

Robust Barron-Loss Tucker Tensor Decomposition

Mahsa Mozaffari and Panos P. Markopoulos

Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology

Email: mmozaffari@mail.rit.edu, panos@rit.edu

Abstract—Tucker decomposition is a standard method for the analysis of high-order tensor data. Standard Tucker decomposition generalizes singular-value decomposition and is formulated as minimization of the L2-norm of the low-rank approximation error. Due to the quadratic emphasis on peripheral data, the L2-norm based formulation of Tucker has been shown to be sensitive against corruptions. L1-norm-based variants (L1-Tucker) are proposed as robust Tucker decomposition methods, and are formulated as subspace estimators by absolute-projection maximization, and have proven effective in outlier resistant subspace computation. However, Tucker decomposition for robust low-rank tensor approximation is not defined/optimized in the literature. In this work, we propose a new formulation for low-rank tensor approximation, with tunable outlier-robustness, and present a unified algorithmic solution framework. This formulation relies on a new generalized robust loss function (Barron loss), which encompasses several well-known loss-functions with variable outlier-resistance. The robustness of the proposed framework is corroborated by the presented numerical studies on synthetic and real data.

I. INTRODUCTION

Across a variety of applications, datasets are multimodal or multi-way and naturally modeled as high-order tensors (e.g., hyper-spectral imagery, video, network/graph relation arrays) [1]–[8]. At the same time, real-world data often contain sporadic highly deviating points (faulty entries) due to errors in data collection/storage or even adversarial data contamination. At the same time, “good data” are of paramount importance in machine learning, especially in applications that cannot afford “big data” (where faulty entries can average out) and performance reliability is most critical, such as medicine and defense [9]. Based on the above, there is a need for robust analysis and refinement of tensor data before providing them as input to machine learning systems.

Tucker decomposition (TD) is a standard method for analysis, compression, and low-rank restoration of tensor data [10], [11]. Standard TD is the high-order extension of singular-value decomposition (SVD) and strives to minimize the L2-norm of the low-rank approximation error. Due to its reliance on the L2-norm (sum of squared entries), which benefits peripheral entries, TD is known to be sensitive against outlier contamination of the processed tensor [12], [13]. Higher Order Singular Value Decomposition (HOSVD), and Higher Order Orthogonal Iterations (HOOI), are two common solvers for standard TD. L1-norm-based variants (such as L1-Tucker [13])

have been proposed in the literature, formulated as subspace estimators based on absolute-projection maximization. Indeed, these methods exhibit strong outlier resistance in the computation of the multilinear data subspaces [13]–[15], but have not been defined/optimized for low-rank tensor approximation.

In this work we propose a new TD formulation, particularly designed for outlier-resistant low-rank tensor approximation. Our formulation relies on a new generalized robust loss function, recently presented in [16]. To solve the proposed TD problem, for any configuration of loss parameters, we develop a novel algorithmic framework, with a particular focus on computational efficiency and ability to handle larger data. Our numerical studies corroborate the outstanding robustness of the proposed framework.

II. BACKGROUND

A. Tensors and Tucker Decomposition

Tensors are a high-order generalization of vectors and matrices. Each entry of an N -way tensor $\mathcal{X} \in \mathbb{R}^{D_1 \times D_2 \times \dots \times D_N}$ is identified by N indices. Groups of entries of a tensor \mathcal{X} that can be addressed by the same $N - 1$ fixed indices are referred to as “fibers” –e.g., vector $[\mathcal{X}]_{:,i_2,\dots,i_N}$ is one of the $D_2 \dots D_N$ mode-1 fibers of \mathcal{X} . \mathcal{X} can be “flattened” into a matrix $\mathcal{X}_{(n)}$ the columns of which are the mode- n fibers of \mathcal{X} . Operator \times_n denotes a mode- n tensor-matrix product so that $[\mathcal{X} \times_n \mathbf{A}]_{(n)} = \mathbf{A} \mathcal{X}_{(n)}$. $[N] = \{1, 2, \dots, N\}$ and $\mathcal{X} \times_{n \in [N]} \mathbf{A}_n = \mathcal{X} \times_1 \mathbf{A}_1 \times_2 \dots \times_N \mathbf{A}_N$ is a multi-way tensor-matrix product (invariant to the order of mode products). $\mathbf{A} \otimes \mathbf{B} \in \mathbb{R}^{p^k \times q^r}$ denotes the Kronecker product of matrices $\mathbf{A} \in \mathbb{R}^{p \times q}$ and $\mathbf{B} \in \mathbb{R}^{k \times r}$.

