = \sum_{n=1}^{r} \lambda_n \, \mathbf{u}_n \otimes \mathbf{u}_n \otimes \mathbf{u}_n^* \otimes \mathbf{u}_n^* \otimes \mathbf{u}_n^* \otimes \mathbf{u}_n. \tag{1.31}$$ CanD-6th-eq

In the CanD above, vectors $\mathbf{u}_n$ are wished to be equal to some normalized column of matrix $\mathbf{A}$ up to a scaling factor.

1.4.1.3 Link with homogeneous polynomials

The linear space of symmetric tensors of order d and dimension P can be bijectively mapped to the space of homogeneous polynomials of degree d in P variables. This property will be useful in Section 1.5.2.1. As pointed out in [17] [54], there exist two different notations in the literature; we recall them below and relate them.

Let $\mathbf{x}$ be an array of unknowns of size P, and $\mathbf{j}$ a multi-index of the same size. One can assume the notation $\mathbf{x}^{\mathbf{j}} \stackrel{\text{def}}{=} \prod_{k=1}^{P} x_k^{j_k}$ and $|\mathbf{j}| \stackrel{\text{def}}{=} \sum_k j_k$. Then for homogeneous monomials of degree d, $\mathbf{x}^{\mathbf{j}}$, we have $|\mathbf{j}| = d$.

Example 1.7 Take the example of cubics in 4 variables to fix the ideas. One can associate every entry T_{ijk} of a 3rd order symmetric tensor with a monomial $T_{ijk}\, x_i x_j x_k$. For instance, T_{114} is associated with $T_{114}\, x_1^2 x_4$, and thus to $T_{114}\, \mathbf{x}^{[2,0,0,1]}$; this means that we have a map $f([1,1,4]) = [2,0,0,1]$.

More generally, the $d-$dimensional vector index $\mathbf{i} = [i,j,k]$ can be associated with a $P-$dimensional vector index $\mathbf{f}(\mathbf{i})$ containing the number of times each variable x_k appears in the associated monomial. Whereas the d entries of $\mathbf{i}$ take their values in $\{1,\ldots,P\}$, the P entries of $\mathbf{f}(\mathbf{i})$ take their values in $\{1,\ldots,d\}$ with the constraint that $\sum_k f_k(\mathbf{i}) = d, \forall \mathbf{i}$.

As a consequence, in order to define the bijection, it suffices to associate every polynomial $p(\mathbf{x})$ with the symmetric tensor $\mathbf{T}$ as:

$$p(\mathbf{x}) = \sum_{|\mathbf{f}(\mathbf{i})|=d} T_{\mathbf{i}}\, \mathbf{x}^{\mathbf{f}(\mathbf{i})} \tag{1.32}$$ bijec-eq

where $T_{\mathbf{i}}$ are the entries of tensor $\mathbf{T}$. The dimension of these spaces is $\binom{P+d-1}{d}$, and one can choose as a basis the set of monomials: $\mathcal{B}(P;d) = \{\mathbf{x}^{\mathbf{j}},\ |\mathbf{j}| = d\}$.

Example 1.8 Let p and q be two homogeneous polynomials in P variables, associated with tensors $\mathbf{P}$ and $\mathbf{Q}$, possibly of different orders. Then, polynomial pq is associated with $\mathbf{P} \otimes \mathbf{Q}$. In fact we have [17]:

$$p(\mathbf{x})\, q(\mathbf{x}) = \sum_{\mathbf{i}} \sum_{\mathbf{j}} P_{\mathbf{i}} Q_{\mathbf{j}} \mathbf{x}^{\mathbf{f}(\mathbf{i})+\mathbf{f}(\mathbf{j})} = \sum_{[\mathbf{i}\,\mathbf{j}]} [\mathbf{P} \otimes \mathbf{Q}]_{[\mathbf{i}\,\mathbf{j}]}\, \mathbf{x}^{\mathbf{f}([\mathbf{i}\,\mathbf{j}])}.$$

uniqueness-sec

1.4.2 Essential uniqueness

Let us, for the sake of convenience, consider the third-order version of (1.27):

$$\mathbf{T} = \sum_{n=1}^{r} \mathbf{u}_n \otimes \mathbf{v}_n \otimes \mathbf{w}_n. \tag{1.33}$$

eq:Candnonnorm

Obviously, the order in which terms enter the sum are not relevant. The consequence is that the triplet of matrices $(\mathbf{U}, \mathbf{V}, \mathbf{W})$ is defined up to a common permutation of columns. Next, if $(\mathbf{U}, \mathbf{V}, \mathbf{W})$ corresponds to a CanD, then so does $(\mathbf{U}\boldsymbol{\Delta}_U, \mathbf{V}\boldsymbol{\Delta}_V, \mathbf{W}\boldsymbol{\Delta}_U^{-1}\boldsymbol{\Delta}_V^{-1})$, for any pair of diagonal invertible matrices $(\boldsymbol{\Delta}_U, \boldsymbol{\Delta}_V)$.

We see that, in the absence of additional assumptions, the best we can do is to calculate one representative of this equivalence class of CanD solutions. In the literature, uniqueness up to scale and permutation is sometimes referred to as *essential uniqueness*. So our first goal will be to identify the conditions under which essential uniqueness is met.

1.4.2.1 Necessary conditions for uniqueness

If tensor $\mathbf{T}$ is of size $I \times J \times K$, and has rank r, then the number of degrees of freedom on the left hand side of (1.33) is IJK, whereas it is equal to $(I + J + K - 2)r$ on the right hand side. So the CanD will not be essentially unique if the number of unknowns is larger than the number of equations, that is, if $IJK \geqslant (I + J + K - 2)r$. More generally, we have the necessary condition:

PROPOSITION 1.7 *A dth order tensor* $\mathbf{T}$ *of dimensions* $P_1 \times P_2 \times \ldots P_d$ *may have an essentially unique CanD only if:*

$$\text{rank}\{\mathbf{T}\} \leqslant \frac{\prod_{k=1}^{d} P_k}{1 - d + \sum_{k=1}^{d} P_k} \stackrel{\text{def}}{=} \rho. \tag{1.34}$$

NecCond-eq

In other words, the rank of $\mathbf{T}$ should not exceed an upper bound ρ. The closest integer larger than ρ, $\lceil \rho \rceil$, is called the *expected rank*. A similar reasoning can be carried out for symmetric tensors, using the normalized CanD expression (1.29), and leads to:

PROPOSITION 1.8 *A dth order symmetric tensor* $\mathbf{T}$ *of dimension* P *may have an essentially unique CanD only if:*

$$\operatorname{rank}\{\mathbf{T}\} \leqslant \frac{\binom{P+d-1}{d}}{P} \stackrel{\text{def}}{=} \rho_s. \tag{1.35}$$

NecCondSym-eq

Two remarks are in order here. The first is that even when $\operatorname{rank}\{\mathbf{T}\} \leqslant \rho$ (equality may occur if ρ is integer), there may be infinitely many solutions. In Section 1.4.2.2 we will present conditions such that there is just one solution, up to permutation and scaling. The second remark is that, for certain dimensions, even in the generic case, more than $\lceil \rho \rceil$ terms might be necessary to form the tensor; in other words, the generic rank might be strictly larger than the expected rank. See [22] [20] for an easily accessible summary of these odd properties.

suf_uniq-sec

1.4.2.2 Sufficient conditions for uniqueness

When studying the arithmetic complexity of the product of two matrices, Kruskal obtained a sufficient condition for the essential uniqueness of the CanD of a 3rd order tensor. This condition involved the notion of *k-rank* of a set of column vectors [48]:

DEFINITION 1.4 (KRUSKAL'S RANK) *A matrix* $\mathbf{A}$ *has k-rank* k_A *if and only if* every subset *of* k_A *columns of* $\mathbf{A}$ *is full column rank, and this does not hold true for* $k_A + 1$*. The k-rank of a matrix* $\mathbf{A}$ *will be denoted by* $\operatorname{krank}\{\mathbf{A}\}$.

Remark 1.5 It is important to distinguish between the k-rank and the rank of a matrix. Remember that in a matrix of rank r, there is *at least* one subset of r linearly independent columns. In a matrix of k-rank k_A, *every* subset of k_A columns is of rank k_A. Note that the $k-$rank is also related to what is sometimes called the kernel of the set of column vectors, or the *spark* of the matrix.

The sufficient condition developed in [48] has been later extended to tensors of arbitrary order [63] [69], and can be stated as follows:

THEOREM 1.9 *A dth order tensor* $\mathbf{T}$ *of dimensions* $P_1 \times P_2 \times \ldots P_d$ *admits an essentially unique CanD, as* $\mathbf{T} = \sum_{n=1}^{\operatorname{rank}\{\mathbf{T}\}} \lambda_n \, \mathbf{a}_n^{(1)} \circledast \mathbf{a}_n^{(2)} \circledast \ldots \mathbf{a}_n^{(d)}$*, if*

$$2\operatorname{rank}\{\mathbf{T}\} + d - 1 \leqslant \sum_{k=1}^{d} \operatorname{krank}\{\mathbf{A}^{(k)}\}. \tag{1.36}$$

KruskalUniqueness-eq

Remark 1.6 For generic $P \times r$ matrices, the k-rank equals $\min(P, r)$. Hence, if matrices $\mathbf{A}^{(k)}$ all have more columns than rows, the sufficient condition (1.36) can be generally simplified to

$$2\operatorname{rank}\{\mathbf{T}\} + d - 1 \leqslant \sum_{k=1}^{d} P_k, \tag{1.37}$$

eq:KruskalUniquenessG

which gives an explicit upper bound on the rank of $\mathbf{T}$.

When we can assume that at least one of the tensor dimensions is "large" (meaning that it is larger than rank$\{\mathbf{T}\}$), a more relaxed sufficient uniqueness condition can be derived. Let us first consider a third-order tensor $\mathbf{T}$ of dimensions $P_1 \times P_2 \times P_3$, with $P_3 \geqslant \text{rank}\{\mathbf{T}\}$. To be able to formulate the condition, we need to introduce the following matrix rank-1 detection tool, which is a variant of the tool introduced in [9]. The proof of the Theorem is given in [27].

prop:mapPhi

THEOREM 1.10 *Consider the mapping* Γ*:* $(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P_1 \times P_2} \times \mathbb{C}^{P_1 \times P_2} \longmapsto \Gamma(\mathbf{X}, \mathbf{Y}) = \mathcal{P} \in \mathbb{C}^{P_1 \times P_2 \times P_1 \times P_2}$ *defined by:*

$$p_{ijkl} = x_{ij}y_{kl} + y_{ij}x_{kl} - x_{il}y_{kj} - y_{il}x_{kj}$$

for all index values. Given $\mathbf{X} \in \mathbb{C}^{P_1 \times P_2}$*,* $\Gamma(\mathbf{X}, \mathbf{X}) = \mathbf{0}$ *if and only if the rank of* $\mathbf{X}$ *is at most one.*

We now have the following sufficient condition for CanD uniqueness [27].

prop:canduniqdetlink1

THEOREM 1.11 *A third order tensor* $\mathbf{T}$ *of dimensions* $P_1 \times P_2 \times P_3$ *admits an essentially unique CanD, as* $\mathbf{T} = \sum_{n=1}^{\text{rank}\{\mathbf{T}\}} \lambda_n \, \mathbf{a}_n^{(1)} \otimes \mathbf{a}_n^{(2)} \otimes \mathbf{a}_n^{(3)}$*, if the following two conditions are satisfied:*

1. $\mathbf{A}^{(3)}$ *is full column rank.*
2. *the tensors* $\Gamma(\mathbf{a}_u^{(1)}\mathbf{a}_u^{(2)\text{T}}, \mathbf{a}_v^{(1)}\mathbf{a}_v^{(2)\text{T}})$*,* $1 \leqslant u < v \leqslant \text{rank}\{\mathbf{T}\}$*, are linearly independent.*

Remark 1.7 The generic version of Theorem 1.11 is that $\mathbf{T}$ admits an essentially unique CanD if (i) $P_3 \geqslant \text{rank}\{\mathbf{T}\}$, and (ii) $\text{rank}\{\mathbf{T}\}(\text{rank}\{\mathbf{T}\} - 1) \leqslant P_1P_2(P_1 - 1)(P_2 - 1)/2$ [27]. The second condition implies that rank$\{\mathbf{T}\}$ is roughly bounded by the product of P_1 and P_2. On the other hand, if $P_3 \geqslant \text{rank}\{\mathbf{T}\}$, then the generic version (1.37) of Kruskal's condition reduces in the third-order case to $\text{rank}\{\mathbf{T}\} + 2 \leqslant P_1 + P_2$, which implies that rank$\{\mathbf{T}\}$ is roughly bounded by the sum of P_1 and P_2. We conclude that, if $\mathbf{A}^{(3)}$ is full column rank, Theorem 1.11 is an order of magnitude more relaxed than Kruskal's condition in the third-order case. Moreover, the proof of Kruskal's condition is not constructive, while the proof of Prop. 1.11 yields an algorithm. This algorithm will be further discussed in Section 1.5.3.

rem:genreal

Remark 1.8 The generic version of Theorem 1.11 depends on the symmetry of $\mathbf{T}$. If $\mathbf{A}^{(1)} = \mathbf{A}^{(2)}$, then we have generic essential uniqueness if $P_3 \geqslant \text{rank}\{\mathbf{T}\}$, and

$$\frac{\text{rank}\{\mathbf{T}\}(\text{rank}\{\mathbf{T}\} - 1)}{2} \leqslant \frac{P_1(P_1-1)}{4}\left(\frac{P_1(P_1-1)}{2} + 1\right) - \frac{P_1!}{(P_1-4)!4!}1_{\{P_1 \geqslant 4\}}, \tag{1.38}$$

where

$$1_{\{P_1 \geqslant 4\}} = \begin{cases} 0 & \text{if} \quad P_1 < 4 \\ 1 & \text{if} \quad P_1 \geqslant 4 \end{cases},$$

as conjectured in [70]. The latter paper also presents an algorithm with which, for any given value of P_1, it can be checked whether expression (1.38) is correct.

