Ann. Appl. Math.
doi: 10.4208/aam.OA-2021-0090

Vol. **38**, No. 4, pp. 385-413
November 2022

Gradient Descent for Symmetric Tensor Decomposition

Jian-Feng Cai[1], Haixia Liu[2,*] and Yang Wang[1]

[1] *Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*
[2] *School of Mathematics and Statistics & Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*

Received 13 December 2021; Accepted (in revised version) 30 August 2022

Abstract. Symmetric tensor decomposition is of great importance in applications. Several studies have employed a greedy approach, where the main idea is to first find a best rank-one approximation of a given tensor, and then repeat the process to the residual tensor by subtracting the rank-one component. In this paper, we focus on finding a best rank-one approximation of a given orthogonally order-3 symmetric tensor. We give a geometric landscape analysis of a nonconvex optimization for the best rank-one approximation of orthogonally symmetric tensors. We show that any local minimizer must be a factor in this orthogonally symmetric tensor decomposition, and any other critical points are linear combinations of the factors. Then, we propose a gradient descent algorithm with a carefully designed initialization to solve this nonconvex optimization problem, and we prove that the algorithm converges to the global minimum with high probability for orthogonal decomposable tensors. This result, combined with the landscape analysis, reveals that the greedy algorithm will get the tensor CP low-rank decomposition. Numerical results are provided to verify our theoretical results.

AMS subject classifications: 65F15, 93B60

Key words: Gradient descent, random initialization, symmetric tensor decomposition, CP decomposition, linear convergence.

*Corresponding author.
Emails: `jfcai@ust.hk` (J. Cai), `liuhaixia@hust.edu.cn` (H. Liu), `yangwang@ust.hk` (Y. Wang)

1 Introduction

Tensor decomposition can be viewed as an extension of the singular value decomposition (SVD) for matrices, which is obviously one of the fundamental tools in numerous applications. Unlike for matrices, the term "decomposition" for tensors can carry very different meanings in different studies. In this paper we focus on one of the most commonly used notions of tensor decomposition: the *canonical polyadic decomposition (CP decomposition, or CPD).*

Before going further we first introduce some notations. Let $\mathcal{A}$ be a tensor, which is an element of $\bigotimes_{j=1}^m \mathbb{R}^{n_j} := \mathbb{R}^{n_1\times n_2\times\cdots\times n_m}$. A *rank-one* tensor $\mathcal{A}$ in $\bigotimes_{j=1}^m \mathbb{R}^{n_j}$ has the form

$$\mathcal{A}=\prod_{k=1}^m \mathbf{v}_k := \mathbf{v}_1\otimes\mathbf{v}_2\otimes\cdots\otimes\mathbf{v}_m,$$

namely $[\mathcal{A}]_{j_1j_2\cdots j_m}=v_{1j_1}v_{2j_2}\cdots v_{mj_m}$. For simplicity we use $\mathbf{j}$ to denote the multi-index $\mathbf{j}:=(j_1,j_2,\cdots,j_m)$, and $[\mathcal{A}]_{\mathbf{j}}$ to denote the $\mathbf{j}$-th entry of $\mathcal{A}$. Given a general tensor $\mathcal{A}\in\bigotimes_{j=1}^m \mathbb{R}^{n_j}$, a *CP decomposition (CPD)* of $\mathcal{A}$ is to decompose it into sum of rank-one tensors, $\mathcal{A}=\mathcal{A}_1+\mathcal{A}_2+\cdots+\mathcal{A}_r$, where each $\mathcal{A}_i$, $i=1,\cdots,r$ is a rank-one tensor. The minimal r is called the *CP rank* of $\mathcal{A}$. A major problem in tensor decomposition is to compute the CP rank and the CP decomposition of a tensor.

A particularly important class of tensors are the so-called super symmetric tensors, or simply *symmetric tensors*. A tensor $\mathcal{A}\in\bigotimes^m\mathbb{R}^n$ is called a symmetric tensor if for any multi-index $\mathbf{j}\in\{1,2,\cdots,n\}^m$ and any permutation $\mathbf{i}$ of $\mathbf{j}$ we have $[\mathcal{A}]_{\mathbf{i}}=[\mathcal{A}]_{\mathbf{j}}$. It is easy to see that a rank-one symmetric tensor $\mathcal{A}$ must have the form

$$\mathcal{A}=\lambda\mathbf{v}^{\otimes m} := \lambda\underbrace{\mathbf{v}\otimes\cdots\otimes\mathbf{v}}_{m},$$

where $\mathbf{v}\in\mathbb{R}^n$ and $\lambda\in\mathbb{R}$ are both nonzero. For a general symmetric tensor $\mathcal{A}$, a *symmetric CP decomposition (symmetric CPD)* is

$$\mathcal{A}=\mathcal{A}_1+\mathcal{A}_2+\cdots+\mathcal{A}_r,$$

where each $\mathcal{A}_i$, $i=1,\cdots,r$ is a symmetric rank-one tensor. The minimal r is called the *symmetric CP rank* of $\mathcal{A}$. Like general tensor decomposition, symmetric CP decomposition of a symmetric tensor is a major (and challenging) problem in the study of tensors. Although finding the symmetric CP rank of a symmetric tensor and its symmetric CP decomposition are generally very challenging, they are very useful in applications. For example, symmetric tensors appear as higher order derivatives or moments and cumulants of random vectors, which are often used in source extraction, mobile communications, machine learning, factor analysis of m-way arrays, biomedical engineering, psychometrics, and chemometrics [4,6–9,16,25].