Given low ranks d_1, d_2, \dots, d_N , TD approximates a tensor $\mathcal{X} \in \mathbb{R}^{D_1 \times D_2 \times \dots \times D_N}$ by $\mathcal{X} \approx \mathcal{G} \times_{n \in [N]} \mathbf{U}_n$; $\mathcal{G} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_N}$ is the “core” of the decomposition and $\{\mathbf{U}_i\}_{i \in [N]} \in \mathbb{S}(D_i, d_i)$ are the “factor matrices” (or bases). In standard TD, $(\mathcal{G}, \{\mathbf{U}_n\}_{n \in [N]})$ are jointly chosen to minimize the L2-norm approximation error $\|\mathcal{X} - \mathcal{G} \times_{n \in [N]} \mathbf{U}_n\|_F^2$. In some applications, core and factor matrices are rotated (with invariant approximation metric) so that $\{\mathbf{U}_n\}_{n \in [N]}$ constitute orthonormal bases of the multi-linear subspaces captured by \mathcal{X} . Hence, L2-norm error-minimization formulation of standard TD is equivalent to maximization of the L2-norm tensor projection, as

$$\max_{\{\mathbf{U}_i \in \mathbb{S}(D_i, d_i)\}_{i \in [N]}} \left\| \mathcal{X} \times_{i \in [N]} \mathbf{U}_i^\top \right\|_F^2. \quad (1)$$

This research was supported in part by the National Science Foundation under Award OAC-1808582 and in part by the Air Force Office of Scientific Research under Award FA9550-20-1-0039.

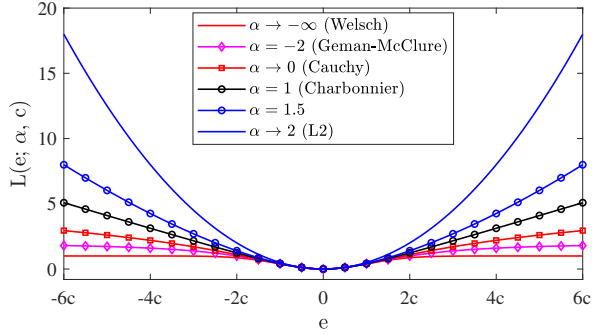


Fig. 1: Generalized loss function, for different values of robustness parameter α [16].

HOOI, and HOSVD are two approximate solvers of (1). HOSVD, disjointly computes each \mathbf{U}_i as the leading left singular vectors of $\mathbf{X}_{(n)}$. HOOI, is an iterative method, that jointly updates each \mathbf{U}_n by the leading left singular vectors of $[\mathbf{X} \times_{i \neq n} \mathbf{U}_i^T]_{(n)}$, until convergence. Both algorithms have exhibited success in low-rank approximation in absence of sparse corruptions, with HOOI being superior to HOSVD, due to joint computation of factor matrices. However, due to the squared emphasis on the peripheral entries of the data, in the presence of outlying tensor entries/fibers/slabs, L2-norm-based solvers favor the outlying entries over the nominal ones. L1-norm-based variants are proposed for subspace estimation, formulated as seeking orthonormal subspaces identified by $\{\mathbf{U}_n\}_{n \in [N]}$ on which the L1-norm of the projection of \mathcal{X} is maximized. L1-HOSVD, and L1-HOOI are two robust solvers of this formulation, and have shown documented success in robust subspace estimation. Nevertheless, the problem of robust low-rank Tucker tensor approximation needs further exploration.

