THE CANONICAL DECOMPOSITION AND BLIND IDENTIFICATION WITH MORE INPUTS THAN OUTPUTS: SOME ALGEBRAIC RESULTS

Lieven De Lathauwer

ETIS (ENSEA, UCP, CNRS)
Cergy-Pontoise
France
delathau@ensea.fr

ABSTRACT

In this paper we link the blind identification of a MIMO Moving Average (MA) system to the calculation of the Canonical Decomposition (CD) in multilinear algebra. This conceptually allows for the blind identification of systems that have many more inputs than outputs. We also derive a new theorem guaranteeing uniqueness of a high-rank CD and an algebraic algorithm for its computation.

1. INTRODUCTION

Blind identification of a linear system is the identification of that system without knowing its inputs. Typically one assumes that the inputs are spatially and temporally independent.

In this contribution we exploit the structure of the higher-order cumulant of the outputs. We show that a part of it satisfies a CD model, in which the components yield the unknown Markov parameters. The CD of a higher-order tensor is the decomposition of that tensor in a minimal number of rank-1 terms; a rank-1 tensor is an outer product of a number of vectors, i.e., an Nth-order tensor $\mathcal{A}$ has rank 1 when there exist N vectors $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$ such that:

$$a_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N}$$

for all values of the indices. A good tutorial on the CD is [1]. A general discussion and a brief overview of the state-of-the-art can also be found in [6]. Uniqueness of the decomposition has been investigated in [2, 3, 9, 10]. For an $(I \times J \times K \times L)$ tensor, the number of terms in an essentially unique CD can be much higher than $\min(I, J, K, L)$ (unlike the situation for matrices), enabling us to blindly identify systems that have many more inputs than outputs.

The link between the CD and blind identification is explained in Section 2. In Section 3 we derive two important algebraic results: (i) a new uniqueness theorem for the CD, based on relatively weak assumptions, and (ii) a powerful technique for the computation of the decomposition via a simultaneous diagonalization. Section 4 elaborates further on the blind identification problem. Section 5 is the conclusion.

Our formulation is in terms of real-valued data; the generalization to the complex case is straightforward.

2. THE CD AND BLIND IDENTIFICATION

Consider the following data model:

$$Y(n) = \sum_{l=0}^{L-1} \mathbf{H}(l) X(n-l) + N(n),$$

in which $X(n) \in \mathbb{R}^J$ are the unkown inputs, which are assumed to be mutually independent in space and time and at least fourth-order stationary, $Y(n) \in \mathbb{R}^I$ are the observed outputs, $\mathbf{H}(l) \in \mathbb{R}^{I \times J}$, $l = 0, \ldots, L-1$ are the unknown Markov parameters, and $N(n) \in \mathbb{R}^I$ is Gaussian noise.

Let us consider the $(I \times I^3)$ matrices $\mathbf{C}(\tau_1, \tau_2, \tau_3, \tau_4)$, defined by

$$(\mathbf{C}(\tau_1, \tau_2, \tau_3, \tau_4))_{i_1, (i_2-1)I^2 + (i_3-1)I + i_4} = \text{Cum}[y_{i_1}(n-\tau_1), y_{i_2}(n-\tau_2), y_{i_3}(n-\tau_3), y_{i_4}(n-\tau_4)].$$

Define also

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}(L-1)^T & \ldots & \mathbf{H}(0)^T \end{pmatrix}^T .$$

Then we have:

$$\tilde{\mathbf{T}} = \begin{pmatrix} \mathbf{C}(0,0,L-1,0) & \ldots & \mathbf{C}(0,0,L-1,L-1) \\ \mathbf{C}(1,0,L-1,0) & \ldots & \mathbf{C}(1,0,L-1,L-1) \\ \vdots & \ldots & \vdots \\ \mathbf{C}(L-1,0,L-1,0) & \ldots & \mathbf{C}(L-1,0,L-1,L-1) \end{pmatrix} = \mathbf{H} \cdot \text{diag}(\kappa_1, \ldots, \kappa_J) \cdot (\mathbf{H}(L-1) \odot \mathbf{H}(0) \odot \mathbf{H})^T ,$$

