Image and Video Completion by Using Bayesian Tensor Decomposition

Lihua Gui*, Xuyang Zhao*♯, Qibin Zhao♭♯ and Jianting Cao*

*** Graduate School of Engineering, Saitama Institute of Technology**
Fukaya, Japan
♯ Tensor Learning Unit, RIKEN AIP
Tokyo, Japan
♭ School of Automation, Guangdong University of Technology
Guangzhou, China

Abstract

Reconstruction of image and video from sparse observations attract a great deal of interest in the filed of image/video compression, feature extraction and denoising. Since the color image and video data can be naturally expressed as a tensor structure, many methods based on tensor algebra have been studied together with promising predictive performance. However, one challenging problem in those methods is tuning parameters empirically which usually requires computational demanding cross validation or intuitive selection. In this paper, we introduce Bayesian Tucker decomposition to reconstruct image and video data from incomplete observation. By specifying the sparsity priors over factor matrices and core tensor, the tensor rank can be automatically inferred via variational bayesian, which greatly reduce the computational cost for model selection. We conduct several experiments on image and video data, which shows that our method outperforms the other tensor methods in terms of completion performance.
Keywords: *Image completion, Tensor completion, Bayesian Tucker decomposition.*

1. Introduction

Image or video completion, which is to reconstruct a full image/video from only sparsely observed information, plays an important role in image processing field. Image data can be naturally expressed as a 3rd-order tensor of size $height \times width \times color\ channel$, while the video data can be represented as 4th-order tensor of size $height \times width \times color \times time$. Tensor is an extension of vectors and matrices to the high order case, which enables us to represent the structured data. The traditional way to handle such data is firstly transforming the data into the vector or matrix form, then many existing algorithms based on matrix analysis can be employed for image and video processing. However, the adjacent structure information of original data will be lost [1] due to the matricization operations. To overcome this limitation, tensor analysis methods are the emerging technology, which attracts a great deal of attentions in recent years. By using multilinear algebra, tensor decomposition can efficiently exploit the intrinsic high order structure information within the data and provide better interpretability. The most popular models of tensor decomposition are Tucker decomposition [2] and CANDECOMP/PARAFAC (CP) decomposition [3–5]. Moreover, tensor method has been applied in various research field such as: image completion [6–14], signal processing [15–21], brain machine interface (BMI) [22–24], image classification [25, 26], face recognition [27], machine learning [28], etc. Basically, there are two type of methods for tensor completion. One type is based on minimization of the convex relaxation function of tensor rank by using nuclear norm of tensor. The nuclear norm can be defined in several different ways related to the different tensor decomposition models. By applying the appropriate optimization algorithm, we can find the optimal low-rank tensor as the approximation of full tensor. Another type is based on tensor decomposition of incomplete tensor. The specific algorithm must be developed to find latent factors under the specific tensor decomposition model by using partially observed entries. It is necessary to predefine the tensor rank, which is considered as a model selection problem. Although cross-validation can be used to determine an optimal tensor rank, it is quite computational demanding. Especially, when the Tucker decomposition is considered, the number of possibilities of tensor rank increases exponentially to the order of tensor.

To overcome these limitations, we introduce a Bayesian tensor decomposition method to perform image and video com-

Qibin Zhao and Jianting Cao are the corresponding authors.

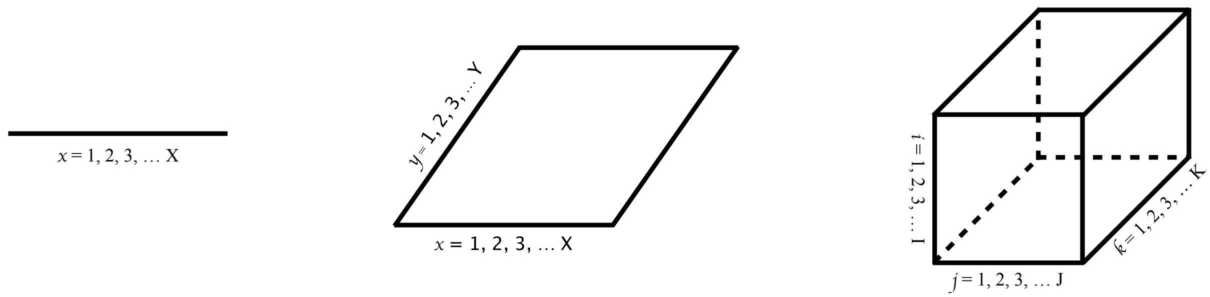

Fig. 1: First, second, thrid-order tensor

pletion. Our methods can automatically adapt model complexity and infer an optimal multilinear rank by the principle of maximum lower bound of model evidence. Experimental results and comparisons on image and video data demonstrate remarkable performance of our models for recovering the groundtruth of multilinear rank and missing pixels.