B. Generalized Robust Loss

For given error e between a desired and an approximated quantity, Barron in [16] proposed the loss function

$$L(e; \alpha, c) = \frac{\alpha - 2}{\alpha} \left(\left(\frac{(e/c)^2}{|\alpha - 2|} + 1 \right)^{\alpha/2} \right). \quad (2)$$

In (2), α is a shape parameter that controls the robustness and $c > 0$ is a scale parameter that tunes the size of the quadratic bowl of $L(e, \alpha, c)$ near $e = 0$. This two-parameter loss function generalizes various standard loss functions such as the Cauchy ($\alpha = 0$), Welsch ($\alpha \rightarrow -\infty$), Geman-McClure ($\alpha = -2$), Charbonnier/pseudo-Huber/L1-L2 ($\alpha = 1$), and L2 ($\alpha \rightarrow 2$), as shown in Fig. 1. Clearly, tuning α to values lower than 2, the loss introduces increasing robustness against high error values. Based on its smoothness (appropriate for gradient-descent-type optimization) and its flexibility, Barron loss function in (2) has already found use in many machine learning applications [17] since first introduced in 2019 [16]. In this work, we employ this loss function for the first time for the formulation of a robust TD, as shown below.

Algorithm 1. Proposed algorithm.

Input: Data $\mathcal{X} \in \mathbb{R}^{D_1 \times D_2 \times \dots \times D_N}$ and target low-ranks $\{d_n\}_{n \in [N]}$.
1: Initialize $\mathbf{U}_n, \forall n$.
2: for $n = 1, 2, \dots, N$:
3: $\mathbf{U}_n \leftarrow$ solve convex problem in (5).
4: $\mathcal{G} \leftarrow$ solve equation (6).
Return: $\hat{\mathcal{X}}, \{\hat{\mathbf{U}}_n\}_{n \in [N]}$, and $\hat{\mathcal{G}}$.

Fig. 2: Pseudocode of proposed algorithm.

III. PROPOSED FORMULATION AND ALGORITHM

Consider an N -way data tensor $\mathcal{X} \in \mathbb{R}^{D_1 \times D_2 \times \dots \times D_N}$. Using the Barron loss, we formulate the proposed TD

$$\min_{\mathcal{G}, \{\mathbf{U}_n\}_{n \in [N]}} F(\mathcal{G}, \{\mathbf{U}_n\}_{n \in [N]}; \mathcal{X}, a, c), \quad (3)$$

$$F(\mathcal{G}, \{\mathbf{U}_n\}_{n \in [N]}; \mathcal{X}, a, c) = \sum_{i_1=1}^{D_1} \dots \sum_{i_N=1}^{D_N} L([\mathcal{X} - \mathcal{G} \times_{i \in [N]} \mathbf{U}_i]_{i_1, \dots, i_N}; a, c). \quad (4)$$

In this work we are mostly interested in robust versions of (3) for $\alpha < 2$. Nonetheless, we present a complete algorithmic framework for solving (3), for any (α, c) configuration. These parameters can be tuned by the user, per application, considering the robustness/accuracy ratio of tuning α and c .

To solve (3), we develop a Block Coordinate Descent (BCD) approach, partitioning the problem into smaller convex sub-problems which we solve iteratively. Each sub-problem is defined by fixing all the block-variables in $(\mathcal{G}, \{\mathbf{U}_i\}_{i \in [N]})$ except for the one that we want to optimize. For instance, to optimize \mathbf{U}_n , we define $\mathbf{A}_n = \mathbf{U}_N \otimes \dots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \dots \otimes \mathbf{U}_1 \in \mathbb{R}^{J_n \times j_n}$, where $J_n = \prod_{m \neq n} D_m$ and $j_n = \prod_{m \neq n} d_m$, and solve the convex problem

$$\mathbf{U}_n \leftarrow \min_{\mathbf{U}} \sum_{j=1}^{J_n} \sum_{d=1}^{D_n} L([\mathcal{X}_{(n)} - \mathbf{U} \mathcal{G}_{(n)} \mathbf{A}_n^T]_{d,j}; a, c). \quad (5)$$

Then, after each basis update, we define the fixed $\tilde{\mathbf{U}} = \mathbf{U}_N \otimes \dots \otimes \mathbf{U}_1$ and optimize the core as

$$\mathcal{G} \leftarrow \min_{\mathcal{G}} \sum_{d=1}^{D_1 \dots D_N} L([\mathcal{X}_{(N+1)} - \mathcal{G}_{(N+1)} \tilde{\mathbf{U}}^T]_d; a, c). \quad (6)$$

Upon termination of the BCD iterations, the algorithm returns estimates $(\hat{\mathcal{G}}, \{\hat{\mathbf{U}}_i\}_{i \in [N]})$ and $\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_{n \in [N]} \hat{\mathbf{U}}_n$. The outline of the BCD algorithm is shown in Fig. 2.

