AN AUGMENTED LAGRANGIAN ALGORITHM FOR DECOMPOSITION OF SYMMETRIC TENSORS OF ORDER-4

Anh-Huy Phan[1], *Masao Yamagishi*[2] *and Andrzej Cichocki*[1]

[1]Brain Science Institute, RIKEN, Wakoshi, Japan
[2]Tokyo Institute of Technology, Japan

ABSTRACT

Decomposition of symmetric tensors has found numerous applications in blind sources separation, blind identification, clustering, and analysis of social interactions. In this paper, we consider fourth order symmetric tensors, and its symmetric tensor decomposition. By imposing unit-length constraints on components, we resort the optimisation problem to the constrained eigenvalue decomposition in which eigenvectors are represented in form of rank-1 matrices. To this end, we develop an augmented Lagrangian algorithm with simple update rules. The proposed algorithm has been compared with the Trust-Region solver over manifold, and achieved higher success rates. The algorithm is also validated for blind identification, and achieves more stable results than the ALSCAF algorithm.

Index Terms— symmetric tensor decomposition, spherical quadratic programming, augmented Lagrangian algorithm

1. INTRODUCTION

Symmetric tensor is a tensor with particular structure, invariant under any permutation of its indices, i.e. tensor permutation. The concept of symmetric tensor is extended from symmetric matrix, and occurs widely in engineering, physics and mathematics. In signal processing, symmetric tensors can be generated as cumulant tensors [1–3], or using characteristic generating function in blind source or identification [1, 4–6]. The symmetric tensor can also represent similarity or interaction between groups of identities, e.g., differences between patches in images, the number of emails exchanging between members, shared-publications between researchers [7–9]. With these representation, decomposition of the symmetric tensors can be used to find common structures between samples, and is useful for clustering [10,11]. In analogy with the theory of symmetric matrices, one can compute eigenvalues and eigenvectors of the symmetric tensors [12].

In this paper, we consider a particular case of the symmetric tensors of order-4, and develop a novel algorithm for decomposition of the tensors into rank-1 symmetric tensors. The tensors can be fourth-order cumulant tensors of the mixture of a linear mixing system in BSS problem [1]. We show that the joint diagonalization of symmetric matrices can be converted to the best rank-1 approximation of symmetric tensor of order-4. For such decompositions, one can apply the Higher-order power method [13], or its symmetric version [14]. The decomposition is also related to the higher order INDSCAL tensor decomposition. For this problem, one can employ the damped Gauss-Newton algorithm [15] which updates all parameters at a time by exploiting the structure of the Hessian. Alternatively, Wang and Qi [16] proposed a successive decomposition method.

In [17], by converting the symmetric tensor to the corresponding homogeneous polynomial, the symmetric tensor decomposition reduces to the decomposition of homogeneous polynomial as a sum of powers of linear forms (Waring's problem). From which the authors deduced the decomposition by solving a simple eigenvalue problem, by means of linear algebra manipulations. The results are often complex-valued.

In this paper, our algorithm first focuses on finding best symmetric rank-1 tensor. Decomposition of a symmetric tensor into high symmetric rank-1 terms can be resorted to best rank-1 symmetric tensor approximation to the residue between the data and the other rank-1 terms [18]. By introducing additional parameters, we convert the best rank-1 symmetric tensor approximation to the constrained eigenvalue decomposition in which eigenvectors are represented in form of rank-1 matrices. Finally, an augmented Lagrangian algorithm for the constrained optimisation problem has been developed. Simulation results show that compared with the Trust-Region algorithm which minimises the problem over sphere, our algorithm achieves higher success rates.

2. DECOMPOSITION OF SYMMETRIC MATRICES

The joint diagonalization of symmetric matrices is known as one of popular methods for blind source separation [19]. The symmetric matrices can be covariance matrices [20], or second derivatives of the cumulant generating function [21], or pairwise distance between samples varying over times as in INDSCAL [7, 15]. We shall present link between the joint diagonalization of symmetric matrices and the decomposition of order-4 symmetric tensor.

