

L1-NORM HIGHER-ORDER ORTHOGONAL ITERATIONS FOR ROBUST TENSOR ANALYSIS

Dimitris G. Chachlakis[‡], Ashley Prater-Bennette[†], and Panos P. Markopoulos^{‡}*

[‡]Dept. of Electrical and Microelectronic Engineering, Rochester Institute of Technology,
Rochester, NY 14623, USA

E-mail: dimitris@mail.rit.edu, panos@rit.edu

[†]Air Force Research Laboratory, Rome, NY 13441, USA

E-mail: ashley.prater.3@us.af.mil

ABSTRACT

Standard Tucker tensor decomposition seeks to maximize the L2-norm of the compressed tensor; thus, it is very responsive to outlying/high-magnitude entries among the processed data. To counteract the impact of outliers in tensor data analysis, we propose L1-Tucker: a reformulation of standard Tucker decomposition, resulting by simple substitution of the outlier-responsive L2-norm by the sturdier L1-norm. Then, we propose the L1-norm Higher Order Orthogonal Iterations (L1-HOOI) algorithm for the approximate solution to L1-Tucker. Our numerical studies on data reconstruction and classification corroborate that L1-HOOI exhibits sturdy resistance against outliers compared to standard counterparts.

1. INTRODUCTION

Tensor analysis finds important applications in areas such as signal processing [1], computer vision [2], and data mining [3], to name a few. Tucker decomposition is a popular tensor factorization approach, typically carried out by the Higher-Order Singular-Value Decomposition (HOSVD) and Higher-Order Orthogonal Iterations (HOOI) [4, 5] algorithms.

In applications of specific interest (e.g., classification of multimodal data [6]), Tucker decomposition of an N -way tensor is reformulated to Tucker2 decomposition which treats the processed tensor as a collection of $(N - 1)$ -way tensor-measurements and strives to jointly Tucker-analyze these measurements. Generalized Low-Rank Approximations of Matrices (GLRAM) [7] is a special case of Tucker2 for $(N = 3)$ -way tensors. Tucker and Tucker2 are both considered multiway generalizations of the standard Principal Component Analysis (PCA) [8] of matrices, carried out by Singular-Value Decomposition (SVD). The focal point of this work will be general-order Tucker tensor decomposition.

Tucker decomposition, owing to its L2-norm formulation, is highly sensitive to erroneous entries (outliers) within

the processed tensor [9–11] which often appear in modern datasets. PCA, a special case of Tucker, has also been known to exhibit sensitivity towards outliers. L1-norm PCA (L1-PCA) [12] has illustrated significant robustness against outliers in applications such as image recovery, video surveillance, and classification of micro-Doppler radar signatures, to name a few [13–15]. Motivated by the success of L1-PCA, L1-norm formulations of higher-order Tucker2 were studied in [10, 11, 16, 17]. More recently, [18] introduced an L1-norm reformulation of general-order tensor Tucker decomposition (L1-Tucker) and offered a first algorithm for its solution, namely L1-HOSVD.

L1-HOSVD [18, 19] exhibited remarkable corruption resistance in data reconstruction and classification, in sharp contrast with HOSVD and HOOI. However, L1-HOSVD decomposes each tensor mode individually and, thus, it does not leverage inter-mode dependencies. In this work, we extend L1-HOSVD and present L1-HOOI: a novel iterative algorithm for the solution to L1-Tucker. The proposed algorithm is accompanied by numerical studies which illustrate that: (i) L1-HOOI attains similar performance to standard Tucker solvers when the processed data are outlier free; (ii) L1-HOOI is significantly more outlier-resistant than standard Tucker solvers; (iii) L1-HOOI iterations converge at a solution that outperforms L1-HOSVD in the L1-Tucker metric.

