The best rank-(R_1, R_2, R_3) approximation of tensors by means of a geometric Newton method

Mariya Ishteva[*], Lieven De Lathauwer[*,†], P.-A. Absil[**] and Sabine Van Huffel[*]

[*] *ESAT/SCD, Katholieke Universiteit Leuven, Belgium*
[†] *Subfaculty Science and Technology, Katholieke Universiteit Leuven Campus Kortrijk, Belgium*
[**] *INMA, Université catholique de Louvain, Belgium*

Abstract. In the matrix case, the best low-rank approximation can be obtained directly from the truncated singular value decomposition. However, the best rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors cannot be computed in a straightforward way. An additional difficulty comes from an invariance property of the cost function by the action of the orthogonal group. This means that the zeros of the cost function are not isolated and the plain Newton method might have difficulties. In this paper, we derive a geometric Newton method using the theory of quotient manifolds. Our algorithm has fast convergence in a neighborhood of the solution and is useful for a large class of problems. We mention some applications.

Keywords: higher-order tensor, rank-$(R_1, R_2, \ldots, R_N)$ reduction, quotient manifold, differential-geometric optimization
PACS: 02.10.Xm, 02.40.Sf, 02.40.Vh

INTRODUCTION

Higher-order tensors are generalizations of vectors (order 1) and matrices (order 2) to order 3 or higher. Their columns are called mode-1 vectors and their rows are called mode-2 vectors. In general, the mode-*n* vectors $(n = 1, 2, \ldots, N)$ are the vectors, obtained by varying the *n*-th index of a *N*-th order tensor, while keeping the other indices fixed. A generalization of the column and row ranks of a matrix is the mode-*n* rank. It is given by the maximal number of linearly independent mode-*n* vectors. Different mode-*n* ranks are not necessarily equal to each other unlike the column and row ranks of a matrix. The theory of higher-order tensors is called multilinear algebra.

The best rank-$(R_1, R_2, \ldots, R_N)$ approximation of a tensor [1, 2] is another tensor, as close as possible to the original one and having mode-*n* ranks bounded by $R_1, R_2, \ldots, R_N$, respectively. It can be used as a tool for dimensionality reduction. Independent component analysis applications, such as electro-encephalography, magneto-encephalography, and nuclear magnetic resonance, often involve high-dimensional data in which only a few sources have significant contributions [3]. Dimensionality reduction based on the best rank-$(R_1, R_2, \ldots, R_N)$ approximation of tensors can be applied in order to reduce the dimensionality of the problem from the number of observation channels to the number of sources [3]. It is also useful to reduce the dimensionality before computation of the parallel factor decomposition (PARAFAC) [4], also called canonical decomposition (CANDECOMP) [5]. This is a decomposition of higher-order tensors in rank-1 terms. It is widely used in chemometrics [6] and wireless communication [7, 8]. Epileptic seizure onset localisation [9, 10, 11] can also make use of PARAFAC since only one of the rank-1 components is related to the seizure activity. Other applications of the best rank-$(R_1, R_2, \ldots, R_N)$ approximation of tensors can be found in image synthesis, analysis and recognition [12], signal processing applications [13, 14], scientific computing [15], and fuzzy control [16].

In this paper, we propose a Newton algorithm for the best rank-(R_1, R_2, R_3) approximation of a tensor. It makes use of differential-geometric techniques for quotient manifolds and takes into account the structure of the cost function. It is based on the differential-geometric Newton method for Oja's vector field [17]. In the literature, an alternating least-squares algorithm has been proposed in [18]. Recently, a Newton-Grassmann method has been suggested [19]. In [20], quasi-Newton methods on a product of Grassmannians have been developed. General theory on optimization on manifolds is discussed in [21, 22]. Finally, specific algorithms for the best rank-1 approximation have been discussed in [18, 23]. Here, for simplicity, we only consider third-order tensors. The generalization to tensors of order higher than three is straightforward.

PROBLEM FORMULATION

The problem of computing the best rank-(R_1,R_2,R_3) approximation of a third-order tensor $A \in \mathbb{R}^{I_1\times I_2\times I_3}$ can be formulated as follows. Find $\hat{A} \in \mathbb{R}^{I_1\times I_2\times I_3}$ such that it minimizes the least-squares cost function $f : \mathbb{R}^{I_1\times I_2\times I_3} \to \mathbb{R}$,

$$f : \hat{A} \mapsto \|A - \hat{A}\|^2 \tag{1}$$

under the constraints $\text{rank}_1(\hat{A}) \leq R_1$, $\text{rank}_2(\hat{A}) \leq R_2, \text{rank}_3(\hat{A}) \leq R_3$, where $\text{rank}_n(.)$ stands for the mode-n rank and $\|.\|$ is the Frobenius norm. The truncated singular value decomposition (SVD) [24] gives the best low-rank approximation of a matrix. However, truncation of the higher-order equivalent of the SVD (HOSVD) [25, 26, 27] usually results in a suboptimal tensor rank-(R_1,R_2,R_3) approximation. This approximation can be refined by iterative algorithms.

