Alternating Mahalanobis Distance Minimization for CP Tensor Decomposition

Navjot Singh, Edgar Solomonik

Abstract

Tensors generalize matrices by representing data in more than two dimensions. Tensor decompositions are mathematical constructs used to efficiently represent, approximate, and manipulate tensors. Tensor decompositions have applications in various fields such as in image analysis [2], in signal processing [6], in quantum chemistry [5], in chemometrics [4] and many more. Finding the most accurate low rank tensor decomposition of a tensor is an NP-hard problem [1] in most cases. Consequently, numerical optimization algorithms are used to compute a low rank approximation efficiently. In this talk, we present a novel alternating optimization algorithm for CP tensor decomposition (CPD) [7].

CP Decomposition. The CPD problem, for an order 3 tensor \mathcal{T} is formulated as following,

$$\min_{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}} \frac{1}{2} \left\| \boldsymbol{\mathcal{T}} - [\![\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}]\!] \right\|_{F}^{2},$$
$$([\![\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}]\!])_{ijk} = \sum_{r} a_{ir} b_{jr} c_{kr}.$$

The most used algorithm to solve the above problem is alternating least squares (ALS). ALS solves for one factor matrix at a time which results in least squares equation. Solving for factor matrix A results in the following least squares equation,

$$\boldsymbol{A}(\boldsymbol{C}\odot\boldsymbol{B})^T\cong\boldsymbol{T}_{(1)},$$

where \odot denotes the Khatri-Rao product. ALS solves these equations via normal equations where the solution is given as

$$\boldsymbol{A} = \boldsymbol{T}_{(1)} \big(\boldsymbol{C} \odot \boldsymbol{B} \big)^{\dagger T} = \boldsymbol{T}_{(1)} \big(\boldsymbol{C} \odot \boldsymbol{B} \big) (\boldsymbol{B}^T \boldsymbol{B} * \boldsymbol{C}^T \boldsymbol{C})^{\dagger},$$

where * denotes the Hadamard product, and \dagger denotes the Moore-Penrose inverse. The normal equations result in a solution that is optimal in Frobenius norm and matrix 2–norm. We propose an update that solves the least squares equations for factor A as

$$\boldsymbol{A} = \boldsymbol{T}_{(1)}(\boldsymbol{C}^{\dagger T} \odot \boldsymbol{B}^{\dagger T}).$$

We prove that the above update leads to an alternating minimization algorithm which has a **local superlinear convergence rate** for exact CP rank problems, when rank is smaller than the dimensions. The above update is an optimal solution of the least squares equations in Mahalanobis norm. The algorithm corresponding to these alternating updates is called alternating Mahalanobis distance minimization (AMDM) [7]. The update for factor \boldsymbol{A} is derived by minimizing the following objective function

$$\min_{\boldsymbol{A}} \frac{1}{2} \operatorname{vec}(\boldsymbol{\mathcal{T}} - \llbracket \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \rrbracket)^T \boldsymbol{M} \operatorname{vec}(\boldsymbol{\mathcal{T}} - \llbracket \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \rrbracket),$$

where M is a Kronecker structured positive definite matrix. M is defined as

$$egin{aligned} & oldsymbol{M} = ig(oldsymbol{M}^{(oldsymbol{A})}ig)^{-1} \otimes ig(oldsymbol{M}^{(oldsymbol{B})}ig)^{-1} \otimes ig(oldsymbol{M}^{(oldsymbol{C})}ig)^{-1}, \ & oldsymbol{M}^{(oldsymbol{B})} = oldsymbol{B}oldsymbol{B}^T + (oldsymbol{I} - oldsymbol{B}oldsymbol{B}^\dagger), \end{aligned}$$

and similarly defined for $M^{(A)}$ and $M^{(C)}$. Mahalanobis norm [3] is a generalization of Frobenius norm by using covariance or ground metric matrices. For exact rank problems, the minima for any Mahalanobis norm corresponds to the minima of Frobenius norm. However, for approximation of a tensor with low CP rank, the stationary point of the AMDM algorithm may not be optimal in Frobenius norm metric that is the most used metric for assessing quality of the decomposition.

We empirically show that changing the metric M from I (which corresponds to the ALS update) to the proposed AMDM metric leads to a well-conditioned decomposition for approximation problems. A well-conditioned decomposition is useful for separation of components, clustering using CPD factors, and stability of application of the operator when an operator is approximated via CPD. We also show that by interpolating between AMDM and ALS updates, we obtain a hybrid algorithm that leads to better fitness as compared to ALS while maintaining a the quality of decomposition.

References

- C. J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. J. ACM, 60(6):45:1–45:39, Nov. 2013.
- [2] M. Jouni, M. D. Mura, and P. Comon. Hyperspectral image classification based on mathematical morphology and tensor decomposition. *Mathematical Morphology-Theory and Applications*, 4(1):1–30, 2020.
- [3] P. Mahalanobis. On the generalised distance in statistics (reprint, 2018). Sankhya A, 80(1):1–7, 1936.
- [4] K. R. Murphy, C. A. Stedmon, D. Graeber, and R. Bro. Fluorescence spectroscopy and multiway techniques. PARAFAC. Analytical Methods, 5(23):6557–6566, 2013.
- [5] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill. Tensor hypercontraction.
 ii. least-squares renormalization. *The Journal of chemical physics*, 137(22), 2012.
- [6] 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 Transactions on Signal Processing*, 65(13):3551–3582, 2017.
- [7] N. Singh and E. Solomonik. Alternating mahalanobis distance minimization for accurate and well-conditioned cp decomposition. SIAM Journal on Scientific Computing, 45(6):A2781– A2812, 2023.