2. BACKGROUND

2.1. Tucker Decomposition

We consider N -way tensor $\mathcal{X} \in \mathbb{R}^{D_1 \times D_2 \times \dots \times D_N}$ and define $[N] = \{1, 2, \dots, N\}$. $\mathcal{X}$ is a structured multiway array of $M = \prod_{n \in [N]} D_n$ scalar entries, each of which is accessible by N indices. For every mode index $n \in [N]$, $\mathcal{X}$ can be treated as a collection of $M_n = \prod_{m \in [N] \setminus n} D_m$ length- D_n vectors arranged across its n -th mode. These vectors are known as the mode- n fibers of $\mathcal{X}$. That is, for any fixed set of indices $\{i_m\}_{m \in [N] \setminus n}$, a mode- n fiber

*Corresponding author.

of $\mathcal{X}$ is vector $\mathcal{X}(i_1, \dots, i_{n-1}, :, i_{n+1}, \dots, i_N) \in \mathbb{R}^{D_n}$. A matrix with columns the mode- n fibers of tensor $\mathcal{X}$ is known as the mode- n matrix flattening of $\mathcal{X}$, commonly denoted by $[\mathcal{X}]_{(n)} \in \mathbb{R}^{D_n \times M_n}$. Noticing that the mode- n fibers of $\mathcal{X}$ can be organized in multiple different orders, in this work we specify that tensor element $(i_1, i_2, \dots, i_N)$ is mapped to the mode- n flattening element (i_n, z) for $z = 1 + \sum_{h \in [N] \setminus n} (i_h - 1)Z_h$ and $Z_h = \prod_{k \in [h-1] \setminus n} D_k$, for every $h \in [N]$ [4].

Tucker decomposition of $\mathcal{X}$ seeks to maximize the L2-norm of the core tensor which results by the multi-projection of $\mathcal{X}$ with a set of N orthonormal-basis factors. Defining the Stiefel manifold $\mathbb{S}(D, d) = \{\mathbf{R} \in \mathbb{R}^{D \times d}; \mathbf{R}^\top \mathbf{R} = \mathbf{I}_d\}$, $d < D$, Tucker decomposition is formulated as

$$\underset{\{\mathbf{R}_n \in \mathbb{S}(D_n, d_n)\}_{n \in [N]}}{\text{maximize}} \quad \|\mathcal{X} \times_{n \in [N]} \mathbf{R}_n^\top\|_F^2, \quad (1)$$

where $\times_n$ denotes the mode- n tensor-to-matrix product [4], L2-norm operator $\|\cdot\|_F^2$ returns the summation of the squared entries of its input argument, and $\times_{n \in [N]} \mathbf{R}_n^\top$ compactly denotes the multiway product $\times_1 \mathbf{R}_1^\top \times_2 \mathbf{R}_2^\top \dots \times_N \mathbf{R}_N^\top$.

2.2. L1-Tucker Decomposition

Outliers often appear in modern large datasets due to a number of causes, including sensor malfunctions, errors in data storage/transfer, and, in some cases, intentional dataset contamination [20]. Researchers from the fields of signal processing, data analysis, and machine learning have long observed that PCA and its multiway generalization, Tucker, are significantly misled by outliers in the processed data, even in a very small fraction [9–12]. Consequently, applications the performance of which depends on PCA/Tucker (e.g., classification, clustering) can be heavily affected. The outlier sensitivity of PCA/Tucker is, in part, attributed to their L2-norm based formulation, which places squared emphasis on each data point, favoring outliers in the data periphery. Various robust reformulations of Tucker have been proposed over time, in order to suppress the undesired impact of outliers in data analysis [21–23]. Arguably, the most straightforward approach is to replace the outlier-responsive L2-norm in PCA by the L1-norm which places linear emphasis on each data point, thus suppressing the undesired effects of outliers. For data matrix $\mathbf{X} \in \mathbb{R}^{D_1 \times D_2}$ and $d_1 < \text{rank}(\mathbf{X})$, this modification results to the L1-PCA formulation [12]

$$\underset{\mathbf{R} \in \mathbb{S}(D_1, d_1)}{\text{maximize}} \quad \|\mathbf{R}^\top \mathbf{X}\|_1, \quad (2)$$

where $\|\cdot\|_1$ denotes the L1-norm operator which returns the summation of the absolute entries of its input argument. L1-PCA in (2) has been extensively studied over the past few years and many solvers (both exact and approximate) have been presented for its solution [12, 24, 25]. Even more recently, L1-PCA was extended to L1-norm-based Tucker2 formulation, specifically for 3-way tensors [11, 16, 17].

