



Robust Multilinear Decomposition of Low Rank Tensors

Xu Han^{1,2,4}, Laurent Albera^{1,2,4}(✉), Amar Kachenoura^{1,2,4}, Huazhong Shu^{3,4},
and Lotfi Senhadji^{1,2,4}

¹ LTSI, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France

² INSERM, U 1099, 35402 Rennes Cedex, France

³ LIST, Southeast University, 2 Sipailou, Nanjing 210096, China

⁴ Centre de Recherche en Information Biomédicale Sino-Français, Rennes, France

hanxu.list@gmail.com, shu.list@seu.edu.cn,

{laurent.albera, amar.kachenoura, lotfi.senhadji}@univ-rennes1.fr

Abstract. Although several methods are available to compute the multilinear rank of multi-way arrays, they are not sufficiently robust with respect to the noise. In this paper, we propose a novel Multilinear Tensor Decomposition (MTD) method, namely R-MTD (Robust MTD), which is able to identify the multilinear rank even in the presence of noise. The latter is based on sparsity and group sparsity of the core tensor imposed by means of the l_1 norm and the mixed-norm, respectively. After several iterations of R-MTD, the underlying core tensor is sufficiently well estimated, which allows for an efficient use of the minimum description length approach and an accurate estimation of the multilinear rank. Computer results show the good behavior of R-MTD.

Keywords: Multilinear tensor rank · MTD · Low rank
Sparse · Mixed-norm · l_1 norm

1 Introduction

The rank estimation problem has been considered for several decades. The noisy matrix case was firstly considered. Methods based on the computation of singular values and the use of either Akaike's information criterion [1], the Bayesian Information Criterion (BIC) [2] or the Minimum Description Length (MDL) principle [3] were proposed. The singular values of a low-rank noiseless matrix can be sorted in decreasing order (see the red diamonds in Fig. 1). Then the estimated rank R^{est} can be computed by searching the breaking point of the singular value curve, which minimizes the MDL criterion [3]:

$$R^{est} = \arg \min_r - 2 \log \left\{ \frac{\prod_{i=r+1}^I \lambda_i^{1/(I-r)}}{\frac{1}{I-r} \sum_{i=r+1}^I \lambda_i} \right\}^{J(I-r)} + r(2I - r) \log(J) \quad (1)$$

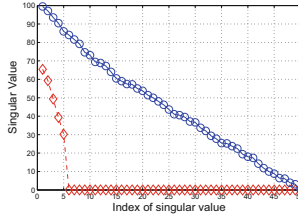


Fig. 1. Singular values of a matrix (red diamonds) and those of its noisy version for an SNR value of 0 dB (blue circles). (Color figure online)

where λ_i is the i -th highest singular value and where $(I \times J)$ is the size of the considered matrix. Unfortunately, when the matrix is strongly noisy the MDL approach has difficulty finding the breaking point (see the blue circles in Fig. 1).

The rank estimation problem was also considered for arrays with more than two entries, commonly named tensors. Contrarily to the matrix case, there are several definitions of tensor rank. In the following, we will consider the definition related to the Multilinear Tensor Decomposition (MTD) model [4]:

$$\mathcal{F} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{i=1}^{R_1} \sum_{j=1}^{R_2} \sum_{k=1}^{R_3} \mathcal{G}_{i,j,k} \mathbf{A}_{:,i} \circ \mathbf{B}_{:,j} \circ \mathbf{C}_{:,k}, \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{I_1 \times R_1}$, $\mathbf{B} \in \mathbb{R}^{I_2 \times R_2}$ and $\mathbf{C} \in \mathbb{R}^{I_3 \times R_3}$ are the loading matrices and where $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is the core tensor of a third order tensor \mathcal{F} . Note that $\mathbf{X}_{:,j}$ denotes the j -th column of \mathbf{X} and \circ denotes the vector outer product operator. The multilinear rank of \mathcal{F} is given by the minimal values of R_1 , R_2 and R_3 for which the equality (2) holds. The classical Higher-Order Singular Value Decomposition (HOSVD) [5] is a special MTD, which is a direct extension of the matrix SVD to tensors. While the multilinear rank can be derived from the rank of the unfolding matrices of the considered tensor, the latter matrix ranks are difficult to estimate as explained previously. Thus authors proposed to minimize the nuclear norm of each unfolding matrix in order to find the minimal MTD [8] and consequently the multilinear rank. More recently, the SCORE algorithm [7], based on the modified eigenvalues of the considered noisy tensor, was proposed too. Unfortunately, such methods are not sufficiently robust with respect to the presence of noise.

