Local minima of the best low multilinear rank approximation of tensors

Mariya Ishteva, P.-A. Absil, Sabine Van Huffel and Lieven De Lathauwer

***Abstract*— Higher-order tensors are generalizations of vectors and matrices to third- or even higher-order arrays of numbers. We consider a generalization of column and row rank of a matrix to tensors, called multilinear rank. Given a higher-order tensor, we are looking for another tensor, as close as possible to the original one and with multilinear rank bounded by prespecified numbers. In this paper, we give an overview of recent results pertaining the associated cost function. It can have a number of local minima, which need to be interpreted carefully. Convergence to the global minimum cannot be guaranteed with the existing algorithms. We discuss the conclusions that we have drawn from extensive simulations and point out some hidden problems that might occur in real applications.**

I. INTRODUCTION

Higher-order tensors are generalizations of vectors and matrices, i.e., they are arrays of numbers indexed with more than two indices. A mode-n vector of an Nth-order tensor, $n = 1, 2, \ldots, N$, is a vector, obtained by varying the nth index of the tensor and fixing the rest of the indices. This is a straightforward generalization of a column or a row vector of a matrix. As in matrix algebra, the concept of rank plays an essential role in tensor algebra. The mode-n rank of the tensor is the number of linearly independent mode-n vectors. The mode-n ranks may be different for each n. The n-tuple of mode-n ranks is called the multilinear rank of the tensor [5]. It is a generalization of column and row rank of a matrix. In this sense, the best low multilinear rank approximation of a higher-order tensor is a tensor, as close as possible to the original one and such that it has multilinear rank bounded by given numbers. This approximation is used for dimensionality reduction and signal subspace estimation in an increasing number of applications, including higher-order statistics, biomedical signal processing, telecommunications and many other fields.

The best low multilinear rank approximation of a tensor is a generalization of the best low-rank approximation of a matrix. Truncation of the singular value decomposition (SVD) yields the unique optimal solution in the matrix case. However, in the tensor case, the corresponding minimization problem often has several local minima. Moreover, truncation of the higher-order singular value decomposition (HOSVD) [3], [10], [11] leads to a suboptimal solution. The latter is often a good starting point for iterative algorithms.

II. PROBLEM FORMULATION

A transparent way to define the problem in mathematical terms is to look for a solution of the minimization problem

$$\min \|\mathcal{A} - \hat{\mathcal{A}}\|^2, \tag{1}$$

where $\mathcal{A}$ is the original tensor, $\hat{\mathcal{A}}$ has bounded multilinear rank and the considered norm is the Frobenius norm (the square root of the sum of squares of all elements). However, this formulation is inconvenient for applying optimization algorithms directly. Equivalently [9], [4], for a third-order tensor $\mathcal{A}$, we look for column-wise orthonormal matrices $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$, such that

$$\max \|\mathcal{A} \bullet_1 \mathbf{U}^T \bullet_2 \mathbf{V}^T \bullet_3 \mathbf{W}^T\|^2 \tag{2}$$

is achieved, where "$\bullet_i$", $i = 1, 2, 3$ stands for the mode-i product of a tensor with a matrix, see [3]. The optimal tensor $\hat{\mathcal{A}}$ is then derived from

$$\hat{\mathcal{A}} = \mathcal{A} \bullet_1 \mathbf{U}\mathbf{U}^T \bullet_2 \mathbf{V}\mathbf{V}^T \bullet_3 \mathbf{W}\mathbf{W}^T .$$

Similar expressions can be obtained for tensors of order higher than three. We consider only third-order tensors for simplicity.

A traditional algorithm for solving the latter problem is the higher-order orthogonal iterations (HOOI) [4], [9]. Other algorithms have recently been proposed in the literature, see [6] and the references therein. Note that it is enough to find the column spaces of the $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$ matrices. The particular entries of $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$ are not essential since multiplying any of the three matrices from the right by an orthogonal matrix leads to the same final approximation $\hat{\mathcal{A}}$. On the other hand, standard optimization algorithms face a difficulty caused by this invariance property. There are infinitely many equivalent solutions whereas numerical algorithms have proven convergence properties if the solutions are isolated. The invariance can be removed by working on quotient matrix manifolds [1].

Research supported by: (1) The Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007–2011), (2) Communauté française de Belgique - Actions de Recherche Concertées, (3) Research Council K.U.Leuven: GOA-AMBioRICS, GOA-MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), (4) F.W.O. project G.0427.10N "Integrated EEG-fMRI". (5) "Impulsfinanciering Campus Kortrijk (2007–2012)(CIF1)" and STRT1/08/023. Part of this research was carried out while M. Ishteva was with K.U.Leuven, supported by OE/06/25, OE/07/17, OE/08/007, OE/09/004. The scientific responsibility rests with the authors.