Following the paradigms of L1-PCA and L1-Tucker2, a recent work presented L1-Tucker decomposition [18], as the outlier-resistant counterpart of general-order Tucker, formulated as

$$\underset{\{\mathbf{R}_n \in \mathbb{S}(D_n, d_n)\}_{n \in [N]}}{\text{maximize}} \quad \|\mathcal{X} \times_{n \in [N]} \mathbf{R}_n^\top\|_1. \quad (3)$$

Moreover, [18] offered the first solver for L1-Tucker, namely L1-HOSVD, which we briefly review below.

2.3. L1-HOSVD [18]

For every $n \in [N]$, L1-HOSVD approximates $\mathbf{R}_n$ by L1-PCA of the mode- n matrix flattening of $\mathcal{X}$. That is, L1-HOSVD approximates the solution to (3) by the set $\{\mathbf{R}_n^{\text{L1-hosvd}}\}_{n \in [N]}$, such that $\mathbf{R}_n^{\text{L1-hosvd}}$ is a maximizing argument of

$$\underset{\mathbf{R} \in \mathbb{S}(D_n, d_n)}{\text{maximize}} \quad \|\mathbf{R}^\top [\mathcal{X}]_{(n)}\|_1. \quad (4)$$

A solution to (4) can be computed both exactly and approximately with any of the existing L1-PCA solvers [12, 24, 26]. For instance, [18] approximated the solution to (4) by means of alternating optimization iterations of [26] as follows. According to [12], the maximization in (4) can be equivalently written as

$$\underset{\substack{\mathbf{R} \in \mathbb{S}(D_n, d_n) \\ \mathbf{B} \in \{\pm 1\}^{M_n \times d_n}}}{\text{maximize}} \quad \text{Tr}(\mathbf{R}^\top [\mathcal{X}]_{(n)} \mathbf{B}). \quad (5)$$

Moreover, for any tall matrix $\mathbf{A} \in \mathbb{R}^{D \times d}$ that admits SVD $\mathbf{A} = \mathbf{W} \mathbf{S}_{d \times d} \mathbf{Q}^\top$, operator $\Omega(\mathbf{A}) = \mathbf{W} \mathbf{Q}^\top$ solves $\underset{\mathbf{R} \in \mathbb{S}(D, d)}{\text{maximize}} \text{Tr}(\mathbf{R}^\top \mathbf{A})$ [27]. Thus, for any fixed $\mathbf{B}$, maximization in (5) is attained by $\mathbf{R} = \Omega([\mathcal{X}]_{(n)} \mathbf{B})$. On the other hand, for any fixed $\mathbf{R}$, maximization in (5) is attained by $\mathbf{B} = \text{sgn}([\mathcal{X}]_{(n)}^\top \mathbf{R})$, where $\text{sgn}(\cdot)$ returns the ± 1 signs of the entries of its argument. Thus, the solution to (4) can be pursued as

$$\mathbf{R}_t = \Omega([\mathcal{X}]_{(n)} \text{sgn}([\mathcal{X}]_{(n)}^\top \mathbf{R}_{t-1})), \quad (6)$$

for $t = 1, 2, \dots$, and arbitrary initialization $\mathbf{R}_0 \in \mathbb{S}(D_n, d_n)$. In [18] it was shown that the metric of (4) increases monotonically across t , which combined with the fact that the metric of (4) is upper bounded by the exact L1-PCA solution, obtainable by the optimal algorithms of [12], guarantees that the iterations in (6) converge. In practice, iterations can be terminated when the L1-PCA objective metric $\|\mathbf{R}_t^\top [\mathcal{X}]_{(n)}\|_1$ stops increasing, or when t exceeds an upper limit of allowed iterations. Upon convergence/termination, L1-HOSVD returns $\mathbf{R}_n^{\text{L1-hosvd}} = \mathbf{R}_t$.

