Optimal Matrix-Mimetic Tensor Algebras

<u>Elizabeth Newman</u>, Katherine Keegan

Abstract

With the explosion of big data, the need for explainable data analysis tools, efficient representations, and structure-exploiting operations has exploded as well. Many data and operators are naturally multiway, and as a result, multilinear or tensor methods have revolutionized the interpretability of feature extraction, the compressibility of large-scale data, and the computational efficiency of multiway operations. Despite numerous successes, many tensor frameworks suffer from a so-called "curse of multidimensionality;" that is, that fundamental linear algebra properties break down in higher dimensions, particularly the notion of optimality. Recent advances in matrix-mimetic tensor frameworks have made it possible to preserve linear algebraic properties for multilinear analysis and, as a result, obtain optimal representations of multiway data.

Matrix mimeticity arises from interpreting tensors as operators that can be multiplied, factorized, and analyzed analogously to matrices. Underlying the tensor operation is an algebraic framework parameterized by an invertible linear transformation. Specifically, consider a third-order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$; i.e., a multiway arrays with rows, columns, and depth indices. We can view \mathcal{A} as an $n_1 \times n_2$ matrix where each entry is a $1 \times 1 \times n_3$ tube. We multiply tubes $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{1 \times 1 \times n_3}$ using the $\star_{\mathbf{M}}$ -product [5] (the prefix is pronounced "star-M") via

$$\mathbf{a} \star_{\mathbf{M}} \mathbf{b} = \operatorname{vec}^{-1}(\mathbf{R}_{\mathbf{M}}[\mathbf{a}]\operatorname{vec}(\mathbf{b})) \quad \text{where} \quad \mathbf{R}_{\mathbf{M}}[\mathbf{a}] = \mathbf{M}^{-1}\operatorname{diag}(\mathbf{M}\operatorname{vec}(\mathbf{a}))\mathbf{M}, \quad (1)$$

vec : $\mathbb{R}^{1 \times 1 \times n_3} \to \mathbb{R}^n_3$ is a bijective map that vectorizes tubes and diag : $\mathbb{R}^{n_3} \to \mathbb{R}^{n_3 \times n_3}$ forms a diagonal matrix from the entries of a vector. We say the action **a** on **b** under the $\star_{\mathbf{M}}$ -product is equivalent to left multiplication by the structured matrix $\mathbf{R}_{\mathbf{M}}[\mathbf{a}]$. A given invertible matrix **M** thereby induces a matrix subalgebra that equips the vector space of tubes with a bilinear operation given by $\mathbf{R}_{\mathbf{M}}[\cdot]$; the term *tensor algebra* refers to this operation.

We define tensor-tensor products analogously to matrix-matrix products by replacing scalar with tubal multiplication given by (1). Using MATLAB indexing notation, the tubal entrywise definition of the tensor-tensor product of $\mathcal{A} \in \mathbb{R}^{n_1 \times m \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{m \times n_2 \times n_3}$ is

$$(\mathcal{A} \star_{\mathbf{M}} \mathcal{B})_{i_1, i_2,:} = \sum_{k=1}^{m} \mathcal{A}_{i_1, k,:} \star_{\mathbf{M}} \mathcal{B}_{k, i_2,:}$$
(2)

for $i_1 = 1, \ldots, n_1$ and $i_2 = 1, \ldots, n_2$. Under the algebraically-consistent $\star_{\mathbf{M}}$ -product, we obtain matrix-mimetic generalizations of $\star_{\mathbf{M}}$ -rank, -orthogonality, -transposition, more [6]. Notably, we can define a tensor singular value decomposition that satisfies an Eckart-Young-like theorem, resulting in optimal, low-rank approximations of multiway data [7].