Now let us turn to fourth-order tensors. We first introduce a third-order tensor rank-1 detection tool [27].

prop:map3 THEOREM 1.12 *Consider the mappings* $\Omega_1 : (\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P_1 \times P_2 \times P_3} \times \mathbb{C}^{P_1 \times P_2 \times P_3} \rightarrow \Omega_1(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P_1 \times P_1 \times P_2 \times P_2 \times P_3 \times P_3}$, $\Omega_2 : (\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P_1 \times P_2 \times P_3} \times \mathbb{C}^{P_1 \times P_2 \times P_3} \rightarrow \Omega_2(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P_1 \times P_1 \times P_2 \times P_2 \times P_3 \times P_3}$ *and* $\Omega : (\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P_1 \times P_2 \times P_3} \times \mathbb{C}^{P_1 \times P_2 \times P_3} \rightarrow \Omega(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P_1 \times P_1 \times P_2 \times P_2 \times P_3 \times P_3 \times 2}$, *defined by*

$$\begin{aligned} (\Omega(\mathbf{X}, \mathbf{Y}))_{ijklmn1} &= (\Omega_1(\mathbf{X}, \mathbf{Y}))_{ijklmn} \\ &= x_{ikm}y_{jln} + y_{ikm}x_{jln} - x_{jkm}y_{iln} - y_{jkm}x_{iln} \quad (1.39) \\ (\Omega(\mathbf{X}, \mathbf{Y}))_{ijklmn2} &= (\Omega_2(\mathbf{X}, \mathbf{Y}))_{ijklmn} \\ &= x_{ikm}y_{jln} + y_{ikm}x_{jln} - x_{ilm}y_{jkn} - y_{ilm}x_{jkn}. \quad (1.40) \end{aligned}$$

Then we have $\Omega(\mathbf{X}, \mathbf{X}) = \mathbf{0}$ *if and only if* $\mathbf{X}$ *is at most rank-1.*

We now have the following sufficient condition for CanD uniqueness [27].

prop:canduniqdetlink2 THEOREM 1.13 *A fourth order tensor* $\mathbf{T}$ *of dimensions* $P_1 \times P_2 \times P_3 \times P_4$ *admits an essentially unique CanD, as* $\mathbf{T} \sum_{n=1}^{\text{rank}\{\mathbf{T}\}} \lambda_n \, \mathbf{a}_n^{(1)} \otimes \mathbf{a}_n^{(2)} \otimes \mathbf{a}_n^{(3)} \otimes \mathbf{a}_n^{(4)}$, *if the following two conditions are satisfied:*

1. $\mathbf{A}^{(4)}$ *is full column rank.*
2. *the tensors* $\{\Omega(\mathbf{a}_t^{(1)} \otimes \mathbf{a}_t^{(2)} \otimes \mathbf{a}_t^{(3)}, \mathbf{a}_u^{(1)} \otimes \mathbf{a}_u^{(2)} \otimes \mathbf{a}_u^{(3)})\}_{1 \leqslant t < u \leqslant \text{rank}\{\mathbf{T}\}}$ *are linearly independent.*

Remark 1.9 The generic version of Theorem 1.13 is that $\mathbf{T}$ admits an essentially unique CanD if (i) $P_4 \geqslant \text{rank}\{\mathbf{T}\}$, and (ii) $\text{rank}\{\mathbf{T}\}(\text{rank}\{\mathbf{T}\} - 1) \leqslant P_1P_2P_3(3P_1P_2P_3 - P_1P_2 - P_2P_3 - P_3P_1 - P_1 - P_2 - P_3 + 3)/4$ [27]. The second condition implies that $\text{rank}\{\mathbf{T}\}$ is roughly bounded by the product of P_1, P_2 and P_3. On the other hand, if $P_4 \geqslant \text{rank}\{\mathbf{T}\}$, then the generic version (1.37) of Kruskal's condition roughly bounds $\text{rank}\{\mathbf{T}\}$ by the sum of P_1, P_2 and P_3. We conclude that, if $\mathbf{A}^{(4)}$ is full column rank, Theorem 1.13 is two orders of magnitude more relaxed than Kruskal's condition in the fourth-order case. Moreover, the proof of Theorem 1.13 is constructive.

In Theorems 1.11 and 1.13 we assumed that at least one of the tensor dimensions is larger than $\text{rank}\{\mathbf{T}\}$. One can work in a similar way if the rank of one matrix representation of $\mathbf{T}$ is larger than $\text{rank}\{\mathbf{T}\}$ (see Section 1.5.1). For an example, we refer to the discussion of the FOOBI algorithm in Section 1.5.4.

Remark 1.10 The uniqueness properties discussed in this section, apply to exact data. In practice, the number of mixing vectors that can be handled, is limited by the number of available samples, the noise level and the condition of the mixture. Moreover, in certain antenna array applications, the characteristics of the antennas and the geometry of the array may induce a structure in the data that by itself bounds the number of sources that can effectively be dealt with [15, 14, 34].

1.4.3 Computation

comp-sec

As explained above, underdetermined mixtures are estimated in this chapter by formulating the problem in terms of a CanD of a (partially) symmetric higher-order tensor. This decomposition may in principle be computed by means of any general-purpose optimization method that minimizes the norm of the difference between the given tensor and its approximation by a sum of rank-1 terms. We mention the popular Alternating Least Squares (ALS) algorithm [42] [47] [66] [77], conjugate gradient [59], Levenberg-Marquardt [73] and minimization of the least absolute error [75]. It is interesting to note that, due to the multilinear structure of the problem, the size of the optimal step in a given search direction may be computed explicitly [57] [62]. The symmetric CanD of symmetric $(2\times 2\times \ldots \times 2)$ tensors can be found by means of an algorithm developed by Sylvester, see Section 1.5.2.1. Overdetermined CanD may be computed by means of the algebraic algorithms proposed in [32] [51] [74]. These algorithms can also be used in the core iteration of the algebraic algorithm for underdetermined CanD presented in [27]. One of the advantages of algebraic algorithms is that they will find the global optimum when the data are exact. Strictly speaking, there are some iterative procedures inside (e.g., the Singular Value Decomposition (SVD)), but these are well mastered.

Variants of these algorithms, designed for underdetermined ICA, are discussed in Section 1.5.

Concerning fitting a sum of rank-1 terms to a higher-order tensor, a comment is in order. Because the set of tensors of rank N is not closed, the error does in general not admit a minimum but only an infimum [52, 21]. In other words, in general there is no best solution for the canonical factors unless the rank of $\mathbf{T}$ is exactly N. In cases where there is no minimum but only an infimum, some entries in the factor matrices will tend to infinity as the algorithm approaches the infimum. This has been called *CP-degeneracy* by some authors [49] [60] [72] [22]. This phenomenon has been the object of study the last few years [21] [33] [46] [67] [68]. The algebraic algorithms in [32] [27] [51] [74] do not suffer from degeneracy.

1.5 Tensor-based algorithms

tensorAlgo-sec

1.5.1 Vector and matrix representations

mapping-sec

Matrices of size $M \times N$ can be stored in 1-dimensional arrays of size MN. We adopt the following conventions relating $\mathbf{X} \in \mathbb{C}^{M\times N}$ and $\mathbf{x} \in \mathbb{C}^{MN}$:

$$\mathbf{x} = \mathrm{vec}(\mathbf{X}) \quad \text{and} \quad \mathbf{X} = \mathrm{unvec}(\mathbf{x}) \quad \Leftrightarrow \quad (\mathbf{x})_{(m-1)N+n} = (\mathbf{X})_{mn}, \quad 1 \leqslant m \leqslant M, \quad 1 \leqslant n \leqslant N.$$

Similarly, tensors can be stored in lower order tensors, e.g., in matrices or vectors. For later use, we now define two such arrangements. First, 4th order

cumulant tensors (1.23)

$$\mathbf{C}_4^{\mathbf{x}} \stackrel{\text{def}}{=} \text{mat}(\mathcal{Q}^{\mathbf{x}}) \tag{1.41}$$

are stored in a $P^2 \times P^2$ Hermitian matrix, sometimes called the *quadricovariance* of $\mathbf{x}$. More precisely, if $\mathcal{Q}_{i\ell}^{\mathbf{x}\,jk}$ is defined as in (1.23), matrix $\mathbf{C}_4^{\mathbf{x}} = \text{mat}(\mathcal{Q})$ is defined as follows:

$$(\mathbf{C}_4^{\mathbf{x}})_{(i-1)P+j,(k-1)P+\ell} = \mathcal{Q}_{i\ell}^{\mathbf{x}\,jk}.$$

Second, 6th order cumulant tensors (1.26) can also be stored in a in a $P^3 \times P^3$ Hermitian matrix, sometimes called *hexacovariance* of $\mathbf{x}$:

$$\mathbf{C}_6^{\mathbf{x}} \stackrel{\text{def}}{=} \text{mat}(\mathcal{H}^{\mathbf{x}}). \tag{1.42}$$

hexacovDef-eq

In that case, matrix $\mathbf{C}_6^{\mathbf{x}} = \text{mat}(\mathcal{H})$ is defined as

$$(\mathbf{C}_6^{\mathbf{x}})_{P(P(i-1)+j-1)+n,P(P(\ell-1)+m-1)+k} = \mathcal{H}_{ijk}^{\ell mn}.$$

Consider two matrices $\mathbf{A}$ and $\mathbf{B}$, of dimensions $I \times J$ and $K \times L$, respectively. The *Kronecker product*, denoted $\mathbf{A} \otimes \mathbf{B}$, is the $IK \times JL$ matrix defined by:

$$\mathbf{A} \otimes \mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix} A_{11}\,\mathbf{B} & A_{12}\,\mathbf{B} & \dots \\ A_{21}\,\mathbf{B} & A_{22}\,\mathbf{B} & \dots \\ \vdots & \vdots & \end{pmatrix}.$$

It is clear from the above that matrix $\mathbf{A} \otimes \mathbf{B}$ contains the coordinates of tensor $\mathbf{A} \otimes \mathbf{B}$; but they should not be confused, since the former is a matrix and the latter a 4th order tensor.

Another useful product that helps in manipulating tensor coordinates in matrix form is the column-wise Kronecker product, also often called the *Khatri-Rao product*. Now let matrices $\mathbf{A}$ and $\mathbf{B}$ have the same number of columns J, and denote $\mathbf{a}_j$ and $\mathbf{b}_j$ their jth column, respectively. Then the Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ is defined as:

KhatriRao

$$\mathbf{A} \odot \mathbf{B} \stackrel{\text{def}}{=} (\mathbf{a}_1 \otimes \mathbf{b}_1,\ \mathbf{a}_2 \otimes \mathbf{b}_2,\ \dots\ \mathbf{a}_J \otimes \mathbf{b}_J).$$

Remark 1.11 As demonstrated in [14] the way cumulants are stored in a matrix array is important, and has an impact on the number of sources that can be localized.