3. PROPOSED L1-HOOI

We present L1-HOOI, a new algorithm for approximating the solution to L1-Tucker in (3). L1-HOOI is an iterative solver,

L1-HOOI algorithm for L1-Tucker tensor decomposition

Input: $\mathcal{X} \in \mathbb{R}^{D_1 \times \dots \times D_N}$, d_n for every $n \in [N]$

- 1: $\mathbf{R}_n \leftarrow \mathbf{R}_n^{\text{L1-hosvd}}$ for every $n \in [N]$
- 2: while $\sum_{n=1}^N \|\mathcal{X}_{(n)}^\top \mathbf{R}_n\|_1$ increases, do
- 3: for $n \in [N]$
- 4: $\mathbf{Z} \leftarrow [\mathcal{X} \times_{m < n} \mathbf{R}_m^\top \times_{k > n} \mathbf{R}_k^\top]_{(n)}$
- 5: while $\|\mathbf{Z}^\top \mathbf{R}_n\|_1$ increases, do
- 6: $\mathbf{R}_n \leftarrow \Omega(\mathbf{Z} \text{sgn}(\mathbf{Z}^\top \mathbf{R}_n))$

Return: $\mathbf{R}_n^{\text{L1-hooi}} \leftarrow \mathbf{R}_n$ for every $n \in [N]$

Fig. 1. The L1-HOOI algorithm for solving L1-Tucker in (3).

arbitrarily initialized to a set of bases $\{\mathbf{R}_{n,0}^{\text{L1-hooi}}\}_{n \in [N]}$, such as the solution of L1-HOSVD. At every iteration $q \geq 1$, L1-HOOI updates $\{\mathbf{R}_{n,q}^{\text{L1-hooi}}\}_{n \in [N]}$ sequentially (across n) while the L1-Tucker metric increases at every single update. Mathematically, at the q -th iteration step, for $n \in [N]$, L1-HOOI defines $\mathbf{Z}_{n,q} = [\mathcal{X} \times_{m < n} \mathbf{R}_{m,q}^{\text{L1-hooi}^\top} \times_{k > n} \mathbf{R}_{k,q-1}^{\text{L1-hooi}^\top}]_{(n)}$, and updates $\mathbf{R}_{n,q}^{\text{L1-hooi}}$ by solving

$$\underset{\mathbf{R} \in \mathbb{S}(D_n, d_n)}{\text{maximize}} \|\mathbf{R}^\top \mathbf{Z}_{n,q}\|_1. \quad (7)$$

Following the L1-PCA alternating optimization solver framework presented above, $\mathbf{R}_{n,q}^{\text{L1-hooi}}$ is optimized by the basis obtained upon convergence of the iteration

$$\mathbf{R}_t = \Omega(\mathbf{Z}_{n,q} \text{sgn}(\mathbf{Z}_{n,q}^\top \mathbf{R}_{t-1})), \quad (8)$$

$t = 1, 2, \dots$, where $\mathbf{R}_0$ is initialized at $\mathbf{R}_{n,q-1}^{\text{L1-hooi}}$. For fixed q , iterations in (8) can practically terminate when the update of $\mathbf{R}_t$ doesn't increase $\|\mathbf{R}_t^\top \mathbf{Z}_{n,q}\|_1$, or when t exceeds a maximum number of allowed iterations. At this point, we observe that, since iterations in (8) increase $\|\mathbf{Z}_{n,q}^\top \mathbf{R}_t\|_1$ across t , it holds that

$$\|\mathbf{R}_{n,q}^{\text{L1-hooi}^\top} \mathbf{Z}_{n,q}\|_1 \geq \|\mathbf{R}_{n,q-1}^{\text{L1-hooi}^\top} \mathbf{Z}_{n,q}\|_1. \quad (9)$$

