

Tensors

[A brief introduction]

Tensor decompositions are at the core of many blind source separation (BSS) algorithms, either explicitly or implicitly. In particular, the canonical polyadic (CP) tensor decomposition plays a central role in the identification of underdetermined mixtures. Despite some similarities, CP and singular value decomposition (SVD) are quite different. More generally, tensors and matrices enjoy different properties, as pointed out in this brief introduction.

MOTIVATION

Originally, BSS exploited mutual statistical independence between sources [20]. Among possible approaches based on the sole hypothesis of source statistical independence, several use cumulants. In fact, when random variables are independent, their cumulant tensor is diagonal [57]. When the source mixture is linear, the decomposition of the data cumulant tensor into a sum of outer products yields the columns of the mixing matrix. This is the first instance of tensor decomposition applied to BSS, even if it is not always explicit. In that case, the tensor is actually symmetric. In the presence of noise, the extraction of sources themselves needs another procedure, based for instance on a spatial matched filter (SMF) [20].

BSS has then been addressed later in different manners. A quite interesting class of approaches consists of exploiting an additional diversity [74]. More precisely, measurements are usually made in two dimensions, generally space and time. But if they are made as a function of three (or more) dimensions, e.g., frequency, polarization, and time repetition, the data are stored in a multiway array. By treating this array as a matrix, information

is lost. Yet, in some real-world applications, it is meaningful to assume a multilinear model for this multiway array, which justifies considering it as a tensor. The decomposition of the latter into a sum of outer products yields not only the columns of the mixture, but also an estimate of the sources. So contrary to the first generation of BSS algorithms, there is no need to resort to an extracting filter. In addition, no statistics are to be estimated, so that the performance is expected to be better for short samples or correlated sources.

Beside numerous books dedicated to applications in physics, there already exist some surveys that can be used in the signal processing field. To begin, some background is presented in [46], i.e., basic engineering tools and a good panel of applications; a more signal processing-oriented tensor overview may be found in [16].

A quite complete digest, more theoretical and oriented toward algebraic geometry, can be found in [49]. This article aims at motivating the signal processing community to dive into the promising world of tensors.

Source Separation and Applications

IMAGE LICENSED BY
INGRAM PUBLISHING

THE WORLD OF TENSORS

Tensors were introduced at the end of the 19th century with the development of the differential calculus. They have then been omnipresent in physics, to express laws independently of coordinate systems. Yet, a tensor is essentially a mapping from a linear space to another, whose coordinates transform multilinearly under a change of bases, as subsequently detailed. For an easier reading, we shall resort to arrays of coordinates, when this indeed eases presentation; interested readers may want to refer to [23] and [49] for a more advanced coordinate-free presentation.

LINEARITY

Linearity expresses the property of a map μ defined on a vector space \mathcal{S} onto another vector space \mathcal{S}' built on the same field \mathbb{K}

that: $\mu(\alpha x + \beta y) = \alpha\mu(x) + \beta\mu(y)$, $\forall x, y \in \mathcal{S}, \alpha, \beta \in \mathbb{K}$. (As far as we are concerned, \mathbb{K} will be either the field of real numbers \mathbb{R} or complex numbers \mathbb{C} .) If \mathcal{S} and \mathcal{S}' are of finite dimension, then this map can be represented by a matrix of coordinates, once the bases of \mathcal{S} and \mathcal{S}' have been fixed. We see that every linear map can be associated with a matrix, say A , so that $\mu(x) = Ax$. On the other hand, every matrix does not uniquely define a map. In fact, a matrix A could, e.g., define a bilinear form from $\mathcal{S} \times \mathcal{S}'$ onto \mathbb{K} , i.e., $f(x_1, x_2) = x_1^T A x_2$. Hence, the correspondence between maps and arrays of coordinates is not one to one.

BILINEARITY

Let's start with a simple example.

EXAMPLE 1

Consider two multidimensional zero-mean random variables \mathbf{z}_1 and \mathbf{z}_2 , and denote the cross-covariance matrix by $\mathcal{G} = \mathbb{E}\{\mathbf{z}_1 \mathbf{z}_2^T\}$. We see that the covariance is linear with respect to \mathbf{z}_1 and \mathbf{z}_2 , which is referred to as *bilinearity*. Now suppose that \mathbf{z}_1 and \mathbf{z}_2 represent two phenomena that are measured in a given coordinate system. \mathcal{G} gives an indication of their correlation. If we change the coordinate system, the covariance matrix changes. More precisely, if $\mathbf{z}'_1 = A\mathbf{z}_1$ and $\mathbf{z}'_2 = B\mathbf{z}_2$, then $\mathcal{G}' = \mathbb{E}\{\mathbf{z}'_1 \mathbf{z}'_2{}^T\}$ can be written $\mathcal{G}' = A\mathcal{G}B^T$. We see that $\mathcal{G}' \neq \mathcal{G}$, whereas the phenomena remain the same. So we must distinguish between the physical phenomena that are coordinate free, and the arrays of measurements we made. And because of bilinearity, we know how to go from one matrix representation to another. We may say that the covariance object is a tensor of order 2, and can be represented by a matrix in any given coordinate system.

What we just saw in Example 1 can be put in more formal terms. Now assume a linear change of coordinates is made in spaces \mathcal{S} and \mathcal{S}' defined by matrices $\{A, B\}$ so that the new coordinates express as $x'_1 = Ax_1$ and $x'_2 = Bx_2$. A tensor \mathcal{G} represented in the original basis with an array \mathcal{G} will be represented (as in Example 1) in the new basis by the new array \mathcal{G}' whose coordinates are:

$$G'_{ij} = \sum_{p,q} A_{ip} B_{jq} G_{pq}$$

This can be compactly denoted by $\mathcal{G}' = (A, B) \cdot \mathcal{G}$. Another notation, equally acceptable, is sometimes used: $\mathcal{G}' = [\mathcal{G}; A, B]$. This will now be extended to orders higher than 2.

MULTILINEARITY

Now assume \mathcal{S}_d are D vector spaces, $1 \leq d \leq D$, and suppose f is a map from $\mathcal{S}_1 \times \dots \times \mathcal{S}_D$ onto \mathbb{K} . Map f is said to be multilinear if $f(x_1, \dots, x_D)$ is linear with respect to every variable x_d , $1 \leq d \leq D$. In other words, $f(x_1, \dots, \alpha x_d + \beta y_d, \dots, x_D) = \alpha f(x_1, \dots, x_d, \dots, x_D) + \beta f(x_1, \dots, y_d, \dots, x_D)$, $\forall d, \forall \alpha, \beta \in \mathbb{K}$. This map is actually a multilinear form. As in the previous section, map f can be represented by an array of coordinates, once the bases of \mathcal{S}_d have been fixed, $1 \leq d \leq D$, and this array needs D indices.

TENSORS

For the sake of simplicity, let us focus on $D = 3$, which is sufficient to give us an idea. Because of multilinearity, special properties are satisfied. For instance, $f(\alpha x_1, x_2, x_3) = f(x_1, x_2, \alpha x_3)$, so that the two triplets of vectors $(\alpha x_1, x_2, x_3)$ and $(x_1, x_2, \alpha x_3)$ have the same image. When dealing with multilinear forms, it is then relevant to consider the equivalence classes defined by the relation $(x, y, z) \sim (x', y', z')$ if there exist $\alpha, \beta, \gamma \in \mathbb{K}$ such that $(x', y', z') = (\alpha x, \beta y, \gamma z)$, with $\alpha\beta\gamma = 1$. Each class may be regarded as a decomposable tensor (decomposable tensors are also called *pure* or *simple*). The space spanned by these classes is denoted as $\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \mathcal{S}_3$, where \otimes is called the *tensor product*. An element of this space is called a *tensor of order 3*. (In physics, the word *rank* is also sometimes used, but we shall avoid it because of the possible confusion with the more standard meaning related to *rank* of a linear operator.) In more mathematical words, one would say that $\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \mathcal{S}_3$ is the quotient space $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3 / \sim$.