In this paper, we propose a novel MTD method, namely R-MTD (Robust MTD), which is able to identify the multilinear rank even in the presence of noise. The latter is based on sparsity and group sparsity of the core tensor imposed by means of the l_1 norm and the mixed-norm, respectively. Note that the mixed-norm was proved to be a convex envelope of rank and used to provide robust canonical polyadic and block term decomposition algorithms [9, 10]. After several iterations of R-MTD, the underlying core tensor is sufficiently well estimated, which allows for an efficient use of the MDL approach and an accurate estimation of the multilinear rank. Computer results show the good behavior of R-MTD.

2 Notions and Preliminaries

A scalar is denoted by an italic letter, e.g., x and I . A vector is denoted by a bold lowercase letter, e.g., $\mathbf{x} \in \mathbb{R}^I$ and a matrix is represented by a bold capital letter, e.g., $\mathbf{X} \in \mathbb{R}^{I \times J}$ and specially, \mathbf{I} is the identity matrix. The vectorization of \mathbf{X} is denoted by $\text{vec}(\mathbf{X}) \in \mathbb{R}^{IJ \times 1}$. \mathbf{X}^\dagger is the pseudo inverse of matrix \mathbf{X} . The nuclear norm of a matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$ with rank r is equal to the sum of its singular values, i.e., $\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i$, and $\|\mathbf{X}\|$ is the largest singular value. The symbol $\text{rank}(\cdot)$ denotes the rank operator. The l_0 norm, l_1 norm and Frobenius norm of $\mathbf{X} \in \mathbb{R}^{I \times J}$ are defined by $\|\mathbf{X}\|_0 = \sum_{i=1}^I \sum_{j=1}^J \mathbf{X}_{i,j} / \mathbf{X}_{i,j}$ with $\mathbf{X}_{i,j} \neq 0$, $\|\mathbf{X}\|_1 = \sum_{i=1}^I \sum_{j=1}^J |\mathbf{X}_{i,j}|$ and $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^I \sum_{j=1}^J \mathbf{X}_{i,j}^2}$, respectively. The mixed-norm pair of $\mathbf{X} \in \mathbb{R}^{I \times J}$ is given by $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^I \sqrt{\sum_{j=1}^J \mathbf{X}_{i,j}^2} = \text{Tr}[\mathbf{X}^\top \Phi \mathbf{X}]$ and $\|\mathbf{X}\|_{1,2} = \sum_{j=1}^J \sqrt{\sum_{i=1}^I \mathbf{X}_{i,j}^2} = \text{Tr}[\mathbf{X} \Psi \mathbf{X}^\top]$, where $\text{Tr}[\cdot]$ is the trace operator, where Φ is a diagonal matrix with $\Phi_{i,i} = 1/\sqrt{\sum_{j=1}^J \mathbf{X}_{i,j}^2}$ denoting the (i, i) -th component of Φ and where Ψ is a diagonal matrix with $\Psi_{j,j} = 1/\sqrt{\sum_{i=1}^I \mathbf{X}_{i,j}^2}$ standing for the (j, j) -th component of Ψ . Note that this trace compact definition is convenient for a convex optimization if Φ and Ψ are fixed as weight matrices. A high-order tensor is denoted by a bold calligraphic letter, e.g., $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. The symbol \otimes denotes the Kronecker product operator. The n -th mode unfolding matrix of the tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $\mathbf{X}^{(n)} \in \mathbb{R}^{I_n \times (I_{n+1} I_{n+2} \dots I_N I_1 I_2 \dots I_{n-1})}$. The sub-tensor $\mathcal{X}_{i_n=k}$ is obtained by fixing the n -th index to k . The scalar product of two tensors $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is defined by $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} \mathcal{X}_{i_1, \dots, i_N} \mathcal{Y}_{i_1, \dots, i_N}$. The Frobenius norm of tensor \mathcal{X} is defined by $\|\mathcal{X}\|_F = \sqrt{\sum_{i_1, \dots, i_N} \mathcal{X}_{i_1, \dots, i_N}^2}$.

3 The R-MTD Method

3.1 Preliminary Tools

First let's recall the definition of a convex envelope which will be used in the following analysis.

Definition 1 (Lower convex envelope). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a real-valued function. The convex envelope of f , represented by \mathcal{C}_f , is the convex pointwise largest function $\mathcal{C}_f : \mathbb{R}^n \rightarrow \mathbb{R}$ which is pointwise less than f . In other words, we have $\mathcal{C}_f = \sup\{g : \mathbb{R}^n \rightarrow \mathbb{R} \mid g \text{ is convex and } g(x) \leq f(x) \text{ for any } x \in \mathbb{R}^n\}$.*

Now let's study the lower convex envelopes of rank. The nuclear norm of \mathbf{X} is one of them in the norm ball, i.e., $\|\mathbf{X}\| \leq 1$, as shown in [12, Theorem 1] and since the low rank constraint involves a NP-hard optimization problem it is better to minimize the nuclear norm. On the other hand, an interesting relationship between the nuclear norm and the mixed-norm was established in [11, Proposition 1] for a thin matrix case, i.e., $\mathbf{X} \in \mathbb{R}^{m \times n}$ with $m > n$. We extended this result by means of the following Theorem 1 and Corollary 1.