The remainder of this paper is organized as follows. Section II briefly overviews the Tucker decomposition and CANDECOMP/PARAFAC (CP) decomposition. Section III describes the Bayesian Tucker decomposition method and the corresponding inference algorithm. Section IV shows the experimental results on real-world image and video data. Finally, we summarize our paper in Section V.

2. Tensor decompositions

Tensor is a multidimensional array which is a generalization of vectors and matrices to higher dimensions. First-order tensor is a vector, second-order tensor is a matrix, and third or higher order tensor is higher-order tensor. First, second, third-order tensor are shown in Fig. 1. Tensor decompositions originated from Hitchcock [29] [30]. Under the work of Tucker [31] [32] [2], Carroll and Chang [3], Harshman [4], Appellof and Davidson [33], the tensor theory and tensor decompositions (factorizations) algorithms have been successfully applied to various fields, examples include signal processing, computer vision and etc.

2.1 Notation

The order of a tensor is the number of dimensions [34]. Tensor of order one (vector) is denoted by boldface lowercase letters, e.g., $\mathbf{a}$, the i-th element of a one-order tensor is denoted by a_i. Tensor of order two (matrix) is denoted by boldface capital letters, e.g., $\mathbf{A}$, the (i, j) element of a two-order tensor is denoted by a_{ij}. Tensor of order three or higher (higher-order tensor) is denoted by boldface Euler script letters, e.g., $\boldsymbol{\mathcal{X}}$, the (i, j, k) element of a three-order tensor is denoted by x_{ijk}. Indices typically range from 1 to their capital version, e.g., $i = 1, ..., I$.

2.2 CP decomposition

CANDECOMP/PARAFAC (CP) decomposition method is proposed by Carroll and Chang [3] and PARAFAC (parallel factors) proposed by Harshman [4]. Usually, we refer to the CANDECOMP/PARAFAC decomposition as CP [5]. CP decomposition is to represent a tensor as the sum of R rank-one tensors. For example, given a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$, we wish to represent it by (1) (2).

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]. \quad (1)$$

$$\begin{gathered} x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}, \\ \forall\, i = 1,..., I, \forall\, j = 1, ..., J, \forall\, k = 1, ..., K. \end{gathered} \quad (2)$$

where $\mathbf{a}_r \in \mathbb{R}^I$, $\mathbf{b}_r \in \mathbb{R}^J$ and $\mathbf{c}_r \in \mathbb{R}^K$, $\forall r = 1, \ldots, R$. The rank of a tensor $\boldsymbol{\mathcal{X}}$, denoted by R = rank($\boldsymbol{\mathcal{X}}$), is define as the smallest number of rank-one tensors that can exactly represent $\boldsymbol{\mathcal{X}}$. The scheme of CP decompositions is illustrated in Fig. 2.

2.3 Tucker decomposition

The Tucker decomposition is proposed firstly in 1963 [35], and refined in subsequent articles by Levin [33] and Tucker [2, 32]. Tucker decomposition can be considered as a form of higher-order PCA (Principal Components Analysis), which

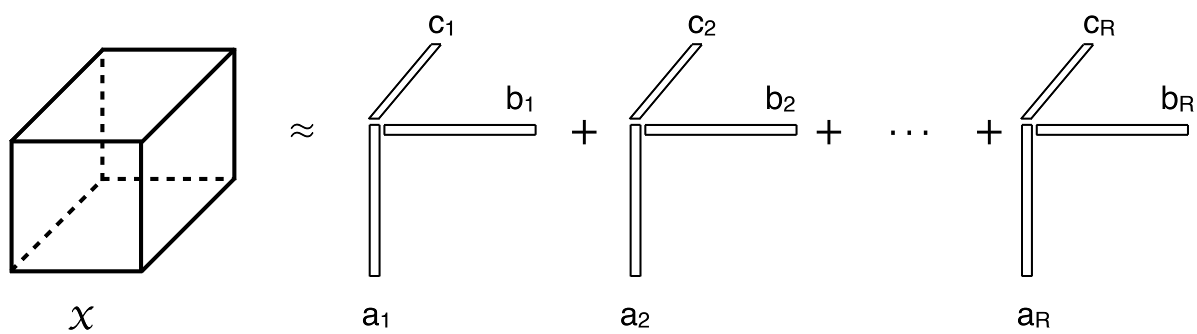

Fig. 2: CP decomposition of a third-order tensor

decomposes a tensor into a core tensor multiplied (or transformed) by several matrices along each mode. For instance, given a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$, we have

$$\begin{aligned}\boldsymbol{\mathcal{X}} &= \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \\ &= \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} g_{pqr} \circ \mathbf{a_p} \circ \mathbf{b_q} \circ \mathbf{c_r} \\ &= [\![\boldsymbol{\mathcal{G}}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!].\end{aligned} \tag{3}$$

Eq. (3) can be also written by the element-wise form, which is

$$\begin{aligned}x_{ijk} \approx \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr}, \\ \forall\, i = 1, ..., I, \forall\, j = 1, ..., J, \forall\, k = 1, ..., K.\end{aligned} \tag{4}$$

Here, $\mathbf{A} \in \mathbb{R}^{I\times P}$, $\mathbf{B} \in \mathbb{R}^{J\times Q}$ and $\mathbf{C} \in \mathbb{R}^{K\times R}$ are the factor matrices, which are usually orthogonal, and can be thought of as the principal components in each mode. Tensor $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{P\times Q\times R}$ is called the *core tensor* and its entries show the level of interaction between the different components. The last equality in (3) uses the shorthand $[\![\boldsymbol{\mathcal{G}}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ which was introduced in [34]. The scheme of Tucker decompositions is illustrated in Fig. 3.