in which $\odot$ is the Katri-Rao or column-wise Kronecker product, and $\kappa_1, \ldots, \kappa_J$ are the cumulants of the inputs. This equation actually correponds to the decomposition of a fourth-order tensor $\mathcal{T} \in \mathbb{R}^{LI \times I \times I \times LI}$, of which $\tilde{\mathbf{T}}$ is a matrix representation, in a sum of rank-1 terms:

$$(\mathcal{T})_{k_1 k_2 k_3 k_4} = (\tilde{\mathbf{T}})_{k_1,(k_2-1)LI^2+(k_3-1)LI+k_4}$$
$$= \sum_{j=1}^{J} \kappa_j \, (\mathbf{H})_{k_1 j} (\mathbf{H}(L-1))_{k_2 j} (\mathbf{H}(0))_{k_3 j} (\mathbf{H})_{k_4 j}. \quad (1)$$

In what follows, we will derive means to compute this decomposition.

3. A CD ALGORITHM

This section is inspired by the techniques derived in [2, 7], which only apply to super-symmetric higher-order tensors (real tensors are super-symmetric when they are invariant under arbitrary index permutations).

Consider an $(I \times J \times K \times L)$ tensor $\mathcal{T}$ of which the CD is given by

$$t_{ijkl} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} d_{lr}, \quad (2)$$

in which $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, $\mathbf{C} \in \mathbb{R}^{K \times R}$, $\mathbf{D} \in \mathbb{R}^{L \times R}$. R is called the rank of $\mathcal{T}$. Trivial indeterminacies of this decomposition are that the R components may be reordered, and that, within the same component, the different factors may be rescaled, as long as the overall component remains the same.

Associate with $\mathcal{T}$ a matrix-to-matrix mapping as follows:

$$(\mathcal{T}(\mathbf{V}))_{ij} = \sum_{kl} t_{ijkl} v_{kl}.$$

Let this mapping be represented by a matrix $\mathbf{T} \in \mathbb{R}^{IJ \times KL}$, and consider a factorization of $\mathbf{T}$ of the form

$$\mathbf{T} = \mathbf{E} \cdot \mathbf{F}^T, \quad (3)$$

with $\mathbf{E} \in \mathbb{R}^{IJ \times R}$ and $\mathbf{F} \in \mathbb{R}^{KL \times R}$ full rank (this factorization may, e.g., be obtained by means of a Singular Value Decomposition (SVD)). The full rank property is generically satisfied if $R \leqslant \min(IJ, KL)$. We call a property generic when it holds everywhere, except for a set of Lebesgue measure 0. Hence we may read the value R as the number of significant singular values of $\mathbf{T}$. Because of (2) and (3), we have:

$$\mathbf{A} \odot \mathbf{B} = \mathbf{E} \cdot \mathbf{W} \qquad (\mathbf{C} \odot \mathbf{D})^T = \mathbf{W}^{-1} \cdot \mathbf{F}^T \quad (4)$$

for some nonsingular $\mathbf{W} \in \mathbb{R}^{R \times R}$. The task is now to find $\mathbf{W}$ such that the columns of $\mathbf{E} \cdot \mathbf{W}$ and $\mathbf{F} \cdot \mathbf{W}^{-T}$ correspond to rank-1 matrices.

Theorem 1. Define a mapping $\Phi : \mathbb{R}^{I \times J} \times \mathbb{R}^{I \times J} \to \mathbb{R}^{I \times J \times I \times J}$ by

$$(\Phi(\mathbf{G}, \mathbf{H}))_{i_1 j_1 i_2 j_2} = g_{i_1 j_1} h_{i_2 j_2} - g_{i_1 j_2} h_{i_2 j_1}. \quad (5)$$

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

Proof: It is easy to check that $\Phi(\mathbf{G}, \mathbf{G}) = 0$ if $\mathbf{G}$ is rank-1. For the "only if" part, let the SVD of $\mathbf{G}$ be given by $\mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T$. We have:

$$g_{i_1 j_1} g_{i_2 j_2} = \sum_{rs} \sigma_r \sigma_s u_{i_1 r} v_{j_1 r} u_{i_2 s} v_{j_2 s}$$
$$g_{i_1 j_2} g_{i_2 j_1} = \sum_{rs} \sigma_r \sigma_s u_{i_1 r} v_{j_1 s} u_{i_2 s} v_{j_2 r}.$$