Let $\mathbf{Y}_t$ be symmetric matrices, and t-th frontal slices of a tensor $\boldsymbol{\mathcal{Y}}$ of size $I \times I \times T$, $t = 1, \ldots, T$. The main aim of joint diagonalization of $\mathbf{Y}_t$ is to find a matrix $\mathbf{A}$ such that its (pseudo-)inverse $\mathbf{A}^\dagger$ jointly diagonalizes $\mathbf{Y}_t$, which is alternatively expressed as a symmetric tensor decomposition as

$$\mathbf{Y}_t = \mathbf{A}\,\mathrm{diag}(\boldsymbol{b}_t)\mathbf{A}^T$$

where $\boldsymbol{b}_t$ are t-th row vectors of a matrix $\mathbf{B}$. The decomposition can be achieved by minimizing a cost function given by

$$\min \quad D = \sum_{t=1}^{T} \|\mathbf{Y}_t - \sum_{r=1}^{R} b_{r,t}\, \boldsymbol{a}_r \boldsymbol{a}_r^T\|_F^2 \tag{1}$$

where $\boldsymbol{a}_r$ are columns of the matrix $\mathbf{A}$, which can be further assumed to have unit-norm. Let $\mathbf{Y}_{r,t} = \mathbf{Y}_t - \sum_{s \neq r} b_{s,t}\, \boldsymbol{a}_s \boldsymbol{a}_s^T$, the above cost function is then rewritten in a quadratic form of $b_{r,t}$

$$\begin{aligned} \min \quad & \sum_{t=1}^{T} \|\mathbf{Y}_{r,t} - b_{r,t}\, \boldsymbol{a}_r \boldsymbol{a}_r^T\|_F^2 \qquad (2)\\ = \quad & \sum_{t=1}^{T} \|\mathbf{Y}_{r,t}\|_F^2 + b_{r,t}^2 - 2\, b_{r,t}\, (\boldsymbol{a}_r^T \mathbf{Y}_{r,t}\, \boldsymbol{a}_r). \end{aligned}$$

Implying that the optimal $b^\star_{r,t} = \boldsymbol{a}_r^T \mathbf{Y}_{r,t}\, \boldsymbol{a}_r$, and the optimisation problem simplifies into a maximisation problem

$$\begin{aligned} \max \quad & \sum_{t=1}^{T} (\boldsymbol{a}_r^T \mathbf{Y}_{r,t}\, \boldsymbol{a}_r)^2 \\ = \quad & (\boldsymbol{a}_r \otimes \boldsymbol{a}_r)^T \left(\sum_{t=1}^{T} \mathrm{vec}(\mathbf{Y}_{r,t})\, \mathrm{vec}(\mathbf{Y}_{r,t})^T \right) (\boldsymbol{a}_r \otimes \boldsymbol{a}_r) \\ = \quad & \mathcal{Q} \bullet (\boldsymbol{a}_r \circ \boldsymbol{a}_r \circ \boldsymbol{a}_r \circ \boldsymbol{a}_r), \end{aligned} \tag{3}$$

where $\mathcal{Q}$ is a symmetric tensor of order-4, "$\otimes$", "$\circ$" and "$\bullet$" represent the Kroncker product, the outer product, and the inner product between two tensors, respectively [22]. It is clear that solving the optimisation (1) or (2) leads to finding the leading eigenvector $\boldsymbol{a}_r$ of the symmetric tensor $\mathbf{Q}$.

3. BEST RANK-1 APPROXIMATION TO SYMMETRIC TENSOR OF ORDER-4

We now consider an order-4 symmetric tensor $\mathcal{Y}$ of size $I \times I \times I \times I$. Decomposition of $\mathcal{Y}$ into R symmetric rank-1 terms is done by minimizing the Frobenius norm of the error

$$\min_{\lambda_r, \boldsymbol{x}_r} \quad \|\mathcal{Y} - \sum_{r=1}^{R} \lambda_r\, \boldsymbol{x}_r \circ \boldsymbol{x}_r \circ \boldsymbol{x}_r \circ \boldsymbol{x}_r\|_F^2 \tag{4}$$