3. Bayesian Tucker decompositions

In this section, we introduce Bayesian Tucker decomposition for tensor completion. Let $\boldsymbol{\mathcal{Y}}$ be an incomplete tensor with missing entries, and $\boldsymbol{\mathcal{O}}$ is a binary tensor which indicates the observation positions. Ω denotes a set of N-tuple indices of observed entries. The value of $\boldsymbol{\mathcal{O}}$ is defined by

$$\begin{cases}\mathcal{O}_{i_1\cdots i_N} = 1 & \text{if } (i_1, \ldots, i_N) \in \Omega, \\ \mathcal{O}_{i_1\cdots i_N} = 0 & \text{if } (i_1, \ldots, i_N) \notin \Omega.\end{cases} \tag{5}$$

$\boldsymbol{\mathcal{Y}}_\Omega$ is a tensor which only include observed entries. The generative model is assumed as

$$\boldsymbol{\mathcal{Y}}_\Omega = \boldsymbol{\mathcal{X}}_\Omega + \varepsilon, \tag{6}$$

where the latent tensor $\boldsymbol{\mathcal{X}}$ is represented exactly by a Tucker model with a low multilinear rank and ε denotes i.i.d. Gaussian noise.

Given an incomplete image tensor, Bayesian Tucker model only considers the observed data, thus the likelihood function can be represented by

$$p\Big(\boldsymbol{\mathcal{Y}}_\Omega\Big) = \prod_{(i_1,i_2,i_3)\in\Omega} \boldsymbol{\mathcal{N}}\Big(\mathcal{Y}_{i_1i_2i_3} \,\Big|\, \mathcal{X}_{i_1i_2i_3},\, \tau^{-1}\Big). \tag{7}$$

Since the latent tensor $\boldsymbol{\mathcal{X}}$ can be decomposed exactly by a Tucker model, we can thus represent the observation model as that $\forall(i_1, i_2, i_3)$,

$$\begin{aligned}\mathcal{Y}_{i_1i_2i_3} \,\Big|\, \Big\{\mathbf{u}_{i_n}^{(n)}\Big\},\, \boldsymbol{\mathcal{G}},\, \tau \sim \\ \boldsymbol{\mathcal{N}}\Big(\Big(\bigotimes_n \mathbf{u}_{i_n}^{(n)T}\Big)\mathrm{vec}(\boldsymbol{\mathcal{G}}),\, \tau^{-1}\Big)^{\mathcal{O}_{i_1i_2i_3}}.\end{aligned} \tag{8}$$

where $n = 1, 2, 3$. $\mathbf{u}_{i_n}^{(n)}$ is the i_n-th row of the factor matrix $\mathbf{U}^{(n)}$, $\boldsymbol{\mathcal{O}}$ is the indicator of missing points. τ is the precision of Gaussian noise.

To employ sparsity priors, we can specify the hierarchical prior distributions by

$$\begin{aligned}\tau &\sim Ga\Big(a_0^\tau,\, b_0^\tau\Big), \\ \mathrm{vec}(\boldsymbol{\mathcal{G}}) \,\Big|\, \Big\{\boldsymbol{\lambda}^{(n)}\Big\},\, \beta &\sim \boldsymbol{\mathcal{N}}\Big\{\mathbf{0},\, \Big(\beta \bigotimes_n \boldsymbol{\Lambda}^{(n)}\Big)^{-1}\Big), \\ \beta &\sim Ga\Big(a_0^\beta,\, b_0^\beta\Big), \\ \mathbf{u}_{i_n}^{(n)} \,\Big|\, \boldsymbol{\lambda}^{(n)} &\sim \boldsymbol{\mathcal{N}}\Big(\mathbf{0},\, \boldsymbol{\Lambda}^{(n)^{-1}}\Big), \quad \forall n, \forall i_n. \\ \boldsymbol{\Lambda}_{r_n}^{(n)} &\sim Ga\Big(a_0^\lambda,\, b_0^\lambda\Big), \quad \forall n, \forall r_n,\end{aligned} \tag{9}$$