Rank-1 terms corresponding to the same $r = s$ cancel out in Eq. (5). However, due to the orthogonality of $\mathbf{U}$ and $\mathbf{V}$, the other terms are linearly independent. So we must have that $\sigma_r \sigma_s = 0$ whenever $r \neq s$; hence, $\mathbf{\Sigma}$ is at most rank-1. ■

Denote the $(I \times J)$ matrices represented by the columns of $\mathbf{E}$ as $\mathbf{E_1}, \ldots, \mathbf{E_R}$, let $\Phi_{\mathbf{st}} = \Phi(\mathbf{E_s}, \mathbf{E_t})$ and let the columns of $\mathbf{A}$, $\mathbf{B}$ be given by $\{A_p\}$, $\{B_p\}$. Due to the bilinearity of Φ, we have

$$\Phi_{\mathbf{st}} = \sum_{pq} (\mathbf{W}^{-1})_{ps} (\mathbf{W}^{-1})_{qt} \Phi(A_p B_p^T, A_q B_q^T). \quad (6)$$

Now consider the matrices $\mathbf{M}$ of which the entries satisfy the following set of homogeneous linear equations:

$$\sum_{st} m_{st} \Phi_{\mathbf{st}} = 0. \quad (7)$$

As will become clear, these matrices form an R-dimensional subspace of the symmetric $(R \times R)$ matrices (under the condition to be specified). Let $\{\mathbf{M_r}\}$ represent a basis of this subspace. If the tensors $\{\Phi(A_p B_p^T, A_q B_q^T)\}_{p \neq q}$ are linearly independent, then substitution of (6) in (7) shows that

$$\sum_{st} (\mathbf{M_r})_{st} (\mathbf{W}^{-1})_{ps} (\mathbf{W}^{-1})_{qt} = (\mathbf{\Lambda_r})_{pq} \delta_{pq} \qquad \forall p, q \quad (8)$$

in which δ is the Kronecker delta. (8) can be rewritten as

$$\begin{aligned} \mathbf{M_1} &= \mathbf{W} \cdot \mathbf{\Lambda_1} \cdot \mathbf{W}^T \\ &\vdots \\ \mathbf{M_R} &= \mathbf{W} \cdot \mathbf{\Lambda_R} \cdot \mathbf{W}^T \end{aligned} \quad (9)$$

in which $\mathbf{\Lambda_1}, \ldots, \mathbf{\Lambda_R}$ are diagonal. $\mathbf{W}$ can be determined from this simultaneous matrix decomposition by means of the algorithms developed in [4, 5, 12, 13, 14].

Working on $\mathbf{F}$ may provide a similar set of equations in $\mathbf{W}^{-1}$ (provided $\{\Phi(C_p D_p^T, C_q D_q^T)\}_{p \neq q}$ are linearly independent). These equations may be combined with (9).

The question is now under which condition on R linear independence of $\{\Phi(A_p B_p^T, A_q B_q^T)\}_{p \neq q}$ can generically be guaranteed. The following theorem establishes sufficient (but not necessary) conditions:

Theorem 2. The CD in Eq. (2) is generically unique, up to the trivial indeterminacies mentioned before, if $R \leqslant \min(IJ, KL)$ and $R(R-1) \leqslant I^2(J^2-J)/2$.

Proof: The first inequality was used in the derivation of Eq. (4). The second inequality generically guarantees linear independence of $\{\Phi(A_p B_p^T, A_q B_q^T)\}_{p \neq q}$. These tensors can be represented in a vector format as

$$\begin{aligned} A_p \otimes A_q \otimes B_p \otimes B_q - A_p \otimes A_q \otimes B_q \otimes B_p \\ = A_p \otimes A_q((\mathbf{I}_{J^2 \times J^2} - \mathbf{P})(B_p \otimes B_q)) \\ = (\mathbf{I}_{I^2 \times I^2} \otimes (\mathbf{I}_{J^2 \times J^2} - \mathbf{P}))(A_p \otimes A_q \otimes B_p \otimes B_q) \end{aligned}$$

in which $\mathbf{P}$ is a specific permutation matrix. $\mathbf{P}$ is such that $\mathbf{I}_{J^2 \times J^2} - \mathbf{P}$ has rank $(J^2 - J)/2$. Hence, the condition is that $R(R-1) \leqslant I^2(J^2-J)/2$. ∎

Note that one may start from a version of $\mathcal{T}$ of which the indices are permuted in the way that leads to the smoothest condition on R.

