SOME RESULTS IN TENSOR RANK DECOMPOSITION

HUNG PHAM

Abstract. This paper addresses the problem of tensor rank, a fundamental concept in multilinear algebra analogous to matrix rank in linear algebra. While matrix rank is a well-understood and solved problem, tensor rank remains an area of active research with many open questions. This paper will provide a comprehensive overview of tensor rank by first defining tensors and tensor rank, then exploring the computational hardness of determining tensor rank. We will present some simple lower and upper bounds, discuss the current state of the art approximation algorithms, and highlight various applications of tensor rank in different fields.

Contents

1. Definitions

There are many ways in the literature to define a tensor (a comprehensive expository survey can be bound in, say, [8]), but this paper will only use the simplest notion. In this paper, we think of a tensor as a multi-dimensional array, in the same sense that a matrix is a 2-dimensional array and a vector is a 1-dimensional array. Because of this simple definition, tensors and tensor methods are as useful in practice as useful in practice as matrices themselves. This paper will focus on tensors of dimension 3, as most results can be extended.

Tensor rank decomposition is a fundamental concept in multilinear algebra. First recall a familiar result from linear algebra, that the outer product uv^T of two

Date: August 30, 2024.

nonzero vectors u and v is always a rank-1 matrix, and that a rank r-matrix M can be decomposed as the sum of r rank-1 matrices via the singular value decomposition

$$M = \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

We can define tensor rank in an entirely analogous way.

Definition 1.1. An $m \times n \times p$ tensor $\mathcal{X}$ is **simple** if it is the **outer product** of three vectors $u \in \mathbb{R}^m, v \in \mathbb{R}^n, w \in \mathbb{R}^p$ i.e.,

$$\mathcal{X}_{j,k,l} = u_j v_k w_l, \ \ \forall\ (j,k,l) \in [m] \times [n] \times [p]$$

where $[n]$ denotes $\{1, 2, \ldots, n\}$ for any integer n. We write

$$\mathcal{X} = u \otimes v \otimes w.$$

Definition 1.2. A **rank-r decomposition** of $\mathcal{X}$ is a way to represent the tensor as the sum of simple tensors:

$$\mathcal{X} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i.$$

The **rank** of a tensor $\mathcal{X}$ is the the smallest r such that $\mathcal{X}$ can be represented as a sum of rank-1 tensors:

$$\mathrm{rk}(\mathcal{X}) = \min\left\{ r : \mathcal{X} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i \right\}.$$

One can think of tensor rank as a measure of "orderliness" of the given tensor, where rank-1 tensors are the most orderly. Thus, the motivation to study tensor decomposition ties directly with the need to understand, process and analyze multi-dimensional data. Some notable applications will be discussed in the last section of the paper.

2. Example of tensor rank

As an introductory example, in this section we will determine the rank of the following tensor:

$$\mathcal{X} = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix} \tag{2.1}$$

Here we are notating a $2 \times 2 \times 2$-shaped tensor as a vector of matrices, where the vector is indexed by the last index of $\mathcal{X}_{jkl}$ and the rows and columns by the first and second indices.

We will call matrices of the form $\mathcal{X}_{j::}$, $\mathcal{X}_{:k:}$, and $\mathcal{X}_{::l}$ **slices** of a tensor, and slices of the form $\mathcal{X}_{::k}$ **frontal slices**. For instance, the frontal slices of $\mathcal{X}$ in (2.1) are given by $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Since there are only 3 non-zero entries, we can write $\mathcal{X}$ as the sum of 3 tensors with only one nonzero entry; such tensors are clearly simple as they are the outer product of standard basis vectors. Therefore $\mathcal{X}$ has rank at most 3. The question is whether or not $\mathcal{X}$ has a rank-2 representation.

We first introduce the following lemma

Lemma 2.2. *Let $\mathcal{X}$ be a tensor with shape $m \times n \times p$. Denote the frontal slices of $\mathcal{X}$ with $M_1, M_2, \dots, M_p$, which are $m \times n$ matrices. Then* $\mathrm{rk}(\mathcal{X})$ *is the size of the smallest set of singular matrices that span* $\mathrm{span}(M_1, M_2, \dots, M_p)$.

