JMLR: Workshop and Conference Proceedings vol 35:1–3, 2014

# Open Problem: Tensor Decompositions: Algorithms up to the Uniqueness Threshold?

**Aditya Bhaskara** BHASKARA@CS.PRINCETON.EDU
*Google Research NYC*

**Moses Charikar** MOSES@CS.PRINCETON.EDU
*Princeton University*

**Ankur Moitra** MOITRA@MIT.EDU
*MIT*

**Aravindan Vijayaraghavan** ARAVINDV@CS.CMU.EDU
*CMU*

## Introduction

*Factor analysis* is a basic tool in statistics and machine learning, where the goal is to take many variables and explain them away using fewer unobserved variables, called factors. It was introduced in a pioneering study by psychologist Charles Spearman, who used it to test his theory that there are fundamentally two types of intelligence – *verbal* and *mathematical*. This study has had a deep influence on modern psychology, to this day. However there is a serious mathematical limitation to this approach, which we describe next.

In its most basic form, we are given a matrix $M = \sum_{i=1}^{R} a_i \otimes b_i$. Our goal is to recover the factors $\{a_i\}_i$ and $\{b_i\}_i$. However this decomposition is only *unique* if we add further assumptions, such as requiring the factors $\{a_i\}_i$ and $\{b_i\}_i$ to be orthonormal. Otherwise we could apply an $R \times R$ rotation to the factors $\{a_i\}_i$ and its transpose to $\{b_i\}_i$ and recover another valid decomposition. This is often called the *rotation problem* and has been a central stumbling block for factor analysis since Spearman's work. To summarize:

> Even if there is a factorization $M = \sum_{i=1}^{R} a_i \otimes b_i$ that has a meaningful interpretation, there is no guarantee that factor analysis finds it!

Tensor methods do not suffer from the same problems: Consider a tensor

$$T = \sum_{i=1}^{R} a_i \otimes b_i \otimes c_i$$

There are many natural conditions on the factors $\{a_i\}_i$, $\{b_i\}_i$ and $\{c_i\}_i$ (as we will describe below) that ensure this decomposition is unique among all decompositions with at most $R$ factors. And so tensor decompositions are unique in a way that standard matrix decompositions are not, and have a broad range of applications in fields as diverse as psychometrics, chemometrics, medicine, signal processing and data mining.

Next we describe the uniqueness conditions that are known. In a result that has been rediscovered many times, Jennrich gave a natural condition under which the above decomposition is unique. Moreover it comes with an algorithm for finding it!

**Theorem 1** *(Harshman, 1970) Suppose the sets of vectors $\{a_i\}_i$ and $\{b_i\}_i$ are linearly independent, and no pair vectors in $\{c_i\}_i$ are scalar multiples of each other, then the decomposition*

$$T = \sum_{i=1}^{R} a_i \otimes b_i \otimes c_i$$

*is unique among all decompositions with at most $R$ factors. Moreover there is a polynomial time algorithm for computing it.*

What makes this theorem so broadly useful is that it does not require that the factors $\{a_i\}_i$, $\{b_i\}_i$ and $\{c_i\}_i$ be orthonormal; it would suffice that they are linearly independent. In many applications, it is much more reasonable to assume that the factors are linearly independent than that they are completely uncorrelated with each other.

This algorithm plays a key role in many applications in statistics and machine learning. The canonical setting is when we are given samples from a high-dimensional mixture model. We collect the moments of the distribution in a tensor $T$, which we can estimate empirically by taking enough samples. (Actually, we only get an approximation to $T$ that is entry-wise close). Our goal is to use these moments to recover the parameters of the mixture which are the factors in the decomposition. In this way the above algorithm has applications to phylogenetic reconstruction, learning mixtures of Gaussians, independent component analysis, multi-view models and learning mixtures of product distributions. There are even more applications in statistics.

Perhaps one of the reasons this algorithm has been rediscovered many times is that it has been more-or-less forgotten in the psychometrics literature, where a stronger uniqueness result is known.