where λ_r are weights of rank-1 tensors, and $\boldsymbol{x}_r$ are unit length vectors, $\boldsymbol{x}_r^T \boldsymbol{x}_r = 1$, for $r = 1, 2, \ldots, R$. In BSS or blind identification, R is the number of sources. The above optimisation problem (4) can be recast as sequential best rank-1 symmetric tensor approximations

$$\min \quad \|\mathcal{Y}_r - \lambda_r\, \boldsymbol{x}_r \circ \boldsymbol{x}_r \circ \boldsymbol{x}_r \circ \boldsymbol{x}_r\|_F^2 \tag{5}$$

where the tensor $\mathcal{Y}_r = \mathcal{Y} - \sum_{s \neq r} \lambda_s\, \boldsymbol{x}_s \circ \boldsymbol{x}_s \circ \boldsymbol{x}_s \circ \boldsymbol{x}_s$ is defined as in the hierarchical alternating update [18, 23]. We will present a novel algorithm for the optimisation in (5).

For simplicity, we write the decomposition of $\mathcal{Y}$ as

$$\min_{\lambda, \boldsymbol{x}} \quad \|\mathcal{Y} - \lambda\, (\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x})\|_F^2 \quad \text{s.t.} \quad \boldsymbol{x}^T \boldsymbol{x} = 1. \tag{6}$$

Because (6) is a quadratic function of λ, the optimal weight $\lambda^\star = \mathcal{Y} \bullet (\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x})$. By replacing λ in (6) by $\lambda^\star$, the optimization (6) becomes

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \|\mathcal{Y} - \lambda^\star (\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x})\|_F^2 \\ = & \|\mathcal{Y}\|_F^2 + (\lambda^\star)^2 \|\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x}\|_F^2 - 2\lambda^\star\, (\mathcal{Y} \bullet (\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x})) \\ = & \|\mathcal{Y}\|_F^2 - (\mathcal{Y} \bullet (\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x}))^2 \end{aligned} \tag{7}$$

subject to $\boldsymbol{x}^T \boldsymbol{x} = 1$. In [14], the authors modified the HOPM algorithm to solve the above problem. In this paper, with a different observation, we interpret the above problem as two separate optimisation problems, which are then reformulated as constrained eigenvalue decompositions. For a positive λ, we maximise the inner product

$$\max \quad \mathcal{Y} \bullet (\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{x}^T \boldsymbol{x} = 1\,, \tag{8}$$

and minimise it for a negative λ

$$\min \quad \mathcal{Y} \bullet (\boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{x}^T \boldsymbol{x} = 1\,. \tag{9}$$

The two optimization problems indeed can be solved in a similar way, e.g., see Step 3 in Algorithm 1. The final solution λ is the one with the larger absolute value. Since $\boldsymbol{x}$ is constrained to be unit-length vector, both optimisation problems over sphere in (8) and (9) can be solved on Riemannian or Stiefel manifold, e.g., using the Trust-Region solver [24]. The above problems are related to eigenvalue decomposition of symmetric tensors [12].

Let $\boldsymbol{z} = \boldsymbol{x} \otimes \boldsymbol{x}$. The minimisation problem in (9) reads

$$\min \quad \boldsymbol{z}^T \mathbf{Q} \boldsymbol{z}\,, \quad \text{s.t.} \;\; \boldsymbol{z} = \boldsymbol{x} \otimes \boldsymbol{x}\,,\;\; \boldsymbol{z}^T \boldsymbol{z} = 1 \tag{10}$$

where $\mathbf{Q}$ is a symmetric matrix, obtained by stacking vectorization of slices $\mathcal{Y}(:, :, i, j)$ into a matrix, i.e., mode-(1,2) matricization of $\mathcal{Y}$. The above optimisation problem is a constrained eigenvalue decomposition, in which the eigenvector $\boldsymbol{z}$ is a rank-1 symmetric matrix after being reshaped into a matrix of size $I \times I$.