Proof. Let $r = \mathrm{rk}(\mathcal{X})$, then

$$\mathcal{X} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i.$$

Fixing the last index, we have:

$$M_l = \sum_{i=1}^{r} u_i \otimes v_i \cdot w_{i,l}.$$

This implies that the set of $u_i \otimes v_i$ span every M_l. Hence, r is at least the size of the smallest set of singular matrices that span $\mathrm{span}(M_1, M_2, \dots, M_p)$

On the other hand, if there is a set $u_1 \otimes v_1, u_2 \otimes v_2, \dots, u_r \otimes v_r$ that spans every M_l. We can then construct w_i's so that:

$$M_l = \sum_{i=1}^{r} u_i \otimes v_i \cdot w_{i,l}, \ \forall \, l$$

which is possible because the M_l are linear combination of $u_i \otimes v_i$'s. Hence, r is at most size of the smallest amount of rank-1 matrices to span all M_l's.

Therefore, there rank of $\mathcal{X}$ is the smallest amount of rank-1 matrices needed to span all M_l's. ☐

Corollary 2.3. *The tensor $\mathcal{X}$ in (2.1) is rank-3*

Proof. Assume for the sake of contradiction that $\mathcal{X}$ has a rank-2 representation. Then, from the lemma, we can find $A = u_1 \otimes v_1$ and $B = u_2 \otimes v_2$ which span $\mathcal{X}_{::1}, \mathcal{X}_{::2}$. The matrices A and B will therefore be linear combinations of $\mathcal{X}_{::1}, \mathcal{X}_{::2}$. However,

$$a\mathcal{X}_{::1} + b\mathcal{X}_{::2} = \begin{bmatrix} b & a \\ a & 0 \end{bmatrix},$$

where a and b are nonzero scalars. This matrix has has determinant a^2. Since both A and B are rank-1, their determinants are 0. Therefore we must have a be nonzero, which is a contradiction. ☐

We can now generate more rank-3 $2 \times 2 \times 2$ tensors, using the notion of equivalent tensors, defined through the **multilinear matrix multiplication** as follow:

Definition 2.4. Let $\mathcal{X} \in \mathbb{R}^{m \times n \times p}, A \in \mathbb{R}^{m' \times m}, B \in \mathbb{R}^{n' \times n}, C \in \mathbb{R}^{p' \times p}$ be tensor and matrices, define

$$[A, B, C] \cdot \mathcal{X} = \mathcal{Y}$$

such that

$$\mathcal{Y}_{xyz} = \sum_{j,k,l} A_{x,j} B_{y,k} C_{z,l} \mathcal{X}_{jkl}.$$

We say $\mathcal{X}$ and $\mathcal{Y}$ are equivalent if A, B and C are invertible.

Theorem 2.5. *Equivalent tensors have the same rank.*

Proof. We will first show that if

$$(A, B, C) \cdot \mathcal{X} = \mathcal{Y}$$

and we have a rank-r representation of $\mathcal{X}$

$$\mathcal{X} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

then

$$\mathcal{Y} = \sum_{i=1}^{r} (Au_i) \otimes (Bv_i) \otimes (Cw_i)$$

To show this, we can calculate each entry of $\mathcal{Y}$ with:

$$\begin{aligned}
\mathcal{Y}_{xyz} &= \sum_{j,k,l} A_{x,j} B_{y,k} C_{z,l} \mathcal{X}_{jkl}. \\
&= \sum_{j,k,l} A_{x,j} B_{y,k} C_{z,l} \sum_{i=1}^{r} u_{i,j} v_{i,k} w_{i,l} \\
&= \sum_{i=1}^{r} \sum_{j,k,l} A_{x,j} B_{y,k} C_{z,l} u_{i,j} v_{i,k} w_{i,l} \\
&= \sum_{i=1}^{r} \sum_{j=1}^{m} A_{x,j} u_{i,j} \sum_{k=1}^{n} B_{y,k} v_{i,k} \sum_{l=1}^{p} C_{z,l} w_{i,l} \\
&= \sum_{i=1}^{r} (Au_i)_x \cdot (Bv_i)_y \cdot (Cw_i)_z
\end{aligned}$$

Which means

$$\mathcal{Y} = \sum_{i=1}^{r} (Au_i) \otimes (Bv_i) \otimes (Cw_i)$$

Now, if we can write $\mathcal{X}$ as the sum of r tensors, we can do so for $\mathcal{Y}$ as well, meaning

$$\text{rk}(\mathcal{X}) \geq \text{rk}(\mathcal{Y}).$$

However, this equivalence relation is reflexive (using the matrices A^{-1}, B^{-1}, C^{-1}), hence

$$\text{rk}(\mathcal{Y}) \geq \text{rk}(\mathcal{X})$$

Therefore, $\text{rk}(\mathcal{X}) = \text{rk}(\mathcal{Y})$. □

Using this, we can create an entire family of rank-3 tensors with shape $2 \times 2 \times 2$ from the tensor in (2.1).