EXAMPLE 2

Let $\mathbf{x}_1 \in \mathcal{S}_1$, $\mathbf{x}_2 \in \mathcal{S}_2$, and $\mathbf{x}_3 \in \mathcal{S}_3$. Tensors $6\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$ and $\mathbf{x}_1 \otimes 2\mathbf{x}_2 \otimes 3\mathbf{x}_3$ are the same, but in $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3$, vectors $(6\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and $(\mathbf{x}_1, 2\mathbf{x}_2, 3\mathbf{x}_3)$ are different.

If a linear change of basis is made in space \mathcal{S}_1 (respectively, \mathcal{S}_2 and \mathcal{S}_3), as $x' = Ax$ (resp. $y' = By$ and $z' = Cz$), then the array \mathcal{T}' defining multilinear form f in the new coordinate system expresses as a function of \mathcal{T} . For so-called contravariant tensors, the relationship is

$$T'_{ijk} = \sum_{pqr} A_{ip} B_{jq} C_{kr} T_{pqr} \quad (1)$$

as in Example 1, or in compact form: $\mathcal{T}' = (A, B, C) \cdot \mathcal{T}$. On the other hand, there also exist covariant tensors for which the inverses of the above matrices are instead involved (cf. Example 4), and even mixed tensors that are partly covariant and partly contravariant [23], [71]. However, we shall concentrate only on contravariant tensors in this article, which follow (1) under a multilinear transformation. Note that (1) holds true for contravariant tensors even if the linear transforms (A, B, C) are not invertible; they can even be rectangular matrices. This property is crucial in BSS when mixtures are underdetermined [20], [83].

EXAMPLE 3

Consider three multidimensional random variables \mathbf{x} , \mathbf{y} , and \mathbf{z} . Then the third-order moment tensor \mathcal{M} is represented by the third-order array $M_{ijk} = \mathbb{E}\{x_i y_j z_k\}$. As in the case of second-order moments, it is a contravariant tensor. In fact, if $\mathbf{x}' = A\mathbf{x}$, $\mathbf{y}' = B\mathbf{y}$, and $\mathbf{z}' = C\mathbf{z}$, then $\mathcal{M}' = (A, B, C) \cdot \mathcal{M}$ as in (1). It turns out that cumulants may also be seen as tensors as pointed out in [57]. Because cross-cumulants of independent random variables are null at any order, they have been extensively used in BSS. For instance, the cumulant tensor of order 2 is nothing else but the covariance matrix, and accounts for the correlation at order 2 only; it is not sufficient to account for statistical independence unless variables are Gaussian.

EXAMPLE 4

The derivatives of order D of a multivariate scalar function can be stored in a covariant tensor of order D . For instance at order 2, if $A_{ij} = \partial^2 f / \partial x_i \partial x_j$, $\mathbf{x}' = \mathbf{M}\mathbf{x}$, and $A'_{pq} = \partial^2 f / \partial x'_p \partial x'_q$, then $\mathbf{A}' = \mathbf{N}^T \mathbf{A} \mathbf{N}$, with $\mathbf{N} = \mathbf{M}^{-1}$. From now on and for the sake of simplicity, we shall only consider contravariant tensors in this article.

More generally, a tensor of order D is an element of $\mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_D$, and can be represented by a D -way array \mathcal{T} once bases of spaces \mathcal{S}_d have been fixed. Under multilinear transforms, these arrays of coordinates change similarly to (1).

EXAMPLE 5

In physics, Hooke's law relates the deformation (strain) of a solid under the action of forces (stress). It states that stress \mathcal{F} is related to strain \mathcal{X} by the elasticity tensor as: $\mathcal{F} = \mathcal{C} \bullet \mathcal{X}$, where \bullet is a contraction operator (see the section "Transformations" for a formal definition). Once bases are fixed in the stress and strain spaces, this relationship can be written in terms of arrays of coordinates

$$F_{ij} = \sum_{p,q} C_{ijpq} X_{pq}.$$

The elasticity tensor \mathcal{C} is of order 4. Strain and stress are tensors of order 2, which are represented by matrices.

As illustrated above, it should be kept in mind that an array of coordinates alone does not suffice to define a tensor: spaces and bases need to be defined. Since we are interested mainly in manipulating arrays, and not so much in the map they may represent, arrays will be subsequently associated with multilinear forms, i.e., maps from a product of spaces to their construction field \mathbb{K} . Even if most results can be stated without introducing arrays of coordinates [49], bases are required in engineering applications because calculations are made with arrays of numbers.

NOTATION

In the literature, indices of D -way arrays are sometimes put in superscripts or in subscripts, depending on the covariant or contravariant character of corresponding subspaces; this notation also allows the use the Einstein summation convention. Because we consider essentially fully contravariant tensors in this article, we do not need to make the distinction.

Throughout the article, arrays of numbers will be printed in boldface. More precisely, one- and two-way arrays will be denoted in bold lowercase and bold uppercase, respectively, like, e.g., \mathbf{v} and \mathbf{M} . Arrays with more than two indices will be denoted by bold calligraphic symbols, as \mathcal{A} . Sets and spaces will be noted in script font, like \mathcal{S} , whereas tensors will be printed in calligraphic font, as \mathcal{T} . Entries of arrays \mathbf{v} , \mathbf{M} , and \mathcal{A} will be noted v_i , M_{ij} , and A_{ijk} , without bold font because they are scalar numbers. In practice, a tensor \mathcal{A} is often assimilated to its array representation \mathcal{A} [16], [21], [46], which is generally not very confusing. Nevertheless, we shall make the distinction in the sequel, to keep the presentation as clear as possible.

TRANSFORMATIONS

The tensor product $\mathcal{A} \otimes \mathcal{B}$ between two tensors $\mathcal{A} \in \mathcal{S}_1 \otimes \mathcal{S}_2$ and $\mathcal{B} \in \mathcal{S}_3 \otimes \mathcal{S}_4$ is a tensor of $\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \mathcal{S}_3 \otimes \mathcal{S}_4$. The consequence is that the orders add up under tensor product.

EXAMPLE 6

Let \mathcal{A} be represented by a three-way array $\mathcal{A} = [A_{ijk}]$ and \mathcal{B} by a four-way array $\mathcal{B} = [B_{lmnp}]$; then tensor $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$ is represented by the seven-way array of components $C_{ijklmnp} = A_{ijk} B_{lmnp}$. With some abuse of notation, the tensor product is often applied to arrays of coordinates, so that notation $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$ may be encountered.

If the tensor product increases the order, the contraction decreases it by two. The contraction consists of a summation over a pair of indices. This operation permits to define the mode- k product between tensors, and can be denoted by \bullet_k , where k indicates which index should be summed.

EXAMPLE 7

If \mathcal{A} and \mathcal{A}' are tensors of order D and D' , the tensor $\mathcal{B} = \mathcal{A} \bullet_k \mathcal{A}'$ is a tensor of order $D + D' - 2$ obtained by summing over the k th index. For instance, if $(D, D', k) = (3, 3, 2)$, this yields $B_{ijpq} = \sum_l A_{ilj} A'_{lpq}$. For $(D, D', k) = (2, 2, 1)$, we would have the matrix product $\mathcal{A} \bullet_1 \mathcal{A}' = \mathbf{A}^T \mathbf{A}'$. However, when the product is between a matrix and a tensor of higher order, it has been the usual practice to always sum over the second matrix index. For instance, if \mathbf{M} is a matrix, $\mathcal{A} \bullet_3 \mathbf{M}$ means that the sum is performed on the third tensor index and the second matrix index.