**Definition 2** *The Kruskal rank of a set of vectors $\{a_i\}_i$ is the largest $k_a$ so that* every *set of $k_a$ vectors are linearly independent.*

Note that given a set of vectors $\{a_i\}_i$ it is hard to compute their Kruskal rank – even in the special case where we want to decide if the vectors are in general position. Nevertheless this notion plays a key role in a stronger uniqueness theorem:

**Theorem 3** *(Kruskal, 1977) Suppose the sets of vectors $\{a_i\}_i$, $\{b_i\}_i$ and $\{c_i\}_i$ have Kruskal ranks of $k_a$, $k_b$ and $k_c$ respectively. Then the decomposition*

$$T = \sum_{i=1}^{R} a_i \otimes b_i \otimes c_i$$

*is unique among all decompositions with at most $R$ factors if $k_a + k_b + k_c \geq 2R + 2$.*

Recall that in Jennrich's theorem we need the factors to be linearly independent. So at most we can set $R = n$ for an $n \times n \times n$ tensor. But the above uniqueness theorem can apply for $R = 3n/2 - O(1)$, in the (optimistic) scenario when $k_a = k_b = k_c = n$. But there is no known algorithmic proof!

**Open Problem 4** *Is there an algorithm for decomposing $T$ under the conditions of Kruskal's uniqueness theorem?*

To put this in context, Jennrich's uniqueness theorem is proven by giving an algorithm for decomposing $T$ and identifying conditions under which it is stable to noise (we need this in applications to learning if we want finite sample complexity bounds) Goyal et al. (2013), Bhaskara et al. (2014). Is there an algorithmic proof of the above uniqueness theorem or its robust analogues Bhaskara et al. (2013)?

If we move to higher order tensors then we can handle $R >> n$ but by flattening the tensor and reducing back to an order three tensor with larger dimensions, and invoking Jennrich's algorithm as depicted in the following equation:

$$T = \sum_{i=1}^{R} \underbrace{a_i \otimes b_i}_{\text{factor}} \otimes \underbrace{c_i \otimes d_i}_{\text{factor}} \otimes \underbrace{e_i}_{\text{factor}}$$

This approach is robust to noise in various settings, including in the model of smoothed analysis Bhaskara et al. (2014). De Lathauwer et al. (2007) gives algorithms for $R = \Omega(n^2)$ for fourth order tensors through an entirely different approach.

But the limits of what should *algorithmically* be possible for order three tensors are far from clear. It could be that it is computationally hard to decompose $T$ even under the conditions of Kruskal's uniqueness theorem. We also remark that if our only condition on the factors involves only their Kruskal rank, then it is known that Kruskal's uniqueness theorem is tight. A random third order tensor generated by choosing its $R$ factors uniformly at random will be unique even for $R = \Omega(n^2)$, but there are analogies with open problems in circuit complexity that lead us to believe that finding a decomposition might be hard in this random setting even though it is unique.

A reward of \$100 is offered for a resolution of this open problem.

## References

Aditya Bhaskara, Moses Charikar, and Aravindan Vijayaraghavan. Uniqueness of tensor decompositions with applications to polynomial identifiability. *CoRR*, abs/1304.8087, 2013.

Aditya Bhaskara, Moses Charikar, Ankur Moitra, and Aravindan Vijayaraghavan. Smoothed analysis of tensor decompositions. In *Symposium on the Theory of Computing (STOC)*, 2014.

L. De Lathauwer, J. Castaing, and J. Cardoso. Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Trans. on Signal Processing*, 55(6):2965–2973, 2007.

Navin Goyal, Santosh Vempala and Ying Xiao. Fourier PCA. In *Symposium on the Theory of Computing (STOC)*, 2014.

Richard A Harshman. Foundations of the parafac procedure: models and conditions for an explanatory multimodal factor analysis. 1970.

Joseph B Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and applications*, 18(2):95–138, 1977.