3. Determining tensor rank is hard

Determining matrix rank is essentially a solved problem and the algorithms to do so are easy and numerous (SVD, QR, Gaussian elimination, etc.). Surprisingly, the opposite has been shown for tensors of dimension 3 or more.

Theorem 3.1 (Hastad)**.** *Given a tensor $\mathcal{X}$ of dimension 3 and an integer r, determining whether $\text{rk}(\mathcal{X}) < r$ is NP-hard.*

The class **NP** denotes the subset of computational problems for which a potential solution can be verified in polynomial time. While verifying might be easy, finding a solution can be exponentially costly. This is the essence of the Millenium Problem $P \stackrel{?}{=} NP$: can we solve NP problems in polynomial time?. The class of **NP-hard** consists of the "hardest" problems in NP, in the sense that we can use an algorithm for an NP-hard problem to solve any NP problem equally efficiently.

The idea of the proof of Theorem 3.1 is as follows. A "solution" to the tensor rank problem refers to a valid rank-r decomposition of $\mathcal{X}$, so verification is easy. The proof of NP-hardness relies on reduction from the 3SAT problem (an archetypal NP-hard problem) using Lemma 2.2. Interested readers can find the full proof in [9].

Theorem 3.1 suggests that in general, there is no algorithm better than brute force to find the rank of a tensor on finite fields. In fact, there is no known finite-runtime algorithm to find the rank on infinite fields either.

4. Some bounds on tensor rank

Given that the problem of deciding the rank of a particular tensor is hard, we hope instead to be able to bound the maximum rank of tensors of a given size. The case for matrices is rather simple; by the rank-nullity theorem, a $m \times n$-matrix has rank at most $\min(m, n)$. The case for tensors is considerably more difficult.

4.1. **Tensors over rings.**

Up to this point we have been considering tensors over $\mathbb{R}$. As a warm-up, we briefly consider tensors over a finite ring. Using a simple counting argument, we can see that there exists tensors of high rank for such tensors.

Theorem 4.1. *On a finite ring R, there exist tensors with shape $m \times n \times p$ with rank at least $mnp/(m + n + p)$.*

Proof. Let q be the number of elements of R. There are q^{mnp} possible tensors, while there are q^{m+n+p} rank-1 tensors. There are at most $q^{t(m+n+p)}$ tensors rank $\leq t$ (since there are q^{m+n+p} ways to choose each component rank-1 tensor). Thus the maximum rank is at least $mnp/(m + n + p)$ □

By counting zeros of a polynomial instead, we can obtain the same bound for an infinite ring R.

Theorem 4.2. *On an infinite commutative ring R, there exist tensors with shape $m \times n \times p$ with rank at least $mnp/(m + n + p)$.*

Proof. Assume that r is the maximum rank of $m \times n \times p$-shaped tensors. Consider mnp polynomials in $r(m + n + p)$ variable:

$$f_{jkl} = \sum_{i=1}^{r} u_{i,j} v_{i,k} w_{i,l}$$

with $i \in [r]$, $j, k, l \in [m] \times [n] \times [p]$.

These are the entries of the tensor

$$\mathcal{X} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i.$$

when we take the $u_{i,j}, v_{i,k}, w_{i,l}$ as variables.

Suppose for the sake of contradiction that $mnp > r(m+n+p)$. Then there are more polynomials than variables. Hence, the polynomials are algebraically dependent, meaning we can then find a nontrivial polynomial $P(x_{111}, x_{112}, \dots, x_{mnp})$ such that:

$$P\big(f_{111}, f_{112}, \dots, f_{mnp}\big) = 0$$

In other words, we can find a non-trivial polynomial that all rank-r tensors satisfy. However, since R is infinite and P is non-trivial, we can find $(x_{111}, x_{112}, \dots, x_{mnp}) \in R^{mnp}$ such that

$$P(x_{111}, x_{112}, \dots, x_{mnp}) \neq 0$$

Choose the x_{jkl} as the entries of $\mathcal{X}$, we have a tensor that does not satisfy P, and thus does not have a rank-r representation. This contradicts the assumption that r is the maximum rank. Therefore, $r \geq mnp/(m+n+p)$. □

The two proofs above are from [2], as Theorem 9 and 11. Additionally, on a infinite field $\mathbb{F}$, we can write $u \cdot v \cdot w = ab(u \cdot (v/a) \cdot (w/b))$, for any $a, b \in \mathbb{F}$. This means we can normalize the vectors that make up a tensor rank decomposition, reducing the right hand side of (**??**) to $r(m+n+p-2)$ variables. Readers can find the proof of this as Theorem 12 in [2] as well.

