Thesis abstract: Tensor-based pattern recognition, data analysis and learning*

S. Hendrikx[1,2], N. Vervliet[1,2], M. Boussé[1,2], and L. De Lathauwer[1,2]

[1] Department of Electrical Engineering (ESAT), KU Leuven, Belgium.
[2] Group Science, Engineering and Technology, KU Leuven Kulak, Belgium.

The aim of this thesis is to develop a scalable algorithm for multilinear regression [1]. Multilinear regression resides between linear and nonlinear regression models, such as neural networks (NN), which are widely used machine learning tools. While linear regression is simple and interpretable, it is less capable of modeling complex phenomena than its nonlinear counterpart. Multilinear regression is a trade-off between both, resulting in an expressive model with more meaningful variables than in NNs, as it is based on multilinear algebra. Unfortunately, the number of coefficients of a multivariate polynomial depends exponentially on its degree. In this thesis, we employ a low-rank tensor decomposition to break this exponential dependency, known as the curse of dimensionality (CoD), allowing us to develop a scalable, optimization-based algorithm.

Tensors, or multiway arrays, are higher-order generalizations of vectors and matrices. Tensor analogues of established matrix decompositions are powerful tools in signal processing, data analysis, and machine learning [2]. The main advantage of using tensor decompositions within this work stems from their ability to break the CoD. The polyadic decomposition (PD) decomposes an Nth-order tensor $\mathcal{T}$ as a sum of rank-1 tensors, where a rank-1 tensor is equal to the outer product, denoted by $\otimes$, of N nonzero vectors:

$$\mathcal{T} = \sum_{r=1}^{R} c_r \mathbf{b}_r^{(1)} \otimes \cdots \otimes \mathbf{b}_r^{(N)} \stackrel{\text{def}}{=} \left[\!\!\left[\mathbf{c}; \mathbf{B}^{(1)}, \cdots, \mathbf{B}^{(N)} \right]\!\!\right]. \tag{1}$$

An Nth-order tensor of size $I \times I \times \cdots \times I$ suffers from the CoD since it contains I^N entries. By approximating such a tensor with a low-rank PD, i.e., a PD with low R, that only has NRI parameters, this curse is broken.

An Nth-degree homogeneous polynomial $p(\mathbf{x})$ with variables $\mathbf{x} \in \mathbb{R}^I$ can be expressed by means of a symmetric[3] tensor of order N and the mode-n product[4]:

$$p(\mathbf{x}) = \mathcal{T} \cdot_1 \mathbf{x}^{\mathrm{T}} \cdot_2 \mathbf{x}^{\mathrm{T}} \cdots \cdot_N \mathbf{x}^{\mathrm{T}}. \tag{2}$$

* This research received funding from the Flemish Government (AI Research Program). This work was supported by the Fonds de la Recherche Scientifique – FNRS and the Fonds Wetenschappelijk Onderzoek – Vlaanderen under EOS Project no 30468160 (SeLMA). KU Leuven Internal Funds: C16/15/059 and IDN/19/014. Nico Vervliet is supported by a Junior postdoctoral fellowship (12ZM220N) from the Research Foundation—Flanders (FWO).

[3] A symmetric tensor is invariant to every possible permutation of its N dimensions.

[4] The mode-n product $\cdot_n$ of $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and $\mathbf{B} \in \mathbb{R}^{J \times I_n}$ is defined as $(\mathcal{A} \cdot_n \mathbf{B})_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} a_{i1 \ldots i_N} b_{j i_n}$.

For example, a degree-two homogenous polynomial can be written as $p(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{T}\mathbf{x} = \mathbf{T} \cdot_1 \mathbf{x}^{\mathrm{T}} \cdot_2 \mathbf{x}^{\mathrm{T}}$. We assume that $\mathcal{T}$ has approximately low rank. This makes sense as real-life data can often be modelled using parsimonious representations due to some inherent structure. For example, measurements of a physical property that are governed by underlying differential equations or the dataset of a recommender system that contains users which behave in similar patterns. Compact models such as low-rank matrix and tensor decompositions are often used for large-scale problems in scientific computing and compressed sensing [3].

Thanks to this low-rank assumption, $\mathcal{T}$ in Eq. (2) can be replaced by a low-rank PD. The symmetry of $\mathcal{T}$ is exploited to obtain an even more compact model by using a symmetric PD, i.e., $\mathbf{B}^{(n)} = \mathbf{B}$ for $1 \leq n \leq N$, which requires fewer parameters than a general PD, namely RI. Therefore this model avoids the CoD since a multivariate polynomial generally has $\frac{(I+N-1)!}{N!(I-1)!}$ coefficients.

In order to derive an optimization-based algorithm, we fit the PD-constrained regression model to a dataset $\mathbf{X} \in \mathbb{R}^{I \times M}, \mathbf{y} \in \mathbb{R}^{M}$ as the following set of linear equations with a structured solution in a compressed-sensing style approach [4]:

$$\mathbf{y} \approx \left(\mathbf{X} \odot^{\mathrm{T}} \mathbf{X} \odot^{\mathrm{T}} \cdots \odot^{\mathrm{T}} \mathbf{X}\right) \mathrm{vec}\left(\llbracket \mathbf{c}; \mathbf{B}, \mathbf{B}, \cdots, \mathbf{B} \rrbracket\right), \tag{3}$$

where $\odot^{\mathrm{T}}$ denotes the row-wise Khatri–Rao product.

To compute the model variables $\mathbf{B}$ and $\mathbf{c}$, we use a Gauss–Newton (GN) algorithm with dogleg trust region to minimize the cost function $\frac{1}{2}\,||\mathbf{r}||_2^2$ in which $\mathbf{r}$ equals the difference between the left and right hand side of Eq. (3). By simultaneously exploiting both the Khatri–Rao and PD structure in Eq. (3) in the derivation of the cost function, gradient, Jacobian and Gramian for the GN algorithm, we obtain a scalable algorithm. Indeed, the overall per-iteration complexity of the algorithm is $\mathcal{O}(MR^2I^2)$, in contrast to $\mathcal{O}\left(MRNI^N\right)$ for a naive algorithm that does not exploit all structure.

To conclude, we have formulated a scalable optimization-based algorithm for multilinear regression through the use of a low-rank symmetric PD. In [5], we demonstrate high accuracy of our model on a materials science dataset.

References

1. Hendrikx, S. Tensorgebaseerde Patroonherkenning, Data-analyse En Machinaal Leren. Leuven: KU Leuven. Faculteit Ingenieurswetenschappen, 2019.
2. Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E. and Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. IEEE Transactions on Signal Processing, 65(13), 3551-3582.
3. Grasedyck L., Kressner D., Tobler C. A literature survey of low-rank tensor approximation techniques. GAMM-Mitteilungen. Feb 2013; 36(1):53–78.
4. Boussé, M., Vervliet, N., Domanov, I., Debals, O. and De Lathauwer, L. (2018). Linear systems with a canonical polyadic decomposition constrained solution: Algorithms and applications. Numerical Linear Algebra with Applications, 25(6), e2190.
5. Hendrikx, S., Boussé, M., Vervliet, N., and De Lathauwer, L. (2019). Algebraic and optimization based algorithms for multivariate regression using symmetric tensor decomposition. In 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (pp. 475-479).