The problem of minimizing (1) is equivalent to maximizing

$$g : (\mathbf{U}, \mathbf{V}, \mathbf{W}) \mapsto \|A \bullet_1 \mathbf{U}^T \bullet_2 \mathbf{V}^T \bullet_3 \mathbf{W}^T\|^2 , \tag{2}$$

over the orthonormal matrices $\mathbf{U} \in \mathbb{R}^{I_1\times R_1}, \mathbf{V} \in \mathbb{R}^{I_2\times R_2}, \mathbf{W} \in \mathbb{R}^{I_3\times R_3}$ [18, 28, 29]. Here, $A \bullet_n \mathbf{M}$, $n = 1,2,3$ is the product of a tensor A and a matrix $\mathbf{M}$ with respect to the n-th mode of the tensor [25]. It is also possible to rewrite g as

$$g(\mathbf{U},\mathbf{V},\mathbf{W}) = \|A \bullet_1 \mathbf{U}^T \bullet_2 \mathbf{V}^T \bullet_3 \mathbf{W}^T\|^2 = \|\mathbf{U}^T (\mathbf{A}_{(1)}(\mathbf{V}\otimes\mathbf{W}))\|^2 ,$$

where $\mathbf{A}_{(1)}$ is a matrix representation of A, obtained by putting its columns one after the other in a specific order [25]. The symbol "$\otimes$" denotes the Kronecker product. Having the optimal $\mathbf{U},\mathbf{V},\mathbf{W}$, the tensor $\hat{A}$ is computed by [18]

$$\hat{A} = B \bullet_1 \mathbf{U} \bullet_2 \mathbf{V} \bullet_3 \mathbf{W}, \quad \text{where } B = A \bullet_1 \mathbf{U}^T \bullet_2 \mathbf{V}^T \bullet_3 \mathbf{W}^T \in \mathbb{R}^{R_1\times R_2\times R_3}.$$

GEOMETRIC NEWTON METHOD

In this section, we formulate the solution of the optimization problem (2) as a zero of a function F. The function F has a symmetry property by the action of the orthogonal group. This causes difficulties when the usual Newton method is applied for computing the zeros of F (see [17]). Based on [17], we propose a remedy for that difficulty in the form of a geometric Newton method. It removes the symmetry by working on a quotient space.

In order to maximize (2), the column space of $\mathbf{U}$ has to be the same as the span of the R_1 dominant left singular vectors of the matrix $\mathbf{A}_{(1)}(\mathbf{V}\otimes\mathbf{W})$. It is then necessary that the column space of $\mathbf{U}$ is an invariant subspace of $\mathbf{A}_{(1)}(\mathbf{V}\otimes\mathbf{W})(\mathbf{V}\otimes\mathbf{W})^T\mathbf{A}_{(1)}^T$, i.e.,

$$\mathbf{U}\,\mathbf{S}_1 = \mathbf{A}_{(1)}(\mathbf{V}\otimes\mathbf{W})(\mathbf{V}\otimes\mathbf{W})^T\mathbf{A}_{(1)}^T\mathbf{U} \tag{3}$$

for some $\mathbf{S}_1 \in \mathbb{R}^{R_1\times R_1}$. Similar equations can be derived for $\mathbf{V}$ and $\mathbf{W}$. Let $\mathbf{X} = \{\mathbf{U},\mathbf{V},\mathbf{W}\}$. We define

