TENSOR COMPLETION THROUGH MULTIPLE KRONECKER PRODUCT DECOMPOSITION

Anh-Huy Phan[1], *Andrzej Cichocki*[1*], *Petr Tichavský*[2†], *Gheorghe Luta*[3], *Austin Brockmeier*[4]

[1]Brain Science Institute, RIKEN, Wakoshi, Japan
[2]Institute of Information Theory and Automation, Prague, Czech Republic
[3]Department of Biostatistics, Bioinformatics, and Biomathematics,
Georgetown University, Washington, D.C., USA
[4]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA.

ABSTRACT

We propose a novel decomposition approach to impute missing values in tensor data. The method uses smaller scale multiway patches to model the whole data or a small volume encompassing the observed missing entries. Simulations on color images show that our method can recover color images using only 5-10% of pixels, and outperforms other available tensor completion methods.

Index Terms— tensor decomposition, tensor completion, Kronecker tensor decomposition (KTD), color image

1. PROBLEM FORMULATION AND RELATED WORK

Tensor decompositions have become a valuable tool for analysis of modern data with multiple modes. The approach has found numerous areas of application such as in chemometrics, telecommunication, analysis of fMRI data, time-varying EEG spectrum, data mining, classification, clustering and compression [1–3]. For real-world and possibly large scale data, missing values may appear during data acquisition and transformation. In practice, a part of data, which does not conform to the generative model, can be treated as missing. For example, in decomposition of fluorescence data [1], data entries corresponding to Rayleigh scatter emission do not conform to the trilinear CANDECOMP/PARAFAC (CP) model. Hence, they are often treated as missing, so that the fitted model is not skewed by them [1]. Moreover, for relatively large data, treating entries as missing can be considered as data sampling, which is similar to using fiber sampling [4] to reduce the data size, but to avoid severely decreasing the quality of the decomposition. There are two major aims when analyzing such incomplete data. One is to extract latent variables through decomposition of the data tensor while discarding the missing values [1, 5]. The other is to predict missing entries based on complete data [6, 7]. Note that the latter problem can be done using the approximate tensor from the former problem. Conversely, the quality of estimated hidden components in the first problem can be improved with data completion.

Low-rank approximations are common approaches to impute missing values [8, 9]. The truncated singular value decomposition (SVD) of the matrix data is widely employed with a soft-threshold and/or additional constraints such as minimization of the nuclear norm [9, 10]. The problem is also related to image inpainting which is often solved using iterative shrinkage/thresholding (IST) algorithm [11, 12]. The considered problem also has connection to compressed sensing [13–15]. For multiway data, CANDECOMP/PARAFAC and Tucker decompositions [16–18] are the most common tensor decompositions. On the basis of estimated factor matrices, missing values can be heuristically imputed [5, 19, 20]. As the algorithms are not exclusively designed for data completion, such decompositions can result in overfitting.

In general, we can apply matrix completion techniques to the unfolded version of the tensor data. However, since this method only exploits spatial dependencies between data entries along two modes, it may produce artifacts. Proper methods should address the structures of tensor data in higher order such as the algorithms proposed in [6, 7, 21]. J. Liu *et al.* solved the problem of low rank tensor completion extended from that for matrix completion [9, 10], by employing block coordinate descent to minimize the tensor trace norm. In [6], S. Gandy *et. al.* considered the n-rank of a tensor as a sparsity measure for the low-rank tensor recovery problem, and minimized the nuclear norm optimization problem based on the Douglas-Rachford splitting technique [22]. In this direction, we propose a novel method to impute missing values using smaller scale multiway patches to model the whole data tensor or a small volume comprising the observed missing entries. Patches are used as dictionary for the construction of the data. Small patches explain details of the tensor, while large patterns capture the background. From the view point of estimation of basis patterns, the decomposition is similar to a dictionary learning problem in which the training data is taken from contiguous non-overlapping patches. The pro-

*Also affiliated with the E.E. Dept., Warsaw University of Technology and with Systems Research Institute, Polish Academy of Science, Poland.

†The work of P. Tichavsky was supported by Grant Agency of the Czech Republic through the project 102/09/1278.