In order to solve the above constrained optimisation problem, we construct the augmented Lagrangian function

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = f(\boldsymbol{z}) + \boldsymbol{y}^T (\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}) + \frac{\gamma}{2} \|\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}\|^2 \tag{11}$$

where $\gamma > 0$, and $f(\boldsymbol{z})$ is the objective function of the minimization of $\boldsymbol{z}^T \mathbf{Q} \boldsymbol{z}$ subject to $\boldsymbol{z}^T \boldsymbol{z} = 1$. Variables $\boldsymbol{x}$, $\boldsymbol{z}$ and $\boldsymbol{y}$ are sequentially updated in the following sequence

$$\begin{aligned} \boldsymbol{z} &= \arg\min \quad f(\boldsymbol{z}) + \boldsymbol{y}^T (\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}) + \frac{\gamma}{2} \|\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}\|^2 \\ &= \arg\min \quad \frac{1}{2} \boldsymbol{z}^T \mathbf{Q} \boldsymbol{z} + (\boldsymbol{y} - \gamma \boldsymbol{x} \otimes \boldsymbol{x})^T \boldsymbol{z} \quad \text{s.t.} \quad \boldsymbol{z}^T \boldsymbol{z} = 1 \qquad (12) \\ \boldsymbol{x} &= \arg\min \quad \boldsymbol{y}^T (\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}) + \frac{\gamma}{2} \|\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}\|^2 \\ &= \arg\min \quad \|\boldsymbol{z} + \frac{\boldsymbol{y}}{\gamma} - \boldsymbol{x} \otimes \boldsymbol{x}\|^2 \qquad (13) \end{aligned}$$

$$\boldsymbol{y} \leftarrow \boldsymbol{y} + \gamma (\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x})\,. \tag{14}$$

3.1. Updating z

The unit-length vector $\boldsymbol{z}$ is a minimiser to a quadratic problem over sphere (12), which indeed can be found in closed-form [25, 26]. Let denote by $\mathbf{Q} = \mathbf{U}\, \mathrm{diag}(\boldsymbol{\sigma}) \mathbf{U}^T$, the eigenvalue decomposition of the matrix $\mathbf{Q}$, where $\boldsymbol{\sigma} = [0 \leq \sigma_1 \leq$

Algorithm 1: Augmented Lagrangian Algorithm for Best rank-1 Tensor Approximation

Input: Order-4 tensor $\mathcal{Y}$ of size $I \times I \times I \times I$
Output: λ and $\boldsymbol{x}$ minimise $\frac{1}{2}\|\mathcal{Y} - \lambda\, \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x})\|$
begin
1 $\mathbf{Q} = [\mathcal{Y}]_{(\{1,2\})}$: mode-(1,2) matricization of $\mathcal{Y}$
2 $\boldsymbol{x}_- = \mathtt{tensor\_eig}(\mathbf{Q})$, $\lambda_- = \mathcal{Y} \bullet (\boldsymbol{x}_- \circ \boldsymbol{x}_- \circ \boldsymbol{x}_- \circ \boldsymbol{x}_-)$
3 $\boldsymbol{x}_+ = \mathtt{tensor\_eig}(-\mathbf{Q})$, $\lambda_+ = \mathcal{Y} \bullet (\boldsymbol{x}_+ \circ \boldsymbol{x}_+ \circ \boldsymbol{x}_+ \circ \boldsymbol{x}_+)$
4 **if** $|\lambda_-| > |\lambda_+|$ **then** $\lambda = \lambda_-$, $\boldsymbol{x} = \boldsymbol{x}_-$
5 **else** $\lambda = \lambda_+$, $\boldsymbol{x} = \boldsymbol{x}_+$