It may be convenient to store D -way arrays in matrices. This transformation is called matrix *unfolding* or *flattening*, and can be performed in different manners, depending on the arbitrarily chosen ordering [27], [46]. Here, the ordering of [46] has been retained, but the choice of [27] would work equally well. In fact, the exact definition is not so important, provided the inverse map is defined consistently. We shall limit ourselves to matrices whose number of rows equals one of the tensor dimensions; this is sometimes referred to as *mode- n unfolding* [46]. Example 8 illustrates how to relate a third-order tensor to its three flattening matrices. But it is also possible to associate a D -way array, $D > 3$, to a multilinear operator of lower order; see, e.g., [9], [29], [64], and [69].

EXAMPLE 8

Let a $2 \times 2 \times 2$ array of coordinates A_{ijk} . Its mode- n unfoldings $\mathbf{A}^{(n)}$ are

$$\begin{aligned} \mathbf{A}^{(1)} &= \begin{bmatrix} A_{111} & A_{121} & A_{112} & A_{122} \\ A_{211} & A_{221} & A_{212} & A_{222} \end{bmatrix} \\ \mathbf{A}^{(2)} &= \begin{bmatrix} A_{111} & A_{211} & A_{112} & A_{212} \\ A_{121} & A_{221} & A_{122} & A_{222} \end{bmatrix} \\ \mathbf{A}^{(3)} &= \begin{bmatrix} A_{111} & A_{211} & A_{121} & A_{221} \\ A_{112} & A_{212} & A_{122} & A_{222} \end{bmatrix}. \end{aligned}$$

Remark that the row number of matrix $\mathbf{A}^{(n)}$ corresponds to the n th index of tensor \mathcal{A} .

The Kronecker product between two matrices \mathbf{A} and \mathbf{B} of size $I \times J$ and $K \times L$, respectively, is the matrix $\mathbf{A} \boxtimes \mathbf{B}$ of size $IK \times JL$ defined blockwise by

$$\mathbf{A} \boxtimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & \dots & A_{1J}\mathbf{B} \\ \vdots & & \vdots \\ A_{I1}\mathbf{B} & \dots & A_{IJ}\mathbf{B} \end{bmatrix}.$$

The Kronecker product is used to represent the tensor product when bases are fixed and when tensors are represented by their array of coordinates unfolded into matrices. It should be borne in mind that the Kronecker product usually applies to matrices (although an extended definition has recently been proposed in [63]), whereas the tensor product is more general and coordinate free. Hence they should not be confused.

SPECIAL TENSORS

A particularly important class of tensors is that of *decomposable* tensors, which are tensor products of vectors. As previously stated in the section “Tensors,” they are of the form $\mathcal{D} = \mathbf{u} \otimes \mathbf{v} \otimes \dots \otimes \mathbf{w}$, and span the whole tensor space. The corresponding array of coordinates is $D_{ij..k} = u_i v_j \dots w_k$. One can view these tensors as a discretization of a multivariate function whose variables separate.

EXAMPLE 9

Take a function of two variables with separated variables: $f(x, y) = u(x)v(y)$. Then its discretization takes the form $f(x_i, y_j) = u(x_i)v(y_j)$, and these numbers can be stored in a rank-1 matrix $\mathbf{D} = \mathbf{u}\mathbf{v}^T$.

A tensor is cubical if all its spaces of construction are identical, with the same basis. (The term *homogeneous* is also used in physics.) A cubical tensor \mathcal{A} is symmetric if its array of coordinates is invariant under permutation of its indices: $A_{\sigma(ij..k)} = A_{ij..k}$, $\forall \sigma$.

EXAMPLE 10

The tensor of moments and the tensor of derivatives, defined in Examples 1, 3, and 4, are symmetric.

The simplest symmetric array is the diagonal one, defined by $\Delta_{ij..k} = 0$ if $(i, j, \dots, k) \neq (i, i, \dots, i)$.

DECOMPOSITIONS AND RANKS

TENSOR RANK

Any tensor \mathcal{T} can be decomposed (nonuniquely) into a linear combination (with coefficients in \mathbb{K}) of decomposable tensors

$$\mathcal{T} = \sum_{r=1}^R \lambda_r \mathcal{D}(r), \quad (2)$$

$\mathcal{D}(r) = \mathbf{a}_r \otimes \mathbf{b}_r \otimes \dots \otimes \mathbf{c}_r$. If tensor spaces are endowed with scalar products, one can impose decomposable tensors $\mathcal{D}(r)$ to be built with unit norm vectors, which permits us to impose $\lambda_r \in \mathbb{R}^+$ if desired. The smallest value R for which (2) holds is called the *tensor rank*. The definition of tensor rank can be traced back to the beginning of the 20th century [38], but it has

been reintroduced in other disciplines under various names [7], [12], [36], [39], [66], [82].

EXAMPLE 11

Let the arrays \mathcal{A} and \mathcal{B} of dimensions $2 \times 2 \times 2$ be defined by their mode-1 unfoldings

$$\mathbf{A}^{(1)} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \mathbf{B}^{(1)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Tensor $\mathcal{A} = [1, 0] \otimes [1, 0] \otimes [1, 1]$ has rank 1. Tensor \mathcal{B} is symmetric and has rank 3, as will be seen with \mathcal{T}_0 in Example 18.

Note that, by definition, a tensor is decomposable if and only if it has rank 1. If the order of a tensor \mathcal{T} is ≥ 3 , the rank may depend on the field, in the sense that a real tensor of rank R may have smaller rank if we allow the decomposition (2) to be complex, as demonstrated in Example 12.

EXAMPLE 12

Take a real symmetric array \mathcal{Y} of dimensions $2 \times 2 \times 2$, defined by its mode-1 unfolding

$$\mathbf{Y}^{(1)} = \begin{bmatrix} 2 & 0 & 0 & -2 \\ 0 & -2 & -2 & 0 \end{bmatrix}.$$

Then, we need three decomposable tensors in \mathbb{R} :

$$\mathcal{Y} = 4 \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} -1 \\ -1 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} -1 \\ 1 \end{bmatrix}^{\otimes 3}$$

but only two in \mathbb{C} , setting $j = \sqrt{-1}$:

$$\mathcal{Y} = \begin{bmatrix} 1 \\ j \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 1 \\ -j \end{bmatrix}^{\otimes 3}.$$

Hence its tensor rank in \mathbb{R} is 3 whereas it is 2 in \mathbb{C} .

Other examples may be found in [18], [46], and [48]. Examples 11 and 12 incidentally show that, unlike matrix rank, tensor rank may exceed all dimensions.

TUCKER DECOMPOSITION

At this stage, it is interesting to make a connection with the matrix SVD. Two important features characterize the SVD of a matrix \mathbf{M} :

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (3)$$

specifically 1) \mathbf{U} and \mathbf{V} have orthonormal columns, and 2) $\mathbf{\Sigma}$ is diagonal. Consider the decomposition below of a three-way array, introduced by Tucker in the 1960s [85]:

$$\mathcal{T}_{ijk} = \sum_p \sum_q \sum_r A_{ip} B_{jq} C_{kr} G_{pqr}, \quad (4)$$

which we shall compactly denote $\mathcal{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{G}$. It is clear that if the number of free parameters in the right-hand side of (4) is smaller than the number of equations, then there will generally be no solution. This happens to be the case if \mathbf{A} , \mathbf{B} , and \mathbf{C} are

orthonormal and \mathcal{G} is diagonal. In the quest for existence, we have to choose: either \mathcal{G} is diagonal, but we have to relax the orthogonality constraint on factor matrices, which will be allowed to have more columns than rows [this corresponds to decomposition (2)], or we keep the orthonormality constraint, but allow \mathcal{G} to have nonzero extra-diagonal entries as elaborated in the next section.