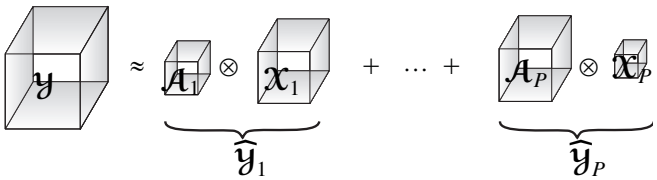


Fig. 1. Illustration of the tensor decomposition of an order-3 tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ into P terms of Kronecker tensor products of $\mathcal{A}_p, \mathcal{X}_p$.

posed method is related to that for matrix data without missing entries in [24] and that for image restoration in [25, 26].

The method is an extension of the tensor decomposition into multiple Kronecker products (KTD) [23] for incomplete tensor data and with additional constraints to control the intensity patterns. Moreover, a shifted decomposition is introduced to efficiently improve the approximation. Simulations on color images show that we can recover data while maintaining perceptually good picture quality, even when using only 5% of the data entries. The proposed algorithm was compared with the LRTC [7] and C-SALSA [12] algorithms.

2. LEARNING STRUCTURAL PATTERNS FOR INCOMPLETE DATA

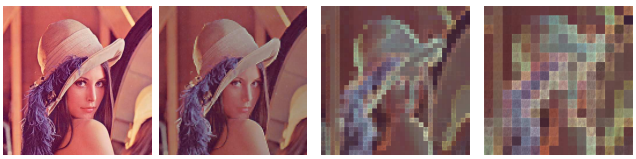
We consider a multiway data $\mathcal{Y}$ of size $I_1 \times I_2 \times \cdots \times I_N$ with missing entries indicated by a tensor $\mathcal{W}$ which is of the same dimension as $\mathcal{Y}$, and with binary values (0 if missing, and 1 if available). The task is to seek self-replicating structures expressed by multiway patches (patterns) tiling the whole or part of the data tensor $\mathcal{Y}$ (as illustrated in Fig. 1) so that the following cost is minimized [23]

$$\varphi(\{\mathcal{A}_p, \mathcal{X}_p\}) = \frac{1}{2} \left\| \mathcal{W} \circledast \left(\mathcal{Y} - \sum_{p=1}^{P} \mathcal{A}_p \otimes \mathcal{X}_p \right) \right\|_F^2, \qquad (1)$$

where $\circledast$ stands for the Hadamard product, $\mathcal{A}_p \otimes \mathcal{X}_p$ denotes the generalized Kronecker product between two tensors $\mathcal{A}_p$ and $\mathcal{X}_p$ (see Defintion 1 and [23]). We term sub-tensors $\mathcal{X}_p$ of size $K_{p1} \times K_{p2} \times \cdots \times K_{pN}$ with $\|\mathcal{X}_p\|_F = 1$ as patterns, while $\mathcal{A}_p$ of dimensions $J_{p1} \times J_{p2} \times \cdots \times J_{pN}$ such that $I_n = J_{pn} K_{pn}$, are called intensities. In Fig. 2, the KTD model is illustrated to approximate the color "Lenna" image of size $512 \times 512 \times 3$ by different pattern sizes 4×4, 16×16 and 32×32. Patterns $\mathcal{X}_p$ tile the entire data in non-overlapping positions. Large patterns capture the background of the image, while small patterns represent the details. In a region of low density of entries, i.e., high density of incomplete values, a large pattern is suggested. The missing values are then imputed on basis of $\mathcal{A}_p$ and $\mathcal{X}_p$.

2.1. Algorithms

In order to minimize (1), the Kronecker unfolding and one major result in [23] are recalled in Appendix. More properties of the Kronecker tensor product are discussed in [23, 27]. We classify patterns $\mathcal{X}_p$ into the same group (say group g) if they have the same dimension, and assume there are G ($G \leq P$)



(a) "Lenna" (b) 4×4 (c) 16×16 (d) 32×32

Fig. 2. "Lenna" image and its approximations by KTD with different pattern sizes.