function $\boldsymbol{x} = \mathtt{tensor\_eig}(\mathbf{Q})$
Input: $\mathbf{Q}$: symmetric of size $I^2 \times I^2$
Output: $\boldsymbol{x}$ minimises $\frac{1}{2}\, \boldsymbol{z}^T \mathbf{Q} \boldsymbol{z}$, s.t., $\boldsymbol{x}^T\boldsymbol{x} = 1$, $\boldsymbol{z} = \boldsymbol{x} \otimes \boldsymbol{x}$
begin
6 Initialize $\boldsymbol{y}$ and $\boldsymbol{z}$ as zero vectors and $\gamma > 0$
repeat
% **Update $\boldsymbol{z}$**
7 $\boldsymbol{z} = \mathtt{spherical\_quadratic\_prog}(\mathbf{Q}, \boldsymbol{y} - \gamma\, \boldsymbol{x} \otimes \boldsymbol{x})$
% **Update $\boldsymbol{x}$**
8 $\mathbf{T} = \mathtt{reshape}(\boldsymbol{z} - \frac{\boldsymbol{y}}{\gamma}, [\mathtt{I} \times \mathtt{I}])$, $\mathbf{T}_s = \frac{1}{2}(\mathbf{T} + \mathbf{T}^T)$
9 $\mathbf{T}_s \approx \sigma \boldsymbol{x}\boldsymbol{x}^T$
% **Update $\boldsymbol{y}$**
10 $\boldsymbol{y} \leftarrow \boldsymbol{y} + \gamma(\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x})$
11 Adjust $\gamma \leftarrow \alpha\gamma$ if the objective function tents to a slow convergence
until *a stopping criterion is met*

$\sigma_2 \le \ldots \le \sigma_K]$ comprises eigenvalues of $\mathbf{Q}$, $\mathbf{U}$ is an orthonormal matrix of size $K \times K$, $K = I^2$, consists of eigenvectors of $\mathbf{Q}$. The vector $\boldsymbol{z}$ is computed as

$$\boldsymbol{z} = \mathbf{U}\,\mathrm{diag}\left(\frac{1}{\lambda - s_1}, \ldots, \frac{1}{\lambda - s_K}\right)\boldsymbol{c} \tag{15}$$

where $\boldsymbol{b} = \boldsymbol{y} - \gamma\boldsymbol{x} \otimes \boldsymbol{x}$, $\boldsymbol{c} = \mathbf{U}^T \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}$, $\boldsymbol{s} = [s_1, \ldots, s_K]$ is a normalized version of the eigenvalues $\boldsymbol{\sigma}$,

$$s_k = \frac{\sigma_k - \sigma_1}{\|\boldsymbol{b}\|} + 1\,, \tag{16}$$

and λ is the unique root in [0, 1] of the following secular equation

$$\sum_{k=1}^{K} \frac{c_k^2}{(\lambda - s_k)^2} = 1\,. \tag{17}$$

3.2. Updating $\boldsymbol{x}$

The vector $\boldsymbol{x}$ in (13) is the eigenvector associated with the largest eigenvalue of the symmetric matrix $\mathbf{T}_s = \frac{1}{2}(\mathbf{T} + \mathbf{T}^T)$ of size $I \times I$, where $\mathbf{T} = \mathbf{Z} + \frac{1}{\gamma}\mathbf{Y}$, and $\boldsymbol{z} = \mathrm{vec}(\mathbf{Z})$ and $\boldsymbol{y} = \mathrm{vec}(\mathbf{Y})$.

3.3. The Proposed Algorithm

The proposed algorithm is summarised in Algorithm 1. The sub-routine `tensor_eig` implements the augmented Lagrangian Algorithm for the constrained optimisation in (10). In step 7, updating $\boldsymbol{z}$ in (15) involves the quadratic programming over sphere, which needs to compute the EVD of the matrix $\mathbf{Q}$ and solve secular equations in (17). Since the quadratic term $\mathbf{Q}$ does not change with iterations, the EVD of $\mathbf{Q}$ is computed only once outside of the loop. The main cost of the algorithm is due to computing the first principal eigenvector of the symmetric matrix $\mathbf{T}_s$ of size $I \times I$.