HOSVD AND MULTILINEAR RANKS

If we impose matrices $\{A, B, C\}$ to have orthogonal and unit-norm columns in the Tucker decomposition (4), then we can make several observations. First, denote by R_n the rank of $\mathbf{T}^{(n)}$, the n th unfolding matrix of \mathcal{T} , $1 \leq n \leq D = 3$. Rank R_n is called *mode- n rank* of \mathcal{T} , or n -rank in short. Then the number of columns of A (respectively, B , C) does not need to exceed R_1 (respectively, R_2 , R_3), and the dimension of the core tensor may be imposed to be $R_1 \times R_2 \times R_3$. In addition R_n cannot exceed the tensor rank R defined in (2), nor the n th dimension. This property is not a surprise, if we view decomposition (2) as a decomposition of the n th unfolding matrix into a sum of rank-1 matrices where rows are imposed to have a special structure. The D -uple of n -ranks is the *multilinear rank* of \mathcal{T} . Another property is less immediate to capture: the core array \mathcal{G} can be imposed to be all-orthogonal, which means that all tensor slices of order $D - 1$ are orthogonal to each other in every mode; when $D = 3$ this means:

$$\sum_{j,k} G_{\alpha j k} G_{\beta j k} = \sum_{i,k} G_{i \alpha k} G_{i \beta k} = \sum_{i,j} G_{i j \alpha} G_{i j \beta} = 0$$

if $\alpha \neq \beta$. See [27] and references therein for more details. It is worth noticing the elementary fact that for tensors of order 2 (i.e., matrices), $R_1 = R_2 = R$, and all equal the matrix rank.

EXAMPLE 13

The multilinear rank of array \mathcal{B} defined in Example 11 is $(2, 2, 2)$, whereas that of \mathcal{A} is $(1, 1, 1)$.

CP DECOMPOSITION

On the contrary, if we keep a diagonal form for \mathcal{G} , we end up with the polyadic decomposition [38], also sometimes called CANDECOMP or PARAFAC because of its rediscovery in the 1970s:

$$\mathcal{T}_{ijk} = \sum_{r=1}^R A_{ir} B_{jr} C_{kr} \lambda_r \quad (5)$$

or, in compact form, $\mathcal{T} = (A, B, C) \cdot \mathcal{L}$, where \mathcal{L} is diagonal. If R is not too large, this decomposition can be unique (cf. the section “Exact Decompositions”) and deserves to be called *canonical polyadic* (CP). Following a practice now adopted in applied mathematics and engineering [5], [42], we shall subsequently use the acronym CP, which can also cleverly stand for CANDECOMP/PARAFAC. After inspection, it may be seen that (5) is nothing else but decomposition (2) in array coordinates. In other words, the CP decomposition reveals the tensor rank.

SYMMETRIC RANK

As already pointed out in the section “Special Tensors,” a tensor \mathcal{T} is symmetric if its coordinate array \mathcal{T} is invariant by permutations of indices. If we impose tensors $\mathcal{D}(r)$ in (2) to be themselves symmetric, then we might end up with a larger value of rank, denoted R_s , which is referred to as the *symmetric rank* of \mathcal{T} . It is clear that $R_s \geq R$ for any symmetric tensor \mathcal{T} , since any constraint on decomposable tensors may increase rank; we have already observed this fact with the real constraint in Example 12. It has been conjectured in [19] that rank and symmetric rank are always equal, but this has not yet been proved in the general case.

NONNEGATIVE RANK

When an array is real nonnegative, one may want to impose rank-1 terms in its CP decomposition to be themselves nonnegative. The minimal number of terms is then called the *nonnegative rank* and is generally strictly larger than the rank in \mathbb{R} . This is already the case for matrices ($D = 2$) as shown in Example 14, due to Herbert E. Robbins. The same phenomenon is observed for tensors, although theoretical results are still lacking.

EXAMPLE 14

The following matrix has rank 3 since vector $[1, -1, -1, 1]$ belongs to its kernel. But it can be proved that its nonnegative rank is four

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

STRUCTURED RANKS

More generally, when matrix factors are imposed to have a special structure, such as banded, van der Monde, Toeplitz, or Hankel, the tensor rank may increase, just as in the nonnegative case. Structure can also have an impact on computational issues [49], [78].

BORDER RANK

A tensor has border rank R if it is the limit of tensors of rank R and not the limit of tensors of smaller rank. Rank and border rank always coincide for matrices, but not for tensors of order larger than two, as shown in the next example.

EXAMPLE 15

Let u and v be fixed real or complex numbers, and ε a small positive number. Then $(1/\varepsilon)[(u + \varepsilon v)^3 - u^3] = 3u^2 v + O(\varepsilon)$. Now if multiplication is not commutative, we have three distinct terms on the right-hand side; this is what happens for the tensor product, so that $\forall \varepsilon > 0$:

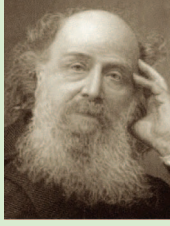
$$\mathcal{T}_\varepsilon = \frac{1}{\varepsilon}[(\mathbf{u} + \varepsilon \mathbf{v})^{\otimes 3} - \mathbf{u}^{\otimes 3}] = \mathcal{T}_0 + O(\varepsilon),$$

$$\mathcal{T}_0 = \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u}$$

hold for any vectors \mathbf{u} and \mathbf{v} . If the latter are not collinear, it can be proved that \mathcal{T}_0 is of rank $R = 3$, but is the limit of tensors $\mathcal{T}_{\varepsilon_i}$, which are all of rank 2. Hence the border rank of \mathcal{T}_0 is $\underline{R} = 2$.



E. Waring
(1736–1798)



J.J. Sylvester
(1814–1897)



A. Clebsch
(1833–1872)

[FIG1] Some contributors to the field of tensors.

The border rank has been defined and utilized by many authors, especially in arithmetic complexity [7], [52], [72], [82]. This concept is crucial in tensor approximation problems, as addressed in the section “Approximate Decompositions.”

RELATION WITH POLYNOMIALS

Homogeneous polynomials are bijectively related to tensors, which allows to transpose existing results of algebraic geometry; see, e.g., [10], [15], [17], [19], [23], [49], [80], and references therein. In fact, one can associate the following polynomial with any array \mathcal{T} :

$$p(\mathbf{x}, \mathbf{y}, \dots, \mathbf{z}) = \sum_{i,j,k} T_{i,j,k} x_i y_j \dots z_k.$$

Conversely, any homogeneous polynomial of degree D and partial degree 1 in every variable can be associated with a (nonsymmetric) tensor \mathcal{T} .

Through this bijection, a decomposable tensor of order D is translated into a product of D linear forms, and the CP decomposition can be translated into a linear combination of such terms:

$$p(\mathbf{x}, \mathbf{y}, \dots, \mathbf{z}) = \sum_{r=1}^R \lambda_r (\mathbf{a}_r^\top \mathbf{x}) (\mathbf{b}_r^\top \mathbf{y}) \dots (\mathbf{c}_r^\top \mathbf{z}). \quad (6)$$

In the case of symmetric tensors, $\mathbf{x} = \mathbf{y} = \dots = \mathbf{z}$. More precisely, a symmetric tensor \mathcal{T} of order D can be identified with the homogeneous polynomial of degree D :

$$p(\mathbf{x}) = \sum_{i,j,\dots,k} T_{i,j,\dots,k} x_i x_j \dots x_k$$

in the indeterminates x_1, \dots, x_n . It can be easily checked that symmetric tensors of rank 1 are associated with a polynomial of the form $a(\mathbf{x})^D$, where $a(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ is a linear form. In other words, they are exactly the D th powers of a homogeneous linear form. The CP decomposition of \mathcal{T} reduces in this case to

$$p(\mathbf{x}) = \sum_{i=1}^{R_s} a_i(\mathbf{x})^D, \quad (7)$$

which has been classically called a Waring decomposition [40]. The minimum number of summands R_s in a Waring decomposition is the symmetric rank of \mathcal{T} , which we defined earlier.