Remark 1. The technique may also be used for the computation of the CD of a third-order tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ of which the rank $R \leqslant \min(IJ, K)$. By starting from the vector-to-matrix mapping

$$(\mathcal{T}(V))_{ij} = \sum_k t_{ijk} v_k,$$

one obtains a simultaneous decomposition of the form (9) (but not a similar set in terms of $\mathbf{W}^{-1}$). Again, the decomposition is unique if $R(R-1) \leqslant I^2(J^2-J)/2$.

Remark 2. For an $(I \times J \times K \times L)$ tensor $\mathcal{T}$ with $L \gg \max(I, J, K)$, one can even allow a higher rank. One starts from the vector-to-tensor mapping

$$(\mathcal{T}(V))_{ijk} = \sum_l t_{ijkl} v_l$$

and imposes the structure

$$\mathbf{A} \odot \mathbf{B} \odot \mathbf{C} = \mathbf{E} \cdot \mathbf{W}.$$

This can be done by resorting to a variant of mapping Φ in Theor. 1.

These results can be applied to, e.g., the problems dealt with in [8, 10, 11], and to the problem sketched in Section 2.

4. THE BLIND IDENTIFICATION PROBLEM

As far as our blind identification problem is concerned, the previous section explicitly leads to the noise-free solution (even when considering only 2 of the decompositions in (9) — these can be combined to an eigenvalue decomposition of the type $\mathbf{M_r} \cdot \mathbf{M_s}^{-1} = \mathbf{W} \cdot \mathbf{\Lambda_r} \cdot \mathbf{\Lambda_s}^{-1} \cdot \mathbf{W}^{-1}$).

However, in the context of Section 2, the results of the previous section should rather be considered as theoretical contributions, intended to set out some marks showing what theoretically can be achieved. We have proved that, using I sensors, one can identify a MIMO system of order L having as much as J inputs, bounded by

$$J(J-1) \leqslant L^2 I^3 (I-1)/2. \tag{10}$$

With respect to application in practice, Section 2 has the drawback that, in Eq. (1), we have actually only exploited the structure of the output cumulant for a very specific set of time lags: $\tau_1 = 0, \ldots, L-1$, $\tau_2 = 0$, $\tau_3 = L-1$, $\tau_4 = 0, \ldots, L-1$. The expressions for the input-output cumulant relations for other sets of time lags are more complex. Assume, without loss of generality, that $\tau_1 = \max\{\tau_1, \tau_2, \tau_3, \tau_4\}$ and $\tau_2 = \min\{\tau_1, \tau_2, \tau_3, \tau_4\}$. Then we have:

$$\mathbf{C}(\tau_1, \tau_2, \tau_3, \tau_4) = \sum_{p=0}^{L-1-\tau_1+\tau_2} \mathbf{H}(p) \cdot \mathrm{diag}(\kappa_1, \ldots, \kappa_J) \cdot$$
$$(\mathbf{H}(\tau_1 - \tau_2 + p) \odot \mathbf{H}(\tau_1 - \tau_3 + p) \odot \mathbf{H}(\tau_1 - \tau_4 + p))^T . \tag{11}$$

One may expect to obtain more accurate results by matching both sides of (11) for a sufficient number of time lags.

5. CONCLUSION

In this paper we have derived a new uniqueness theorem for the CD of a higher-order tensor, involving mild conditions on dimensions and rank. We have shown that, under the conditions specified by the theorem, the canonical components can be computed by means of a simultaneous congruence transformation. These powerful results are important for many higher-order signal processing problems. Our exposition was limited to third- and fourth-order tensors, but can be generalized to arbitrary tensor orders.

We have established a link between the blind identification of a MIMO MA system and the CD of a fourth-order output cumulant. Based on this link, we have proved that it is possible to blindly identify systems of which the number of inputs is roughly proportional to $I^2 L$, in which I is the number of outputs and L the system order.