1.5.2 The 2-dimensional case

algecaf-sec

1.5.2.1 Sylvester's theorem

sylvester-sec

In this section, we concentrate on symmetric tensors of order d and dimension 2. Such tensors are bijectively associated with binary quantics, that is, homogeneous polynomials in two variables.

THEOREM 1.14 (SYLVESTER, 1896) *A binary quantic* $p(x_1, x_2) = \sum_{i=0}^{d} \binom{d}{i} c_i \, x_1^i \, x_2^{d-i}$ *can be written as a sum of* d^{th} *powers of* r *distinct linear forms in* $\mathbb{C}$ *as:*

$$p(x_1, x_2) = \sum_{j=1}^{r} \lambda_j \, (\alpha_j \, x_1 + \beta_j \, x_2)^d, \tag{1.43}$$

cand-eq2

if and only if **(i)** *there exists a vector* $\mathbf{q}$ *of dimension* $r+1$*, with components* q_ℓ*, such that*

$$\begin{bmatrix} c_0 & c_1 & \cdots & c_r \\ \vdots & & & \vdots \\ c_{d-r} & \cdots & c_{d-1} & c_d \end{bmatrix} \mathbf{q} = \mathbf{0}. \tag{1.44}$$

sylves-eq

and **(ii)** *the polynomial* $q(x_1, x_2) = \sum_{i=0}^{r} q_i \, x_1^i \, x_2^{r-i}$ *admits* r *distinct roots, i.e. it can be written as* $q(x_1, x_2) = \prod_{j=1}^{r} (\beta_j^* \, x_1 - \alpha_j^* \, x_2)$.

The proof of this theorem is fortunately constructive [23] [21] and yields Algorithm 1.1, as described in [7] for instance. Given a binary polynomial $p(x_1, x_2)$ of degree d with coefficients $a_i = \binom{d}{i} c_i$, $0 \leqslant i \leqslant d$, define the Hankel matrix $H[r]$ of dimensions $d-r+1 \times r+1$ with entries $H[r]_{ij} = c_{i+j-2}$. Then the following algorithm outputs coefficients λ_j and coefficients of the linear forms $\boldsymbol{\ell}_j^{\mathrm{T}} \mathbf{x}$, for $1 \leqslant j \leqslant \operatorname{rank}\{p\}$.

sylvester-alg

ALGORITHM 1.1 (SYLVESTER)

1. Initialize $r = 0$.

step2-item

2. Increment $r \leftarrow r + 1$.
3. If the column rank of $H[r]$ is full, then go to step 2.
4. Else compute a basis $\{\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_l\}$ of the right kernel of $H[r]$.

step5-item

5. Specialization:

   - Take a generic vector $\mathbf{q}$ in the kernel, e.g. $\mathbf{q} = \sum_i \mu_i \boldsymbol{\ell}_i$ by drawing randomly coefficients μ_i.
   - Compute the roots of the associated polynomial $q(x_1, x_2) = \sum_{i=0}^{r} q_i \, x_1^i \, x_2^{d-i}$. Denote them $(\beta_j, -\alpha_j)$, where $|\alpha_j|^2 + |\beta_j|^2 = 1$.
   - If the roots are not distinct in the projective space, try another specialization. If distinct roots cannot be obtained, go to step 2.
   - Else if $q(x_1, x_2)$ admits r distinct roots then compute coefficients λ_j, $1 \leqslant j \leqslant r$, by solving the linear system below, where a_i denotes $\binom{d}{i} c_i$

$$\begin{bmatrix} \alpha_1^d & \ldots & \alpha_r^d \\ \alpha_1^{d-1}\beta_1 & \ldots & \alpha_r^{d-1}\beta_r \\ \alpha_1^{d-2}\beta_1^2 & \ldots & \alpha_r^{d-1}\beta_r^2 \\ \vdots & \vdots & \vdots \\ \beta_1^d & \ldots & \beta_r^d \end{bmatrix} \boldsymbol{\lambda} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix}.$$

6. The decomposition is $p(x_1, x_2) = \sum_{j=1}^{r} \lambda_j \, \boldsymbol{\ell}_j(\mathbf{x})^d$, where $\boldsymbol{\ell}_j(\mathbf{x}) = (\alpha_j \, x_1 + \beta_j \, x_2)$.

Note that step 5 is a specialization only if the dimension of the right kernel is strictly larger than 1.

sylvcum-sec

1.5.2.2 Sylvester's algorithm applied to cumulants

From (1.32), we know that any symmetric tensor can be associated with a homogeneous polynomial. First, consider in this section the case of 4th order cumulants. Decomposing the 4th order cumulant tensor of 2 random variables is equivalent to decomposing a homogeneous polynomial of degree 4 in 2 variables. As a consequence, we can use Sylvester's theorem described in the previous section.

But for $P = 2$ sensors, the necessary condition (1.35) is never satisfied for under-determined mixtures [23] [17]. That's why it has been proposed in [16, sec. 3.2] [18, sec. III] to fix the indeterminacy remaining in the decomposition of a tensor, whose rank is larger than generic, by using jointly two 4th order tensors of dimension 2, namely (1.23) and (1.25). This is explained in details in the algorithm below. Variants of the algorithm are presented in [30] [31].

ALGECUM-alg

ALGORITHM 1.2 (ALGECUM)

1. Compute an estimate of the two 4th order cumulant arrays $\kappa_{ij}^{\mathbf{x}\,k\ell}$ and $\kappa_{ijk\ell}^{\mathbf{x}}$, defined in Section 1.3.2.4.
2. Compute the 2×4 Hankel matrix defined in (1.44), built on $\kappa_{ijk\ell}^{\mathbf{x}}$.
3. Compute two 4-dimensional vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ forming a basis of its null space.
4. Associate the 4-way array $\kappa_{ij}^{\mathbf{x}\,k\ell}$ with a 4th degree polynomial in 4 real variables; this polynomial lives in a linear space of dimension 35, and can be expressed in a (arbitrarily chosen) basis of the latter.
5. For a finite subset of values $\theta \in [0, \pi)$ and $\varphi \in [0, 2\pi)$, compute vector $\mathbf{g}(\theta, \varphi) \stackrel{\text{def}}{=} \mathbf{v}_1 \cos\theta + \mathbf{v}_2 \sin\theta \, e^{\jmath\varphi}$.
6. For each value (θ, φ), compute the three linear forms $\boldsymbol{\ell}_j(x|\theta, \varphi)$, $1 \leq j \leq 3$ associated with $\mathbf{g}(\theta, \varphi)$.
7. Express $|\boldsymbol{\ell}_j(x|\theta, \varphi)|^4$ in the chosen basis of the 35-dimensional linear space, and devote $\mathbf{u}_j(\theta, \varphi)$ their coordinate vector.
8. Detect the value (θ_o, φ_o) of (θ, φ) for which $\kappa_{ij}^{\mathbf{x}\,k\ell}$ is the closest to the linear space spanned by $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.
9. Set $L_j = \boldsymbol{\ell}_j(\theta_o, \varphi_o)$, and $\mathbf{A} = [L_1, L_2, L_3]$, where L_j are expressed by their two coordinates.

1.5.2.3 Sylvester's algorithm applied to characteristic functions

sylvchar-sec

In the previous section, we have used Sylvester's algorithm in order to decompose the 4th order cumulant tensor of a 2-dimensional random variable. We shall see in this section how to apply the latter theorem to the characteristic function itself.

The starting point is to take any two derivatives of the core equation (1.5), and to combine them so as to cancel one term in the sum in (1.5). More precisely, for any triplet of indices, (n, i, j), $n \leq N$, $i, j \leq P$, define the differential operator below:

$$D_{n,i,j} \stackrel{\text{def}}{=} A_{in} \frac{\partial}{\partial u_j} - A_{jn} \frac{\partial}{\partial u_i}.$$

Then, it is clear that for any triplet (n, i, j), $D_{n,i,j}\Psi_x(\mathbf{u})$ does not depend on n, because that term involving ψ_n vanishes (recall that we denoted $\psi_n \stackrel{\text{def}}{=} \Psi_{s_n}$ for the sake of simplicity).

Thus, by applying such an operator N times for all the successive values of n, one eventually gets zero. Of course, the problem is not yet solved, because we don't know the entries of $\mathbf{A}$, and hence the exact form of operator $D_{n,i,j}$. However, in dimension 2, the pair (i, j) is necessarily kept fixed to $(1, 2)$. As a consequence, after N successive actions of D, we have:

$$\Big\{ \prod_{n=1}^{N} D_{n,i,j} \Big\} \Psi_x(\mathbf{u}) = \sum_{k=0}^{N} q_k \frac{\partial^N \Psi_x(\mathbf{u})}{\partial u_j^{N-k} \partial u_i^k} = 0, \ \forall \mathbf{u} \in \Omega \qquad (1.45)$$

derivN:eq

where q_k is a known function of the unknown matrix $\mathbf{A}$.

By plugging the core equation (1.5) into (1.45), we obtain:

$$\sum_{n=1}^{N} \left[\sum_{k=0}^{N} q_k A_{jn}^{N-k} A_{in}^k \right] \psi_n^{(N)}(\sum_p A_{pn} u_p) = 0 \qquad (1.46)$$

where $\psi_n^{(N)}$ denotes the Nth derivative of ψ_n. Since it is true for any vu in the neighborhood of the origin, we have eventually:

$$\sum_{k=0}^{N} q_k A_{jn}^{N-k} A_{in}^k = 0, \ \forall n. \qquad (1.47)$$

q2A:eq

The latter equation may be seen as a polynomial in A_{in}/A_{jn}, and can be rooted, which would yield the N solutions for this ratio if q_k were known. So let's now concentrate on how to estimate q_k.

As suggested in Section 1.3.1, consider equation (1.45) on a grid $\mathcal{G}$ of L values $\{\mathbf{u}[1], \ldots, \mathbf{u}[L]\}$. One can then build the over-determined linear system $\mathbf{H}[N]\,\mathbf{q} = \mathbf{0}$, where $\mathbf{H}[N]$ is the $K \times N+1$ matrix of Nth order derivatives given

below:

$$H[N] \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial^N \Psi_x(\mathbf{u}[1])}{\partial u_j^N} & \frac{\partial^N \Psi_x(\mathbf{u}[1])}{\partial u_j^{N-1} \partial u_i} & \cdots & \frac{\partial^N \Psi_x(\mathbf{u}[1])}{\partial u_i^N} \\ \frac{\partial^N \Psi_x(\mathbf{u}[2])}{\partial u_j^N} & \frac{\partial^N \Psi_x(\mathbf{u}[2])}{\partial u_j^{N-1} \partial u_i} & \cdots & \frac{\partial^N \Psi_x(\mathbf{u}[2])}{\partial u_i^N} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^N \Psi_x(\mathbf{u}[K])}{\partial u_j^N} & \frac{\partial^N \Psi_x(\mathbf{u}[K])}{\partial u_j^{N-1} \partial u_i} & \cdots & \frac{\partial^N \Psi_x(\mathbf{u}[K])}{\partial u_i^N} \end{pmatrix}. \quad (1.48)$$ defHN:eq

The latter system allows to estimate components q_k of vector $\mathbf{q}$. These results consequently lead us to the following algorithm [25] [24], able to estimate two rows of matrix $\mathbf{A}$ up to a scale factor:

algecaf1-alg **Algorithm 1.3 (Algecaf1)**

1. Fix the number N of sources sought (the algorithm can increment on N, starting with $N = P$).
2. Select two sensor indices $[i, j]$, $1 \leq [i, j] \leq P$.
3. Define a grid $\mathcal{G}$ of L values $\mathbf{u}[\ell]$ in a neighborhood of the origin in $\mathbb{R}^P$, $1 \leq \ell \leq K$.
4. Estimate the Nth order derivatives of the joint second characteristic function of observation $[x_i, x_j]$, $\psi_x(\mathbf{u})$ on this grid, and store them in a matrix $\mathbf{H}[N]$ as defined in (1.48).
5. Compute the right singular vector $\mathbf{q}$ of $\mathbf{H}[N]$ associated with the smallest singular value.
6. Root the Nth degree polynomial whose coefficients are q_k, $0 \leq k \leq N$ in the projective space (that is, include infinity if necessary).
7. Associate each root with the ratio A_{in}/A_{jn}.