EXAMPLE 16

The polynomials associated with tensors \mathcal{A} and \mathcal{B} of Example 11 are, respectively, $a(x_1, x_2, y_1, y_2, z_1, z_2) = x_1 y_1 (z_1 + z_2)$ and $b(x_1, x_2) = 3x_1^2 x_2$.

EXAMPLE 17

Take the polynomial of degree $D = 3$:

$$\begin{aligned} 2x_1^3 - 6x_1 x_2^2 &= (x_1 + jx_2)^3 + (x_1 - jx_2)^3 \\ &= 4(x_1)^3 - (x_1 + x_2)^3 - (x_1 - x_2)^3, \end{aligned}$$

where $j = \sqrt{-1}$. It has complex symmetric rank equal to two and real symmetric rank equal to three. This polynomial is actually associated with tensor \mathcal{Y} given in Example 12.

EXAMPLE 18

Example 15 can be written in terms of polynomials, and is even easier to understand this way. Take $\mathbf{u} = [1, 0]$ and $\mathbf{v} = [0, 1]$. Then $\mathbf{u}^{\otimes 3}$ and $\mathbf{v}^{\otimes 3}$ are associated with polynomials x_1^3 and x_2^3 respectively, whereas $(\mathbf{u} + \epsilon \mathbf{v})^{\otimes 3}$ is associated with $(x_1 + \epsilon x_2)^3$, which can be expanded as $x_1^3 + 3\epsilon x_1^2 x_2 + o(\epsilon)$. This shows that \mathcal{T}_ϵ is associated with $3x_1^2 x_2 + o(\epsilon)$. Hence \mathcal{T}_ϵ tends to \mathcal{T}_0 , because \mathcal{T}_0 is associated with $3x_1^2 x_2$. Moreover, the rank of \mathcal{T}_0 is three because $3x_1^2 x_2$ cannot be written as a sum of fewer than three cubes.

EXACT DECOMPOSITIONS

Now one can ask whether the CP decomposition defined in (2) and (5) is unique or not. First of all, the D -way array associated with a D th order decomposable tensor \mathcal{D} is not uniquely represented by an outer product of vectors: there remain $D - 1$ scaling factors of unit modulus. So we are rather interested in the uniqueness of coefficients $\lambda_r \in \mathbb{R}^+$ and tensors $\mathcal{D}(r)$, which is more meaningful; this is sometimes called *essential uniqueness*. In this section, we will see two ways of assessing uniqueness: almost surely or deterministically.

EXPECTED RANK

A naive approach is to count the number of degrees of freedom on both sides of (6), which is a rewriting of (2) in terms of polynomials, and say that the number of equations should be at least as large as the number of unknowns. To fix the ideas, take a tensor of order D and dimensions $n_1 \times \dots \times n_D$. It is clear that a necessary condition for uniqueness of the CP decomposition is that

$$R \left(\sum_{i=1}^D n_i - D + 1 \right) \leq N, \quad (8)$$

where $N = \prod_i n_i$. We can proceed similarly for symmetric tensors and count equations and unknowns in (7). This leads to

$$R_s n \leq N_s, \quad (9)$$

where $N_s = \binom{n+D-1}{D}$ corresponds to the number of free parameters in a symmetric tensor. Equations (8) and (9) induce an upper bound on rank, which is called the *expected rank*, and is defined as

$$R \leq R^o = \left\lfloor \frac{N}{1 - D + \sum_i n_i} \right\rfloor \quad (10)$$

$$R_s \leq R_s^o = \left\lfloor \frac{N_s}{n} \right\rfloor. \quad (11)$$

When the fraction above is not an integer, there will always be an infinity of solutions, because of too many free parameters.

When it is an integer, the number of unknowns is equal to the number of equations, and we could expect that there is a finite number of solutions. However, things are not so simple, as pointed out by Clebsch in the 19th century (Figure 1). In fact, there are exceptions [1], [3], [23], [58].

EXAMPLE 19

Consider fourth-order symmetric tensors of dimension three. In that case, $N_s/n = \binom{6}{4}/3 = 5$ is an integer. Our hope is not realized since five forms are generally not sufficient in their decomposition. This exception was first noticed by Clebsch (Figure 1) from the polynomial framework: the “generic rank” of ternary quartics is in fact six [33]. This means that most homogeneous polynomials of degree four in three variables in \mathbb{C} can be written as a sum of six linear forms raised to the fourth power, and not fewer with probability 1.

TYPICAL AND GENERIC RANKS

Generic (resp. typical) ranks are the ranks that we encounter with probability one (resp. nonzero probability), when their entries are drawn independently according to a continuous probability distribution, hence their importance. Contrary to the matrix case, they are not maximal; tables of rank values may be found in [24], as well as simple codes to compute numerically the generic rank of a large panel of tensors. (Codes can be downloaded from [94].)

A striking fact is that only one rank occurs with probability one (the so-called generic rank) in \mathbb{C} , whereas several typical ranks may exist in \mathbb{R} . The generic rank in \mathbb{C} is always equal to the smallest typical rank one would find in \mathbb{R} . This problem was first addressed by Sylvester (Figure 1) in the 19th century. The case of real symmetric tensors of dimension two is now well understood [13], [22], [67]. In fact, all the integers between $\lfloor (D+2)/(2) \rfloor$ and D have been shown to be typical ranks [8]. If the tensor rank is smaller than a bound depending on the generic rank [typically $R^o - 1$ as defined in (10) and (11)], there exist almost surely finitely many CP decompositions. See [23] for a survey of recent results on almost sure uniqueness.

UNIQUENESS RESULTS BASED ON LINEAR ALGEBRA

Instead of associating tensors with polynomials and making use of results borrowed from algebraic geometry, uniqueness conditions can be obtained by considering particular factor matrices. However, these conditions are generally only sufficient [41], and often much more restrictive. The most well known is that published by Kruskal [47] and extended later in [73] and [81]; alternate proofs have been derived in [49] and [68]. It requires the following definition: The Kruskal rank of a matrix is the largest number κ such that any subset of κ columns is full rank. By construction, Kruskal’s rank cannot exceed matrix rank.

EXAMPLE 20

The matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

has rank 3, but its Kruskal rank is $\kappa = 2$.

The CP decomposition is unique if the sufficient condition holds:

$$2R + D - 1 \leq \sum_{d=1}^D \kappa_d, \quad (12)$$

where κ_d denotes the Kruskal rank of the d th factor matrix in the CP decomposition. Further recent deterministic results may be found in [25], [31], and [32]. These results do not need algebraic geometry but advanced linear algebra (i.e., compound matrices formed of minors). They are sometimes much more powerful than Kruskal’s bound.

EXACT COMPUTATION

Due to the space restrictions in this article, various existing algorithms will not be described. However, we provide below some guidance to related literature, among many others. In [6], algorithms to compute the symmetric rank of symmetric tensors of small border rank are proposed. When the rank is small, the symmetric CP decomposition can be computed with the help of Sylvester’s algorithm [10]; when it is not unique, one CP decomposition can still be delivered. In [60], approaches based on special eigenvector computations are proposed. Direct computation is proposed in [4] for $2 \times n \times n$ arrays.