Acknowledgements

Lieven De Lathauwer holds a permanent research position with the French Centre National de la Recherche Scientifique (C.N.R.S.) (http://www-etis.ensea.fr); he also holds a honorary research position with the K.U.Leuven (http://www.esat.kuleuven.ac.be/sista-cosic-docarch).

This work is supported by several institutions:

1. Research Council K.U.Leuven: Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology),

2. Flemish Government: the F.W.O. projects G.0240.99 (Multilinear Generalisations of the Singular Value Decomposition and Applications in Signal Processing and System Identification) and G.0256.97 (Numerical Algorithms for Subspace System Identification, Extension to Special Cases), the F.W.O. Research Communities ICCoS (Identification and Control of Complex Systems) and ANMMM (Advanced Numerical Methods for Mathematical Modelling),

3. Belgian Federal Government: Interuniversity Poles of Attraction Programme IUAP IV-02 (1997-2001): Modeling, Identification, Simulation and Control of Complex Systems, and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling.

The scientific responsibility is assumed by the author.

6. REFERENCES

[1] R. Bro, "PARAFAC. Tutorial & Applications", *Chemom. Intell. Lab. Syst.*, Special Issue *2nd Internet Conf. in Chemometrics (INCINC'96)*, No. 38, 1997, pp. 149–171.

[2] J.-F. Cardoso, "Super-Symmetric Decomposition of the Fourth-Order Cumulant Tensor. Blind Identification of More Sources than Sensors." *Proc. ICASSP-91*, Vol. 5, pp. 3109–3112.

[3] P. Comon, B. Mourrain, "Decomposition of Quantics in Sums of Powers of Linear Forms", *Signal Process.*, Special Issue on *Higher Order Statistics*, Vol. 53, 1996, pp. 93–108.

[4] L. De Lathauwer, *Signal Processing Based on Multilinear Algebra*, Ph.D. thesis, K.U.Leuven, E.E. Dept., Belgium, 1997.

[5] L. De Lathauwer, J. Vandewalle, B. De Moor, "Jacobi-Algorithm for Simultaneous Generalized Schur Decomposition in Higher-Order-Only ICA", *Proc. IEEE Benelux Signal Processing Chapter Signal Processing Symposium (SPS'98)*, Leuven, Belgium, March 26–27, 1998, pp. 67–70.

[6] L. De Lathauwer, B. De Moor, J. Vandewalle, "Computation of the Canonical Decomposition by Means of a Simultaneous Generalized Schur Decomposition", Tech. Report No. 01-109, ESAT/SCD-SISTA, K.U.Leuven, submitted to: *SIAM J. Matrix Anal. Appl.*.

[7] L. De Lathauwer, "Independent Component Analysis of Largely Underdetermined Mixtures", *Proc. ICA 2003*.

[8] T. Jiang, N. Sidiropoulos, J. ten Berge, "Almost-Sure Identifiability of Multidimensional Harmonic Retrieval", *IEEE Trans. on Signal Processing*, Vol. 49, 2001, pp. 1849–1859.

[9] J.B. Kruskal, "Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics", *Linear Algebra and its Applications*, No. 18, 1977, pp. 95–138.

[10] N. Sidiropoulos, G. Giannakis, R. Bro, "Blind PARAFAC Receivers for DS-CDMA Systems", *IEEE Trans. on Signal Processing*, Vol. 48, 2000, pp. 810–823.

[11] N. Sidiropoulos, "Generalizing Carathéodory's Uniqueness of Harmonic Parameterization to N Dimensions, Vol. 47, No. 4, 2001, pp. 1687–1690.

[12] A.-J. van der Veen, A. Paulraj, "An Analytical Constant Modulus Algorithm", *IEEE Trans. Signal Processing*, Vol. 44, 1996, pp. 1136–1155.

[13] A.-J. van der Veen, "Joint Diagonalization via Subspace Fitting Techniques", *Proc. IEEE ICASSP*, Salt Lake City (UT), May 2001.

[14] A. Yeredor, "Non-Orthogonal Joint Diagonalization in the Least-Squares Sense With Application in Blind Source Separation", *IEEE Trans. Signal Processing*, Vol. 50, No. 7, July 2002, pp. 1545–1553.