4.2. **Tensors over $\mathbb{R}$.** Using the upper bound of matrix rank on the slices of the tensor, we the following upper bound:

Theorem 4.3. *A $m \times n \times p$ tensor can have rank at most* $\min(mn, np, mp)$

Proof. Given a tensor $\mathcal{X}$, consider its frontal slices $M_1, M_2, \dots, M_p$ (fixing the last index) as $m \times n$ matrices. Then, the rank of $\mathcal{X}$ is at most sum of $\mathrm{rk}(M_i)$, since

$$\mathcal{X} = M_1 \otimes e_1 + M_2 \otimes e_2 + \cdots + M_p \otimes e_p,$$

where e_i are the standard basis vectors. Then from a decomposition of the M_i's, we obtain a decomposition of $\mathcal{X}$ with rank:

$$\sum_{i=0}^{p} \mathrm{rk}(M_i) \leq \sum_{i=0}^{p} \min(m,n) \leq \min(mp, np)$$

Repeat the process for the other directional slices (fixing the first or second index), and we are done. □

It is rather surprising that such a simple result is tight for tall tensors (p much larger than m, n):

Theorem 4.4. *If $p \geq mn$, there is a $m \times n \times p$ tensor with rank mn.*

Proof. Again, consider the frontal slices of a tensor $\mathcal{X}$. If we can write

$$\mathcal{X} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i,$$

then the matrices $u_i \otimes v_i$'s span the frontal slices of $\mathcal{X}$, since:

$$M_k = \sum_{i=1}^{r} w_{i,k} \cdot u_i \otimes v_i.$$

Choosing the frontal slices of $\mathcal{X}$ to be the standard basis matrices, we have now mn linearly independent frontal slices. Applying Lemma 2.2 along with the upper bound of mn in the previous theorem, we are done. □

For more balanced tensors (m, n, p closer to each other), we can improve the bounds by about a factor of 2.

Theorem 4.5. *A $m \times n \times p$ tensor can have rank at most $m + \lfloor p/2 \rfloor n$, if $m \leq n$*

The proof of this can be found in [3]. The idea of the proof is to show that given m arbitrary singular matrices, we only need another n for each additional 2 frontal slices, instead of $2m$ as in Theorem 4.4.

Applying the theorems above on a $n \times n \times n$ tensor, we know that the maximum rank is between $n^3/(3n-2) \approx n^2/3$ and $n + \lfloor n/2 \rfloor n \approx n^2/2$. To our knowledge, these are the current state-of-the art bounds without any additional assumptions.

5. Approximating tensors

Since computing the rank of a tensor is hard, a sensible approach might be to guess the rank of a tensor and see how close the best rank r approximation is.

Definition 5.1. Let APPROX($\mathcal{X}, r$) be the rank-r tensor $\mathcal{Y}$ that minimizes $\|\mathcal{X} - \mathcal{Y}\|$, where $\|\cdot\|$ denotes the Frobenius norm on tensors, that is, the square root of the sum of squared entries.

If we can solve APPROX reasonably fast, a sensible way to estimate the matrix rank of $\mathcal{X}$ might be to begin with $r = 1$, find the best rank-r approximation of $\mathcal{X}$, and increase r until the error is sufficiently small. In the matrix case, APPROX is solved by the following well-known result, the proof of which can be found at [10].

Theorem 5.2 (Eckart-Young)**.** *Let $A \in \mathbb{R}^{m \times n}$ be a matrix with $m \leq n$ with rank r. Let the singular value decomposition of A be given by*

$$A = U\Sigma V^T = \sum_{i=r}^{R} \sigma_i u_i v_i^T$$

Assume that $\sigma_i > \sigma_{i+1}$ for all i. Then:

$$APPROX(A, k) = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

For tensors of dimension 3, we run into some problems.