When one tensor dimension is large compared to its rank and to other dimensions, it is possible to compute the CP decomposition via a joint congruent diagonalization of its matrix slices; this has been first proposed in [50] for two matrix slices. In the presence of errors with more than two slices, such a diagonalization becomes approximate [25] and needs more care (see the next section). In a similar spirit, for low-rank tensors of order larger than three, one can also decrease the order by working jointly on tensor slices of lower orders [29], or by rearranging the original tensor into another of lower order but larger dimensions [64].

APPROXIMATE DECOMPOSITIONS

In practice, measurements are always corrupted by some noise, which almost always has a continuous probability distribution. For this reason, the tensor rank is generic or typical, and the CP decomposition is generally not unique. That’s why a best rank- r approximation must be computed [21], [44]. General-purpose optimization algorithms will generally suffice to solve the problem, e.g., [21], [46], [65], [77], and [84]; they are widely used but their convergence toward a minimum is not guaranteed, because the objective function may have only an infimum.

In fact, low-rank approximations are useful and even unavoidable, but unfortunately ill posed in general [37], [75], except for special cases of tensors under constraints, like non-negativity [54]. Most algorithms presently utilized by engineering communities ignore this fact, which may raise serious practical problems in a small fraction of cases.

Ill posedness comes from the fact that the set of tensors of rank at most R is not closed, as pointed out in section “Border Rank.” Some remedies have been proposed in the literature to face or circumvent this difficulty. In practice, this means that another problem is solved, often by imposing constraints in the CP decomposition.

These include:

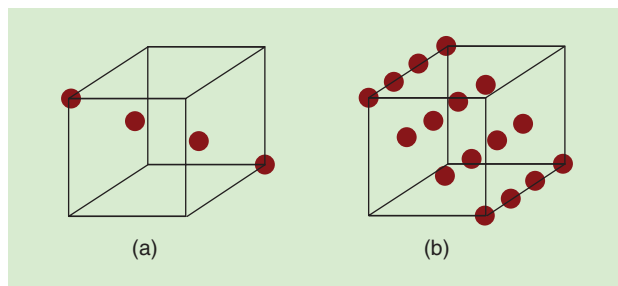
- Impose orthogonality between columns of factor matrices [20]—in BSS, this takes the form of a spatial prewhitening.
- Impose orthogonality between decomposable tensors [45].
- Prevent divergence by bounding coefficients λ_r [54], [61].
- If the tensor is nonnegative, use a nonnegative CP [54].
- Impose a minimal angle between columns of factor matrices [55].
- Compute an exact CP of another tensor, which has undergone a multilinear compression via truncated HOSVD [11], [21]. It may happen that the problem remains ill posed after this type of compression, because reducing the mode-ranks does not necessarily reduce tensor rank.
- Compute another decomposition where the core tensor is block diagonal instead of diagonal [26], [79].
- Compute a joint approximate diagonalization (JAD) of matrix slices, which may be viewed as another decomposition where the core tensor is not diagonal [2], [14], [20], [30], [51], [56], [62], [69], [86], [87], [89], as depicted in Figure 2. The drawbacks of this family of approaches, which become more and more popular, are threefold. First, rank must be smaller than two dimensions; in [25], the latter constraint is nevertheless relaxed. Second, replacing the core tensor by its diagonal yields an approximate CP decomposition whose optimality is not known. Third, a closed subclass of invertible matrices needs to be (arbitrarily) chosen, and indeed varies from one algorithm to another.
- When one dimension is much larger than the others, the optimality of this kind of approach can be significantly improved by imposing a structure in the diagonalization process [25].

Some codes are freely available on the Internet. See, e.g., the home pages of R. Bro, L. De Lathauwer, T. Kolda, A.H. Phan, and P. Comon [90]–[94]. A good site to find applications and related references is the Three-Mode Company's maintained by P. Kroonenberg [95].

THE CASE OF RANK-1 APPROXIMATE

The rank-1 approximation problem is of interest for at least two reasons: first it is always well posed, and second it shows up in the deflation approach of BSS [20]. In addition, it is much easier to compute than a full CP decomposition [28], [43]. This problem may be seen to be related to tensor eigenvalues [17], [35], [53], [59], [88]. It has been proved recently that the best rank-1 approximation of a symmetric tensor is symmetric [34]; a shorter proof can be found in [35], as well as uniqueness issues. So a question deserves to be raised: can the exact or approximate CP decompositions be computed by successive rank-1 approximations? It is already known that this does not generally work.

In fact, attention should be paid to the fact that subtracting the best rank-1 approximate does not decrease tensor rank in general [80], contrary to the matrix case. Simple examples may be found in [18]; similar examples also exist for nonsymmetric or nonnegative tensors. The consequence is that the rank-1 terms appearing in the best rank- k tensor approximation are not the same for different values of k . Hence, it is not possible



[FIG2] Because the optimization criteria are different in JAD and CP decompositions, one does not attempt to zero the same entries. This figure shows the location of the entries that are (a) not minimized in the core tensor in the CP decomposition of a third-order $4 \times 4 \times 4$ tensor, and (b) during the execution of a JAD algorithm. Note that JAD algorithms deliver two factor matrices; the entries of third one remain in the core tensor.

to compute a full CP decomposition by solving successive best rank-1 approximations, contrary to what has been claimed by some authors. This procedure, called *deflation*, works in BSS for other reasons. In fact, BSS does not only reduce to a low-rank tensor approximation, but also includes a regression stage. However, whether deflation works in special cases (such as structured CP decompositions) is still an open question.

EXAMPLE 21

The tensor defined by its mode-1 unfolding

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{bmatrix}$$

is of rank 2. Its best rank-1 approximate is [80]:

$$\mathbf{Y} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}.$$

One checks out that the difference

$$\mathbf{T} - \mathbf{Y} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

is of rank 3. In this example, deflation does not permit to decrease tensor rank.

APPLICATIONS

Applications of tensor decompositions (essentially CP) include arithmetic complexity, separation of variables, blind identification of linear mixtures, BSS, data mining, spectroscopy, antenna array processing, and phylogenetics, among others. Tucker and HOSVD have other application fields in which uniqueness is not requested, like data compression. Due to the space constraints of this article, we shall now detail only one application of the CP decomposition, particularly fluorescence spectroscopy [76], for which very few theoretical results can apply, unfortunately. The reader is invited to consult, e.g., [16], [20], and [46] for pointers to other applications.

An optical excitation applied to a solution produces several effects, including Rayleigh and Raman diffusions, and fluorescence. If the latter effect can be isolated, it may allow to accurately measure the relative concentrations of fluorescent solutes. In fact, at low concentrations and in the presence of R

fluorescent solutes, the Beer–Lambert law can be linearized and takes the form

$$\mathcal{T}(x, y, z) = T_o \sum_{\ell=1}^R \gamma_{\ell}(x) \epsilon_{\ell}(y) c_{\ell}(z),$$

where x , y , and z denote the fluorescence emission wavelength, the excitation wavelength, and the sample number, respectively, \mathcal{T} is the fluorescence intensity measured as a function of the latter variables, $\gamma_{\ell}(x)$ denotes fluorescence emission spectrum of the ℓ th solute, $\epsilon_{\ell}(y)$ its absorbance spectrum (sometimes called excitation spectrum), and $c_{\ell}(z)$ its relative concentration. In practice, only a finite number of samples are available, and measurements are made on discrete values within a limited spectral range, so that variables x , y , and z take a finite number of values. In other words, we deal with a CP decomposition of a finite three-way array, often of rather large dimensions (several hundreds). The particularity of this CP decomposition is that \mathcal{T} is real nonnegative, as well as all the terms involved in its CP decomposition. Hence, R is the nonnegative rank of \mathcal{T} . The good news is that 1) the best low-rank approximate always exists [54], and that 2) there are simple efficient numerical algorithms available for its computation [70]. The bad news is that known uniqueness results, which we have reviewed in this article, are not appropriate for nonnegative CP decompositions. For instance, if nonnegative rank is plugged in place of rank in (12), the obtained sufficient condition is more restrictive, and does not even guarantee that factor matrices are nonnegative. This is the subject of ongoing research.