This algorithm can be made more robust by observing that similar equations can be obtained if derivatives of order higher than N are computed. In fact, (1.45) is still null if further derivatives are taken. For one additional derivative, we have:

$$\frac{\partial}{\partial u_\ell} \sum_{k=0}^{N} q_k \frac{\partial^N \Psi_x(\mathbf{u})}{\partial u_j^{N-k} \partial u_i^k} = 0, \quad (1.49)$$ derivN+1:eq

where $\ell \in \{i, j\}$. Denote $\mathbf{H}[N+1, i]$ and $\mathbf{H}[N+1, j]$ the two $K \times N+1$ matrices built from (1.49), and $\ell \in \{i, j\}$. Then $\mathbf{q}$ satisfies the following linear system:

$$\begin{bmatrix} \mathbf{H}[N] \\ \mathbf{H}[N+1, i] \\ \mathbf{H}[N+1, j] \end{bmatrix} \cdot \mathbf{q} = \mathbf{0} \quad (1.50)$$ Hq:eq

where matrices $\mathbf{H}[N, \ell]$, $\ell \in \{i, j\}$, are defined by (1.49). We then obtain the following:

algecaf2-alg

ALGORITHM 1.4 (ALGECAF2)

1. Run steps 1 to 4 of algorithm ALGECAF1.
2. Build matrices $\mathbf{H}[N+1, i]$ and $\mathbf{H}[N+1, j]$.
3. Compute the right singular vector $\mathbf{q}$ of (1.50).
4. Run steps 6 and 7 of algorithm ALGECAF1.

Remark 1.12 Note that if the grid of values contains only the origin (which implies $L = 1$), only cumulants of order N are required in Algorithm 1.3, and additional cumulants of order $N + 1$ in Algorithm 1.4. There is no need to estimate derivatives of the characteristic function. However, taking a few values in the neighborhood of the origin has shown to improve the results. Computational details are omitted and can be found in [25].

subsect:sobium

1.5.3 SOBIUM family

In this section we assume that the sources are mutually uncorrelated but individually correlated in time. The presentation is in terms of complex data, but the method works also for real data, under somewhat more restrictive conditions on the number of sources, see e.g. Remark 1.8. Up to a noise term, the spatial covariance matrices of the observations satisfy [5] (see also Chapter **??**):

$$\begin{aligned} \mathbf{C}_1 \stackrel{\text{def}}{=} \mathbb{E}\,\mathbf{x}_t \mathbf{x}_{t+\tau_1}^{\mathrm{H}} &= \mathbf{A} \cdot \mathbf{D}_1 \cdot \mathbf{A}^{\mathrm{H}} \\ &\vdots \\ \mathbf{C}_K \stackrel{\text{def}}{=} \mathbb{E}\,\mathbf{x}_t \mathbf{x}_{t+\tau_K}^{\mathrm{H}} &= \mathbf{A} \cdot \mathbf{D}_K \cdot \mathbf{A}^{\mathrm{H}} \end{aligned} \tag{1.51}$$

in which $\mathbf{D}_k \stackrel{\text{def}}{=} \mathbb{E}\,\mathbf{s}_t \mathbf{s}_{t+\tau_k}^{\mathrm{H}}$ is diagonal, $k = 1, \ldots, K$. One of the delays τ_k can be equal to zero. The approach in this section applies to any ICA technique that is based on a simultaneous matrix diagonalization like (1.51). Besides spatial covariance matrices for different time lags, matrices $\{\mathbf{C}_k\}$ may correspond to spatial covariance matrices measured at different time instances, in the case of non-stationary sources subject to a constant mixing [61]. They may also correspond to spatial time-frequency distributions [6]. They could for instance also be Hessian matrices of the second characteric function of the observations, sampled at different working points [76].

Stack the matrices $\mathbf{C}_1, \ldots, \mathbf{C}_K$ in a tensor $\mathbf{T} \in \mathbb{C}^{P \times P \times K}$ as follows: $(\mathbf{T})_{p_1 p_2 k} \stackrel{\text{def}}{=} (\mathbf{C}_k)_{p_1 p_2}$, $p_1 = 1, \ldots, P, p_2 = 1, \ldots, P, k = 1, \ldots, K$. Then (1.51) can be rewritten as follows:

$$\mathbf{T} = \sum_{n=1}^{N} \mathbf{a}_n \otimes \mathbf{a}_n^* \otimes \mathbf{d}_n, \tag{1.52}$$

eq:reform

in which

$$(\mathbf{d}_n)_k \stackrel{\text{def}}{=} (\mathbf{D}_k)_{nn}, \qquad 1 \leqslant n \leqslant N, \quad 1 \leqslant k \leqslant K.$$

For later use, we define $\mathbf{D} \in \mathbb{C}^{K \times N}$ as follows:

$$(\mathbf{D})_{kn} \stackrel{\text{def}}{=} (\mathbf{D}_k)_{nn}, \qquad 1 \leqslant n \leqslant N, \quad 1 \leqslant k \leqslant K.$$

The key observation is that decomposition (1.52) is a CanD of tensor $\mathbf{T}$. The CanD uniqueness properties (see Section 1.4.2) allow one to determine mixtures even in the underdetermined case. One can in principle use any CanD algorithm. This approach is called *SOBIUM*, which stands for *Second-Order Blind Identification of Underdetermined Mixtures* [28].

A powerful technique can be derived if we can assume that $\mathbf{A} \odot \mathbf{A}^*$ and $\mathbf{D}$ are tall (i.e., have at least as many rows as columns) and that they are full column rank. Let us define a matrix representation $\bar{\mathbf{T}} \in \mathbb{C}^{P^2 \times K}$ of $\mathbf{T}$ as follows:

$$(\bar{\mathbf{T}})_{(p_1-1)P+p_2,k} = (\mathbf{T})_{p_1 p_2 k} \qquad p_1 = 1, \ldots, P, \quad p_2 = 1, \ldots, P, \quad k = 1, \ldots, K.$$

Then Eq. (1.52) can be written in a matrix format as:

$$\bar{\mathbf{T}} = (\mathbf{A} \odot \mathbf{A}^*) \cdot \mathbf{D}^{\mathrm{T}}. \tag{1.53}$$ eq:matrix_model1

The full rank property of the two factors in this product implies that N is equal to the rank of $\bar{\mathbf{T}}$. This is an easy way to estimate the number of sources, even in the underdetermined case.

Let the "economy size" SVD of $\bar{\mathbf{T}}$ be given by:

$$\bar{\mathbf{T}} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^{\mathrm{H}}, \tag{1.54}$$ eq:svd_Y

in which $\mathbf{U} \in \mathbb{C}^{P^2 \times N}$ and $\mathbf{V} \in \mathbb{C}^{K \times N}$ are column-wise orthonormal matrices and in which $\mathbf{\Sigma} \in \mathbb{R}^{N \times N}$ is positive diagonal. Combination of (1.53) and (1.54) yields that there exists an a priori unknown nonsingular matrix $\mathbf{F} \in \mathbb{C}^{N \times N}$ that satisfies:

$$\mathbf{A} \odot \mathbf{A}^* = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{F}. \tag{1.55}$$ eq:defin_F

Matrix $\mathbf{F}$ can be found by imposing the Khatri-Rao structure in the left-hand side of this equation. This Khatri-Rao structure is a vectorized form of a matrix rank-1 structure. We work as follows.

Define $\mathbf{H} = \mathbf{U}\mathbf{\Sigma} \in \mathbb{C}^{P^2 \times N}$ and $\mathbf{H}_n = \mathrm{unvec}(\mathbf{h}_n) \in \mathbb{C}^{P \times P}$, $n = 1, \ldots, N$. Eq. (1.55) can now be written as:

$$\mathbf{H}_n = \sum_{t=1}^{N} (\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}}) \, (\mathbf{F}^{-1})_{tn}. \tag{1.56}$$ eq:defH

The rank-1 structure of $\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}}$ is exploited by using the mapping Γ, defined in Theorem 1.10. From the set of matrices $\{\mathbf{H}_n\}$, we construct the set of N^2 tensors $\{\mathbf{P}_{rs} \stackrel{\text{def}}{=} \Gamma(\mathbf{H}_r, \mathbf{H}_s)\}_{1 \leqslant r,s \leqslant N}$. Due to the bilinearity of Γ, we have from (1.56):

$$\mathbf{P}_{rs} = \sum_{t,u=1}^{N} (\mathbf{F}^{-1})_{tr} (\mathbf{F}^{-1})_{us} \Gamma\left(\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}}, \ \mathbf{a}_u \mathbf{a}_u^{\mathrm{H}}\right). \tag{1.57}$$ eq:phist

We now check whether there exists a symmetric matrix $\mathbf{M} \in \mathbb{C}^{N \times N}$ that is a solution to the following set of homogeneous linear equations (it will soon become clear that such a solution indeed exists):

$$\sum_{r,s=1}^{N} M_{rs} \mathbf{P}_{rs} = \mathbf{O}. \tag{1.58}$$ eq:defM

Substitution of (1.57) in (1.58) yields:

$$\sum_{r,s=1}^{N} \sum_{t,u=1}^{N} (\mathbf{F}^{-1})_{tr} (\mathbf{F}^{-1})_{us} M_{rs} \Gamma\left(\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}},\ \mathbf{a}_u \mathbf{a}_u^{\mathrm{H}}\right) = \mathbf{O}. \tag{1.59}$$ eq:substphiM

Using the symmetry of $\mathbf{M}$, the fact that Γ is symmetric in its arguments and Theorem 1.10, (1.59) can be reduced to:

$$\sum_{r,s=1}^{N} \sum_{\substack{t,u=1 \\ t<u}}^{N} (\mathbf{F}^{-1})_{tr} (\mathbf{F}^{-1})_{us} M_{rs} \Gamma\left(\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}},\ \mathbf{a}_u \mathbf{a}_u^{\mathrm{H}}\right) = \mathbf{O}. \tag{1.60}$$ eq:sum_rs_pq

We now assume, like in the second condition of Prop. 1.11, that the tensors $\Gamma\left(\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}},\ \mathbf{a}_u \mathbf{a}_u^{\mathrm{H}}\right)$, $1 \leqslant t < u \leqslant N$, are linearly independent. Then (1.60) can only hold if the coefficients $\sum_{r,s=1}^{N} (\mathbf{F}^{-1})_{tr} (\mathbf{F}^{-1})_{us} M_{rs}$ vanish when $t \neq u$. (If $t = u$, then $\Gamma\left(\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}},\ \mathbf{a}_u \mathbf{a}_u^{\mathrm{H}}\right) = \mathbf{O}$ because of Prop. 1.10.) This can be expressed in matrix terms as follows:

$$\mathbf{M} = \mathbf{F} \cdot \mathbf{\Lambda} \cdot \mathbf{F}^{\mathrm{T}}, \tag{1.61}$$ eq:defM2

in which $\mathbf{\Lambda}$ is diagonal. It is easy to verify that *any* diagonal matrix $\mathbf{\Lambda}$ generates a matrix $\mathbf{M}$ that satisfies Eq. (1.58). Hence, solving (1.58) yields N linearly independent matrices $\{\mathbf{M}_n\}$, which can be decomposed as

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{F} \cdot \mathbf{\Lambda}_1 \cdot \mathbf{F}^{\mathrm{T}} \\ &\vdots \\ \mathbf{M}_N &= \mathbf{F} \cdot \mathbf{\Lambda}_N \cdot \mathbf{F}^{\mathrm{T}}, \end{aligned} \tag{1.62}$$

in which $\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_N$ are diagonal. Note that, if the covariance matrices $\{\mathbf{C}_k\}$ are such that tensor $\mathbf{T}$ satisfies the two conditions in Theorem 1.11, then the exact solution of the underdetermined problem may be found by means of an Eigenvalue Decomposition (EVD). Indeed, $\mathbf{F}$ can be found from the EVD

$$\mathbf{M}_1 \cdot \mathbf{M}_2^{-1} = \mathbf{F} \cdot (\mathbf{\Lambda}_1 \cdot \mathbf{\Lambda}_2^{-1}) \cdot \mathbf{F}^{-1}.$$

If $\mathbf{M}_2$ is singular, or if $\mathbf{M}_1 \cdot \mathbf{M}_2^{-1}$ has coinciding eigenvalues, then one may work with linear combinations of $\{\mathbf{M}_n\}$. In the case of inexact data, it is preferable to take all matrices in (1.62) into account. Eq. (1.62) may be solved by means of any method for joint approximate nonorthogonal matrix diagonalization, such as the algorithms presented in [32] [59] [74] [77] and references therein.