Theorem 1. *Given any matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ and its orthonormal subspace decomposition denoted by $\mathbf{X} = \mathbf{D}\boldsymbol{\alpha}$ and $\mathbf{X} = \boldsymbol{\theta}\mathbf{Z}$, where $\mathbf{D} \in \mathbb{R}^{m \times m}$ and $\mathbf{Z} \in \mathbb{R}^{n \times n}$ are orthonormal matrices, with $\boldsymbol{\alpha} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{\theta} \in \mathbb{R}^{m \times n}$, the mixed-norms of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ are larger than or equal to the nuclear norm of \mathbf{X} , i.e. $\|\boldsymbol{\alpha}\|_{2,1} \geq \|\mathbf{X}\|_*$ and $\|\boldsymbol{\theta}\|_{1,2} \geq \|\mathbf{X}\|_*$.*

Then we can easily derive the following corollary by fixing \mathbf{D} and \mathbf{Z} equal to the identity matrix in Theorem 1.

Corollary 1. *We have $\|\mathbf{X}\|_{2,1} \geq \|\mathbf{X}\|_*$ and $\|\mathbf{X}\|_{1,2} \geq \|\mathbf{X}\|_*$.*

In fact, for a matrix with full column rank or full row rank, the nuclear norm is not the tightest convex envelope of rank and the mixed-norm is more qualified to be a lower convex envelop of rank compared with the nuclear norm based on the Definition 1 and the explanation is given in the following theorem.

Theorem 2. *Given any matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ with linearly independent rows or columns, we have:*

1. *If \mathbf{X} has linearly independent rows, which means that $\text{rank}(\mathbf{X}) = m$ and $m \leq n$, then we get $\text{rank}(\mathbf{X}) \geq \frac{\|\mathbf{X}\|_{2,1}}{\|\mathbf{X}\|} \geq \frac{\|\mathbf{X}\|_*}{\|\mathbf{X}\|}$.*
2. *If \mathbf{X} has linearly independent columns, which means that $\text{rank}(\mathbf{X}) = n$ and $m \geq n$, then we get $\text{rank}(\mathbf{X}) \geq \frac{\|\mathbf{X}\|_{1,2}}{\|\mathbf{X}\|} \geq \frac{\|\mathbf{X}\|_*}{\|\mathbf{X}\|}$.*

Consequently, we will use the mixed-norm in the following subsection in order to impose the low rank property. Besides, we will also minimize the l_1 norm of the over-estimated core tensor in order to vanish the residual nonzero elements for which the absolute values are close to zero.

3.2 Towards the R-MTD Algorithm

In order to guarantee a sufficient group sparsity of each unfolding matrix along the different dimensions, and each unfolding matrix consist with linear independent rows, the proposed method is implemented under the following assumption: $R_i \ll \min\{I_i, \prod_{k=1, k \neq i}^3 I_k\}$ and $R_i \leq \prod_{k=1, k \neq i}^3 R_k$, $i = 1, 2, 3$.

R-MTD for Rank Estimation: The mixed-norm $\|\widehat{\mathbf{G}}^{(i)}\|_{2,1}$ is considered as a lower convex envelope of rank in the objective function and the other mixed-norm $\|\widehat{\mathbf{G}}^{(i)}\|_{1,2}$ is also adopted as a convex upper bound of the nuclear norm for a more robust estimation. The objective function is presented below for a third order multilinear tensor as an example but it is not hard to generalize it to higher orders:

$$\begin{aligned}
 & \min_{\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}, \widehat{\mathcal{G}}} \sum_{i=1}^3 \lambda_i \left(\|\widehat{\mathbf{G}}^{(i)}\|_{2,1} + \|\widehat{\mathbf{G}}^{(i)}\|_{1,2} \right) + \|\widehat{\mathcal{G}}\|_1 \\
 & \text{s.t. } \mathcal{T} = \widehat{\mathcal{G}} \times_1 \widehat{\mathbf{A}} \times_2 \widehat{\mathbf{B}} \times_3 \widehat{\mathbf{C}} + \mathcal{N}, \widehat{\mathcal{G}} \in \mathbb{R}^{\widehat{R}_1 \times \widehat{R}_2 \times \widehat{R}_3}, \\
 & \widehat{\mathbf{A}} \in \mathbb{R}^{I_1 \times \widehat{R}_1}, \widehat{\mathbf{B}} \in \mathbb{R}^{I_2 \times \widehat{R}_2}, \widehat{\mathbf{C}} \in \mathbb{R}^{I_3 \times \widehat{R}_3}, \quad (3)
 \end{aligned}$$