M. Ishteva and P.-A. Absil are with the Department of Mathematical Engineering, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium mariya.ishteva@uclouvain.be, http://www.inma.ucl.ac.be/~absil

S. Van Huffel and L. De Lathauwer are with the Department of Electrical Engineering - ESAT/SCD, K.U.Leuven, 3001 Leuven, Belgium. L. De Lathauwer is also with Group Science, Engineering and Technology, K.U.Leuven Campus Kortrijk, 8500 Kortrijk, Belgium sabine.vanhuffel@esat.kuleuven.be, lieven.delathauwer@kuleuven-kortrijk.be

III. LOCAL MINIMA

The cost function (1) has local nonglobal minima. This is a key observation since the best low-rank approximation of a matrix has a unique minimum, which can be obtained from the truncated SVD. In the case of higher-order tensors, even if the iterative algorithms are initialized with the truncated HOSVD, convergence to the global minimum cannot be guaranteed.

Our simulations [7] indicate that in case of tensors with low multilinear rank, perturbed by a small amount of additive noise, the algorithms converge to a small number of local minima. Increasing the noise level leads to tensors that are less structured and then more local minima are found. This behavior is related to the distribution of the multilinear singular values of the tensor. Let $\mathbf{A}_{(1)}$ be a matrix the columns of which are the mode-1 vectors of the tensor and let R_1 be the mode-1 rank of the noise-free tensor. In case of low noise level, there is a gap between the R_1th and (R_1+1)th singular values of $\mathbf{A}_{(1)}$ and a similar property holds for the other two modes. For high noise levels this property is lost and the approximation problem becomes more difficult.

Another interesting result is that in the above experiment the difference between the cost function values at the different local minima that were found seems to be small. This is good news for applications where the best low multilinear rank approximation is used as a compression tool, since there different local minima would lead to similar compression rates. On the other hand, the column spaces of two matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ from (2) corresponding to two different local minima are very different and the same holds for $\mathbf{V}$ and $\mathbf{W}$ [7]. This difference may have important consequences for applications where the actual subspaces of the matrices are of interest.

Finally, we mention that different algorithms could converge to different local minima, even if the algorithms are initialized in the same way. This fact could be exploited in order to find a larger set of local minima. If the global minimum is required or if some properties of the desired solution are known, all found solutions could be examined in order to find the most suitable one.

IV. PARTICLE SWARM OPTIMIZATION

Instead of performing a number of independent runs, the landscape of the cost function could also be explored using a stochastic population-based technique. Searching for the global minimum of the best low multilinear rank approximation problem, an algorithm based on (guaranteed convergence) particle swarm optimization ((GC)PSO) [8], [12] can be considered. Several points, called particles, are initialized randomly and evolve in the search space following simple rules. Each point is attracted by its individual best position, the global best position of all particles and also tends to continue along its latest direction. The behavior of the points resemble the behaviour of social groups. Convergence to a stationary point is guaranteed with a slight modification of the algorithm [12] concerning the update of the best particle at each iteration. In order to improve the convergence speed, a gradient component could also be taken into account. An adaptation of GCPSO to the best low multilinear rank approximation problem is presented in [2]. Note that the optimization problem takes place on a product space of three Grassmann manifolds and not just in $\mathbb{R}^n$.

Some preliminary results for low multilinear rank tensors affected by additive noise are given in [2]. The algorithm seems to find the desired local minimum for low noise levels. For high noise levels, the landscape of the cost function becomes too intricate and the current version of the algorithm often converges to another stationary point.

Finally, we mention that the PSO algorithm also has niching capabilities that could be explored in order to find different local minima. This will be discussed in future work.

V. CONCLUSIONS

The problem considered in this paper is finding the best low multilinear rank approximation of a higher-order tensor. The first iterative algorithm was proposed in 1980 [9] and a number of algorithms have appeared in the literature since then. However, all of them search for any (local) minimum of the problem and the provided solution is not further analyzed. There are applications where any local minimum could be used without losing much precision. However, the solutions are in general essentially different from each other to the extent that in certain types of applications taking just any minimum may lead to false results.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, January 2008.

[2] P. B. Borckmans, M. Ishteva, and P.-A. Absil. A modified particle swarm optimization algorithm for the best low multilinear rank approximation of higher-order tensors. In *7th International Conference on Swarm Intelligence (ANTS 2010), Brussels, Belgium*, 2010. Accepted.

[3] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278, April 2000.

[4] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, April 2000.

[5] F. L. Hitchcock. Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematical Physics*, 7(1):39–79, 1927.

[6] M. Ishteva. *Numerical methods for the best low multilinear rank approximation of higher-order tensors*. PhD thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, December 2009.

[7] M. Ishteva, P.-A. Absil, S. Van Huffel, and L. De Lathauwer. Tucker compression and local optima. Technical report, UCL-INMA-2010.012, Université catholique de Louvain and ESAT-SISTA-09-247, K.U.Leuven, Belgium, 2010.

[8] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks, 1995*, volume 4, pages 1942–1948, 1995.

[9] P. M. Kroonenberg and J. de Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1):69–97, 1980.

[10] L. R. Tucker. The extension of factor analysis to three-dimensional matrices. In H. Gulliksen and N. Frederiksen, editors, *Contributions to mathematical psychology*, pages 109–127. Holt, Rinehart & Winston, NY, 1964.

[11] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[12] F. van den Bergh and A. P. Engelbrecht. A new locally convergent particle swarm optimiser. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 96–101, 2002.