The latter inequality implies that each basis update across q and n increases the L1-Tucker metric, which in turn, suggests that when initialized at the L1-HOSVD solution, L1-HOOI offers a set of bases that is superior with respect to the L1-Tucker metric. In addition, using standard norm inequalities and the fact that orthogonal projection is a non-expansive mapping in the L2 norm, we can easily show that, for any $\{\mathbf{R}_n \in \mathbb{S}(D_n, d_n)\}_{n \in [N]}$, it holds

$$\|\mathcal{X} \times_{n \in [N]} \mathbf{R}_n^\top\|_1 \leq \sqrt{\prod_{n \in [N]} d_n} \|\mathcal{X}\|_F. \quad (10)$$

Therefore, the L1-HOOI iterations are guaranteed to converge. In practice, iterations across q can terminate when $\sum_{n=1}^N \|\mathbf{R}_{n,q}^{\text{L1-hooi}^\top} [\mathcal{X}]_{(n)}\|_1$ ceases to increase, or when q exceeds a maximum permitted number.

4. EXPERIMENTAL STUDIES

4.1. Tensor Reconstruction – Synthetic Data

We consider ($N = 5$)-way tensor following Tucker structure $\mathcal{X} = \mathcal{G} \times_{n \in [5]} \mathbf{R}_n$, where $D_{n \in [5] \setminus 3} = 10$, $D_3 = 8$, $d_{n \in [5] \setminus 3} = 4$, $d_3 = 5$. The entries of $\mathcal{G}$ are drawn from $\mathcal{N}(0, 12)$. Arbitrary orthonormal-basis factors $\mathbf{R}_n, n \in [5]$ are considered. The entries of $\mathcal{X}$ are corrupted with zero-mean unit-variance additive white Gaussian noise (AWGN), disrupting its low-rank Tucker structure. Moreover, N_o out of the 80,000 entries of $\mathcal{X}$ are additively contaminated with high magnitude outliers drawn from $\mathcal{N}(0, \sigma_o^2)$. Compactly, we obtain $\mathcal{X}^c = \mathcal{X} + \mathcal{N} + \mathcal{O}$, where tensors $\mathcal{N}$ and $\mathcal{O}$ model noise and sparse outliers respectively. The task of interest is to recover $\mathcal{X}$ from the corrupted $\mathcal{X}^c$. To that end, we carry out Tucker decomposition of $\mathcal{X}^c$ by means of HOSVD, HOOI, L1-HOSVD, and the proposed L1-HOOI algorithms. Then, we estimate $\mathcal{X}$ as $\hat{\mathcal{X}} = \mathcal{X}^c \times_{n \in [5]} \hat{\mathbf{R}}_n \hat{\mathbf{R}}_n^\top$ where $\{\hat{\mathbf{R}}_n\}_{n \in [5]}$ are the orthonormal-basis factors returned by the Tucker decomposition algorithms. To measure the success of each algorithm at recovering $\mathcal{X}$, we compute the normalized squared error (NSE) $\|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 / \|\mathcal{X}\|_F^2$. In Fig. 2, we plot the the mean normalized squared error (MNSE), estimated over 1000 independent noise/outlier realizations, versus outlier standard deviation $\sigma_o = 4, 8, \dots, 28$, for $N_o = 70$. In the absence of outliers ($\sigma_o = 0$), all methods exhibit similarly low MNSE. As σ_o increases, the MNSE of all methods increases. We observe that the performances of HOSVD and HOOI significantly deteriorate for $\sigma_o \geq 16$ and $\sigma_o \geq 24$, respectively. L1-HOSVD and L1-HOOI remain robust across the board with L1-HOOI exhibiting the lowest MNSE.