$$F(\mathbf{X}) = (F_1(\mathbf{X}), F_2(\mathbf{X}), F_3(\mathbf{X})), \tag{4}$$

where

$$\begin{aligned}
F_1(\mathbf{X}) &= \mathbf{U}\,\mathbf{R}_1(\mathbf{X})\mathbf{R}_1(\mathbf{X})^T - \mathbf{A}_{(1)}(\mathbf{V}\otimes\mathbf{W})(\mathbf{V}\otimes\mathbf{W})^T\mathbf{A}_{(1)}^T\mathbf{U}, \quad \mathbf{R}_1(\mathbf{X}) &= \mathbf{U}^T\mathbf{A}_{(1)}(\mathbf{V}\otimes\mathbf{W}),\\
F_2(\mathbf{X}) &= \mathbf{V}\,\mathbf{R}_2(\mathbf{X})\mathbf{R}_2(\mathbf{X})^T - \mathbf{A}_{(2)}(\mathbf{W}\otimes\mathbf{U})(\mathbf{W}\otimes\mathbf{U})^T\mathbf{A}_{(2)}^T\mathbf{V}, \quad \mathbf{R}_2(\mathbf{X}) &= \mathbf{V}^T\mathbf{A}_{(2)}(\mathbf{W}\otimes\mathbf{U}),\\
F_3(\mathbf{X}) &= \mathbf{W}\mathbf{R}_3(\mathbf{X})\mathbf{R}_3(\mathbf{X})^T - \mathbf{A}_{(3)}(\mathbf{U}\otimes\mathbf{V})(\mathbf{U}\otimes\mathbf{V})^T\mathbf{A}_{(3)}^T\mathbf{W}, \quad \mathbf{R}_3(\mathbf{X}) &= \mathbf{W}^T\mathbf{A}_{(3)}(\mathbf{U}\otimes\mathbf{V}).
\end{aligned}$$

Thus, (3) as well as the corresponding equations for $\mathbf{V}$ and $\mathbf{W}$ follow, if we find $\mathbf{X} = \{\mathbf{U},\mathbf{V},\mathbf{W}\}$ such that

$$F(\mathbf{X}) = 0.$$

Note that the zeros of F are not isolated because of the following invariance properties

$$F_i(\mathbf{X}\mathbf{Q}) = F_i(\mathbf{X})\,\mathbf{Q}_i,$$

where $\mathbf{X}\mathbf{Q} = \{\mathbf{U}\mathbf{Q}_1, \mathbf{V}\mathbf{Q}_2, \mathbf{W}\mathbf{Q}_3\}$ and $\mathbf{Q}_i \in O_{R_i}, i = 1,2,3$ are orthogonal matrices. Thus,

$$F(\mathbf{X}) = 0 \quad \Longleftrightarrow \quad F(\mathbf{X}\mathbf{Q}) = 0.$$

This is why a Newton method will most probably perform poorly (see, for example, [22, Prop. 2.1.2] or [17]). This problem can be handled by a differential-geometric approach as in [17]. To this end, we consider an equivalence relation $\sim$ on $\mathbb{R}_*^{n\times p}$, defined by

$$\mathbf{Y}\sim\mathbf{Y}_1 \quad\Longleftrightarrow\quad \text{there exists } \mathbf{Q}\in O_p, \text{ s.t. } \mathbf{Y}_1=\mathbf{YQ},$$

where O_p is the orthogonal group (the set of all orthogonal $p\times p$ matrices) and $\mathbb{R}_*^{n\times p}$ is the set of all full-rank $n\times p$ matrices. It can be proved that the quotient $\mathbb{R}_*^{n\times p}/O_p$ is a quotient manifold. It is described in [17]. The general theory on Newton's method on manifolds can be found in [22], in a form that is based on [30]. Here we only mention the properties that we will use later on. First, the vertical space at point $\mathbf{Y}$ is given by the following formula

$$V_{\mathbf{Y}}=\{\mathbf{Y}\Omega : \Omega=-\Omega^T\in\mathbb{R}^{p\times p}\}.$$

The direct sum of the horizontal space $H_{\mathbf{Y}}$ and the vertical space $V_{\mathbf{Y}}$ yields $\mathbb{R}^{n\times p}$. Thus, we can take

$$H_{\mathbf{Y}}=\{\mathbf{YS}+\mathbf{Y}^{\perp}\mathbf{K} : \mathbf{S}=\mathbf{S}^T\in\mathbb{R}^{p\times p}, \mathbf{K}\in\mathbb{R}^{(n-p)\times p}, \mathbf{Y}^{\perp}\in\mathbb{R}_*^{n\times(n-p)}, \mathbf{Y}^T\mathbf{Y}=0\}.$$

Finally, the projection onto $H_{\mathbf{Y}}$ along $V_{\mathbf{Y}}$ is computed as

$$P_{\mathbf{Y}}^h\mathbf{Z} \quad=\quad \mathbf{Z}-\mathbf{Y}\,\mathrm{skew}((\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{Z}), \quad \text{where} \quad \mathrm{skew}(\mathbf{B})=(\mathbf{B}-\mathbf{B}^T)/2.$$