The problems in (5) and (6) are regression-type problems in terms of the Barron loss and can be solved by an array of standard solvers, including gradient descent. In the experiments of this paper, gradient descent optimization is adopted for optimizing (5) and (6). Gradient of the loss in (5) is computed by

$$\frac{\partial F}{\partial \mathbf{U}_n} = \mathcal{Y}_{(n)} \mathbf{A}_n \mathcal{G}_{(n)}^T = [\mathcal{Y} \times_{i \neq n} \mathbf{U}_i^T]_{(n)} \mathcal{G}_{(n)}^T, \quad (7)$$

where \mathcal{Y} is the derivative of the entrywise loss function with respect to the residual error of the low-rank approximation.

In other words, $[\mathcal{Y}_{(n)}]_{d,j} = \frac{\partial L}{\partial \mathbf{e}}([\mathcal{X}_{(n)} - \mathbf{U}\mathcal{G}_{(n)}\mathbf{A}_n^T]_{d,j}; a, c)$. Similarly, the gradient of the loss in (6) is

$$\frac{\partial F}{\partial \mathcal{G}} = \mathcal{Y} \times_{i \in [N]} \mathbf{U}_i^\top. \quad (8)$$

Handling missing data: With the proposed formulation/framework we can easily handle low-rank approximation in the presence of missing data. To that end, we rewrite the loss function at (4), as

$$F(\mathcal{G}, \{\mathbf{U}_n\}_{n \in [N]}; \mathcal{X}, a, c) = \sum_{i_1=1}^{D_1} \cdots \sum_{i_N=1}^{D_N} L([\mathcal{W} \circ (\mathcal{X} - \mathcal{G} \times_{i \in [N]} \mathbf{U}_i)]_{i_1, \dots, i_N}; a, c), \quad (9)$$

where \mathcal{W} is an indicator tensor, such that $[\mathcal{W}]_{i_1, \dots, i_N} = 1$ if $[\mathcal{X}]_{i_1, \dots, i_N}$ is not missing, and 0 otherwise. The gradient calculations, must account for the \mathcal{W} , accordingly.

IV. NUMERICAL STUDIES

A. Synthetic data

To demonstrate the effectiveness of the proposed TD formulation and the proposed algorithm against high levels of outlier corruption, we conducted a numerical experiment on a synthetic data. We set $N = 3$, $D_1 = D_2 = D_3 = 10$, $d_1 = d_2 = d_3 = 3$, and generated low-rank nominal tensor $\mathcal{X} = \mathcal{G} \times_{n \in [3]} \mathbf{U}_n$, where the entries of \mathcal{G} are drawn independently from a zero-mean Gaussian distribution with standard deviation 3, and arbitrary orthonormal factor matrices. We corrupt the low-rank tensor by adding i.i.d. Gaussian noise drawn from $\mathcal{N}(0, 1^2)$, to all entries. Then, we additively corrupt $N_o = 50$ arbitrary entries of the tensor (frequency of corruption occurrence is 5%) by adding non-zero outlying values drawn independently from $\mathcal{N}(0, \sigma_o^2)$. We estimate the nominal low-rank tensor by TD on the corrupted tensor $\mathcal{X}_{\text{CORR}}$, implemented by means of the proposed formulation/algorithm ($\hat{\mathcal{X}}_{\text{PROP}}$), as well as standard HOOI ($\hat{\mathcal{X}}_{\text{HOOI}}$) and HOSVD ($\hat{\mathcal{X}}_{\text{HOSVD}}$). The tensor reconstruction performance for each estimate $\hat{\mathcal{X}}$ is then measured by the Mean Normalized Squared Error (MNSE)

$$\text{MNSE} = \|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 / \|\mathcal{X}\|_F^2. \quad (10)$$