ACKNOWLEDGMENTS

This work has been funded by the European Research Council under the European Community's Seventh Framework Programme FP7/2007–2013 grant agreement number~320594.

AUTHOR

Pierre Comon (p.comon@ieee.org) graduated in 1982 and received the doctorate degree in 1985 both from the University of Grenoble, France. He later received the Habilitation to Lead Research in 1995, from the University of Nice, France. His research interests include high-order statistics (HOS), BSS, statistical signal and array processing, tensor decompositions, multiway factor analysis and data mining, with applications to biomedical end environment. He was an associate editor of *IEEE Transactions on Signal Processing* (1995–1998), the EURASIP journal *Signal Processing* (2006–2011), and a member of the French National Committee of Scientific Research (1995–2000). Between 2001 and 2004, he was an associate editor of *IEEE Transactions on Circuits and Systems I*. He is a member of the editorial board of *SIAM Journal on Matrix Analysis and Applications* since January 2011. He is a Fellow of the IEEE.

REFERENCES

- [1] H. Abo, G. Ottaviani, and C. Peterson, "Induction for secant varieties of Segre varieties," *Trans. Amer. Math. Soc.*, vol. 361, pp. 767–792, 2009.
- [2] L. Albera, A. Ferreol, P. Chevalier, and P. Comon, "ICAR, a tool for blind source separation using fourth order statistics only," *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3633–3643, Oct. 2005.

- [3] J. Alexander and A. Hirschowitz, "Polynomial interpolation in several variables," *J. Algebr. Geom.*, vol. 4, no. 2, pp. 201–222, 1995.
- [4] 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*, vol. 56, no. 4, pp. 631–636, 1991.
- [5] J. M. F. Ten Berge, "Simplicity and typical rank of three-way arrays," *J. Chemomet.*, vol. 18, no. 1, pp. 17–21, 2004.
- [6] A. Bernardi, A. Gimigliano, and M. Ida, "Computing symmetric rank for symmetric tensors," *J. Symb. Computat.*, vol. 46, no. 1, pp. 34–53, 2011.
- [7] D. Bini, M. Capovani, F. Romani, and G. Lotti, "O(n^{2.77}) complexity for nxn approximate matrix multiplication," *Inform. Process. Lett.*, vol. 8, no. 5, pp. 234–235, 1979.
- [8] G. Blekherman, "Typical real ranks of binary forms," *Found. Computat. Math.*, to be published. DOI 10.1007/s10208-013-9174-8. arXiv:1205.3257
- [9] M. M. Boizard, G. Ginolhac, F. Pascal, S. Miron, and P. Forster, "Numerical performance of a tensor MUSIC algorithm based on HOSVD," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Marrakech, Morocco, Dec. 13–15, 2013.
- [10] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, "Symmetric tensor decomposition," *Linear Algebr. Appl.*, vol. 433, nos. 11–12, pp. 1851–1872, Dec. 2010.
- [11] R. Bro and C. A. Andersson, "Improving the speed of multiway algorithms. Part II: Compression," *Chemomet. Intell. Lab. Syst.*, vol. 42, nos. 1–2, pp. 105–113, 1998.
- [12] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-Young decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sept. 1970.
- [13] A. Causa and R. Re, "On the maximum rank of a real binary form," *Annali di Matematica Pura e Applicata*, vol. 190, no. 1, pp. 55–59, 2011.
- [14] G. Chabriel, M. Kleinstaubert, E. Moreau, H. Shen, P. Tichavsky, and A. Yeredor, "Joint matrices decompositions and blind source separation," *IEEE Signal. Processing Mag.*, vol. 31, no. 3, pp. 34–43, May 2014.
- [15] L. Chiantini and G. Ottaviani, "On generic identifiability of 3-tensors of small rank," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 3, pp. 1018–1037, 2012.
- [16] A. Cichocki, D. Mandic, A.-H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing applications," *IEEE Signal. Processing Mag.*, to be published.
- [17] P. Comon, "Tensor decompositions, state of the art and applications," in *Mathematics in Signal Processing V*, J. G. McWhirter and I. K. Proudler, Eds. Oxford, U.K.: Clarendon Press, 2002, pp. 1–24.
- [18] P. Comon, "Tensors, usefulness and unexpected properties," in *Proc. 15th IEEE Workshop on Statistical Signal Processing (SSP'09)*, Cardiff, U.K., 31 Aug.–3 Sept., 2009, pp. 781–788.
- [19] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensors and symmetric tensor rank," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1254–1279, Sept. 2008.
- [20] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Oxford U.K.: Academic, 2010.
- [21] P. Comon, X. Luciani, and A. L. F. De Almeida, "Tensor decompositions, alternating least squares and other tales," *J. Chemomet.*, vol. 23, pp. 393–405, Aug. 2009.
- [22] P. Comon and G. Ottaviani, "On the typical rank of real binary forms," *Linear Multilinear Algebr.*, vol. 60, no. 6, pp. 657–667, May 2012.
- [23] P. Comon and G. Ottaviani, "Tensor decompositions, a geometric viewpoint," submitted for publication.
- [24] P. Comon, J. M. F. Ten Berge, L. De Lathauwer, and J. Castaing, "Generic and typical ranks of multi-way arrays," *Linear Algebr. Appl.*, vol. 430, nos. 11–12, pp. 2997–3007, June 2009.
- [25] L. De Lathauwer, "A link between canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.
- [26] L. De Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-(Lr,Lr,1) terms," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 145–1474, 2011.
- [27] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, Apr. 2000.
- [28] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(R1,R2,...,RN) approximation of high-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, Apr. 2000.
- [29] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Independent component analysis and (simultaneous) third-order tensor diagonalization," *IEEE Trans. Signal Processing*, vol. 49, no. 10, pp. 2262–2271, Oct. 2001.
- [30] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 2, pp. 295–327, 2004.
- [31] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part I: Basic results and uniqueness of one factor matrix," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 855–875, 2013.
- [32] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: Overall uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 876–903, 2013.