In [17], the following Newton equation has been considered

$$\left|\begin{array}{l} P_{\mathbf{X}}^h\,\mathrm{D}(P_{\mathbf{X}}^hF)(\mathbf{X})[\Delta]=-P_{\mathbf{X}}^hF(\mathbf{X}), \quad \Delta\in H_{\mathbf{X}} \\ \mathbf{X}_+=\mathbf{X}+\Delta, \end{array}\right.$$

where $\mathrm{D}(P_{\mathbf{X}}^hF)(\mathbf{X})[\Delta]$ is the derivative of $P_{\mathbf{X}}^hF(\mathbf{X})$ along Δ and $\mathbf{X}_+$ stands for the next iterate. In the case of function (4), we need to work on the product manifold

$$\mathbb{R}_*^{I_1\times R_1}/O_{R_1}\times\mathbb{R}_*^{I_2\times R_2}/O_{R_2}\times\mathbb{R}_*^{I_3\times R_3}/O_{R_3}$$

instead of the manifold $\mathbb{R}_*^{n\times p}/O_p$. This is a nontrivial generalization of the algorithm proposed in [17], which involves more difficult computations. Its theoretical part can be summarized as follows.

The **geometric Newton method**:

1. Find initial orthonormal matrices $\mathbf{U}_0\in\mathbb{R}^{I_1\times R_1}$, $\mathbf{V}_0\in\mathbb{R}^{I_2\times R_2}$, and $\mathbf{W}_0\in\mathbb{R}^{I_3,\times R_3}$, e.g., from the truncated HOSVD, and set $\mathbf{X}_0=\{\mathbf{U}_0,\mathbf{V}_0,\mathbf{W}_0\}$.
2. For $k=0,1,2,\ldots$, solve the linear system of equations:

$$\left|\begin{array}{l} P_{\mathbf{U}}^h\,\mathrm{D}(P_{\mathbf{U}}^hF_1)(\mathbf{X}_k)[\Delta_k]=-P_{\mathbf{U}}^hF_1(\mathbf{X}_k) \\ P_{\mathbf{V}}^h\,\mathrm{D}(P_{\mathbf{V}}^hF_2)(\mathbf{X}_k)[\Delta_k]=-P_{\mathbf{V}}^hF_2(\mathbf{X}_k), \quad \Delta_k=\{\Delta_{k1},\Delta_{k2},\Delta_{k3}\}\in H_{\mathbf{U}}\times H_{\mathbf{V}}\times H_{\mathbf{W}}. \\ P_{\mathbf{W}}^h\,\mathrm{D}(P_{\mathbf{W}}^hF_3)(\mathbf{X}_k)[\Delta_k]=-P_{\mathbf{W}}^hF_3(\mathbf{X}_k) \\ \mathbf{X}_{k+1}=\mathbf{X}_k+\Delta_k \end{array}\right. \tag{5}$$

3. With the converged triplet $\mathbf{X}=\{\mathbf{U},\mathbf{V},\mathbf{W}\}$ from step 2, compute

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

Notice that (5) is a linear system of equations in the unknown Δ_k. It can be solved by using Matlab's GMRES solver. Numerical experiments performed in this way have shown quadratic convergence of the algorithm in the neighborhood of the nondegenerate zeros of $P_{\mathbf{X}}^hF(\mathbf{X})$.

CONCLUSIONS

This paper presents a geometric Newton method for the best rank-(R_1,R_2,R_3) approximation of a higher-order tensor. It is based on quotient manifold techniques. An advantage of the algorithm is its fast convergence in the neighborhood of the solution. Although global convergence can not be guaranteed, starting from a good initial point can improve the chances of converging to a solution. More details can be found in the full paper [31].

ACKNOWLEDGMENTS

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. Research supported by: (1) Research Council K.U.Leuven: GOA-Ambiorics, CoE EF/05/006 Optimization in Engineering (OPTEC), (2) F.W.O.: (a) project G.0321.06, (b) Research Communities ICCoS, ANMMM and MLDM, (3) the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007–2011), (4) EU: ERNSI. M. Ishteva is supported by a K.U.Leuven doctoral scholarship (OE/06/25, OE/07/17, OE/08/007), L. De Lathauwer is supported by "Impulsfinanciering Campus Kortrijk (2007-2012)(CIF1)".