different pattern sizes $K_{g1} \times K_{g2} \times \cdots \times K_{gN}$ ($g = 1, 2, \ldots, G$). We denote by $\mathcal{I}_g$ the set of pattern indices in group g, that is, $\mathcal{I}_g = \{p : \mathcal{X}_p \in \mathbb{R}^{K_{g1} \times K_{g2} \times \cdots \times K_{gN}}\} = \{p_1^{(g)}, p_2^{(g)}, \ldots, p_{P_g}^{(g)}\}$, where $\mathtt{card}\{\mathcal{I}_g\} = P_g$, $\sum_{g=1}^{G} P_g = P$, and denote by

$$\mathcal{Y}^{(g)} = \sum_{p_g \in \mathcal{I}_g} \mathcal{A}_{p_g} \otimes \mathcal{X}_{p_g}, \quad g = 1, 2, \ldots, G, \qquad (2)$$

the approximate tensor to $\mathcal{Y}$ by the group g. Hence, according to Lemma 1, Kronecker unfoldings $\mathbf{Y}^{(g)}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)}$ with $\boldsymbol{K}_g = [K_{g1}, K_{g2}, \ldots, K_{gN}]$, $\boldsymbol{J}_g = [J_{g1}, J_{g2}, \ldots, J_{gN}]$ are rank-P_g matrices, that is, $\mathbf{Y}^{(g)}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} = \mathbf{A}_g \mathbf{X}_g^T$, where $\mathbf{A}_g = \left[\mathrm{vec}\left(\mathcal{A}_{p_g}\right)\right]_{p_g \in \mathcal{I}_g}$, $\mathbf{X}_g = \left[\mathrm{vec}\left(\mathcal{X}_{p_g}\right)\right]_{p_g \in \mathcal{I}_g}$. We apply the approach successively to estimate patterns and intensity tensors in the group g while fixing all other groups. The cost function is rewritten as

$$\varphi(\{\mathcal{A}_p, \mathcal{X}_p\}) = \frac{1}{2} \left\| \mathbf{W}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} \circledast \left(\mathbf{Y}^{(-g)}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} - \mathbf{A}_g \mathbf{X}_g^T \right) \right\|_F^2, \qquad (3)$$

where $\mathcal{Y}^{(-g)} = \mathcal{Y} - \sum_{h \neq g} \mathcal{Y}^{(h)}$, and $\mathbf{W}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)}$ is a Kronecker unfolding of $\mathcal{W}$ (see Definition 2). The new optimization problem indicates that seeking patterns in the group g is equivalent to low-rank approximation to the Kronecker unfolding of the tensor $\mathcal{Y}^{(-g)}$ in which truncated SVD with a threshold is suggested [9, 10]. Note that the approach is not simply a low-rank matrix approximation to the unfolded tensor, but to multiple Kronecker unfoldings with different pattern dimensions.

In this paper, we restrict the problem to nonnegative data (such as color images), and nonnegative basis patterns. With additional constraints to control intensity tensors $\mathcal{A}_p$, the optimization problem is rewritten as

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \| \mathbf{W}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} \circledast \left(\mathbf{Y}^{(-g)}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} - \mathbf{A}_g \mathbf{X}_g^T \right) \|_F^2 + \\ & \frac{1}{2} \lambda \sum_{p}^{P} \|\mathcal{A}_p\|_F^2 + \frac{1}{2} \gamma \sum_{p}^{P} \|\mathcal{A}_p \times_1 \mathbf{L}_1 \cdots \times_N \mathbf{L}_N\|_F^2, \end{aligned}$$

where $\lambda \geq 0$ and $\gamma \geq 0$ are regularization parameters, "$\times_n$" denotes mode-n tensor-matrix product [2], $\mathbf{L}_n$ of size $J_{p_n} \times J_{p_n}$ is the second-order Laplace operator which measures the smoothness by estimating the differences between neighboring samples along mode-n [1, 28, 29]

$$\mathbf{L}_n = \begin{pmatrix} -2 & 2 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 2 & -2 \end{pmatrix}.$$

Fig. 3. Sample color images were used in the simulations.