Once matrix $\mathbf{F}$ has been found from (1.62), the mixing matrix $\mathbf{A}$ can be found from (1.55). Define $\tilde{\mathbf{A}} = \mathbf{A} \odot \mathbf{A}^* \in \mathbb{C}^{P^2 \times N}$ and $\tilde{\mathbf{A}}_n = \text{unvec}(\tilde{\mathbf{a}}_n) \in \mathbb{C}^{P \times P}$, $n = 1, \ldots, N$. Then the rank of $\tilde{\mathbf{A}}_n$ is theoretically equal to 1: $\tilde{\mathbf{A}}_n = \mathbf{a}_n \mathbf{a}_n^{\text{H}}$. Consequently, $\mathbf{a}_n$ can, up to an irrelevant scaling factor, be determined as the left singular vector associated with the largest singular value of $\tilde{\mathbf{A}}_n$, $n = 1, \ldots, N$.

alg:sobium

ALGORITHM 1.5 (SOBIUM, CASE $N \leqslant K$)

1. Estimate the covariance matrices $\mathbf{C}_1, \ldots, \mathbf{C}_K$. Define $\bar{\mathbf{T}} = [\text{vec}(\mathbf{C}_1) \ \cdots \ \text{vec}(\mathbf{C}_K)]$.
2. Compute the SVD $\bar{\mathbf{T}} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^{\text{H}}$. $\mathbf{H} = \mathbf{U} \cdot \mathbf{\Sigma}$. The number of sources N equals $\text{rank}(\bar{\mathbf{T}})$.
3. Compute $\mathbf{P}_{st} = \Gamma(\mathbf{H}_s, \mathbf{H}_t)$, $1 \leqslant s \leqslant t \leqslant N$.
4. Compute N linearly independent symmetric matrices $\mathbf{M}_n$ that (approximately) satisfy $\sum_{r,s=1}^{N} M_{rs} \mathbf{P}_{rs} = \mathbf{O}$.
5. Compute nonsingular $\mathbf{F}$ that best simultaneously diagonalizes the matrices $\mathbf{M}_n$.
6. Compute $\tilde{\mathbf{A}} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{F}$.
7. Estimate mixing vector $\mathbf{a}_n$ as the dominant left singular vector of $\tilde{\mathbf{A}}_n = \text{unvec}(\tilde{\mathbf{a}}_n)$, $n = 1, \ldots, N$.

1.5.4 FOOBI family

subsect:foobi

In this section we work with the quadricovariance, defined in Sections 1.3.2.5 and 1.5.1. The presentation is in terms of complex data, but the method works also for real data, under somewhat more restrictive conditions on the number of sources. Eq. (1.23) is a CanD of the fourth-order cumulant tensor. The CanD uniqueness properties (see Section 1.4.2) allow one to determine mixtures even in the underdetermined case. This approach is called *FOOBI*, which stands for *Fourth-Order-Only Blind Identification* [9] [29].

Eq. (1.23) can be written in a matrix format as follows:

$$\mathbf{C}_4^{\mathbf{x}} = (\mathbf{A} \odot \mathbf{A}^*) \cdot \mathbf{\Delta}_4 \cdot (\mathbf{A} \odot \mathbf{A}^*)^{\text{H}}, \tag{1.63}$$

eq:CANDsym2m

where $\mathbf{\Delta}_4$ is the $N \times N$ diagonal matrix containing the 4th order source marginal cumulants. We have the following theorem [29].

theor:EVDSS4

THEOREM 1.15 *Consider a tensor $\mathbf{T} \in \mathbb{C}^{P \times P \times P \times P}$, satisfying the symmetries $t_{klij} = t^*_{ijkl}$ and $t_{jilk} = t^*_{ijkl}$, and its matrix representation $\bar{\mathbf{T}} = mat(\mathbf{T}) \in \mathbb{C}^{P^2 \times P^2}$. The matrix $\bar{T}$ can be eigen-decomposed as*

$$\mathbf{T} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^{\text{H}}, \tag{1.64}$$

eq:propeig2

in which $\mathbf{E} \in \mathbb{C}^{P^2 \times N}$ is column-wise orthonormal, with $e_{(p_1-1)P+p_2,n} = e^*_{(p_2-1)P+p_1,n}$, $p_1, p_2 = 1, \ldots, P$, $n = 1, \ldots, N$, and in which $\mathbf{\Lambda} \in \mathbb{C}^{N \times N}$ is a diagonal matrix of which the diagonal elements are real and nonzero. N is the rank of $\bar{\mathbf{T}}$.

From this theorem we have that $\mathbf{C}_4^{\mathbf{x}}$ can be decomposed as

$$\mathbf{C}_4^{\mathbf{x}} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^{\mathrm{H}}. \tag{1.65}$$ eq:tensevd2

This matrix EVD may be easily computed. Note that N is equal to the rank of $\mathbf{C}_4^{\mathbf{x}}$. This is an easy way to estimate the number of sources, even in the underdetermined case.

We now assume that all sources have strictly positive kurtosis. The more general situation will be addressed in Remark 1.13. From Eq. (1.63) it follows that $\mathbf{C}_4^{\mathbf{x}}$ is positive (semi)definite. We have the following theorem [29].

theor:EVDPARAFAC THEOREM 1.16 *Let $\mathbf{C}_4^{\mathbf{x}}$ be positive (semi)definite and assume that it can be decomposed as in (1.63) and (1.65). Then we have:*

$$(\mathbf{A} \odot \mathbf{A}^*) \cdot (\mathbf{\Delta}_4)^{1/2} = \mathbf{E} \cdot \mathbf{\Lambda}^{1/2} \cdot \mathbf{V}, \tag{1.66}$$ eq:squareroot

in which $\mathbf{V}$ is real $(N \times N)$ orthogonal.

Eq. (1.66) is analogous to the SOBIUM equation (1.55). The matrix $\mathbf{V}$ can be determined by exploiting the Khatri-Rao structure of $\mathbf{A} \odot \mathbf{A}^*$. The following rank-1 detection tool was used in the original FOOBI algorithm [9] [29]. $\mathbb{H}^{P \times P}$ represents the space of $(P \times P)$ Hermitean matrices.

theor:rank1a THEOREM 1.17 *Consider the mapping $\tilde{\Gamma} : (\mathbf{X}, \mathbf{Y}) \in \mathbb{H}^{P \times P} \times \mathbb{H}^{P \times P} \rightarrow \tilde{\Gamma}(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{P \times P \times P \times P}$ defined by*

$$(\tilde{\Gamma}(\mathbf{X}, \mathbf{Y}))_{ijkl} = x_{ij} y_{kl}^* + y_{ij} x_{kl}^* - x_{ik} y_{jl}^* - y_{ik} x_{jl}^*. \tag{1.67}$$ eq:rank1det1

Then we have that $\tilde{\Gamma}(\mathbf{X}, \mathbf{X}) = \mathbf{0}$ if and only if $\mathbf{X}$ is at most rank-1.

Analogous to SOBIUM, we define a matrix $\mathbf{H} \stackrel{\text{def}}{=} \mathbf{E} \cdot \mathbf{\Lambda}^{1/2}$, Hermitean matrices $\mathbf{H}_s \stackrel{\text{def}}{=} \text{unvec}(\mathbf{h}_s)$ and we construct a set of N^2 fourth-order tensors $\{\mathbf{P}_{rs} \stackrel{\text{def}}{=} \tilde{\Gamma}(\mathbf{H}_r, \mathbf{H}_s)\}_{1 \leqslant r,s \leqslant N}$. Then we look for a real symmetric matrix $\mathbf{M} \in \mathbb{R}^{N \times N}$ that is a solution to the following set of homogeneous linear equations:

$$\sum_{r,s=1}^{P} M_{rs} \mathbf{P}_{rs} = \mathbf{O}. \tag{1.68}$$ eq:defMfoobi

We assume that the tensors $\tilde{\Gamma}(\mathbf{a}_t \mathbf{a}_t^{\mathrm{H}}, \mathbf{a}_u \mathbf{a}_u^{\mathrm{H}})$, $1 \leqslant t < u \leqslant N$, are linearly independent. This corresponds to the second condition of Prop. 1.11. It turns out that, under this condition, (1.68) has N linearly independent solutions, which can be decomposed as:

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{V} \cdot \mathbf{D}_1 \cdot \mathbf{V}^{\mathrm{T}} \\ &\vdots \\ \mathbf{M}_N &= \mathbf{V} \cdot \mathbf{D}_N \cdot \mathbf{V}^{\mathrm{T}}, \end{aligned} \tag{1.69}$$

in which $\mathbf{D}_1, \ldots, \mathbf{D}_N \in \mathbb{R}^{N \times N}$ are diagonal. The difference with (1.62) is that all matrices are real, even when $\mathbf{C}_4^{\mathbf{x}}$ is complex, and that $\mathbf{V}$ is orthogonal. Note that, in the case of exact data, the solution of the underdetermined problem may be found by means of an EVD. Indeed, every equation in (1.69) is an EVD of a real symmetric matrix. In the case of inexact data, it is preferable to take all matrices in (1.69) into account. The joint approximate orthogonal matrix diagonalization problem may be solved by means of the Jacobi iteration presented in [11] [12]. Once matrix $\mathbf{V}$ has been found from (1.69), the mixing matrix $\mathbf{A}$ can be found from (1.66) in a similar way as in SOBIUM.

alg:foobi

ALGORITHM 1.6 (FOOBI)

1. Estimate the quadricovariance $\mathbf{C}_4^{\mathbf{x}}$ of the data.
2. Compute the EVD $\mathbf{C}_4^{\mathbf{x}} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^{\mathrm{H}}$. $\mathbf{H} = \mathbf{E} \cdot \mathbf{\Lambda}^{1/2}$. The number of sources N equals $\mathrm{rank}(\mathbf{C}_4^{\mathbf{x}})$. Normalize the eigenvectors such that $\mathbf{H}_n = \mathrm{unvec}(\mathbf{h}_n)$, $1 \leqslant n \leqslant N$, is Hermitean.
3. Compute $\mathbf{P}_{st} = \tilde{\Gamma}(\mathbf{H}_s, \mathbf{H}_t)$, $1 \leqslant s \leqslant t \leqslant N$.
4. Compute N linearly independent real symmetric matrices $\mathbf{M}_n$ that (approximately) satisfy $\sum_{r,s=1}^{N} M_{rs} \mathbf{P}_{rs} = \mathbf{O}$.
5. Compute orthogonal $\mathbf{V}$ that best simultaneously diagonalizes the matrices $\mathbf{M}_n$.
6. Compute $\mathbf{F} = \mathbf{E} \cdot \mathbf{\Lambda}^{1/2} \cdot \mathbf{V}$.
7. Estimate mixing vector $\mathbf{a}_n$ as the dominant left singular vector of $\mathrm{unvec}(\mathbf{f}_n)$, $n = 1, \ldots, N$.

rem:mixedsign

Remark 1.13 In the derivation above, we have assumed that all sources are strictly leptokurtic. If all the sources all have strictly negative kurtosis, then we simply process $-\mathcal{C}_4^{\mathbf{x}}$. When not all kurtosis values have the same sign, (1.66) can be replaced by

$$\mathbf{A} \odot \mathbf{A}^* = \mathbf{E} \cdot \mathbf{V}, \tag{1.70}$$

eq:squareroot2

in which now $\mathbf{V}$ is real nonsingular instead of orthogonal. Eq. (1.69) then becomes a nonorthogonal joint diagonalization problem, like in SOBIUM.

In the remainder of this section, we assume that all the sources are strictly leptokurtic. (If all the sources are strictly platykurtic, then we process $-\mathbf{C}_4^{\mathbf{x}}$ instead of $\mathbf{C}_4^{\mathbf{x}}$).