REFERENCES

1. F. L. Hitchcock, *Journal of Mathematical Physics* **6**, 164–189 (1927).
2. F. L. Hitchcock, *Journal of Mathematical Physics* **7**, 39–79 (1927).
3. L. De Lathauwer, and J. Vandewalle, *Linear Algebra Appl.* **391**, 31–55 (2004), special Issue on Linear Algebra in Signal and Image Processing.
4. R. A. Harshman, *UCLA Working Papers in Phonetics* **16**, 1–84 (1970).
5. J. Carroll, and J. Chang, *Psychometrika* **35**, 283–319 (1970).
6. A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the Chemical Sciences*, John Wiley and Sons, Chichester, U.K., 2004.
7. N. Sidiropoulos, R. Bro, and G. Giannakis, *IEEE Trans. Signal Process.* **48**, 2377–2388 (2000).
8. L. De Lathauwer, and J. Castaing, *Signal Processing* **87**, 322–336 (2007), special Issue on Tensor Signal Processing.
9. M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini, and W. Van Paesschen, *NeuroImage* **37**, 844–854 (2007).
10. M. De Vos, L. De Lathauwer, B. Vanrumste, S. Van Huffel, and W. Van Paesschen, *Journal of Computational Intelligence and Neuroscience* **2007**, 1–10 (2007), special Issue on EEG/MEG Signal Processing.
11. E. Acar, C. A. Bingol, H. Bingol, R. Bro, and B. Yener, *ISMB 2007 Conference Proceedings, Bioinformatics* **23**, i10–i18 (2007).
12. M. A. O. Vasilescu, and D. Terzopoulos, "Multilinear Subspace Analysis for Image Ensembles," in *Proc. Computer Vision and Pattern Recognition Conf. (CVPR '03), Madison, WI*, 2003, vol. 2, pp. 93–99.
13. J.-M. Papy, L. De Lathauwer, and S. Van Huffel, *Numer. Linear Algebra Appl.* **12**, 809–826 (2005).
14. J.-M. Papy, L. De Lathauwer, and S. Van Huffel, Exponential data fitting using multilinear algebra: The decimative case, Technical report 04-70, ESAT-SISTA, K.U.Leuven (2004), submitted.
15. B. N. Khoromskij, and V. Khoromskaia, *Central European Journal of Mathematics* **5**, 523–550 (2007).
16. P. Baranyi, A. R. Varkonyi-Koczy, Y. Yam, and R. J. Patton, *IEEE Trans. Instrumentation and Measurement* **54**, 52–60 (2005).
17. P.-A. Absil, M. Ishteva, L. De Lathauwer, and S. Van Huffel, A geometric Newton method for Oja's vector field, Tech. Rep. UCL-INMA-2008.013, Université catholique de Louvain, Département d'ingénierie mathématique, Av. G. Lemaître 4, 1348 Louvain-la-Neuve, Belgium (2008), arXiv:0804.0989.
18. L. De Lathauwer, B. De Moor, and J. Vandewalle, *SIAM J. Matrix Anal. Appl.* **21**, 1324–1342 (2000).
19. L. Eldén, and B. Savas, A Newton–Grassmann method for computing the best multi-linear rank-(r_1, r_2, r_3) approximation of a tensor, Tech. Rep. LITH-MAT-R-2007-6-SE, Department of Mathematics, Linköpings Universitet (2007).
20. B. Savas, and L.-H. Lim, Best multilinear rank approximation of tensors with quasi-Newton methods on Grassmannians, Tech. Rep. LITH-MAT-R-2008-01-SE, Department of Mathematics, Linköpings Universitet (2008).
21. A. Edelman, T. A. Arias, and S. T. Smith, *SIAM J. Matrix Anal. Appl.* **20**, 303–353 (1998).
22. P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.
23. T. Zhang, and G. H. Golub, *SIAM J. Matrix Anal. Appl.* **23**, 534–550 (2001).
24. G. H. Golub, and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 1996, 3rd edn.
25. L. De Lathauwer, B. De Moor, and J. Vandewalle, *SIAM J. Matrix Anal. Appl.* **21**, 1253–1278 (2000).
26. L. R. Tucker, "The extension of factor analysis to three-dimensional matrices," in *Contributions to mathematical psychology*, edited by H. Gulliksen, and N. Frederiksen, Holt, Rinehart & Winston, NY, 1964, pp. 109–127.
27. L. R. Tucker, *Psychometrika* **31**, 279–311 (1966).
28. P. M. Kroonenberg, *Applied Multiway Data Analysis*, Wiley, 2008.
29. P. M. Kroonenberg, and J. de Leeuw, *Psychometrika* **45**, 69–97 (1980).
30. R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, *IMA J. Numer. Anal.* **22**, 359–390 (2002).
31. M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel, Differential-geometric Newton method for the best rank-(R_1, R_2, R_3) approximation of tensors, Internal Report 08-111, ESAT-SISTA, K.U.Leuven (Leuven, Belgium) (2008).