In Fig. 3, we set $\sigma_o = 25$ and plot the MNSE versus the number of outliers $N_o = 0, 20, \dots, 100$. We notice that in the absence of outliers ($N_o = 0$) all methods exhibit low MNSE. As the number of outliers increases, HOSVD and HOOI tensor estimates deviate significantly from $\mathcal{X}$. For example, for $N_o = 80$ outliers (about 0.1% of the entries of $\mathcal{X}$ are corrupted) HOSVD and HOOI are clearly misled; on the other hand, L1-HOSVD and L1-HOOI exhibit very low MNSE.

4.2. Classification – AT&T Face Dataset

We conduct a classification study on the AT&T face dataset [28] which consists of size $(D_1 = 112) \times (D_2 = 92)$ image-samples. We consider $P = 3$ training samples from each of the first $C = 5$ classes. Following the approach of [6], we organize all training data in tensor $\mathcal{X} \in \{0, 1, \dots, 255\}^{D_1 \times D_2 \times 15}$ and decompose modes 1 and 2 with $d_1 = d_2 = 25$ (compression ratio 16.48) by means of HOSVD, HOOI, and L1-HOOI (i.e., we fix $\mathbf{U}_3 = \mathbf{I}_{15}$). Then, we obtain 15 training vectors in $\mathbf{G} = [\mathbf{G}_1 \ \mathbf{G}_2 \ \dots \ \mathbf{G}_5] = [\mathcal{X} \times_1 \mathbf{R}_1^\top \times_2 \mathbf{R}_2^\top]_{(3)}^\top \in \mathbb{R}^{625 \times 15}$, where $\mathbf{G}_c$ has as columns the feature-vectors of class c . These vectors are used to build a k -NN classifier. A testing point $\mathbf{T} \in \mathbb{R}^{D_1 \times D_2}$ is

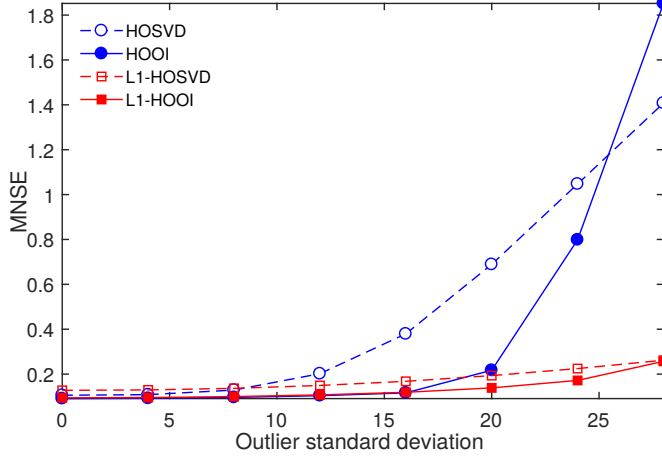


Fig. 2. MNSE versus outlier standard deviation σ_o , for HOSVD, HOOI, L1-HOSVD, and L1-HOOI. $N_o = 70$ (about 0.0875% of the total number of entries of $\mathcal{X}$ are outlier corrupted).

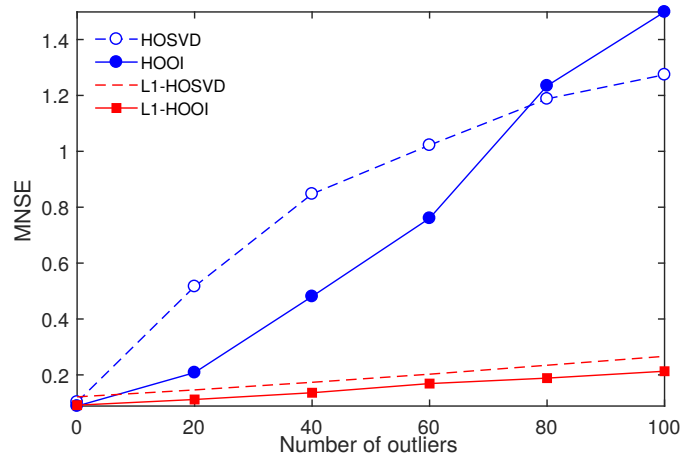


Fig. 3. MNSE versus number of outliers N_o , for HOSVD, HOOI, L1-HOSVD, and L1-HOOI. $\sigma_o = 25$.