Generically, as long as $N \leqslant P^2$ (complex case) or $N \leqslant P(P+1)/2$ (real case), the number of sources corresponds to the rank of $\mathbf{C}_4^{\mathbf{x}}$ and Theorem 1.16 still applies. We now introduce a new rank-1 detecting tool [29].

theor:rank1c

THEOREM 1.18 *Consider the mapping* $\Theta : (\mathbf{X}, \mathbf{Y}) \in \mathbb{H}^{P \times P} \times \mathbb{H}^{P \times P} \rightarrow \Theta(\mathbf{X}, \mathbf{Y}) \in \mathbb{H}^{P \times P}$ *defined by*

$$\Theta(\mathbf{X}, \mathbf{Y}) = \mathbf{X}\mathbf{Y} - \mathit{trace}(\mathbf{X})\mathbf{Y} + \mathbf{Y}\mathbf{X} - \mathit{trace}(\mathbf{Y})\mathbf{X}. \tag{1.71}$$

eq:rank1det3

Then we have that $\Theta(\mathbf{X}, \mathbf{X}) = \mathbf{0}$ *if and only if* $\mathbf{X}$ *is at most rank-1.*

Define $\mathbf{H} \in \mathbb{C}^{P^2 \times N}$ and $\mathbf{H}_s = \mathrm{mat}(\mathbf{h}_s) \in \mathbb{C}^{P \times P}$, $s = 1, \ldots, N$, as above. Also define symmetric matrices $\mathbf{B}_{p_1 p_2} \in \mathbb{C}^{N \times N}$ by

$$(\mathbf{B}_{p_1 p_2})_{st} = (\Theta(\mathbf{H}_s, \mathbf{H}_t))_{p_1 p_2}, \quad 1 \leqslant p_1, p_2 \leqslant P, \quad 1 \leqslant s, t \leqslant N.$$

The following theorem leads to a new algorithm for the computation of $\mathbf{V}$ [29].

theor:simoff THEOREM 1.19 *The matrix* $\mathbf{V}$ *in Eq. (1.66) satisfies*

$$\begin{aligned} &\mathit{diag}(\mathbf{V}^{\mathrm{T}} \cdot \mathit{Real}(\mathbf{B}_{p_1 p_2}) \cdot \mathbf{V}) = \mathbf{0}, \quad 1 \leqslant p_1 \leqslant p_2 \leqslant P, \\ &\mathit{diag}(\mathbf{V}^{\mathrm{T}} \cdot \mathit{Imag}(\mathbf{B}_{p_1 p_2}) \cdot \mathbf{V}) = \mathbf{0}, \quad 1 \leqslant p_1 < p_2 \leqslant P. \end{aligned} \tag{1.72}$$

This theorem shows that the matrix $\mathbf{V}$ can be computed by means of simultaneous off-diagonalization of a number of real symmetric matrices. One can use a simple variant of the Jacobi algorithm derived in [11, 12], the difference being that one should chose in each step the Jacobi rotation that minimizes (instead of maximizes) the sum of the squared diagonal entries. Simultaneous orthogonal off-diagonalization was also used in [4].

alg:foobi2 ALGORITHM 1.7 (FOOBI-2)

1. Estimate the quadricovariance $\mathbf{C}_4^{\mathbf{x}}$ of the data.
2. Compute the EVD $\mathbf{C}_4^{\mathbf{x}} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^{\mathrm{H}}$. $\mathbf{H} = \mathbf{E} \cdot \mathbf{\Lambda}^{1/2}$. The number of sources N equals $\mathrm{rank}(\mathbf{C}_4^{\mathbf{x}})$. Normalize the eigenvectors such that $\mathbf{H}_n = \mathrm{unvec}(\mathbf{h}_n)$, $1 \leqslant n \leqslant N$, is Hermitean.
3. Compute $\Theta(\mathbf{H}_s, \mathbf{H}_t)$, $1 \leqslant s \leqslant t \leqslant N$. Stack the results in $\mathbf{B}_{p_1 p_2}$, $1 \leqslant p_1 \leqslant p_2 \leqslant P$.
4. Compute orthogonal $\mathbf{V}$ that best simultaneously off-diagonalizes the matrices $\{\mathrm{Real}(\mathbf{B}_{p_1 p_2})\}$ and $\{\mathrm{Imag}(\mathbf{B}_{p_1 p_2})\}$.
5. Compute $\mathbf{F} = \mathbf{E} \cdot \mathbf{\Lambda}^{1/2} \cdot \mathbf{V}$.
6. Estimate mixing vector $\mathbf{a}_n$ as the dominant left singular vector of $\mathrm{unvec}(\mathbf{f}_n)$, $n = 1, \ldots, N$.

1.5.5 BIOME family

subsect:biome

.

The algorithm we describe in this section, BIRTH, is using 6th order cumulants as in [2], but other orders can be considered as shown in [3], where a general family of algorithms called BIOME$-2q$ is described. We shall restrict our attention to the former, which actually corresponds to BIOME-6.

The basic idea is that the hexacovariance $\mathbf{C}_6^{\mathbf{y}}$ defined in (1.42) enjoys the property below

$$\mathbf{C}_6^{\mathbf{y}} = \mathbf{A}^{\odot 3} \, \mathbf{\Delta}_6 \, \mathbf{A}^{\odot 3\mathrm{H}}$$

where $\mathbf{\Delta}_6$ is the $N \times N$ diagonal matrix containing the 6th order source marginal cumulants. In fact, this is another way of writing the multi-linearity property 1.5, with the help of the Khatri-Rao product. If we define a full rank square root

matrix of $\mathbf{C}_6^{\mathbf{y}}$, denoted $\mathbf{C}_6^{\mathbf{y}\,1/2}$, then it is related to $\mathbf{A}^{\odot\,3}$ up to an (unknown) unitary matrix $\mathbf{V}$:

$$\mathbf{C}_6^{\mathbf{y}\,1/2} = \mathbf{A}^{\odot\,3}\mathbf{\Delta}_6^{1/2}\,\mathbf{V}.$$

The problem is that none of the three matrix factors in the right hand side is known. But there is a redundancy in the left-hand side, which can be exploited. This is done in the algorithm below, by rewriting the above equation for each $P^2 \times N$ submatrix $\mathbf{\Gamma}[n]$ of $\mathbf{C}_6^{\mathbf{y}\,1/2}$, as:

$$\mathbf{\Gamma}[n] = (\mathbf{A} \odot \mathbf{A}^{\mathrm{H}})\,\mathbf{D}[n]\mathbf{\Delta}_6^{1/2}\mathbf{V}$$

where $\mathbf{D}[n]$ is the diagonal matrix containing te nth row of $\mathbf{A}$, $1 \leqslant n \leqslant P$. Hence matrices $\mathbf{\Gamma}[n]$ share the same common right singular subspace defined by $\mathbf{V}$. Yet, if we notice that we can get rid of the unknown factor $(\mathbf{A} \odot \mathbf{A}^{\mathrm{H}})$ by computing the product between a pseudo-inverse $\mathbf{\Gamma}^{\dagger}[m]$ and $\mathbf{\Gamma}[n]$, then we see that $\mathbf{V}$ can be eventually obtained via the EVD of the Hermitian matrix

$$\mathbf{\Theta}_{m,n} \stackrel{\text{def}}{=} \mathbf{\Gamma}_m^{\dagger}\mathbf{\Gamma}_n = \mathbf{V}^{\mathrm{H}}\,\left(\mathbf{D}[m]^{-1}\mathbf{D}[n]\right)\,\mathbf{V}.$$

This can be done for any pair of indices (m, n), $m \neq n$. But it is more stable to do it for all of them simultaneously. Once matrix $\mathbf{V}$ has been obtained this way, matrix $\mathbf{A}^{\odot\,3}$ is obtained up tio a scale factor with the product $\mathbf{C}_6^{\mathbf{y}\,1/2}\mathbf{V}^{\mathrm{H}}$. As a conclusion, the whole BIRTH algorithm can be described as follows

alg:birth

ALGORITHM 1.8 (BIRTH)

1. Estimate 6th order cumulants of observation $\mathbf{y}$ and store them in the hexacovariance matrix $\mathbf{C}_6^{\mathbf{y}}$, of size $P^3 \times P^3$ as defined in (1.42).
2. Compute a square root $\mathbf{C}_6^{\mathbf{y}\,1/2}$ of $\mathbf{C}_6^{\mathbf{y}}$, of size $P^3 \times N$, e.g. via an EVD.
3. Cut $\mathbf{C}_6^{\mathbf{y}\,1/2}$ into P blocks $\mathbf{\Gamma}_n$ each of size P^2N.
4. Compute the $P(P-1)/2$ products $\mathbf{\Theta}_{m,n} \stackrel{\text{def}}{=} \mathbf{\Gamma}_m^{\dagger}\mathbf{\Gamma}_n$ and jointly diagonalize them (approximately), so that $\mathbf{\Theta}_{m,n} \approx \mathbf{V}\mathbf{\Lambda}[m,n]\mathbf{V}^{\mathrm{H}}$, where $\mathbf{V}$ is $N \times N$ unitary.
5. Compute $\widehat{\mathbf{A}^{\odot\,3}} = (\mathbf{C}_6^{\mathbf{y}})^{1/2}\mathbf{V}^{\mathrm{H}}$.
6. Store each of the N columns of $\widehat{\mathbf{A}^{\odot\,3}}$ in a vector $\mathbf{b}_n$ of size P^3.
7. Transform each vector $\mathbf{b}_n$ in a family of P matrices of size $P \times P$. Compute the common dominant eigenvector, $\hat{\mathbf{a}}_n$, of these Hermitian matrices.
8. Stack column vectors $\hat{\mathbf{a}}_n$ to form matrix $\hat{\mathbf{A}}$.

Several variants are described in [2], each allowing to obtain vectors $\hat{\mathbf{a}}_n$ with better accuracy levels, to the price of an increased computational complexity.

alescaf-sec

1.5.6 ALESCAF and LEMACAF

In this section, we elaborate on the calculation of the solution of (1.7). The goal is to minimize the fitting error [25] [24]:

$$\Upsilon = ||\mathbf{T} - \sum_{n=1}^{\text{rank}\{\mathbf{T}\}} \lambda_n \, \mathbf{a}_n \otimes \mathbf{a}_n \otimes \mathbf{a}_n \otimes \mathbf{b}_n||^2 \tag{1.73}$$

optimCAF-eq

where $\mathbf{\Lambda}$ is a $N \times N \times N \times N$ diagonal tensor containing ones in its diagonal.

Beside the fact that this tensor has a partial symmetry (in fact in the first three modes), this minimization problem is similar to that for general tensors:

$$\Upsilon = ||\mathbf{T} - \sum_{n=1}^{\text{rank}\{\mathbf{T}\}} \lambda_n \, \mathbf{a}_n \otimes \mathbf{b}_n \otimes \mathbf{c}_n \otimes \mathbf{d}_n||^2. \tag{1.74}$$

optimGene-eq

So let's look at this slightly more general problem first and solve it by means of an ALS scheme.

ALESCAF-alg

ALGORITHM 1.9 (ALESCAF) [25]

1. Define a grid $\mathcal{G}$ containing L points of $\mathbb{C}^P$.
2. For every point $\mathbf{u}$ of $\mathcal{G}$, compute the derivative tensor $\mathbf{T}(\mathbf{u})$.
3. Initialize, possibly randomly, quantites $\mathbf{a}(k)$. Denote $\mathbf{A}[1]$, $\mathbf{A}[2]$ and $\mathbf{A}[3]$ matrices containing initial values.
4. Execute the ALS algorithm from $t = 1$ and until stopping criterion is reached:
   $\mathbf{B}^{\mathrm{T}}[3t] = (\mathbf{A}[3t] \odot \mathbf{A}[3t-1] \odot \mathbf{A}[3t-2])^{-1} \, \mathbf{T}_{N^3 \times L}$
   $\mathbf{A}^{\mathrm{T}}[3t+1] = (\mathbf{A}[3t] \odot \mathbf{A}[3t-1] \odot \mathbf{B}[3t])^{-1} \, \mathbf{T}_{N^2 L \times N}$
   $\mathbf{A}^{\mathrm{T}}[3t+2] = (\mathbf{A}[3t+1] \odot \mathbf{A}[3t] \odot \mathbf{B}[3t])^{-1} \, \mathbf{T}_{N^2 L \times N}$
   $\mathbf{A}^{\mathrm{T}}[3t+3] = (\mathbf{A}[3t+2] \odot \mathbf{A}[3t+1] \odot \mathbf{B}[3t])^{-1} \, \mathbf{T}_{N^2 L \times N}$.