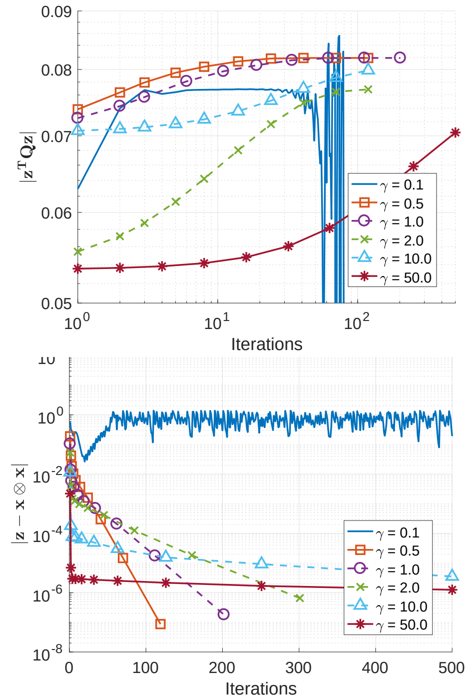
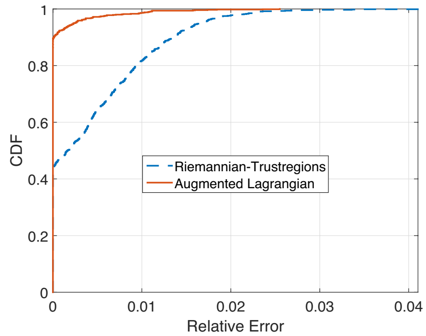


Fig. 1. Convergence behaviour and error of the rank-1 constraint $\|\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}\|$ of the proposed algorithm for various γ in solving the optimization (8).

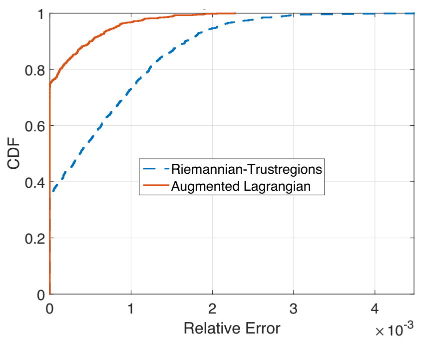
The vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ are initialised as zeros. The regularized parameter γ enforces the rank-1 constraint onto $\boldsymbol{z}$. When running the algorithm with a high value of γ, $\boldsymbol{z}$ quickly holds the rank-1 constraint, but the objective function converges slowly. This is illustrated in Fig. 1 for the case when $\gamma = 50$ and tensors of size $10 \times 10 \times 10 \times 10$. Setting γ to a relatively small value can make the algorithm unstable after several to dozen of iterations. For example, see convergence of the algorithm when $\gamma = 0.1$ in Fig. 1. In order to obtain a good setting, we should run the algorithm in a few iterations with various values of γ, then choose the setting which gives a good convergence result.

4. SIMULATIONS

Example 1 ***(Best rank-1 tensor approximation to symmetric tensor of order-4)*** This example compares performance of our proposed algorithm for the best rank-1 tensor approximation for symmetric tensors of order-4, and the algorithm using the Riemannian trust-region solver in the Manopt toolbox [24]. We generated 1000 random tensors of size $I \times I \times I \times I$, where $I = 10$ or 20, then matricized them so that they were symmetric tensors of order-4. The tensors were normalized to have unit Frobenius norm. For each run, the



(a) $I = 10$



(b) $I = 20$

Fig. 2. Empirical cumulative distribution functions of the relative errors returned by the augmented Lagrangian algorithm and the Riemannian trust-region method.

best approximation error $\varepsilon^\star$ was defined as the smaller error among approximation errors of the two methods: Augmented Lagrangian method (Algorithm 1) and the Riemannian trust-region

$$\varepsilon = \|\mathcal{Y} - \lambda \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x}\|_F^2 = 1 - \lambda^2 . \tag{18}$$

Relative errors to the best approximation error $\frac{\varepsilon - \varepsilon^\star}{\varepsilon^\star}$ is used to assess success rate of the considered approximation.