Intensity of corruption with respect to the added noise is quantified by the Outlier-to-Noise Ratio (ONR)

$$\text{ONR} = \frac{\mathbb{E}\{\|\mathcal{O}\|_F^2\}}{\mathbb{E}\{\|\mathcal{N}\|_F^2\}} = \frac{N_o}{\prod_{n \in [N]} D_n} \frac{\sigma_o^2}{\sigma_n^2}. \quad (11)$$

In this experiments the outlier standard deviation σ_o is set to 0, 2, ..., 16. Fig. 3 shows the performance curve of MNSE vs. ONR over 1000 realizations of noise/outliers. The scale parameter of the proposed TD's loss function is set to 1 ($c = 1$), and the proposed formulation/algorithm's performance is shown for $\alpha = 0, 1, 2$. For $\text{ONR} \leq 0.2$ (corresponding to $\sigma_o \leq 2$), all algorithms show minuscule error in low-rank tensor restoration. With the increase in corruption intensity, HOOI, and HOSVD, and TD with a shape parameter $\alpha = 2$ quickly break down. On the contrary, the two robust configurations of the TD maintain robustness across the board. For $\text{ONR} > 1$,

TD with $\alpha = 0$ outperforms TD with shape parameter $\alpha = 1$, and attains the smallest approximation error among all, for every ONR value.

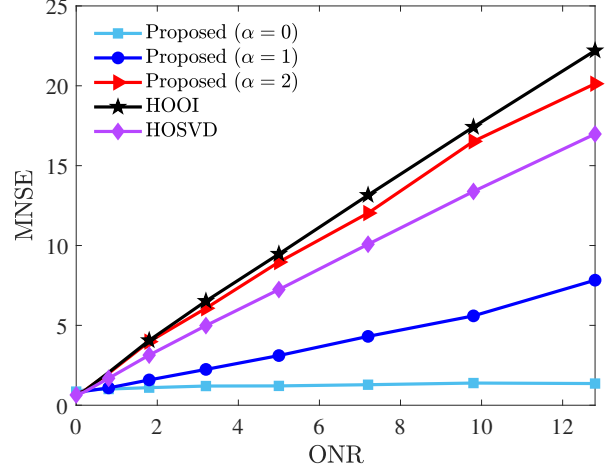


Fig. 3: Performance comparison of three configurations of the proposed method versus standard solvers (HOSVD, HOOI) of the L2-norm TD formulation, for $N_o = 50$.

B. Real data: Uber Pickup dataset

In this experiment, we perform a similar performance evaluation on "Uber Pickups" dataset [18], which is a 4-way tensor of size $183 \times 24 \times 1140 \times 1717$. The (i, j, k, l) -th entry of this tensor represents the number of total Uber pickups in j -th hour of day i , in a location identified by k -th longitude, and l -th latitude coordinates, within New York city. In this experiment, we reduced the original tensor to a 3-way tensor $\mathcal{X} \in \mathbb{R}^{125 \times 125 \times 14}$, by aggregating over the 24 hours of the first 14 days, and summing over the coordinate blocks. The first two modes of the obtained tensor represent location coordinates, and the third mode represent days.

In this study, the clean tensor is additively corrupted as $\mathcal{X}_{\text{corr}} = \mathcal{X} + \mathcal{O}$, where \mathcal{O} is a random outlier tensor, in which $\beta = 20\%$ of the entries within a single arbitrarily chosen slab are random positive integers drawn from a $\text{unif}(v/10, v)$, and the rest of the entries are 0. We decompose the tensor along the first two modes, while fixing the third mode (sample mode) at $\mathbf{U}_3 = \mathbf{I}_{125 \times 125}$, and obtain the approximation tensors $\hat{\mathcal{X}}_{\text{PROP}}$, $\hat{\mathcal{X}}_{\text{HOOI}}$, and $\hat{\mathcal{X}}_{\text{HOSVD}}$.