where $\widehat{\mathcal{G}}$ is the over-estimated core tensor, where $\widehat{\mathbf{G}}^{(i)}$ are the unfolding matrices along the different dimensions and where $\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}$ are the corresponding over-estimated loading matrices. Here, we should note that \widehat{R}_i should be larger than R_i . The minimization problem (3) is solved by means of the Alternating Directions Method of Multiplier (ADMM) [14]:

$$\begin{aligned} \min_{\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}, \widehat{\mathcal{G}}} \quad & \sum_{i=1}^3 \lambda_i \left(\|\widehat{\mathbf{G}}^{(i)}\|_{2,1} + \|\widehat{\mathbf{G}}^{(i)}\|_{1,2} \right) + \|\widehat{\mathcal{P}}\|_1 \\ & + \frac{u}{2} \|\mathcal{T} - \widehat{\mathcal{G}} \times_1 \widehat{\mathbf{A}} \times_2 \widehat{\mathbf{B}} \times_3 \widehat{\mathbf{C}}\|_F^2 \\ \text{s.t.} \quad & \widehat{\mathcal{G}} - \widehat{\mathcal{P}} = \mathbf{0}. \end{aligned} \quad (4)$$

The augmented Lagrangian function \mathcal{L} of the variables $\widehat{\mathcal{G}}, \widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}, \widehat{\mathcal{P}}, \widehat{\mathcal{Y}}, \beta, \Phi^{(i)}$ and $\Psi^{(i)}$ can be written as follows using the mixed-norm definition given in Sect. 2:

$$\begin{aligned} \mathcal{L} = \sum_{i=1}^3 \lambda_i & \left[\text{Tr}(\widehat{\mathbf{G}}^{(i)\top} \Phi^{(i)} \widehat{\mathbf{G}}^{(i)}) + \text{Tr}(\widehat{\mathbf{G}}^{(i)} \Psi^{(i)} \widehat{\mathbf{G}}^{(i)\top}) \right] \\ & + \|\widehat{\mathcal{P}}\|_1 + \frac{u}{2} \|\mathcal{T} - \widehat{\mathcal{G}} \times_1 \widehat{\mathbf{A}} \times_2 \widehat{\mathbf{B}} \times_3 \widehat{\mathbf{C}}\|_F^2 \\ & + \langle \widehat{\mathcal{G}} - \widehat{\mathcal{P}}, \widehat{\mathcal{Y}} \rangle + \frac{\beta}{2} \|\widehat{\mathcal{G}} - \widehat{\mathcal{P}}\|_F^2, \end{aligned} \quad (5)$$

where the parameters λ_i, u and β are penalty coefficients and where $\widehat{\mathcal{Y}}$ is the Lagrangian multiplier tensor. Now let's derive the update rule of each variable. The mode-1 unfolding matrix of the core tensor $\widehat{\mathcal{G}}$ is computed from the following gradient equation:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \widehat{\mathbf{G}}^{(1)}} = 2\lambda_1 \Phi^{(1)} \widehat{\mathbf{G}}^{(1)} + 2\lambda_1 \widehat{\mathbf{G}}^{(1)} \Psi^{(1)} + \beta \widehat{\mathbf{G}}^{(1)} + u \widehat{\mathbf{A}}^\top \widehat{\mathbf{A}} \widehat{\mathbf{G}}^{(1)} (\widehat{\mathbf{B}} \otimes \widehat{\mathbf{C}})^\top (\widehat{\mathbf{B}} \otimes \widehat{\mathbf{C}}) \\ + \widehat{\mathcal{Y}}^{(1)} - \beta \widehat{\mathcal{P}}^{(1)} - u \widehat{\mathbf{A}}^\top \mathbf{T}^{(1)} (\widehat{\mathbf{B}} \otimes \widehat{\mathbf{C}}) = 0. \end{aligned} \quad (6)$$

The solution of (6) can be calculated by Encapsulating Sum in [15], the $(k+1)$ -th iteration of $\widehat{\mathbf{G}}^{(1)}$ is given by:

$$\begin{aligned} \text{vec}(\widehat{\mathbf{G}}_{k+1}^{(1)}) = \left\{ \mathbf{I} \otimes (2\lambda_1 \Phi_k^{(1)}) + (2\lambda_1 \Psi_k^{(1)} + \beta_k \mathbf{I}) \otimes \mathbf{I} + [(\widehat{\mathbf{B}}_k \otimes \widehat{\mathbf{C}}_k)^\top (\widehat{\mathbf{B}}_k \otimes \widehat{\mathbf{C}}_k)] \right. \\ \left. \otimes (u \widehat{\mathbf{A}}_k^\top \widehat{\mathbf{A}}_k) \right\}^{-1} \text{vec} \left[\beta_k \widehat{\mathcal{P}}_k^{(1)} - \widehat{\mathcal{Y}}_k^{(1)} + u \widehat{\mathbf{A}}_k^\top \mathbf{T}^{(1)} (\widehat{\mathbf{B}}_k \otimes \widehat{\mathbf{C}}_k) \right]. \end{aligned} \quad (7)$$