The parameter γ was chosen among values [0.1, .2, .5, 1,2, 10, 50]. In Fig. 1, we illustrate the convergence behaviour of the proposed algorithm with various selection of γ. The algorithm completely failed when $\gamma = 0.1$. The objective function diverged, while the rank-1 condition $\|\boldsymbol{z} - \boldsymbol{x} \otimes \boldsymbol{x}\|$ did not preserve. When $\gamma = 10$, the algorithm converged, but the error of the constrained reduced slowly. For this decomposition, $\gamma = 0.5$ or 1 is a good setting.

In Fig. 2, we plot empirical cumulative distribution functions of 1000 relative errors. The results indicate that our algorithm achieved higher success rate. For example, for the case when $I = 20$, our algorithm attained an error less than 0.001 with a rate of 96.8%, whereas the trust-region solver achieved a rate of 73.1% for the same error range. When $I = 10$, the augmented Lagrangian algorithm had a success rate of 92.5% for a similar accuracy of 0.001, while the trust-region algorithm had a quite low rate of 47.4%.

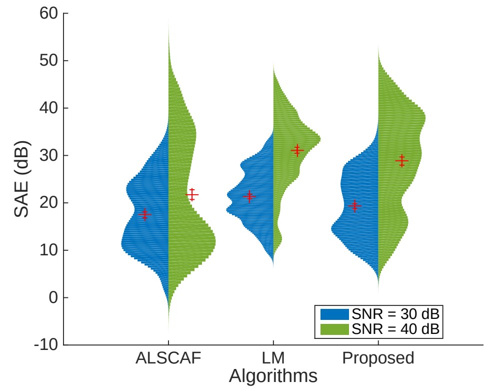


Fig. 3. Comparison of squared angular errors of ALSCAP, LM and the proposed algorithm in Example 2. Red crosses show mean ± standard error of the mean.

Example 2 *(***Blind identification of 4 sources from 3 mixtures.)** In this example, we illustrate the proposed algorithm for a blind identification problem for 4 sources and 3 mixtures, $\mathbf{X} = \mathbf{HS} + \mathbf{E}$. The sources $\mathbf{S}$ were synthesized real-valued 2-QAM signals, the mixing matrix $\mathbf{H}$ of size 3×4 was randomly generated, and the Gaussian noise was added into $\mathbf{S}$ such the Signal-Noise-Ratios SNR = 30 or 40 dB. More specifically, we estimated $\mathbf{H}$ from 100 second-order derivatives of the cumulant generating function of the observations [21], by sequentially solving the maximisation in (3), i.e., the case in (8). We verified performances of the ALSCAF algorithm [21], the Levenberg-Marquardt algorithm [15], and the proposed algorithm over 200 independent runs. Distributions of the squared angular errors $(SAE(\boldsymbol{h}, \hat{\boldsymbol{h}}) = -20 \log_{10} \arccos \frac{\boldsymbol{h}^T \hat{\boldsymbol{h}}}{\|\boldsymbol{h}\|_2 \|\hat{\boldsymbol{h}}\|_2})$ of the considered algorithms are compared in Fig. 3. The results indicate that our algorithm achieved better performance than the ALSCAF algorithm for both cases when SNR = 30 and 40 dB, and was at most comparable to the LM algorithm.

5. CONCLUSIONS

Different from other existing algorithms for symmetric tensor decompositions, we have interpreted the problem as two constrained eigenvalue decompositions in which eigenvectors are rank-1 matrices. In our proposed augmented Lagrangian algorithm, parameters are updated in closed-form, one vector $\boldsymbol{z}$ is updated based on the quadratic programming over sphere, and the vector $\boldsymbol{x}$ is the principal eigenvector of a symmetric matrix. Simulation results have confirmed convergence of our proposed algorithm, and its superior over the Trust-Region solver over manifold. The algorithm can be extended to nonnegative symmetric tensor factorization, or symmetric tensors of higher order.

6. REFERENCES

[1] J. F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. blind identification of more sources than sensors," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP91)*, 1991, vol. 5, pp. 3109–3112.

[2] S.A. Cruces, L. Castedo, and A. Cichocki, "Robust blind source separation algorithms using cumulants," *Neurocomputing*, vol. 49, pp. 87–118, Dec. 2002.