- [33] R. Ehrenborg and G. C. Rota, "Apolarity and canonical forms for homogeneous polynomials," *Eur. J. Combinatorics*, vol. 14, no. 3, pp. 157–181, May 1993.
- [34] S. Friedland, "Best rank one approximation of real symmetric tensors can be chosen symmetric," *Front. Math. China*, vol. 8, no. 1, pp. 19–40, 2013.
- [35] S. Friedland and G. Ottaviani, "The number of singular vector tuples and uniqueness of best rank one approximation of tensors," submitted for publication. arXiv:1210.8316.
- [36] R. A. Harshman, "Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Working Pap. Phonet.*, vol. 16, pp. 1–84, Dec. 1970.
- [37] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," submitted for publication. arXiv:0911.1393.
- [38] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *J. Math. Phys.*, vol. 6, no. 1, pp. 165–189, 1927.
- [39] T. D. Howell, "Global properties of tensor rank," *Linear. Algebr. Applicat.*, vol. 22, pp. 9–23, Dec. 1978.
- [40] A. Iarrobino, "Inverse system of a symbolic power. II. The Waring problem for forms," *J. Algebr.*, vol. 174, no. 3, pp. 1091–1110, June 1995.
- [41] T. Jiang and N. Sidiropoulos, "Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2625–2636, Sept. 2004.
- [42] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemomet.*, vol. 14, no. 3, pp. 105–122, June 2000.
- [43] E. Kofidis and P. Regalia, "One the best rank-1 approximation of higher-order symmetric tensors," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 3, pp. 863–884, 2002.
- [44] E. Kofidis and P. A. Regalia, "Tensor approximation and signal processing applications," in *Structured Matrices in Mathematics, Computer Science, and Engineering*, V. Olshevsky, Ed. American Math. Soc.: Providence, RI, 2001.
- [45] T. G. Kolda, "Orthogonal tensor decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 243–255, 2001.
- [46] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Sept. 2009.
- [47] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions," *Linear Algebr. Applicat.*, vol. 18, no. 2, pp. 95–138, 1977.
- [48] J. B. Kruskal, "Rank of N-way arrays and the geometry of $2 \times 2 \times 2$ arrays," *Tech. Rep.*, AT&T Bell Labs, Feb. 1985.
- [49] J. M. Landsberg, "Tensors: Geometry and applications," in *Graduate Studies in Mathematics*. American Math. Soc.: Providence, RI, 2012, vol. 128.
- [50] S. Leurgans, R. T. Ross, and R. B. Abel, "A decomposition for three-way arrays," *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1064–1083, Oct. 1993.
- [51] X. L. Li and X. D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1803–1814, 2007.
- [52] T. Lickteig, "Typical tensorial rank," *Linear Algebr. Applicat.*, vol. 69, pp. 95–120, Aug. 1985.
- [53] L.-H. Lim, "Singular values and eigenvalues of tensors: A variational approach," in *Proc. IEEE Int. Workshop on Computational Advances Multi-Sensor Adaptive Processing*, Puerto Vallarta, Mexico, Dec. 13–15, 2005, pp. 129–132.
- [54] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *J. Chemomet.*, vol. 23, pp. 432–441, Aug. 2009.
- [55] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inform. Theory*, vol. 60, no. 2, 2014, pp. 1260–1280.
- [56] X. Luciani and L. Albera, "Semi-algebraic canonical decomposition of multiway arrays and joint eigenvalue decomposition," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2011)*, Prague, Mar. 12–15, 2011, pp. 4104–4107.
- [57] P. McCullagh, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability. London, U.K.: Chapman & Hall, 1987.
- [58] R. Miranda, "Linear systems of plane curves," *Not. AMS*, vol. 46, no. 2, pp. 192–202, Feb. 1999.
- [59] G. Ni and Y. Wang, "On the best rank-1 approximation to higher-order symmetric tensors," *Math. Comput. Modell.*, vol. 46, no. 9–10, pp. 1345–1352, Nov. 2007.
- [60] L. Oeding and G. Ottaviani, "Eigenvectors of tensors and algorithms for Waring decomposition," *J. Symb. Comput.*, vol. 54, pp. 9–35, July 2013.
- [61] P. Paatero, "Construction and analysis of degenerate Parafac models," *J. Chemomet.*, vol. 14, no. 3, pp. 285–299, May/June 2000.
- [62] D. T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136–1152, 2001.
- [63] A. H. Phan, A. Cichocki, P. Tichavsky, G. Luta, and A. Brockmeier, "Tensor completion through multiple Kronecker product decomposition," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2013)*, Vancouver, Canada, May 26–31, pp. 3233–3237.
- [64] A. H. Phan, P. Tichavsky, and A. Cichocki, "Candecomp/Parafac decomposition of high-order tensors through tensor reshaping," *IEEE Trans. Signal Processing*, vol. 61, pp. 4847–4860, Oct. 2013.
- [65] M. Rajih, P. Comon, and R. Harshman, "Enhanced line search: A novel method to accelerate Parafac," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1148–1171, Sept. 2008.
- [66] B. Reznick, "Sums of even powers of real linear forms," *Mem. AMS*, vol. 96, no. 463, pp. 1–155, Mar. 1992.
- [67] B. Reznick, "Laws of inertia in higher degree binary forms," *Proc. Amer. Math. Soc.*, vol. 138, no. 3, pp. 815–826, June 2010.
- [68] J. A. Rhodes, "A concise proof of Kruskal's theorem on tensor decomposition," *Linear Algebr. Applicat.*, vol. 432, no. 7, pp. 1818–1824, Mar. 2010.
- [69] F. Roemer and M. Haardt, "A semi-algebraic framework for approximate CP decompositions via simultaneous matrix diagonalizations (SECSI)," *Signal Process.*, vol. 93, no. 9, pp. 2722–2738, Sept. 2013.
- [70] J.-P. Royer, N. Thirion-Moreau, and P. Comon, "Computing the polyadic decomposition of nonnegative third order tensors," *Signal Process.*, vol. 91, no. 9, pp. 2159–2171, Sept. 2011.
- [71] J. R. Ruiz-Tolosa and E. Castillo, *From Vectors to Tensors*, Universitext. Berlin, Heidelberg: Springer, 2005.
- [72] A. Schönhage, "Partial and total matrix multiplication," *SIAM J. Computat.*, vol. 10, no. 3, pp. 434–455, 1981.
- [73] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemomet.*, vol. 14, no. 3, pp. 229–239, May/June 2000.
- [74] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Processing*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
- [75] 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.*, vol. 30, no. 3, pp. 1084–1127, 2008.
- [76] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*. Chichester, U.K.: Wiley, 2004.
- [77] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-s terms and a new generalization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 695–720, Apr. 2013.
- [78] M. Sorensen and P. Comon, "Tensor decompositions with banded matrix factors," *Linear Algebr. Applicat.*, vol. 438, no. 1, pp. 919–941, Jan. 2013.
- [79] 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.*, vol. 30, no. 4, pp. 1614–1638, 2009.
- [80] A. Stegeman and P. Comon, "Subtracting a best rank-1 approximation does not necessarily decrease tensor rank," *Linear Algebr. Applicat.*, vol. 433, no. 7, pp. 1276–1300, Dec. 2010.
- [81] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition," *Linear Algebr. Applicat.*, vol. 420, no. 2–3, pp. 540–552, Jan. 2007.
- [82] V. Strassen, "Rank and optimal computation of generic tensors," *Linear Algebr. Applicat.*, vol. 52, pp. 645–685, July 1983.
- [83] P. Tichavsky and Z. Koldovsky, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 1037–1047, Mar. 2011.
- [84] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the Parafac model," *Comp. Stat. Data Anal.*, vol. 50, pp. 1700–1734, 2006.
- [85] L. R. Tucker, "Some mathematical notes for three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sept. 1966.
- [86] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Processing*, vol. 54, pp. 3270–3278, Sept. 2006.
- [87] A. Yeredor, "Non-orthogonal joint diagonalization in the LS sense with application in blind source separation," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1545–1553, 2002.
- [88] X. Zhang, C. Ling, and L. Qi, "The best rank-1 approximation of a symmetric tensor and related spherical optimization problems," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 3, pp. 806–821, 2012.
- [89] R. Ziehe, G. Nolte, and K. R. Müller, "A fast algorithm for joint diagonalization with non orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, Dec. 2004.
- [90] R. Bro. Home Page. [Online]. Available: <http://www.models.life.ku.dk/users/rasmus>
- [91] L. De Lathauwer. Home Page. [Online]. Available: <http://homes.esat.kuleuven.be/~delathau>
- [92] T. Kolda. Home Page. [Online]. Available: <http://www.sandia.gov/~tgkolda>
- [93] A. H. Phan. Home Page. [Online]. Available: <http://www.bsp.brain.riken.jp/~phan>
- [94] P. Comon. Home Page. [Online]. Available: <http://www.gipsa-lab.grenoble-inp.fr/~pierre.comon>
- [95] P. Kroonenberg. Three-Mode Company. [Online]. Available: <http://www.gipsa-lab.grenoble-inp.fr/~pierre.comon>