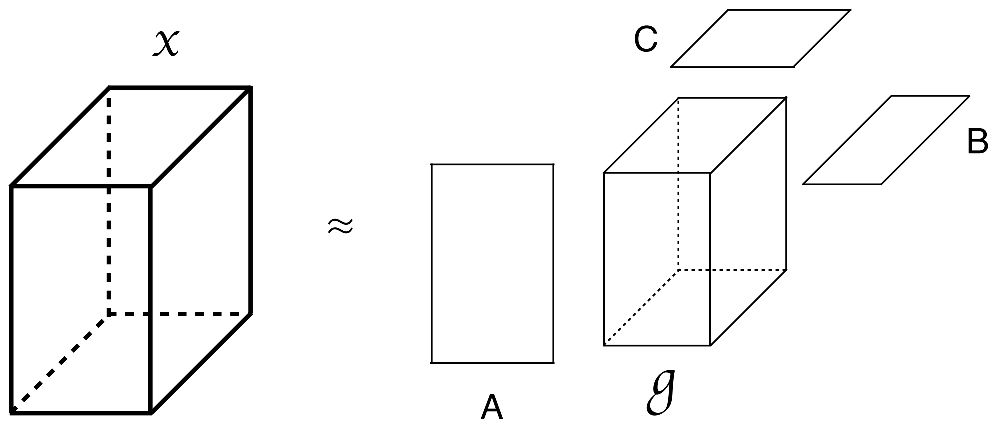

Fig. 3: Tucker decomposition of a third-order tensor

where β is a scale parameter related to the magnitude of $\boldsymbol{\mathcal{G}}$, on which a hyperprior can be placed. The hyperprior for $\boldsymbol{\lambda}^{(n)}$ play a key role for different sparsity inducing priors. We propose the hierarchical prior corresponding to the Student-t distribution for group sparsity. Note that $\boldsymbol{\Lambda}^{(n)} = \mathrm{diag}(\boldsymbol{\lambda}^{(n)})$.

For Tucker decomposition of an incomplete tensor, the problem is ill-conditioned and has infinite solutions. The low-rank assumption play an key role for successful tensor completion, which implies that the determination of multilinear rank significantly affects the predictive performance. However, standard model selection strategies, such as cross-validation, cannot be applied for finding the optimal multilinear rank because it varies dramatically with missing ratios. Therefore, the inference of multilinear rank is more challenging when missing values occur.

As shown in (9), we employ a hierarchical group sparsity prior over the factor matrices and core tensor with aim to seek the minimum multilinear rank automatically, which is more efficient and elegant than the standard model selections by repeating many times and selecting one optimum model. By combining likelihood model in (8), we propose a Bayesian Tucker Completion (BTC) method, which enables us to infer the minimum multilinear rank as well as the noise level solely from partially observed data without requiring the tuning parameters.

To learn the BTC model, we employ the VB inference framework under a fully Bayesian treatment. In this section, we present only the main solutions. As can be derived, the variational posterior distribution over the core tensor $\boldsymbol{\mathcal{G}}$ is given by

$$q(\boldsymbol{\mathcal{G}}) = \boldsymbol{\mathcal{N}}\Big(\mathrm{vec}(\boldsymbol{\mathcal{G}}) \,\Big|\, \mathrm{vec}(\widetilde{\boldsymbol{\mathcal{G}}}),\ \Sigma_G\Big), \tag{10}$$

where the posterior parameters can be updated by

$$\mathrm{vec}(\widetilde{\boldsymbol{\mathcal{G}}}) = \mathbb{E}[\tau]\Sigma_G \sum_{(i_1,i_2,i_3)\in\Omega} \Big(\mathcal{Y}_{i_1 i_2 i_3} \bigotimes_{n=1}^{3} \mathbb{E}\Big[\mathbf{u}_{i_n}^{(n)}\Big]\Big). \tag{11}$$

$$\Sigma_G = \Big\{\mathbb{E}[\beta] \bigotimes_{n} \mathbb{E}\Big[\boldsymbol{\Lambda}^{(n)}\Big] + \mathbb{E}[\tau] \sum_{(i_1,i_2,i_3)\in\Omega} \bigotimes_{n=1}^{3} \mathbb{E}\Big[\mathbf{u}_{i_n}^{(n)} \mathbf{u}_{i_n}^{(n)T}\Big]\Big\}^{-1}. \tag{12}$$