Consecutively, $\widehat{\mathbf{G}}^{(2)}$ and $\widehat{\mathbf{G}}^{(3)}$ at the $(k+1)$ -th iteration can be computed as follows:

$$\begin{aligned} \text{vec}(\widehat{\mathbf{G}}_{k+1}^{(2)}) = \left\{ \mathbf{I} \otimes (2\lambda_2 \Phi_k^{(2)}) + (2\lambda_2 \Psi_k^{(2)} + \beta_k \mathbf{I}) \otimes \mathbf{I} + [(\widehat{\mathbf{C}}_k \otimes \widehat{\mathbf{A}}_k)^\top (\widehat{\mathbf{C}}_k \otimes \widehat{\mathbf{A}}_k)] \right. \\ \left. \otimes (u \widehat{\mathbf{B}}_k^\top \widehat{\mathbf{B}}_k) \right\}^{-1} \text{vec} \left[\beta_k \widehat{\mathcal{P}}_k^{(2)} - \widehat{\mathcal{Y}}_k^{(2)} + u \widehat{\mathbf{B}}_k^\top \mathbf{T}^{(2)} (\widehat{\mathbf{C}}_k \otimes \widehat{\mathbf{A}}_k) \right], \end{aligned} \quad (8)$$

$$\text{vec}(\widehat{\mathbf{G}}_{k+1}^{(3)}) = \left\{ \mathbf{I} \otimes (2\lambda_i \boldsymbol{\Phi}_k^{(3)}) + (2\lambda_i \boldsymbol{\Psi}_k^{(3)} + \beta_k \mathbf{I}) \otimes \mathbf{I} + [(\widehat{\mathbf{A}}_k \otimes \widehat{\mathbf{B}}_k)^\top (\widehat{\mathbf{A}}_k \otimes \widehat{\mathbf{B}}_k)] \right. \\ \left. \otimes (u \widehat{\mathbf{C}}_k^\top \widehat{\mathbf{C}}_k) \right\}^{-1} \text{vec} \left[\beta_k \widehat{\mathbf{P}}_k^{(3)} - \widehat{\mathbf{Y}}_k^{(3)} + u \widehat{\mathbf{C}}_k^\top \mathbf{T}^{(3)} (\widehat{\mathbf{A}}_k \otimes \widehat{\mathbf{B}}_k) \right]. \quad (9)$$

Note that, the weighting matrices $\boldsymbol{\Phi}_k^{(1)}$ and $\boldsymbol{\Psi}_k^{(1)}$ are derived from $\widehat{\mathbf{G}}_k^{(1)}$ as the definition in Sect. 2. For the computation of $\boldsymbol{\Phi}_k^{(2)}$ and $\boldsymbol{\Psi}_k^{(2)}$, they are calculated using the 2-th mode unfolding matrix of the updated core tensor $\widehat{\mathbf{G}}_{k+1}^{(1)}$. Identically, $\boldsymbol{\Phi}_k^{(3)}$ and $\boldsymbol{\Psi}_k^{(3)}$ are given using the updated core tensor $\widehat{\mathbf{G}}_{k+1}^{(2)}$. Then, updating factor matrices by using tensor $\widehat{\mathcal{G}}_{k+1}$, i.e., $\widehat{\mathbf{G}}_{k+1}^{(3)}$. The sub-problem function about factor matrix $\widehat{\mathbf{A}}$ in $(k+1)$ -th iteration is given as following:

$$\min_{\widehat{\mathbf{A}}_{k+1}} \left\| \mathbf{T}^{(1)} - \widehat{\mathbf{A}}_{k+1} \widehat{\mathbf{G}}_{k+1}^{(1)} (\widehat{\mathbf{B}}_k \otimes \widehat{\mathbf{C}}_k)^\top \right\|_F^2. \quad (10)$$

The updated $\widehat{\mathbf{A}}_{k+1}$ matrix at the $(k+1)$ -th iteration is given by:

$$\widehat{\mathbf{A}}_{k+1} = \mathbf{T}^{(1)} \left[\widehat{\mathbf{G}}_{k+1}^{(1)} (\widehat{\mathbf{B}}_k \otimes \widehat{\mathbf{C}}_k)^\top \right]^\dagger. \quad (11)$$