first compressed as $\mathbf{t} = \text{vec}(\mathbf{R}_1^T \mathbf{T} \mathbf{R}_2) \in \mathbb{R}^{625 \times 1}$ and then classified by the k -NN; that is, for $k = 1$, $\mathbf{t}$ is classified to class $c^* = \text{argmin}_{c \in [C]} \{\min_{j \in [P]} \|\mathbf{t} - [\mathbf{G}_c]_{:,j}\|_2^2\}$. To test the corruption resistance of the compared methods, we corrupt each training and each testing sample, with probability $\alpha = 0.1$. Each pixel of a corrupted sample is substituted by salt ('255') or pepper ('0'), with probability β . For each training-data configuration, we test the performance of the methods on 7 testing points from each class. We repeat this study on 2000 configurations of training data, testing data, and corruption, and calculate the average classification accuracy of plain k -NN, HOSVD, HOOI, and L1-HOOI. In Fig. 4, we plot the performance of these methods, versus β . Once again, we observe that, even though all methods perform sim-

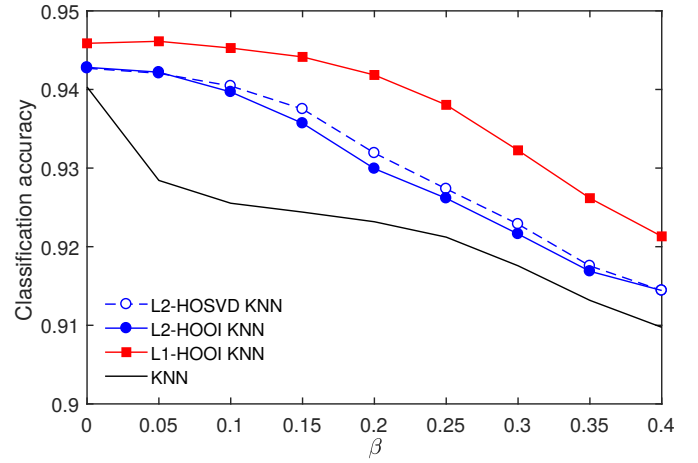


Fig. 4. AT&T classification study: Classification accuracy versus β , for $\alpha = 0.1$ and $d = 25$.

ilarly well, L1-HOOI exhibits superior performance than all its counterparts, for every value of β .

5. CONCLUSIONS

We introduced L1-HOOI, an iterative algorithm for the solution of L1-Tucker tensor decomposition. We provided numerical studies on tensor reconstruction and classification that compare L1-HOOI with standard counterparts. Our studies showed that L1-HOOI outperforms L1-HOSVD in the L1-Tucker metric and that, in contrast to HOSVD and HOOI, it exhibits sturdy resistance against outliers in the tensor data.

6. ACKNOWLEDGMENTS

This research was supported in part by the Air Force Office of Scientific Research under grant 18RICOR029 and the National Science Foundation under grant OAC-1808582.

References

- [1] 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, Jul. 2017.
- [2] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proc. Eur. Conf. Comput. Vision (ECCV)*, Copenhagen, Denmark, May 2002, pp. 447–460.
- [3] 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. Technol.*, vol. 8, pp. 16:1–16:44, Jan. 2017.