Since the variational posterior distribution over $\{\mathbf{U}^{(n)}\}$ can be factorized as

$$q(\mathbf{U}^{(n)}) = \prod_{i_n} \boldsymbol{\mathcal{N}}\Big(\mathbf{u}_{i_n}^{(n)} \,\Big|\, \widetilde{\mathbf{u}}_{i_n}^{(n)}, \Psi_{i_n}^{(n)}\Big), \quad n = 1, \ldots, 3. \tag{13}$$

the posterior parameters are updated by

$$\widetilde{\mathbf{u}}_{i_n}^{(n)} = \mathbb{E}[\tau]\Psi_{i_n}^{(n)} \mathbb{E}\big[\mathbf{G}_{(n)}\big] \sum_{(i_1,i_2,i_3)\in\Omega} \Big(\mathcal{Y}_{i_1 i_2 i_3} \bigotimes_{k\neq n} \mathbb{E}\Big[\mathbf{u}_{i_k}^{(k)}\Big]\Big). \tag{14}$$

$$\Psi_{i_n}^{(n)} = \Big\{\mathbb{E}[\boldsymbol{\Lambda}^{(n)}] + \mathbb{E}[\tau]\mathbb{E}\Big[\mathbf{G}_{(n)} \boldsymbol{\Phi}_{i_n}^{(n)} \mathbf{G}_{(n)}^T\Big]\Big\}^{-1}. \tag{15}$$

$$\boldsymbol{\Phi}_{i_n}^{(n)} = \sum_{(i_1,\ldots,i_N)\in\Omega} \bigotimes_{k\neq n} \mathbf{u}_{i_k}^{(k)} \mathbf{u}_{i_k}^{(k)T}. \tag{16}$$

The summation is performed over the observed data locations whose mode-n index is fixed to i_n. In other words, $\boldsymbol{\Phi}_{i_n}^{(n)}$ represents the statistical information of mode-k ($k \neq n$) latent factors that interact with $\mathbf{u}_{i_n}^{(n)}$. In (15), the complex posterior expectation can be computed efficiently by

$$\mathrm{vec}\left\{ \mathbb{E}\left[\mathbf{G}_{(n)}\boldsymbol{\Phi}_{i_n}^{(n)}\mathbf{G}_{(n)}^T\right]\right\} = \mathbb{E}\Big[\mathbf{G}_{(n)} \otimes \mathbf{G}_{(n)}\Big]\mathrm{vec}\Big(\boldsymbol{\Phi}_{i_n}^{(n)}\Big). \tag{17}$$

The variation posterior distribution over $\{\boldsymbol{\lambda}^{(n)}\}$ is i.i.d. Gamma distributions due to the conjugate priors, which is $\forall n = 1, \ldots, 3$,

$$q(\boldsymbol{\lambda}^{(n)}) = \prod_{r_n=1}^{R_n} Ga\Big(\lambda_{r_n}^{(n)} \mid \tilde{a}_{r_n}^{(n)},\, \tilde{b}_{r_n}^{(n)}\Big). \tag{18}$$

where the posterior parameters can be updated by

$$\begin{aligned} \tilde{a}_{r_n}^{(n)} &= a_0^\lambda + \frac{1}{2}\Big(I_n + \prod_{k\neq n} R_k\Big), \\ \tilde{b}_{r_n}^{(n)} &= b_0^\lambda + \frac{1}{2}\mathbb{E}\Big[\mathbf{u}_{\cdot r_n}^{(n)T}\mathbf{u}_{\cdot r_n}^{(n)}\Big] \\ &\quad + \frac{1}{2}\mathbb{E}[\beta]\mathbb{E}\Big[\mathrm{vec}(\boldsymbol{\mathcal{G}}^2_{\cdots r_n \cdots})^T\Big]\bigotimes_{k\neq n}\mathbb{E}\big[\boldsymbol{\lambda}^{(k)}\big]. \end{aligned} \tag{19}$$

Finally, the predictive distributions over missing entries, given observed entries, can be approximated by using variational posterior distributions $q(\Theta)$ as follows

$$\begin{aligned} p(\mathcal{Y}_{i_1 i_2 i_3} \mid \boldsymbol{\mathcal{Y}}_\Omega) &= \int p(\mathcal{Y}_{i_1 i_2 i_3} \mid \Theta) p(\Theta \mid \boldsymbol{\mathcal{Y}}_\Omega)\, \mathrm{d}\Theta \\ &\approx \boldsymbol{\mathcal{N}}\Big(\mathcal{Y}_{i_1 i_2 i_3} \,\Big|\, \tilde{\mathcal{Y}}_{i_1 i_2 i_3},\, \mathbb{E}[\tau]^{-1} + \sigma^2_{i_1 i_2 i_3}\Big). \end{aligned} \tag{20}$$

where the posterior parameters can be obtained by

$$\begin{aligned} \tilde{\mathcal{Y}}_{i_1 i_2 i_3} &= \Big(\bigotimes_n \mathbb{E}\Big[\mathbf{u}_{i_n}^{(n)T}\Big]\Big)\mathbb{E}\big[\mathrm{vec}(\boldsymbol{\mathcal{G}})\big], \\ \sigma^2_{i_1 i_2 i_3} &= \mathrm{Tr}\Big(\mathbb{E}\Big[\mathrm{vec}(\boldsymbol{\mathcal{G}})\mathrm{vec}(\boldsymbol{\mathcal{G}})^T\Big]\bigotimes_n \mathbb{E}\Big[\mathbf{u}_{i_n}^{(n)}\mathbf{u}_{i_n}^{(n)T}\Big]\Big) \\ &\quad - \mathbb{E}\big[\mathrm{vec}(\boldsymbol{\mathcal{G}})\big]^T\Big(\bigotimes_n \mathbb{E}\Big[\mathbf{u}_{i_n}^{(n)}\Big]\mathbb{E}\Big[\mathbf{u}_{i_n}^{(n)T}\Big]\Big)\mathbb{E}\big[\mathrm{vec}(\boldsymbol{\mathcal{G}})\big]. \end{aligned} \tag{21}$$

Therefore, our model can provide not only predictions over missing entries, but also the uncertainty of predictions, which is quite important for some specific applications.