The multiplicative update rules derived from the above cost function have the form [30–33]

$$\mathbf{A}_g \leftarrow \mathbf{A}_g \circledast \Big[\bar{\mathbf{Y}}^{(-g)}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} \mathbf{X}_g\Big]_+ \oslash \Big(\bar{\mathbf{Y}}^{(g)}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} \mathbf{X}_g + \lambda \mathbf{A}_g + \gamma \mathbf{S}_g\Big),$$

$$\mathbf{X}_g \leftarrow \mathbf{X}_g \circledast \Big[\bar{\mathbf{Y}}^{(-g)T}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} \mathbf{A}_g\Big]_+ \oslash \Big(\bar{\mathbf{Y}}^{(g)T}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)} \mathbf{A}_g\Big),$$

where $\oslash$ denotes the Hadamard division, $\bar{\mathcal{Y}}^{(-g)} = \mathcal{W} \circledast \mathcal{Y}^{(-g)}$, $\bar{\mathcal{Y}}^{(g)} = \mathcal{W} \circledast \mathcal{Y}^{(g)}$, and $\mathbf{S}_g = \Big[\text{vec}\Big(\mathcal{A}_p \times_1 \mathbf{L}_1^T \mathbf{L}_1 \times_2 \mathbf{L}_2^T \mathbf{L}_2 \cdots \times_N \mathbf{L}_N^T \mathbf{L}_N\Big)\Big]_{p \in \mathcal{I}_g}$ of size $\Big(\prod_n J_{gn}\Big) \times P_g$. The rectifier $[x]_+ = \max\{x, \varepsilon\}$ forces a negative value to zero or a small enough value (e.g., 10^{-8}). It prevents $\mathcal{A}_p$ and $\mathcal{X}_p$ from having negative values which could appear when $\mathcal{Y}^{(-g)} = \mathcal{Y} - \mathcal{Y}^{(g)}$ includes negative values.

Patterns $\mathcal{X}_p$, i.e., columns of $\mathbf{X}_g$, are normalized to unit norm after updating, that is, $\mathcal{X}_p \leftarrow \frac{\mathcal{X}_p}{\|\mathcal{X}_p\|_F^2}$, and the intensity tensors $\mathcal{A}_p$ are scaled accordingly. The tensors $\mathcal{A}_p$ and $\mathcal{X}_p$ in group g are initialized using leading singular components of Kronecker unfoldings $\mathbf{Y}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)}$. The regularization parameters λ, γ can be chosen from within the range $[10^{-5}, 10^{-1}]$. Note that the optimization problems with nonnegative constraints can be solved by other robust algorithms for nonnegative matrix factorizations which are discussed in [2].

2.2. Shifted KTD

Pattern tensors $\mathcal{X}_p$ tile the entire data in non-overlapping positions. Hence, large blocks may cause blocking artifact. To this end, the objects can overlap the tile boundaries. Unfortunately, such convolutive decomposition cannot be imposed on the Kronecker tensor product. A simple method is to decompose the shifted tensors $\overset{\boldsymbol{s}\rightarrow}{\mathcal{Y}}$

$$\overset{\boldsymbol{s}\rightarrow}{\mathcal{Y}}(i_1, i_2, \ldots, i_N) = \mathcal{Y}(i_1 - s_1, i_2 - s_2, \ldots, i_N - s_N),$$

where $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]$ denotes the shift indices along all modes. Fibers shifted in from outside the tensor are set to zero or replicated from the data. The final imputed values are averaged over the estimated ones of all shifted tensors.

3. SIMULATIONS

Although the proposed model and algorithms can work on general high order tensor data, in this section, we evaluate the proposed approach on the standard test images of size $512 \times 512 \times 3$ which are: "Lenna", "Tiffany", "Pepper", "Barbara" and "House" images as illustrated respectively in Fig. 3. 'Dead' pixels were randomly generated with missing ratios from 40-95%. For the KTD model, pattern sizes were chosen according to the density of incomplete values. Basically, the decomposition should have patterns with small size to represent details, and larger patterns for flat regions. When the missing ratio is lower than 80%, we extracted patterns with sizes of 2×1 , 4×2, 2×4, 4×8, 8×4, 8×8. For higher ratios ($\geq 90\%$), we added an additional pattern group of size 16×16. The number of patterns P_g was set to the maximum possible rank of Kronecker unfoldings $\mathbf{Y}^{(g)}_{(\boldsymbol{J}_g \times \boldsymbol{K}_g)}$. For all the experiments, we set $\lambda = 0.01$. Moreover, patterns were also estimated from the shifted tensors with $s_n \in [-4{:}4]$. Note that the standard median filter which is commonly used for impulse noise does not work for these simulations due to high density of missing entries.