- [4] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [5] L. De Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 1253–1278, 2000.
- [6] A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *IEICE Nonlinear Theory Appl.*, vol. 1, no. 1, pp. 37–68, 2010.
- [7] J. Ye, "Generalized low rank approximations of matrices," *Mach. Learn.*, vol. 61, pp. 167–191, Nov. 2005.
- [8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.
- [9] X. Fu, K. Huang, W. K. Ma, N. D. Sidiropoulos, and R. Bro, "Joint tensor factorization and outlying slab suppression with applications," *IEEE Trans. Signal Process.*, vol. 63, pp. 6315–6328, Dec. 2015.
- [10] X. Cao, X. Wei, Y. Han, and D. Lin, "Robust face clustering via tensor decomposition," *IEEE Trans. Cybern.*, vol. 45, pp. 2546–2557, Nov. 2015.
- [11] P. P. Markopoulos, D. G. Chachlakis, and E. E. Papalexakis, "The exact solution to rank-1 L1-norm Tucker2 decomposition," *IEEE Trans. Signal Process. Lett.*, vol. 25, no. 4, pp. 511–515, Jan. 2018.
- [12] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Optimal algorithms for L1-subspace signal processing," *IEEE Trans. Signal Process.*, vol. 62, pp. 5046–5058, Oct. 2014.
- [13] P. P. Markopoulos, S. Kundu, and D. A. Pados, "L1-fusion: Robust linear-time image recovery from few severely corrupted copies," Quebec City, Canada, Sep. 2015, pp. 1225–1229.
- [14] Y. Liu and D. A. Pados, "Compressed-sensed-domain L1-PCA video surveillance," *IEEE Trans. on Multimedia*, vol. 18, pp. 351–363, Mar. 2016.
- [15] P. P. Markopoulos and F. Ahmad, "Indoor human motion classification by L1-norm subspaces of micro-Doppler signatures," in *Proc. IEEE Radar Conf.*, Seattle, WA, May 2017, pp. 1807–1810.
- [16] D. G. Chachlakis and P. P. Markopoulos, "Robust decomposition of 3-way tensors based on L1-norm," in *Proc. SPIE Defense and Commercial Sens.*, Orlando, FL, Apr. 2018, pp. 1 065 807:1–1 065 807:15.
- [17] —, "Novel algorithms for exact and efficient L1-norm-based Tucker2 decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, Canada, Apr. 2018, pp. 6294–6298.
- [18] P. P. Markopoulos, D. G. Chachlakis, and A. Prater-Bennette, "L1-norm higher-order singular-value decomposition," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Anaheim, CA, Nov. 2018, pp. 1353–1357.
- [19] D. G. Chachlakis, M. Dhanaraj, A. Prater-Bennette, and P. P. Markopoulos, "Options for multimodal classification based on L1-Tucker decomposition," in *Proc. SPIE Defense and Commercial Sens.*, Baltimore, MD, Apr. 2019, pp. 109 890O:1–109 890O:13.
- [20] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *Nature Sci. Rev.*, vol. 1, pp. 293–314, Jun. 2014.
- [21] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.
- [22] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," Vancouver, Canada, Dec. 2010, pp. 2496–2504.
- [23] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, Jun. 2016, pp. 5249–5257.
- [24] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados, "Efficient L1-norm principal-component analysis via bit flipping," *IEEE Trans. Signal Process.*, vol. 65, pp. 4252–4264, Aug. 2017.
- [25] N. Tsagkarakis, P. P. Markopoulos, G. Sklivanitis, and D. A. Pados, "L1-norm principal-component analysis of complex data," *IEEE Trans. Signal Process.*, vol. 66, pp. 3256–3267, Jun. 2018.
- [26] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy L1-norm maximization," in *Proc. Int. Joint Conf. Artif. Intell.*, Barcelona, Spain, Jul. 2011.
- [27] J. C. Gower and G. B. Dijkstra, *Procrustes Problems*. Oxford, UK: Oxford University Press, 2004.
- [28] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *IEEE Workshop Appl. Comput. Vision*, Sarasota, FL, Dec. 1994, pp. 138–142.