4. Experiments Results

We verified the proposed method experimentally and compared it with related methods, i.e., high accuracy low rank tensor completion (HaLRTC) [8]. Alternating Direction Method of Multipliers (ADMM) [36] algorithm, developed in the 1970s, was employed by HaLRTC to solve the nuclear norm optimization problems with multiple non-smooth terms. HaLRTC algorithm using ADMM framework is based on simple low rank tensor completion (SiLRTC) algorithm [8]. By replacing the dummy matrices $M_i s$ by their tensor versions, the algorithm is shown in Algorithm (1).

Algorithm 1 HaLRTC Algorithm

Input: $\boldsymbol{\mathcal{X}}$ with $\boldsymbol{\mathcal{X}}_\Omega = \boldsymbol{\mathcal{T}}_{\Omega,p}$ and K
Output: $\boldsymbol{\mathcal{X}}$
Set $\boldsymbol{\mathcal{X}}_\Omega = \boldsymbol{\mathcal{T}}_\Omega$ and $\boldsymbol{\mathcal{X}}_{\overline{\Omega}} = 0$
for $k = 0$ to K **do**
  for $i = 1$ to n **do**
    $\boldsymbol{\mathcal{M}}_i = fold_i\left[D_{\frac{\alpha_i}{\rho}}\left(\boldsymbol{\mathcal{X}}_{(i)} + \frac{1}{\rho}\boldsymbol{\mathcal{Y}}_{i(i)}\right)\right]$
  end for
  $\boldsymbol{\mathcal{X}}_\Omega = \frac{1}{n}\left(\sum_{i=1}^n \boldsymbol{\mathcal{M}}_i - \frac{1}{\rho}\boldsymbol{\mathcal{Y}}_i\right)_{\overline{\Omega}}$
  $\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{Y}}_i - \rho(\boldsymbol{\mathcal{M}}_i - \boldsymbol{\mathcal{X}})$
end for

4.1 MRI Completion

Magnetic resonance imaging (MRI) is a medical imaging and widely used in the clinical diagnosis [37]. We evaluate our method by using MRI data (http://brainweb.bic.mni.mcgill.ca/brainweb), this dateset contains a set of realistic MRI data volumes produced by an MRI simulator. Because MRI data is high-dimensional, the completion from sparse observations becomes very challenging. So we separate the high-dimensional tensor data into low-dimensional small tensors. Hence, our method can be applied to small tensors completion. In experiment, we use the size of small tensors in $50 \times 50 \times 50$.

We use missing ratio (20% - 50%) and consider the noises in MRI data, and evaluation the algorithms using Peak Signal to Noise Ratio (PRSN) and RRSE. The result are shown in Table 1, and the visual quality is shown in Fig. 4. As we can see that the proposed method can effectively recover the missing values with high performance.

4.2 Video Completion

The video data is natural representation by a tensor as shown in Fig. 5. We evaluate the performance of the proposed method on a video sequence corrupted by additive Gaussian noise. The video sequence is downloaded from the benchmark data in [38]. We consider the noise standard deviation

Table 1: The Performance of MRI Completion Evaluated by PSNR and RRSE

Methods	Missing 50%				Missing 40%				Missing 30%				Missing 20%			
	Original		Noisy		Original		Noisy		Original		Noisy		Original		Noisy	
	PSNR	RRSE	PSNR	RRSE	PSNR	RRSE	PSNR	RRSE	PSNR	RRSE	PSNR	RRSE	PSNR	RRSE	PSNR	RRSE
BTC-T	27.27	0.11	26.42	0.12	27.84	0.10	27.12	0.11	28.12	0.10	27.55	0.11	28.38	0.10	27.83	0.10
HaLRTC	24.19	0.16	23.17	0.18	26.73	0.12	25.00	0.14	29.57	0.085	26.69	0.12	32.93	0.057	28.22	0.099

of 0.03, 0.15, 0.27 and missing ratio of 20% - 50%. The results are shown in Table 2.