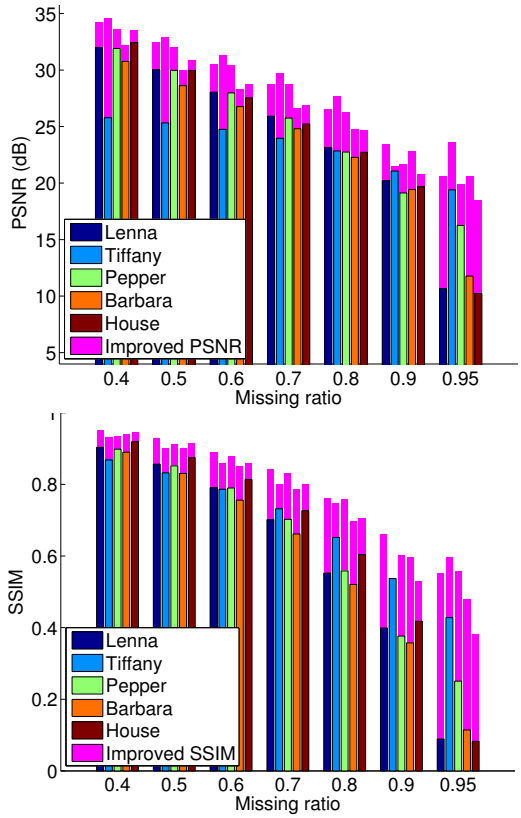
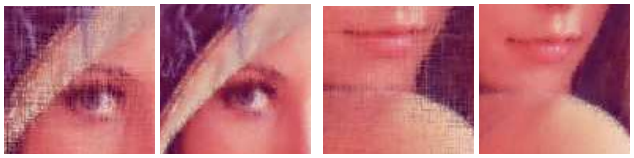


Fig. 4. Performance comparison of image quality.

The low-rank tensor completion algorithm in [7] using the default parameters was compared to our algorithm. In some simulations, the C-SALSA algorithm [12] for image inpainting was applied with its default parameters to separately recover each R,G,B layer of the color images. The objective performance of the two algorithms was compared in term of the structural similarity index (SSMI) [34] and PSNR as illustrated in Fig. 4. Note that although the quantity of peak signal-to-noise ratio (PSNR) is commonly used in practice, this index is inconsistent with human eye perception [34].

As seen in Fig. 4, through all the simulations, KTD achieved higher PSNR and SSMI than LRTC. When the missing ratio was lower than 50%, both methods reconstructed well the images with PSNR $\geq$ 30 dB. For higher missing ratios ($>$ 60%), KTD significantly outperformed LRTC, especially, when there were 90-95% incomplete pixels. Moreover, LRTC produced blocking artifacts on both

(a) Lenna. (b) LRTC, 25.91 dB. (c) KTD, 28.74 dB.



(d) (left)-LRTC, (right)-KTD. (e) (left)-LRTC, (right)-KTD.

Fig. 5. Reconstruction of "Lenna" image with 70% dead pixels.

(a) Tiffany. (b) LRTC, 22.84 dB. (c) KTD, 27.69 dB.

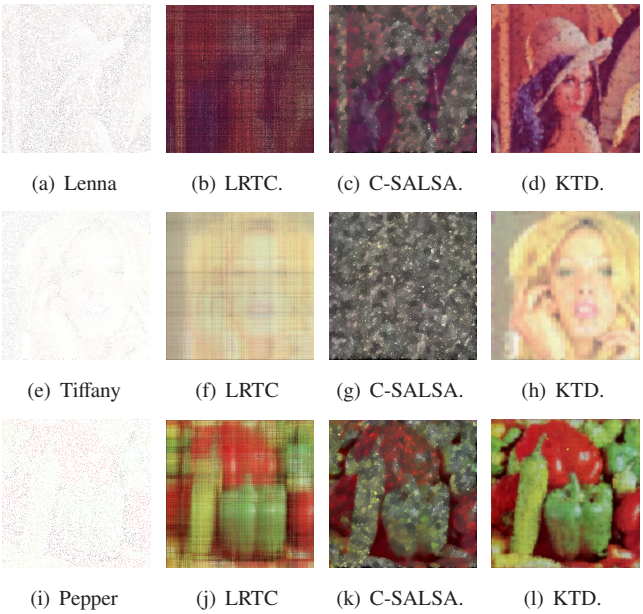
Fig. 6. Reconstruction of "Tiffany" image with 80% dead pixels.

detail and flat regions, for example, the eye and shoulder regions of the "Lenna" image illustrated in Fig. 5. Such defects lowered the quality of the reconstructed image in display although its PSNR = 25.91 dB. Another comparison is for the "Tiffany" image with 80% 'dead' pixels where Fig. 6 confirms the high quality of the proposed approach.