[3] L. de Lathauwer, J. Castaing, and J.F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. on Signal Processing*, vol. 55, no. 6, pp. 2965–2973, 2007.

[4] N. Yuen and B. Friedlander, "Asymptotic performance analysis of blind signal copy using fourth order cumulant," *Int. Journal of Adapative Control Signal Processing*, vol. 10, no. 2–3, pp. 239–265, 1996.

[5] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1260–1280, 2014.

[6] A. L. F. de Almeida, X. Luciani, A. Stegeman, and P. Comon, "CONFAC decomposition approach to blind identification of underdetermined mixtures based on generating function derivatives," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5698–5713, 2012.

[7] J.D. Carroll and J.J. Chang, "Analysis of individual differences in multidimensional scaling via an *n*-way generalization of Eckart–Young decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[8] B.W. Bader, R.A. Harshman, and T.G. Kolda, "Temporal analysis of semantic graphs using ASALSAN," in *ICDM 2007: Proceedings of the 7th IEEE International Conference on Data Mining*, October 2007, pp. 33–42.

[9] A.-H. Phan, A. Cichocki, and T. V. Dinh, "Nonnegative DEDICOM based on tensor decompositions for social networks exploration," *Austr. J. Intelligent Information Processing Systems*, vol. 12, no. 1, 2010.

[10] A. Shashua, R. Zass, and T. Hazan, "Multi-way clustering using super-symmetric non-negative tensor factorization," in *European Conference on Computer Vision (ECCV)*, Graz, Austria, May 2006.

[11] D. Muti and S. Bourennane, "Multiway filtering based on fourth order cumulants," *Applied Signal Processing EURASIP*, vol. 7, pp. 1147–1159, 2005.

[12] L. Qi, "Eigenvalues of a real supersymmetric tensor," *Journal of Symbolic Computation*, vol. 40, no. 6, pp. 1302 – 1324, 2005.

[13] L. De Lathauwer, P. Comon, B. De Moor, and J. Vandewalle, "Higher-order power method – application in independent component analysis," in *Proceedings of 1995 International Symposium on Nonlinear Theory and its Applications (NOLTA-95)*, 1995, pp. 91–96.

[14] E. Kofidis, Phillip, and A. Regalia, "On the best rank-1 approximation of higher-order supersymmetric tensors," *SIAM J. Matrix Anal. Appl*, vol. 23, pp. 863–884, 2002.

[15] Z. Koldovský, P. Tichavský, and A.-H. Phan, "Stability analysis and fast damped Gauss-Newton algorithm for INDSCAL tensor decomposition," in *Statistical Signal Processing Workshop (SSP), IEEE*, 2011, pp. 581–584.

[16] Y. Wang and L. Qi, "On the successive supersymmetric rank-1 decomposition of higher-order supersymmetric tensors," *Numerical Linear Algebra with Applications*, vol. 14, no. 6, pp. 503–519, 2007.

[17] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, "Symmetric tensor decomposition," *Linear Algebra and its Applications*, vol. 433, no. 11, pp. 1851 – 1872, 2010.

[18] A. Cichocki and A.-H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Transactions*, vol. 92-A, no. 3, pp. 708–721, 2009.

[19] G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavský, and A. Yeredor, "Joint matrices decompositions and blind source separation," *IEEE Signal Processing Magazine*, vol. 31, pp. 34–43, 2014.

[20] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and É. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.

[21] X. Luciani, A. L. F. de Almeida, and P. Comon, "Blind identification of underdetermined mixtures based on the characteristic function: The complex case," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 540–553, 2011.

[22] A. Cichocki, D. P. Mandic, A.-H. Phan, C. Caifa, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing applications. from two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.

[23] A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, Chichester, 2009.

[24] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *Journal of Machine Learning Research*, vol. 15, pp. 1455–1459, 2014.

[25] W. Gander, G. H. Golub, and U. von Matt, "A constrained eigenvalue problem," *Linear Algebra and its Applications*, vol. 114, pp. 815 – 839, 1989.

[26] W. W. Hager, "Minimizing a quadratic over a sphere," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 188–208, 2001.