Fig. 5: Tensor representation of a video sequence

Table 2: The Performance of Video Completion Evaluated by RRSE

	Missing				
	60%	50%	40%	30%	20%
Noise	RRSE	RRSE	RRSE	RRSE	RRSE
0.03	0.645	0.559	0.476	0.397	0.325
0.15	0.646	0.561	0.480	0.402	0.336
0.27	0.650	0.564	0.483	0.408	0.344

5. Conclusion

In this paper, we proposed a image completion method based on Bayesian Tucker decomposition. By using variational bayesian inference, we can avoids the computational demanding rank selection procedure. We apply the proposed method to image and video with 20-50 % missing voxels, the experimental results demonstrate that our method can effectively recover the whole data with a high predictive performance.

Acknowledgment

This work was supported by JSPS KAKENHI (Grant No. 17K00326 and 18K04178) and NSFC China (Grant No. 61773129).

References

[1] A. Shashua and T. Hazan, “Non-negative tensor factorization with applications to statistics and computer vision,” in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 792–799.

[2] L. R. Tucker, “Some mathematical notes on three-mode factor analysis,” *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[3] J. D. Carroll and 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.

[4] R. A. Harshman, “Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis,” *Foundations of the PARAFAC procedure*, 1970.

[5] H. A. Kiers, “Towards a standardized notation and terminology in multiway analysis,” *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 14, no. 3, pp. 105–122, 2000.

[6] L. Gui, Q. Zhao, and J. Cao, “Brain image completion by bayesian tensor decomposition,” in *Digital Signal Processing (DSP), 2017 22nd International Conference on*. IEEE, 2017, pp. 1–4.

[7] L. Yuan, Q. Zhao, and J. Cao, “High-order tensor completion for data recovery via sparse tensor-train optimization,” *arXiv preprint arXiv:1711.02271*, 2017.

[8] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor completion for estimating missing values in visual data,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

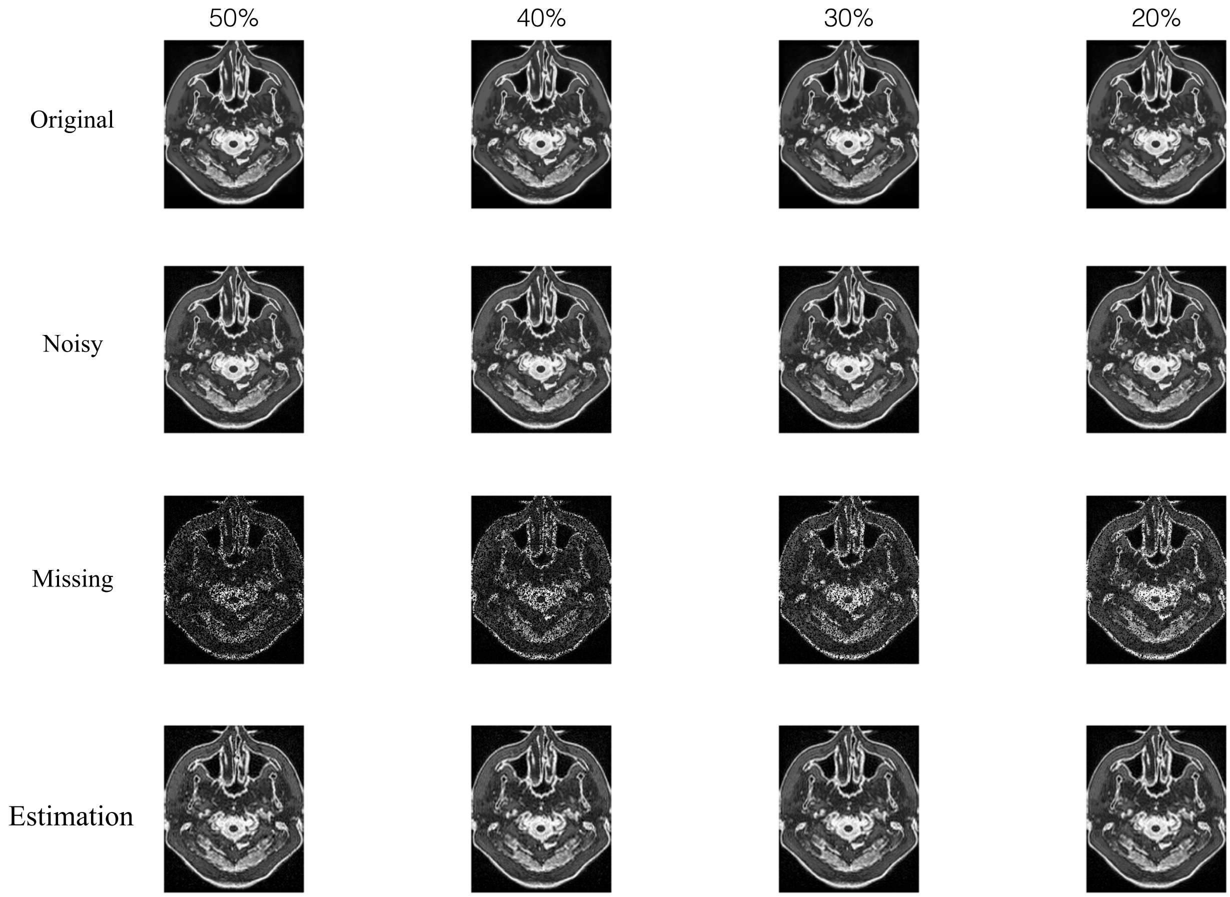

Fig. 4: Visualization of MRI data completion obtained by BTC

[9] L. Yuan, Q. Zhao, and J. Cao, "Completion of high order tensor data with missing entries via tensor-train decomposition," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 222–229.