The choice of linear mapping \mathbf{M} and associated tensor algebra is crucial to approximation quality. Traditionally, \mathbf{M} is chosen heuristically to leverage expected correlations in the data. However, in many cases, these correlations are unknown and common heuristic mappings lead to suboptimal performance. This presentation, based on the work in [8], introduces $\star_{\mathbf{M}}$ -optimization, an algorithm to learn optimal linear transformations and corresponding optimal tensor representations (e.g., low- $\star_{\mathbf{M}}$ -rank) simultaneously. The new framework explicitly captures the coupling between the transformation and representation by solving the bilevel optimization problem

$$\min_{\mathbf{M}\in\mathcal{O}_{n_3}} \Phi(\mathbf{M}, \boldsymbol{\mathcal{X}}(\mathbf{M})) \qquad \text{s.t.} \qquad \boldsymbol{\mathcal{X}}(\mathbf{M}) \in \arg\min_{\boldsymbol{\mathcal{X}}\in\mathcal{X}} \Phi(\mathbf{M}, \boldsymbol{\mathcal{X}}). \tag{3}$$

Here, \mathcal{X} is the desired representation belonging to feasible set \mathcal{X} , and $\mathcal{X}(\mathbf{M})$ is an optimal representation for a given transformation, \mathbf{M} . Our goal is to learn an invertible \mathbf{M} , which we guarantee by optimizing over the orthogonal group of $n_3 \times n_3$ matrices, \mathcal{O}_{n_3} . The objective function $\Phi : \mathcal{O}_{n_3} \times \mathcal{X} \to \mathbb{R}$ measures the quality of the representation. We solve (3) for \mathbf{M} using Riemannian optimization over the orthogonal group [2, 1, 3].

A key innovation of $\star_{\mathbf{M}}$ -optimization is the use of variable projection to form $\mathcal{X}(\mathbf{M})$, which eliminates the variable \mathcal{X} via partial optimization [4]. We heavily leverage the optimality of $\star_{\mathbf{M}}$ representations to guarantee the existence of an optimal $\mathcal{X}(\mathbf{M})$; other comparable tensor approaches typically only guarantee quasi-optimality.

In the talk, we will highlight the generality of the $\star_{\mathbf{M}}$ -optimization framework by considering two prototype problems for fitting tensor data and for finding compressed representations. We will present new theoretical results regarding the uniqueness and invariances of the $\star_{\mathbf{M}}$ -operator and convergence guarantees of $\star_{\mathbf{M}}$ -optimization. We will demonstrate the efficacy of learning the transformation and provide interpretable insight into $\star_{\mathbf{M}}$ -optimization behavior through several numerical examples, including image compression and reduced order modeling.

References

- P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, NJ, 2008.
- [2] N. Boumal, An introduction to optimization on smooth manifolds, Cambridge University Press, 2023, https://doi.org/10.1017/9781009166164.
- [3] A. Edelman, T. A. Arias, and S. T. Smith, The geometry of algorithms with orthogonality constraints, SIAM Journal on Matrix Analysis and Applications, 20 (1998), pp. 303–353, https://doi.org/10.1137/S0895479895290954.
- [4] G. H. Golub and V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, SIAM Journal on Numerical Analysis, 10 (1973), pp. 413– 432, https://doi.org/10.1137/0710036.
- [5] E. Kernfeld, M. Kilmer, and S. Aeron, Tensor-tensor products with invertible linear transforms, Linear Algebra and its Applications, 485 (2015), pp. 545–570, https://doi.org/10.1016/j.laa.2015.07.021.
- [6] Misha E. Kilmer, Karen Braman, Ning Hao, and Randy C. Hoover. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. SIAM Journal on Matrix Analysis and Applications, 34(1):148–172, 2013, https://doi.org/10.1137/110837711.
- [7] M. E. Kilmer, L. Horesh, H. Avron, and E. Newman, Tensor-tensor algebra for optimal representation and compression of multiway data, Proceedings of the National Academy of Sciences of the United States of America, 118 (2021), https://doi.org/10.1073/pnas. 2015851118.
- [8] Elizabeth Newman and Katherine Keegan. Optimal matrix-mimetic tensor algebras via variable projection, 2024, https://arxiv.org/abs/2406.06942.