One could also regularize the optimization criterion by adding a penalty term involving the norm of the columns, as in [58], which yields the new objective:

$$\Upsilon_R = ||\mathbf{T} - \sum_{p=1}^{\text{rank}\{\mathbf{T}\}} \lambda_p \, \mathbf{a}_p \otimes \mathbf{b}_p \otimes \mathbf{c}_p \otimes \mathbf{d}_p||^2 + \eta \sum_{p=1}^{N} ||\mathbf{a}_p||^2 + ||\mathbf{b}_p||^2 + ||\mathbf{c}_p||^2 + ||\mathbf{d}_p||^2 \tag{1.75}$$

optimRegul-eq

if we decide to assign the same weight η to every penalty term.

Another way is to leave the diagonal tensor $\mathbf{\Lambda}$ take free values, but to constraint loading matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ to have a unit norm. It is easy to see that the two latter approaches are the same, as pointed out in [52].

Now let's go back to our original problem (1.73) and give the example of the Levenberg-Marquardt implementation presented in [22].

LEMACAF-alg

ALGORITHM 1.10 (LEMACAF)

1. Define a grid $\mathcal{G}$ containing L points of $\mathbb{C}^P$.
2. For every point $\mathbf{u}$ of $\mathcal{G}$, compute the derivative tensor $\mathbf{T}(\mathbf{u})$.
3. Initialize, possibly randomly, quantities $\boldsymbol{\beta}_k(\mathbf{u})$ and $\mathbf{a}(k)$. Denote $\mathbf{B}[0]$ and $\mathbf{A}[0]$ the matrices containing these initial values.
4. Arrange all values of $\boldsymbol{\beta}_k(\mathbf{u})$ and $\mathbf{a}(k)$ in a vector $\mathbf{p}$ of size $(P+L)N$. The penalty term in (1.75) is now just $\eta\,||\mathbf{p}||^2$. For t=1 and until stop, do:
   - Compute the gradient $\mathbf{g}[t]$ of Υ_R at $\mathbf{p}[t]$.
   - Compute the Jacobian $\mathbf{J}[t]$ of Υ_R at $\mathbf{p}[t]$.
   - Compute $\mathbf{p}[t+1] = \mathbf{p}[t] - [\mathbf{J}[t]^{\mathrm{H}}\mathbf{J}[t] + \lambda[t]\,\mathbf{I}]^{-1}\,\mathbf{g}[t]$.

The delicate issue in such algorithms is the choice of parameter $\lambda[t]$. We refer to optimization textbooks for this important question [53] [39].

1.5.7 Other algorithms

We have reviewed algorithms that exploit the mutual statistical independence between sources. All these algorithms are based on the fact that the problem can be formulated in terms of a CanD of a tensor that enjoys symmetries. In some cases there exist closed form solutions.

Other algorithms have not been mentioned. In particular, the decomposition of symmetric tensors in the complex field has been addressed in [7] for dimensions larger than 2. Bayesian approaches are addressed in Chapter **??**. Algorithms exploiting sparseness properties or nonnegativity constraints are studied in Chapters **??** and **??**, respectively.

An important class of approaches, which we consider as quite promising, consists of working directly on the data arranged in tensor format. This avoids to resort to statistics of the data, and hence does not need the assumption of source independence. On the other hand, the availability of some diversity in the measurements is required. To build a data tensor from measurements is often the actual challenge, but it has been successfully addressed in a number of applications [64] [65] [66].

In this context, several *deterministic* algorithms devoted to blind identification have been developed. They range from descent or Newton algorithms, Alternating Least Squares (ALS) [42] [66] [22], Levenberg-Marquardt [56] [22] [73] or conjugate gradient [59], to global line search techniques, *e.g.*, the so-called Enhanced Line Search (ELS) algorithm [62] [57]. Note that uniqueness conditions, already tackled in Section 1.4.2, and reported in [22], can sometimes be improved [27] [26].

1.6 Appendix: expressions of complex cumulants

cumulants-app

In this appendix, we only give expressions of cumulants for zero-mean complex variables that are distributed symmetrically with respect to the origin. However,

they do not need to be circularly distributed. Below, cumulants are denoted with κ and moments with μ. As before, superscripts correspond to variables that are complex conjugated, and McCullagh's bracket notation [54] is used, with an obvious extension to the complex case [1]; note that in [10] the cumulant expressions are incomplete, and assume that random variables are circularly distributed. The notation $[\bar{2}]$ means twice the real part of the argument. We have for orders 4 and 6:

$$\begin{aligned}
\kappa_{ijk\ell} &= \mu_{ijk\ell} - [3]\mu_{ij}\mu_{k\ell}, \\
\kappa_{ijk}^{\ell} &= \mu_{ijk}^{\ell} - [3]\mu_{ij}\mu_{k}^{\ell}, \\
\kappa_{ij}^{k\ell} &= \mu_{ij}^{k\ell} - [2]\mu_{i}^{k}\mu_{j}^{\ell} - \mu_{ij}\mu^{k\ell},
\end{aligned}$$

$$\begin{aligned}
\kappa_{ijk\ell mn} &= \mu_{ijk\ell mn} - [15]\mu_{ijk\ell}\mu_{mn} + 2[15]\mu_{ij}\mu_{k\ell}\mu_{mn}, \\
\kappa_{ijk\ell m}^{n} &= \mu_{ijk\ell m}^{n} - [5]\mu_{ijk\ell}\mu_{m}^{n} - [10]\mu_{ijk}^{n}\mu_{\ell m} \\
&\quad +2[15]\mu_{ij}\mu_{k\ell}\mu_{m}^{n}, \\
\kappa_{ijk\ell}^{mn} &= \mu_{ijk\ell}^{mn} - \mu_{ijk\ell}\mu^{mn} - [8]\mu_{ijk}^{m}\mu_{\ell}^{n} - [6]\mu_{ij}^{mn}\mu_{k\ell} \\
&\quad +[6]\mu_{ij}\mu_{k\ell}\mu^{mn} + 2[12]\mu_{ij}\mu_{k}^{m}\mu_{\ell}^{n}, \\
\kappa_{ijk}^{\ell mn} &= \mu_{ijk}^{\ell mn} - [3]\mu_{ijk}^{\ell}\mu^{mn} - [9]\mu_{ij}^{\ell m}\mu_{k}^{n} - [3]\mu_{ij}\mu_{k}^{\ell mn} \\
&\quad +2[9]\mu_{ij}\mu_{k}^{\ell}\mu^{mn} + 2[6]\mu_{i}^{\ell}\mu_{j}^{m}\mu_{k}^{n},
\end{aligned}$$

and eventually for order 8:

$$\begin{aligned}
\kappa_{ijk\ell mnpq} &= \mu_{ijk\ell mnpq} - [28]\mu_{ijk\ell mn}\mu_{pq} - [35]\mu_{ijk\ell}\mu_{mnpq} \\
&\quad +2[210]\mu_{ijk\ell}\mu_{mn}\mu_{pq} - 6[105]\mu_{ij}\mu_{k\ell}\mu_{mn}\mu_{pq}, \\
\kappa_{ijk\ell mnp}^{q} &= \mu_{ijk\ell mnp}^{q} - [7]\mu_{ijk\ell mn}\mu_{p}^{q} - [21]\mu_{ijk\ell m}^{q}\mu_{np} \\
&\quad -[35]\mu_{ijk\ell}\mu_{mnp}^{q} + 2[105]\mu_{ijk}^{q}\mu_{\ell m}\mu_{np} \\
&\quad +2[105]\mu_{ijk\ell}\mu_{mn}\mu_{p}^{q} - 6[105]\mu_{ij}\mu_{k\ell}\mu_{mn}\mu_{p}^{q}, \\
\kappa_{ijk\ell mn}^{pq} &= \mu_{ijk\ell mn}^{pq} - \mu_{ijk\ell mn}\mu^{pq} - [12]\mu_{ijk\ell m}^{p}\mu_{n}^{q} \\
&\quad -[15]\mu_{ijk\ell}^{pq}\mu_{mn} - [15]\mu_{ijk\ell}\mu_{mn}^{pq} - [20]\mu_{ijk}^{p}\mu_{\ell mn}^{q} \\
&\quad +2[15]\mu_{ijk\ell}\mu_{mn}\mu^{pq} + 2[30]\mu_{ijk\ell}\mu_{m}^{p}\mu_{n}^{q} \\
&\quad +2[120]\mu_{ijk}^{p}\mu_{\ell m}\mu_{n}^{q} + 2[45]\mu_{ij}^{pq}\mu_{k\ell}\mu_{mn} \\
&\quad -6[15]\mu_{ij}\mu_{k\ell}\mu_{mn}\mu^{pq} - 6[90]\mu_{ij}\mu_{k\ell}\mu_{m}^{p}\mu_{n}^{q},
\end{aligned}$$

$$\begin{aligned}
\kappa_{ijk\ell m}^{npq} &= \mu_{ijk\ell m}^{npq} - [3]\mu_{ijk\ell m}^{n}\mu^{pq} - [15]\mu_{ijk\ell}^{np}\mu_{m}^{q} \\
&\quad -[10]\mu_{ijk}^{npq}\mu_{\ell m} - [5]\mu_{ijk\ell}\mu_{m}^{npq} - [30]\mu_{ijk}^{n}\mu_{\ell m}^{pq} \\
&\quad +2[15]\mu_{ijk\ell}\mu_{m}^{n}\mu^{pq} + 2[30]\mu_{ijk}^{n}\mu_{\ell m}\mu^{pq} \\
&\quad +2[60]\mu_{ijk}^{n}\mu_{\ell}^{p}\mu_{m}^{q} + 2[90]\mu_{ij}^{np}\mu_{k\ell}\mu_{m}^{q} \\
&\quad +2[15]\mu_{i}^{npq}\mu_{jk}\mu_{\ell m} \\
&\quad -6[45]\mu_{ij}\mu_{k\ell}\mu_{m}^{n}\mu^{pq} - 6[60]\mu_{ij}\mu_{k}^{n}\mu_{\ell}^{p}\mu_{m}^{q}, \\
\kappa_{ijk\ell}^{mnpq} &= \mu_{ijk\ell}^{mnpq} - [6]\mu_{ijk\ell}^{mn}\mu^{pq} - [6]\mu_{ij}^{mnpq}\mu_{k\ell} - [16]\mu_{ijk}^{mnp}\mu_{\ell}^{q} \\
&\quad -[16]\mu_{ijk}^{m}\mu_{\ell}^{npq} - [18]\mu_{ij}^{mn}\mu_{k\ell}^{pq} - \mu_{ijk\ell}\mu^{mnpq} \\
&\quad +2\,[\bar{2}]([3]\mu_{ijk\ell}\mu^{mn}\mu^{pq} + [48]\mu_{ijk}^{m}\mu_{\ell}^{n}\mu^{pq}) \\
&\quad +2[36]\mu_{ij}^{mn}\mu_{k\ell}\mu^{pq} + 2[72]\mu_{ij}^{mn}\mu_{k}^{p}\mu_{\ell}^{q} \\
&\quad -6[72]\mu_{ij}\mu_{k}^{m}\mu_{\ell}^{n}\mu^{pq} - 6[24]\mu_{i}^{m}\mu_{j}^{n}\mu_{k}^{p}\mu_{\ell}^{q} \\
&\quad -6[9]\mu_{ij}\mu_{k\ell}\mu^{mn}\mu^{pq}.
\end{aligned}$$

Bibliography

AlbeC02:sam [1] L. Albera and P. Comon, *Asymptotic performance of contrast-based blind source separation algorithms*, in Second IEEE Sensor Array and Multichannel Signal Processing Workshop, Rosslyn, VA, 4–6 August 2002.

AlbeCCF04:montreal [2] L. Albera, P. Comon, P. Chevalier, and A. Ferreol, *Blind identification of underdetermined mixtures based on the hexacovariance*, in ICASSP'04, vol. II, Montreal, May 17–21 2004, pp. 29–32.

AlbeFCC04:laa [3] L. Albera, A. Ferreol, P. Comon, and P. Chevalier, *Blind identification of overcomplete mixtures of sources (BIOME)*, Lin. Alg. Appl., 391 (2004), pp. 1–30.

belouchraniicassp01 [4] A. Belouchrani, K. Abed-Meraim, M. G. Amin, and A. M. Zoubir, *Joint antidiagonalization for blind source separation*, in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-01), vol. 5, Salt Lake City, May 2001, pp. 2789–2792.

belouchrani1 [5] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, *A blind source separation technique using second order statistics*, IEEE Trans. Signal Processing, 45 (1997), pp. 434–444.

belouchrani3 [6] A. Belouchrani and M. G. Amin, *Blind source separation based on time-frequency signal representations*, IEEE Trans. Signal Processing, 46 (1998), pp. 2888–2897.