When images had only 5% of pixels as seen in Fig. 7, KTD was still able to reconstruct major structures of the images, while LRTC did not succeed in this case. Its reconstructed images contained many rows and columns that had not been filled out. Although PSNR of the reconstructed "Tiffany" image by LRTC was relatively high and comparable to PSNR values in the case of a missing ratio of 90%, this subjective assessment did not accurately reflect the quality of this image which is illustrated in Fig. 7(f). For KTD, smoothness constraints were considered for this case with $\gamma = 0.001$. Fig. 7 also shows that the C-SALSA algorithm [12] cannot reveal structures of the original images. The C-SALSA algorithm worked well for lower missing ratios ($\leq$ 80%), but this algorithm failed for this test case. One possible explanation is that the default parameters of C-SALSA were not the best choice for our data. Closer inspection of the images by zooming in reveals the differences in the details.

4. CONCLUSION

We have proposed a novel approach to impute incomplete values in high order data tensors. While other existing tensor decompositions seek common component vectors along modes



(a) Lenna (b) LRTC. (c) C-SALSA. (d) KTD.

(e) Tiffany (f) LRTC (g) C-SALSA. (h) KTD.

(i) Pepper (j) LRTC (k) C-SALSA. (l) KTD.

Fig. 7. Reconstructed images when there were 95% dead pixels.

of the data tensor, the proposed method retrieves a set of multiway structured patterns as a dictionary for data reconstruction. This explains why our approach can work even when there are only 5% of entries. The performance of the proposed approach can be further improved through multi-level and multiscale processing with additional procedures to correct deficiencies due to improper imputation. Matlab code that implements the new method is available upon request.

Appendix. Kronecker product and unfolding

Definition 1 (Kronecker tensor product [23]). *Let $\mathcal{A} = [a_j]$ and $\mathcal{B} = [b_k]$ be two order-N tensors of size $J_1 \times J_2 \times \cdots \times J_N$ and $K_1 \times K_2 \times \cdots \times K_N$, respectively, $\boldsymbol{j} = [j_1, j_2, \ldots, j_N]$, $1 \leq j_n \leq J_n$ and $\boldsymbol{k} = [k_1, k_2, \ldots, k_N]$, $1 \leq k_n \leq K_n$. A Kronecker tensor product of $\mathcal{A}$ and $\mathcal{B}$ is defined as an order-N tensor $\mathcal{C} = [c_i] \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $\boldsymbol{i} = [i_1, i_2, \ldots, i_N]$, $I_n = J_n K_n$ such that $c_i = a_j\, b_k$, $i_n = k_n + (j_n - 1)K_n$, and is expressed as $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$.*

Definition 2 (Kronecker unfolding [23]). *An $(\boldsymbol{J} \times \boldsymbol{K})$ Kronecker unfolding of $\mathcal{C} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with $I_n = J_n K_n$, for all n, is a matrix $\mathbf{C}_{(\boldsymbol{J} \times \boldsymbol{K})}$ of the size $\prod_{n=1}^{N} J_n \times \prod_{n=1}^{N} K_n$ whose entries (j, k) are given by*

$$\mathbf{C}_{(\boldsymbol{J} \times \boldsymbol{K})}(j, k) = \mathcal{C}(\boldsymbol{i}),$$

for all $\boldsymbol{j} = [j_1, \ldots, j_N]$, $j_n = 1, \ldots, J_n$, $\boldsymbol{k} = [k_1, \ldots, k_N]$, $k_n = 1, \ldots, K_n$, $n = 1, \ldots, N$ and $j = \text{ivec}(\boldsymbol{j}, \boldsymbol{J})$, and $k = \text{ivec}(\boldsymbol{k}, \boldsymbol{K})$, $\boldsymbol{i} = [i_1, \ldots, i_N]$, $i_n = k_n + (j_n - 1)K_n$.