For the purpose of visual illustration, two sample 125×125 tensor slabs $\mathcal{X}_{:::, i_{\text{clean}}}$, and $\mathcal{X}_{:::, i_{\text{corr}}}$, at a single realization of the above experiment are visualized in Figure 6. After tensor decomposition/restoration, the restored corrupted slabs of the proposed (for $\alpha = 0$), HOOI, and HOSVD algorithms are shown respectively in 4b, 4c, 4d. Corruption-free slabs of the $\hat{\mathcal{X}}_{\text{HOSVD}}$, $\hat{\mathcal{X}}_{\text{HOOI}}$, and $\hat{\mathcal{X}}_{\text{PROP}} (\alpha = 0)$ are shown respectively in Fig. 5a, 5b, 5c. The figures exhibit the HOSVD, and HOOI's poor performance in restoring the original entries of both corrupted and non-corrupted slabs. In contrast to the L2-norm-based methods, the proposed algorithm for $\alpha = 0$, demonstrates robustness and outlier suppression.

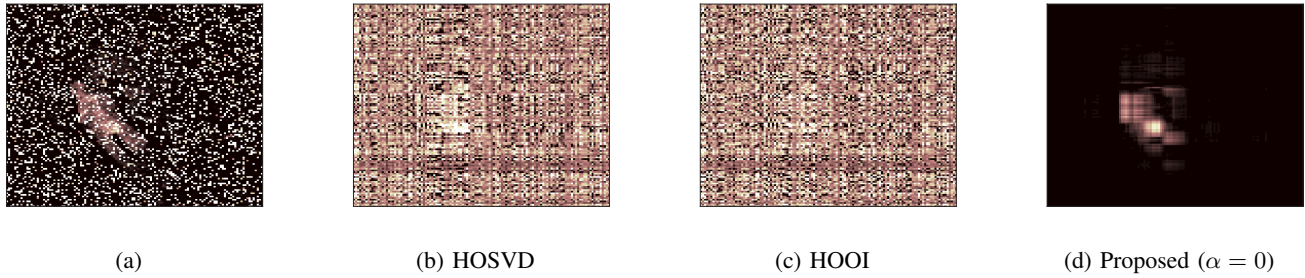


Fig. 4: Corrupted frontal slab of the tensor, and the corresponding approximated slab by the three algorithms, for rank $d_1 = d_2 = 6$, $d_3 = 14$ TD; (a) the corrupted outlying slab.

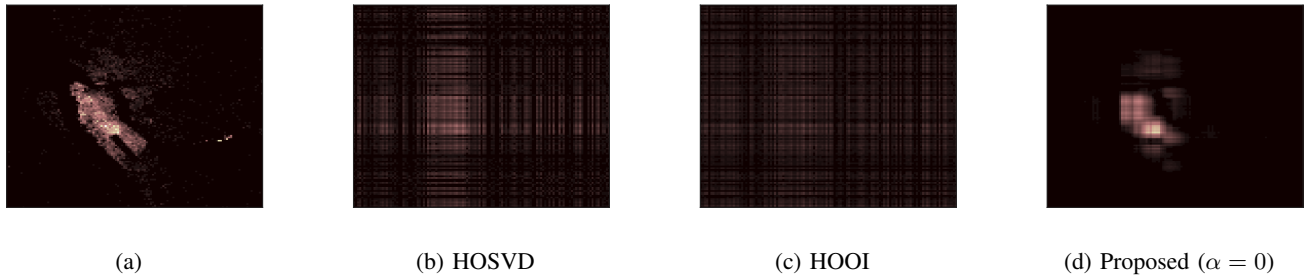


Fig. 5: A clean frontal slab of the tensor, and the corresponding approximated slab by the three algorithms, for rank $d_1 = d_2 = 6$, $d_3 = 14$ TD; (a) the clean slab.

In Fig. 6 we plot the performance curve of MNSE vs. v , for $0 \leq v \leq 500$, over 100 realizations of outliers, for the competing algorithms. The figure demonstrates that, as corruption magnitude increases, the HOSVD, and HOOI break down quickly, while the robust configuration of the proposed algorithm exhibits outstanding outlier resistance even for very high corruption levels.