[10] L. Yuan, J. Cao, Q. Wu, and Q. Zhao, "Higher-dimension tensor completion via low-rank tensor ring decomposition," *arXiv preprint arXiv:1807.01589*, 2018.

[11] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian cp factorization of incomplete tensors with automatic rank determination," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1751–1763, 2015.

[12] M. Filipović and A. Jukić, "Tucker factorization with missing data with application to low-n-rank tensor completion," *Multidimensional systems and signal processing*, vol. 26, no. 3, pp. 677–692, 2015.

[13] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, "Rank minimization on tensor ring: A new paradigm in scalable tensor decomposition and completion," *arXiv preprint arXiv:1805.08468*, 2018.

[14] L. Yuan, Q. Zhao, L. Gui, and J. Cao, "High-dimension tensor completion via gradient-based optimization under tensor-train format," *arXiv preprint arXiv:1804.01983*, 2018.

[15] L. De Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1096–1105, 2008.

[16] L. Gui, Q. Zhao, and J. Cao, "Tensor denoising using bayesian cp factorization," in *Information Science and Technology (ICIST), 2016 Sixth International Conference on*. IEEE, 2016, pp. 49–54.

[17] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "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.

[18] L. Gui, G. Cui, Q. Zhao, D. Wang, A. Cichocki, and J. Cao, "Video denoising using low rank tensor decomposition," in *Ninth International Conference on Machine Vision (ICMV 2016)*, vol. 10341. International Society for Optics and Photonics, 2017, p. 103410V.

[19] D. Muti and S. Bourennane, "Multidimensional filtering based on a tensor approach," *Signal Processing*, vol. 85, no. 12, pp. 2338–2353, 2005.

[20] L. De Lathauwer and B. De Moor, "From matrix to tensor: Multilinear algebra and signal processing," in *Institute of Mathematics and Its Applications Conference Series*, vol. 67. Citeseer, 1998, pp. 1–16.

[21] Y.-l. Zhang and G. Li, "An efficient algorithm to estimate mixture matrix in blind source separation using tensor decomposition," *International Journal of Computer Science Issues (IJCSI)*, vol. 11, no. 1, p. 19, 2014.

[22] Y. Liu, M. Li, H. Zhang, H. Wang, J. Li, J. Jia, Y. Wu, and L. Zhang, "A tensor-based scheme for stroke patients? motor imagery EEG analysis in BCI-FES rehabilitation training," *Journal of neuroscience methods*, vol. 222, pp. 238–249, 2014.

[23] J. Mocks, "Topographic components model for event-related potentials and some biophysical considerations," *IEEE transactions on biomedical engineering*, vol. 35, no. 6, pp. 482–484, 1988.

[24] Y. Zhang, Q. Zhao, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Removal of eeg artifacts for bci applications using fully bayesian tensor completion," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 819–823.

[25] A. Shashua and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–I.

[26] H. Hota, S. Shukla, and G. K. Kiran, "Review of intelligent techniques applied for classification and preprocessingof medical imagedata," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 1, p. 267, 2013.

[27] X. Geng, K. SmithMiles, Z. Zhou, and L. Wang, "Face image modeling by multilinear subspace analysis with missing values," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 3, pp. 881–892, 2011.

[28] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S. Amari, "Bayesian robust tensor factorization for incomplete multiway data," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 4, pp. 736–748, 2016.

[29] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematics and Physics*, vol. 6, no. 1-4, pp. 164–189, 1927.

[30] ——, "Multiple invariants and generalized rank of a p-way matrix or tensor," *Journal of Mathematics and Physics*, vol. 7, no. 1-4, pp. 39–79, 1928.

[31] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," *Problems in measuring change*, vol. 15, pp. 122–137, 1963.

[32] ——, "The extension of factor analysis to three-dimensional matrices," *Contributions to mathematical psychology*, vol. 110119, 1964.

[33] C. J. Appellof and E. R. Davidson, "Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents," *Analytical Chemistry*, vol. 53, no. 13, pp. 2053–2056, 1981.

[34] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[35] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," in *Problems in measuring change.*, C. W. Harris, Ed. Madison WI: University of Wisconsin Press, 1963, pp. 122–137.

[36] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[37] R. Srivaramangai, P. Hiremath, and A. S. Patil, "Preprocessing mri images of colorectal cancer," *International Journal of Computer Science Issues (IJCSI)*, vol. 14, no. 1, p. 48, 2017.

[38] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.