Using $\widehat{\mathbf{A}}_{k+1}$, $\widehat{\mathbf{B}}_{k+1}$ at the $(k+1)$ -th iteration can be computed as follows:

$$\widehat{\mathbf{B}}_{k+1} = \mathbf{T}^{(2)} \left[\widehat{\mathbf{G}}_{k+1}^{(2)} (\widehat{\mathbf{C}}_k \otimes \widehat{\mathbf{A}}_{k+1})^\top \right]^\dagger. \quad (12)$$

Similarly, $\widehat{\mathbf{C}}_{k+1}$ at the $(k+1)$ -th iteration is given by:

$$\widehat{\mathbf{C}}_{k+1} = \mathbf{T}^{(3)} \left[\widehat{\mathbf{G}}_{k+1}^{(3)} (\widehat{\mathbf{A}}_{k+1} \otimes \widehat{\mathbf{B}}_{k+1})^\top \right]^\dagger. \quad (13)$$

Regarding variable $\widehat{\mathcal{P}}$, we have to minimize:

$$\min_{\widehat{\mathcal{P}}} \left\| \widehat{\mathcal{P}} \right\|_1 + \left\langle \widehat{\mathcal{G}} - \widehat{\mathcal{P}}, \widehat{\mathcal{Y}} \right\rangle + \frac{\beta}{2} \left\| \widehat{\mathcal{G}} - \widehat{\mathcal{P}} \right\|_F^2. \quad (14)$$

An equivalent problem of (14) is given by:

$$\min_{\widehat{\mathcal{P}}} \left\| \widehat{\mathcal{P}} \right\|_1 + \frac{\beta}{2} \left\| \widehat{\mathcal{G}} - \widehat{\mathcal{P}} + \frac{1}{\beta} \widehat{\mathcal{Y}} \right\|_F^2. \quad (15)$$

The solution of $\widehat{\mathcal{P}}$ in problem (15) has been solved in [13] as follows:

$$\widehat{\mathcal{P}}_{k+1} = \mathbb{S}_{\frac{1}{\beta_k}} \left(\widehat{\mathcal{G}}_{k+1} + \frac{1}{\beta_k} \widehat{\mathcal{Y}}_k \right), \quad (16)$$

where $\mathbb{S}_\tau(x) = \text{sign}(x)(|x| - \tau, 0)$, $x \in \mathbb{R}$ is the soft-thresholding operator and it is used elementwise. The multiplier tensor $\widehat{\mathcal{Y}}$ in $(k+1)$ -th iteration is calculated based on the following update rule:

$$\widehat{\mathcal{Y}}_{k+1} = \widehat{\mathcal{Y}}_k + \beta_k (\widehat{\mathcal{G}}_{k+1} - \widehat{\mathcal{P}}_{k+1}). \quad (17)$$

The parameter β in $(k + 1)$ -th iteration is updated as:

$$\beta_{k+1} = \rho\beta_k, \rho > 1. \quad (18)$$

Finally, the proposed algorithm R-MTD stops when the convergence criterion is reached, i.e.,

$$\frac{\text{error}_{k+1} - \text{error}_k}{\text{error}_k} < \textit{tolerance} \quad (19)$$

with $\text{error}_{k+1} = \|\mathcal{T} - \widehat{\mathcal{G}}_{k+1} \times_1 \widehat{\mathbf{A}}_{k+1} \times_2 \widehat{\mathbf{B}}_{k+1} \times_3 \widehat{\mathbf{C}}_{k+1}\|_F$ and $\text{error}_k = \|\mathcal{T} - \widehat{\mathcal{G}}_k \times_1 \widehat{\mathbf{A}}_k \times_2 \widehat{\mathbf{B}}_k \times_3 \widehat{\mathbf{C}}_k\|_F$. Eventually, the values R_i^{est} are estimated using the MDL criterion (2) on the singular values derived from the SVD of the unfolding matrices $\widehat{\mathbf{G}}_{k+1}^{(i)}$, $i = 1, 2, 3$.

R-MTD for Denoising: Once the multilinear rank is estimated, we can use it to replace the over-estimated rank in (3) and the penalty function for denoising is rewritten as follows:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{G}} \quad & \sum_{i=1}^3 \lambda_i \left(\|\mathbf{G}^{(i)}\|_{2,1} + \|\mathbf{G}^{(i)}\|_{1,2} \right) + \gamma \|\mathcal{G}\|_1 \\ \text{s.t.} \quad & \mathcal{T} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} + \mathcal{N}, \mathcal{G} \in \mathbb{R}^{R_1^{est} \times R_2^{est} \times R_3^{est}}, \\ & \mathbf{A} \in \mathbb{R}^{I_1 \times R_1^{est}}, \mathbf{B} \in \mathbb{R}^{I_2 \times R_2^{est}}, \mathbf{C} \in \mathbb{R}^{I_3 \times R_3^{est}}. \end{aligned} \quad (20)$$

Here, the weight parameters of the constraints are set to zero, i.e., $\lambda_i = \gamma = 0$, $i = 1, 2, 3$. With these zero weight parameters, we can repeat the same procedure as described previously.