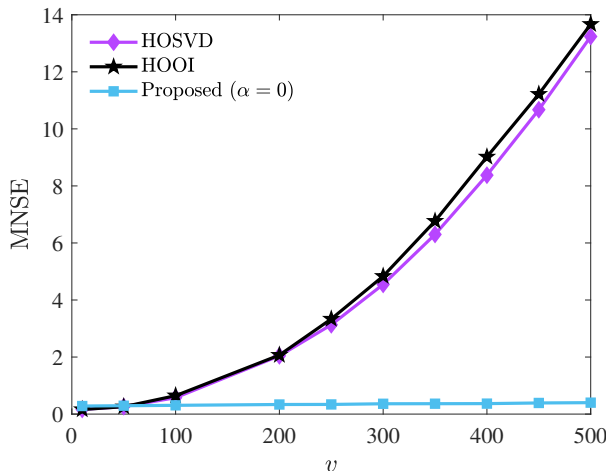


Fig. 6: Performance comparison of three configurations of the proposed method vs. standard solvers (HOSVD, HOOI) of the L2-norm TD formulation "Uber Pickups" dataset.

V. CONCLUSION

In this work we proposed a new formulation for low-rank tensor approximation, based on a generalized robust loss function, and developed for it a novel algorithmic framework. The new loss function allows for tunable robustness of the Tucker decomposition while it can also handle the missing data. Additionally, the smoothness and differentiability of the loss function allows for gradient-descent-type optimization. Our numerical studies on synthetic and real data corroborate the robustness of the proposed formulation/algorithm.

REFERENCES

- [1] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Trans. Intell. Syst. Tech.*, vol. 8, no. 2, pp. 1–44, 2016.
- [2] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [3] M. Ashraphijuo and X. Wang, "Union of low-rank tensor spaces: Clustering and completion," *J. Mach. Learn. Res.*, vol. 21, no. 69, pp. 1–36, 2020.
- [4] K. Gilman and L. Balzano, "Online tensor completion and free submodule tracking with the T-SVD," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020, pp. 3282–3286.
- [5] M. S. Asif and A. Prater-Bennette, "Low-rank tensor ring model for completing missing visual data," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020, pp. 5415–5419.
- [6] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proc. Eur. Conf. Comput. Vision (ECCV)*, Copenhagen, Denmark, 2002, pp. 447–460.
- [7] X. Fu, S. Ibrahim, H.-T. Wai, C. Gao, and K. Huang, "Block-randomized stochastic proximal gradient for low-rank tensor factorization," *IEEE Trans. Signal Process.*, vol. 68, pp. 2170–2185, 2020.

- [8] K. Tountas, D. A. Pados, and M. J. Medley, "Conformity evaluation and L1-norm principal-component analysis of tensor data," in *Proc. SPIE Big Data: Learn. Analytics Appl.*, vol. 10989, Baltimore, MD, 2019, p. 109890P.
- [9] D. D. Miller, "The medical ai insurgency: what physicians must know about data to practice with intelligent machines," *NPJ digital medicine*, vol. 2, no. 1, pp. 1–5, 2019.
- [10] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [11] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [12] P. P. Markopoulos, D. G. Chachlakis, and E. E. Papalexakis, "The exact solution to rank-1 L1-norm TUCKER2 decomposition," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 511–515, 2018.
- [13] D. G. Chachlakis, A. Prater-Bennette, and P. P. Markopoulos, "L1-norm Tucker tensor decomposition," *IEEE Access*, vol. 7, pp. 178 454–178 465, 2019.
- [14] D. G. Chachlakis, M. Dhanaraj, A. Prater-Bennette, and P. P. Markopoulos, "Dynamic L1-norm Tucker tensor decomposition," *IEEE J. Selected Topics Signal Process.*, vol. 15, no. 3, pp. 587–602, 2021.
- [15] D. G. Chachlakis, A. Prater-Bennette, and P. P. Markopoulos, "L1-norm higher-order orthogonal iterations for robust tensor analysis," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. Barcelona, Spain: IEEE, 2020, pp. 4826–4830.
- [16] J. T. Barron, "A general and adaptive robust loss function," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4331–4339.
- [17] R. Garg, N. Wadhwa, S. Ansari, and J. T. Barron, "Learning single camera depth estimation using dual-pixels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, Korea, 2019, pp. 7628–7637.
- [18] S. Smith, J. W. Choi, J. Li, R. Vuduc, J. Park, X. Liu, and G. Karypis, "FROSTT: The formidable repository of open sparse tensors and tools," 2017, [Online]. Available: <http://frostt.io/>.