Lemma 1 (Rank-P Factorization [23]). *Let a tensor $\mathcal{C}$ be expressed as a sum of P Kronecker products $\mathcal{C} = \sum_{p=1}^{P} \mathcal{A}_p \otimes \mathcal{B}_p$, where $\mathcal{A}_p \in \mathbb{R}^{J_1 \times \cdots \times J_N}$ and $\mathcal{B}_p \in \mathbb{R}^{K_1 \times \cdots \times K_N}$, $p = 1, 2, \ldots, P$. Then the Kronecker unfolding of $\mathcal{C}$ is a matrix of rank-P, such that*

$$\mathbf{C}_{(\boldsymbol{J} \times \boldsymbol{K})} = \sum_{p=1}^{P} \text{vec}\left(\mathcal{A}_p\right) \text{vec}\left(\mathcal{B}_p\right)^T. \tag{4}$$

5. REFERENCES

[1] R. Bro, *Multi-way Analysis in the Food Industry - Models, Algorithms, and Applications*, Ph.D. thesis, University of Amsterdam, Holland, 1998.

[2] 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.

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

[4] M.W. Mahoney, M. Maggioni, and P. Drineas, "Tensor-CUR decompositions and data applications," *SIAM J. Matrix Anal. Appl.*, vol. 30, pp. 957–987, 2008.

[5] G. Tomasi and R. Bro, "PARAFAC and missing values," *Chemometr. Intell. Lab.*, vol. 75, no. 2, pp. 163–180, 2005.

[6] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-*n*-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, pp. 025010, 2011.

[7] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," in *ICCV*, 2009, pp. 2114–2121.

[8] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, 2012.

[9] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *J. Mach. Learn. Res.*, vol. 11, pp. 2287–2322, Aug. 2010.

[10] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.

[11] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[12] M.V. Alfonso, J.M. Bioucas-Dias, and M.A.T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345 –2356, sept. 2010.

[13] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2005.

[14] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[15] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[16] R.A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[17] 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.

[18] L.R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.

[19] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometr. Intell. Lab.*, vol. 106, no. 1, pp. 41–56, 2011.

[20] P. Tichavský, A.-H. Phan, and Z. Koldovský, "Cramér-Rao-Induced Bounds for CANDECOMP/PARAFAC tensor decomposition," *IEEE Trans. Signal Process.*, accepted for publication, 2013.

[21] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 99, no. preprints, Jan. 2012.

[22] P.L. Combettes and J.-C. Pesquet, "A Douglas Rachford splitting approach to nonsmooth convex variational signal recovery," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 564 –574, dec. 2007.

[23] A.-H. Phan, A. Cichocki, P. Tichavský, D. P. Mandic, and K. Matsuoka, "On revealing replicating structures in multiway data: A novel tensor decomposition approach," in *Latent Variable Analysis and Signal Separation*, vol. 7191 of *LNCS*, pp. 297–305. Springer Berlin Heidelberg, 2012.

[24] C. Van Loan and N. Pitsianis, "Approximation with kronecker products," in *Linear Algebra for Large Scale and Real Time Applications*. 1993, pp. 293–314, Kluwer Publications.

[25] J. G. Nagy and M. E. Kilmer, "Kronecker product approximation for preconditioning in three-dimensional imaging applications," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 604–613, 2006.

[26] A. Bouhamidi and K. Jbilou, "A kronecker approximation with a convex constrained optimization method for blind image restoration," *Optimization Letters*, pp. 1–14, 10.1007/s11590-011-0370-7.

[27] S. Ragnarsson, *Structured Tensor Computations: Blocking, Symmetries And Kronecker Factorizations*, Ph.D. thesis, Cornell University, 2012.

[28] M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers," *SIAM Journal on Numerical Analysis*, vol. 40, no. 3, pp. 965–994, 2002.

[29] 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.

[30] M. E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorthm suitable for volume ECT," *IEEE Trans. Med. Imag.*, vol. 5, pp. 61–66, 1986.

[31] H. Lantéri, R. Soummmer, and C. Aime, "Comparison between ISRA and RLA algorithms: Use of a Wiener filter based stopping criterion," *Astronomy and Astrophysics Supplemantary Series*, vol. 140, pp. 235–246., 1999.

[32] D. D. Lee and H.S. Seung, "Learning of the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[33] C. J. Lin, "Projected gradient methods for non-negative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, October 2007.

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.