4 Experiment Results

In this section, the noisy simulated data is utilized to test the performance of R-MTD on rank estimation and signal denoising. A comparative study of our method with the recent and efficient SCORE algorithm [7] is also provided. For the proposed method, the selection of parameter λ_i , $i = 1, 2, 3$ should be larger when the size of underlying core tensor is smaller than the over-estimated core tensor size. Typically, these coefficients are set as $\lambda_i = 5$, $i = 1, 2, 3$ and the other parameters μ , β and ρ are fixed as 0.025, 0.3 and 20, respectively. The selected parameters of SCORE method are chosen as the paper [7] advised. All results reported in this section are averaged over 100 Monte Carlo (MC) realizations.

Experimental Setting. The size of \mathcal{F} is fixed with $(I_1 = I_2 = I_3 = 100)$ and we consider different sizes of the core tensor \mathcal{G} : $(R_1 = R_2 = R_3 = 3)$, $(R_1 = R_2 = R_3 = 4)$ and $(R_1 = R_2 = R_3 = 5)$. All the elements of factor matrices and the core tensor follow a Gaussian distribution. To implement our method, the over-estimated core size is set to $\widehat{R}_1 = \widehat{R}_2 = \widehat{R}_3 = 10$. The noise

tensor $\mathcal{E} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ also follows Gaussian distribution with different Signal-to-Noise Ratio (SNR). Finally, the noisy tensor \mathcal{T} is obtained as follows:

$$\mathcal{T} = \mathcal{F} + \sigma \frac{\|\mathcal{F}\|_F}{\|\mathcal{E}\|_F} \mathcal{E},$$

where the parameter σ controls the SNR, i.e. $\text{SNR} = -20\log_{10}(\sigma)$.

Rank Estimation Performance. To evaluate the performance of all methods on rank estimation, we propose two criteria. The first one named Accuracy Rate (AR) measures the good estimate rate of all ranks, i.e. $R_1 = R_1^{est}$, $R_2 = R_2^{est}$ and $R_3 = R_3^{est}$, and is defined as the following:

$$\text{AR} : \text{Times of } R_i = R_i^{est}, i = 1, 2, 3.$$

The second criterion, called Average Rank Estimation Error (AREE), measures the deviation between the estimated ranks and exact ranks. It is given by:

$$\text{AREE} : \frac{1}{100} \sum_{\text{times}=1}^{100} \left(\sum_{i=1}^3 |R_i - R_i^{est}| \right).$$

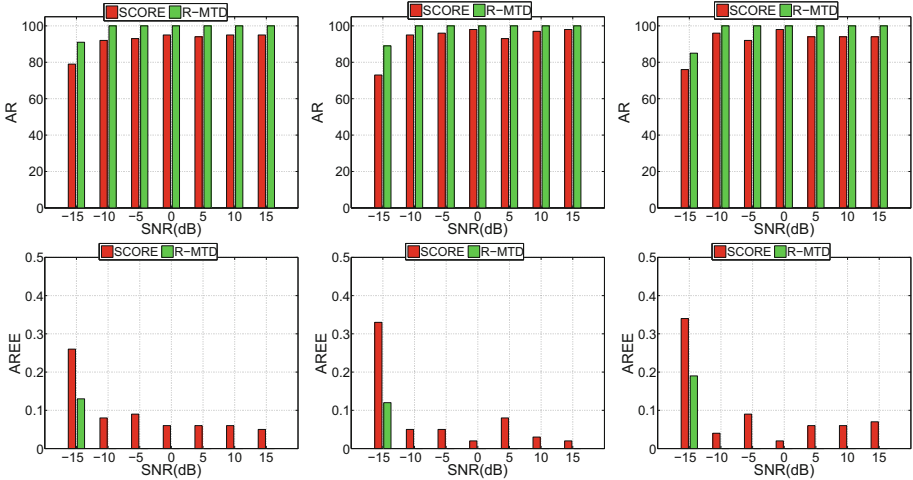


Fig. 2. The Accuracy Rate (AR) and the Average Rank Estimation Error (AREE) criteria for different sizes of the core tensor as a function of SNR.