BracCMT09:laa [7] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, *Symmetric tensor decomposition*, Lin. Alg. Appl., (2009). submitted. hal:inria-00355713, arXiv:0901. 3706.

CaoL96:ieeesp [8] X. R. Cao and R. W. Liu, *General approach to blind source separation*, IEEE Trans. Sig. Proc., 44 (1996), pp. 562–570.

cardoso7 [9] J.-F. Cardoso, *Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors*, in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-91), Toronto, 1991, pp. 3109–3112.

CardM95:ieeesp [10] J.-F. Cardoso and E. Moulines, *Asymptotic performance analysis of direction-finding algorithms based on fourth-order cumulants*, IEEE Trans. Signal Processing, 43 (1995), pp. 214–224.

cardoso11 [11] J.-F. Cardoso and A. Souloumiac, *Blind beamforming for non-Gaussian signals*, IEE Proc.-F, 140 (1994), pp. 362–370. [Matlab code available at http://www.tsi.enst.fr/∼cardoso/stuff.html].

cardososimax96 [12] ———, *Jacobi angles for simultaneous diagonalization*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 161–164.

carroll3 [13] J. Carroll and J. Chang, *Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition*, Psychometrika, 35 (1970), pp. 283–319.

ChevAFC05:ieeesp [14] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, *On the virtual array concept for higher order array processing*, IEEE Trans. Signal Processing, 53 (2005), pp. 1254–1271.

chevalier1 [15] P. Chevalier and A. Ferréol, *On the virtual array concept for the fourth-order direction finding problem*, IEEE Trans. Signal Processing, 47 (1999), pp. 2592–2595.

Como98:spie [16] P. Comon, *Blind channel identification and extraction of more sources than sensors*, in SPIE Conference, San Diego, July 19-24 1998, pp. 2–13. keynote address.

Como02:oxford [17] ——, *Tensor decompositions, state of the art and applications*, in Mathematics in Signal Processing V, J. G. McWhirter and I. K. Proudler, eds., Clarendon Press, Oxford, UK, 2002, pp. 1–24. arXiv:0905. 0454v1.

Como04:ieeesp [18] ——, *Blind identification and source separation in* 2×3 *under-determined mixtures*, IEEE Trans. Signal Processing, 52 (2004), pp. 11–22.

Como04:ijacsp [19] ——, *Contrasts, independent component analysis, and blind deconvolution*, Int. Journal Adapt. Control Sig. Proc., 18 (2004), pp. 225–243. special issue on Signal Separation: http://www3. interscience. wiley. com/cgi-bin/jhome/4508. Preprint: I3S Research Report RR-2003-06.

Como09:cardiff [20] ——, *Tensors, usefulness and unexpected properties*, in IEEE Workshop on Statistical Signal Processing (SSP'09), Cardiff, UK, Aug. 31 - Sep. 3 2009. keynote.

ComoGLM08:simax [21] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1254–1279.

ComoLA09:jchemo [22] P. Comon, X. Luciani, and A. L. F. De Almeida, *Tensor decompositions, alternating least squares and other tales*, J. Chemometrics, 23 (2009).

ComoM96:SP [23] P. Comon and B. Mourrain, *Decomposition of quantics in sums of powers of linear forms*, Signal Processing, 53 (1996), pp. 93–107. special issue on High-Order Statistics.

ComoR04:barcelona [24] P. Comon and M. Rajih, *Blind identification of under-determined complex mixtures of 2 independent sources*, in 3rd IEEE SAM Workshop, Barcelona, Spain, July 18–21 2004.

ComoR06:SP [25] ——, *Blind identification of under-determined mixtures based on the characteristic function*, Signal Processing, 86 (2006), pp. 2271–2281.

ComoTDC09:laa [26] P. Comon, J. M. F. ten Berge, L. De Lathauwer, and J. Castaing, *Generic and typical ranks of multi-way arrays*, Lin. Alg. Appl., 430 (2009), pp. 2997–3007.

Dela06:simax [27] L. De Lathauwer, *A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 642–666.

sobium [28] L. De Lathauwer and J. Castaing, *Blind identification of underdetermined mixtures by simultaneous matrix diagonalization*, IEEE Trans. Signal Processing, 56 (2008), pp. 1096–1105.

DelaCC07:ieeesp [29] L. De Lathauwer, J. Castaing, and J.-F. Cardoso, *Fourth-order cumulant-based identification of underdetermined mixtures*, IEEE Trans. Signal Processing, 55 (2007), pp. 2965–2973.

delathauwerHOS99 [30] L. De Lathauwer, P. Comon, and B. De Moor, *ICA algorithms for 3 sources and 2 sensors*, in Proc. IEEE Signal Processing Workshop on Higher-Order Statistics (HOS'99), Caesarea, Israel, June 14–16 1999, pp. 116–120.

delathauwerEus00 [31] L. De Lathauwer, B. De Moor, and J. Vandewalle, *Algebraic ICA algorithm for 3 sources and 2 sensors*, in Proc. Xth European Signal Processing Conference (EUSIPCO 2000), Tampere, Finland, Sept. 5–8 2000.

delathauwercand [32] ——, *Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 295–327.

DesiL08:simax [33] V. De Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.

dogan [34] M. C. Dogan and J. M. Mendel, *Applications of cumulants to array processing — Part I: Aperture extension and array calibration*, IEEE Trans. Signal Processing, 43 (1995), pp. 1200–1216.

DonoET06:ieeeit [35] D. Donoho, M. Elad, and N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Information Theory, 52 (2006), pp. 6–18.

ErikK04:spl [36] J. Eriksson and V. Koivunen, *Identifiability, separability and uniquness of linear ICA models*, IEEE Sig. Proc. Letters, (2004), pp. 601–604.

ErikK06:ieeeit [37] ——, *Complex random vectors and ICA models: Identifiability, uniqueness and separability*, IEEE Trans. Information Theory, 52 (2006), pp. 1017–1029.

Fell66 [38] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. II, Wiley, 1966, 1971. 2nd ed.

GillM81 [39] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, 1981.

GrelCMT02:ieeesp [40] O. Grellier, P. Comon, B. Mourrain, and P. Trebuchet, *Analytical blind channel identification*, IEEE Trans. Signal Processing, 50 (2002), pp. 2196–2207.

GribB03:ieeesp [41] R. Gribonval and E. Bacry, *Harmonic decomposition of audio signals with matching pursuit*, IEEE Trans. Signal Processing, 51 (2003), pp. 101–111.

harshman2 [42] R. A. Harshman, *Foundations of the Parafac procedure: model and conditions for an "explanatory" multi-mode factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.

Hitc27a:jmp [43] F. L. Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. and Phys., 6 (1927), pp. 165–189.

KagaLR73 [44] A. M. Kagan, Y. V. Linnik, and C. R. Rao, *Characterization Problems in Mathematical Statistics*, Probability and Mathematical Statistics, Wiley, New York, 1973.

KendS77 [45] M. Kendall and A. Stuart, *The Advanced Theory of Statistics, Distribution Theory*, vol. 1-3, C. Griffin, 1977.

KrijDS08:psy [46] W. P. Krijnen, T. K. Dijkstra, and A. Stegeman, *On the non-existence of optimal solutions and the occurrence of degeneracy in the Candecomp/Parafac model*, Psychometrika, 73 (2008), pp. 431–439.

Kroo08 [47] P. M. Kroonenberg, *Applied Multiway Data Analysis*, Wiley, 2008.

Krus77:laa [48] J. B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions*, Lin. Alg. Appl., 18 (1977), pp. 95–138.

KrusHL89:mda [49] J. B. Kruskal, R. A. Harshman, and M. E. Lundy, *How 3-MFA data can cause degenerate Parafac solutions, among other relationships*, in Multiway Data Analysis, R. Coppi and S. Bolasco, eds., Elsevier Science, North-Holland, 1989, pp. 115–121.

LeeS99:nature [50] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization.*, Nature, 401 (1999), pp. 788–791.

LeurRA93:simax [51] S. Leurgans, R. T. Ross, and R. B. Abel, *A decomposition for three-way arrays*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1064–1083.

LimC09:jchemo [52] L.-H. Lim and P. Comon, *Nonnegative approximations of nonnegative tensors*, J. Chemometrics, 23 (2009).

MadsNT04 [53] K. Madsen, H. B. Nielsen, and O. Tingleff, *Methods for Non-Linear Least Squares Problems*, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2004.

Mccu87 [54] P. Mc Cullagh, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability, Chapman and Hall, 1987.

nikias4 [55] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

NionDL08 [56] D. Nion and L. De Lathauwer, *Block component model based blind DS-CDMA receivers*, IEEE Trans. Signal Processing, 56 (2008), pp. 5567–5579.

NionD08:SP [57] D. Nion and L. De Lathauwer, *An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA*, Signal Processing, 88 (2008), pp. 749–755.

Paat97:cils [58] P. Paatero, *A weighted non-negative least squares algorithm for three-way Parafac factor analysis*, Chemometrics Intell. Lab. Syst., 38 (1997), pp. 223–242.

Paat99:jcgs [59] ——, *The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model*, Journal of Computational and Graphical Statistics, 8 (1999), pp. 854–888.

Paat00:jchemo [60] ——, *Construction and analysis of degenerate Parafac models*, J. Chemometrics, 14 (2000), pp. 285–299.

pham [61] D.-T. Pham and J.-F. Cardoso, *Blind separation of instantaneous mixtures of non-stationary sources*, IEEE Trans. Signal Processing, 49 (2001), pp. 1837–1848.

RajiCH08:simax [62] M. Rajih, P. Comon, and R. A. Harshman, *Enhanced line search: A novel method to accelerate Parafac*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1148–1171.

SidiB00:jchemo [63] N. D. Sidiropoulos and R. Bro, *On the uniqueness of multilinear decomposition of N-way arrays*, J. Chemometrics, 14 (2000), pp. 229–239.

SidiBG00:ieeesp [64] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, *Parallel factor analysis in sensor array processing*, IEEE Trans. Signal Processing, 48 (2000), pp. 2377–2388.

SidiGB00:ieeesp [65] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, *Blind Parafac receivers for DS-CDMA systems*, IEEE Trans. Signal Processing, 48 (2000), pp. 810–823.

SmilBG04 [66] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*, Wiley, 2004.

Steg07:psy [67] A. Stegeman, *Degeneracy in Candecomp/Parafac and Indscal explained for several three-sliced arrays with a two-valued typical rank*, Psychometrika, 72 (2007), pp. 601–619.

stegeman09 [68] A. Stegeman and L. De Lathauwer, *A method to avoid diverging components in the Candecomp / Parafac model for generic $I \times J \times 2$ arrays*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1614–1638.

StegS07:laa [69] A. Stegeman and N. D. Sidiropoulos, *On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition*, Lin. Alg. Appl., 420 (2007), pp. 540–552.

StegtBDela06:psy [70] A. Stegeman, J. M. F. ten Berge, and L. De Lathauwer, *Sufficient conditions for uniqueness in Candecomp/Parafac and Indscal with random component matrices*, Psychometrika, 71 (2006), pp. 219–229.

TaleJ99:phoenix [71] A. Taleb and C. Jutten, *On underdetermined source separation*, in ICASSP99, Phoenix, Arizona, March 15-19 1999.

Tenb91:psy [72] J. M. F. ten Berge, *Kruskal's polynomial for $2 \times 2 \times 2$ arrays and a generalization to $2 \times n \times n$ arrays*, Psychometrika, 56 (1991), pp. 631–636.

tomasi [73] G. Tomasi and R. Bro, *A comparison of algorithms for fitting the Parafac model*, Comp. Stat. & Data Anal., 50 (2006), pp. 1700–1734.

vanderveen2 [74] A.-J. van der Veen and A. Paulraj, *An analytical constant modulus algorithm*, IEEE Trans. Signal Processing, 44 (1996), pp. 1136–1155.

vorobyov [75] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, *Robust iterative fitting of multilinear models*, IEEE Trans. Signal Processing, (2005), pp. 2678–2689.

yeredor2000 [76] A. Yeredor, *Blind source separation via the second characteristic function*, Signal Processing, 80 (2000), pp. 897–902.

yeredor2002 [77] ———, *Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation*, IEEE Trans. Signal Processing, 50 (2002), pp. 1545–1553.