Figure 2, displays the obtained results of the rank estimation for different SNR and different sizes of the core tensor. We can see that the AR criterion obtained for R-MTD and SCORE is increasing as the SNR increases (first line of Fig. 2). We also observed that, for all cases of the size of the core tensor, the robustness to noise of the R-MTD outperforms the SCORE whatever the SNR is. The second

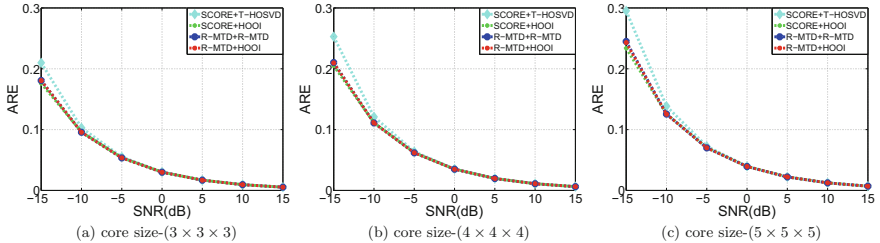


Fig. 3. The Average Relative Error (ARE) criterion for different sizes of the core tensor as a function of SNR.

line of Fig. 2 shows the AREE criterion. Clearly, the rank estimation errors of R-MTD are always smaller than those of the SCORE method for all scenarios and SNR.

Signal Denoising Efficiency. In this section, we consider the effectiveness and robustness of the proposed method on signal denoising. To do so, four combination cases of rank estimation methods and tensor decomposition ones are considered: R-MTD+R-MTD, R-MTD+HOOI [6], SCORE+T-HOSVD [5] and SCORE+HOOI. The reconstruction error is defined as the Average Relative Error (ARE), given by:

$$\text{ARE} : \frac{1}{100} \sum_{\text{times}=1}^{100} \frac{\|\mathcal{F}^{est} - \mathcal{F}\|_F}{\|\mathcal{F}\|_F}.$$

Figure 3 gives the effectiveness of the four compared methods. For SNR higher or equal to -5 dB, all methods seem to have similar performances. Regarding the lower SNR, we can see that the performances of the SCORE+HOOI, R-MTD+R-MTD and R-MTD+HOOI are better than the SCORE+T-HOSVD. This can be explained by the fact that the T-HOSVD method only extract the first several components corresponding to the estimated rank without any denoising process, it cannot gives excellent denoising result for lower SNR cases. Besides, although AR results of the R-MTD is more effective than the SCORE, the denoising efficiency of SCORE+HOOI is almost the same as the efficiency of R-MTD+R-MTD and R-MTD+HOOI because of the smaller difference of AREE between the SCORE and the R-MTD.

5 Conclusion

We have proposed a new MTD method, named R-MTD, that permits to identify the multilinear rank of noisy data. To do so, a sparsity and group sparsity of the core tensor are imposed by means of the l_1 norm and the mixed-norm, respectively. More precisely, we first over-estimate the core tensor, then an optimal core tensor is estimated from noisy multilinear tensor after several iterations of

the R-MTD algorithm. We also assessed the importance of a good estimation of the rank of the core tensor to signal denoising. Different experiments conducted using simulated data demonstrated the effectiveness of the R-MTD algorithm both for rank estimation and for data denoising.

References

1. Wax, M., Kailath, T.: Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech Signal Process.* **33**(2), 387–392 (1985)
2. Schwarz, G.: Estimating the dimension of a model. *Ann. Stat.* **6**(2), 461–464 (1978)
3. Rissanen, J.: Modeling by shortest data description. *Automatica* **14**(5), 465–471 (1978)
4. Tucker, L.R.: Implications of factor analysis of three-way matrices for measurement of change. In: *Problems in Measuring Change*, pp. 122–137 (1963)
5. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21**(4), 1253–1278 (2000)
6. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank- (R_1, R_2, \dots, R_n) approximation of higher-order tensors. *SIAM J. Matrix Anal.* **21**(4), 1324–1342 (2000)
7. Yokota, T., Lee, N., Cichocki, A.: Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria. *IEEE Trans. Signal Process.* **65**(5), 1196–1206 (2017)
8. Xie, Q., Zhao, Q., Meng, D., Xu, Z.: Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Trans. Pattern Anal. Mach. Intell.* (2017)
9. Han, X., Albera, L., Kachenoura, A., Senhadji, L., Shu, H.: Low rank canonical polyadic decomposition of tensors based on group sparsity. In: *IEEE 25th European Signal Processing Conference (EUSIPCO)*, pp. 668–672 (2017)
10. Han, X., Albera, L., Kachenoura, A., Senhadji, L., Shu, H.: Block term decomposition with rank estimation using group sparsity. In: *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)* (2017)
11. Shu, X., Porikli, F., Ahuja, N.: Robust orthonormal subspace learning: efficient recovery of corrupted low-rank matrices. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3874–3881 (2014)
12. Fazel, M.: Matrix rank minimization with applications. Ph.D. dissertation, Ph.D. thesis, Stanford University (2002)
13. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **90**(432), 1200–1224 (1995)
14. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends[®] Mach. Learn.* **3**(1), 1–122 (2011)
15. Petersen, K.B., Pedersen, M.S., et al.: *The Matrix Cookbook*. Technical University of Denmark, pp. 7–15 (2008)