Theorem 5.3. *There exists $2 \times 2 \times 2$ tensors of rank 3 that can be approximated arbitrarily well with rank $\leq$ 2 tensors. In other words, there exists a rank-3 tensor $\mathcal{X}$ such that for all $\varepsilon > 0$, there exists a rank-2 tensor $\mathcal{Y}$ such that $\|\mathcal{X} - \mathcal{Y}\| < \varepsilon$.*

Proof. Given 6 arbitrary vectors u, v, w, a, b, c, consider the following tensor of rank at most 3:

$$\mathcal{X} = a \otimes v \otimes w + u \otimes b \otimes w + u \otimes v \otimes c \tag{5.4}$$

For $n \in \mathbb{N}$, consider also the tensors of rank at most 2 given by

$$\mathcal{Y}_n = n\Big(u + \frac{a}{n}\Big) \otimes \Big(v + \frac{b}{n}\Big) \otimes \Big(w + \frac{c}{n}\Big) - nu \otimes v \otimes w,$$

We claim $\lim_{n\to\infty} \mathcal{Y}_n = \mathcal{X}$. Observe that if we expand the first term of $\mathcal{Y}_n$, we get

$$\begin{aligned} &n\Big(u + \frac{a}{n}\Big) \otimes \Big(v + \frac{b}{n}\Big) \otimes \Big(w + \frac{c}{n}\Big) \\ &= n\left(u \otimes v \otimes w + \frac{1}{n}\left(a \otimes v \otimes w + u \otimes b \otimes w + u \otimes v \otimes c\right) + O(\frac{1}{n^2})\right) \\ &= nu \otimes v \otimes w + \mathcal{X} + O(\frac{1}{n}). \end{aligned}$$

So $\mathcal{Y}_n \to \mathcal{X}$ as $n \to \infty$.

If we choose u, v, w and a, b, c so that $\mathcal{X}$ is a rank-3 tensor, we are done. We can do this by considering the example tensor in (2.1) and representing it in the form of (5.4) with u, v, w, a, b, c being appropriate choices of unit basis vectors. □

The claim of the theorem is true for $n \times n \times n$ tensors as well, interested readers can find it in [12].

This poses a problem for us, because if $\mathcal{X}$ is the tensor in (2.1), APPROX($\mathcal{X}$, 2) has no best approximation to return! Moreover, the gap between this approximated rank and the actual rank can be arbitrarily small, obtained by combining tensors of the form (5.4). Their ratio can also be made arbitrarily large, as shown in [11]. This suggests that determining the rank of a tensor is highly ill-posed. Because of this, the following notion of rank is useful.

Definition 5.5. The **border rank** of $\mathcal{X}$ is the smallest number of rank-1 tensors needed to approximate $\mathcal{X}$ arbitrarily well.

$$\underline{\mathrm{rk}}(\mathcal{X}) = \min\{r \mid \forall\, \varepsilon > 0,\ \exists \text{ rank-}r \text{ tensor } \mathcal{Y} \text{ s.t. } \|\mathcal{X} - \mathcal{Y}\| < \varepsilon\}$$

For $n \times n \times n$ shaped tensors, we have the following relationship between rank and border rank.

Theorem 5.6. *The rank of almost all (everything except a set of Lebesgue measure 0) $n \times n \times n$ tensors, which is also their maximum border rank, is exactly*

$$\left\lceil \frac{n^3}{3n-2} \right\rceil.$$

The intuition behind typical rank and maximum border rank being equal is that if we can approximate every tensor with a rank-r tensor, then almost all tensors should be rank-r. The full proof can be found at [6].

Because of the difference between border rank and actual rank, along with others detailed in [12], finding a best approximation is frequently impossible. Two notable exceptions are matrices, as well as finding the best rank-1 approximation of tensors of dimension 3. The former is the Eckart-Young theorem mentioned above, and as an example for the latter, we will compute the rank-1 $APPROX$ of the following tensor:

Lemma 5.7. *Let*

$$\mathcal{X} = \begin{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{bmatrix}$$

then

$$APPROX(\mathcal{X}, 1) = \mathcal{S} = \begin{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix}$$

Proof. Let $\mathcal{Y}$ be a rank-1 tensor, i.e.,

$$\mathcal{Y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \otimes \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \otimes \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

We wish to choose $\mathcal{Y}$ to minimize $\|\mathcal{X} - \mathcal{Y}\|$. If $\mathcal{Y} = \mathcal{S}$, then the tensor $\mathcal{X} - \mathcal{Y}$ has only 3 non-zero entries (all being 1s), meaning $\|\mathcal{X} - \mathcal{Y}\| = \sqrt{3}$. We will show that it cannot get smaller.

If either $z_1 = 0$ or $z_2 = 0$, the problem becomes finding the best rank-1 approximation of one of the frontal slices of $\mathcal{X}$, which we can do with Eckart-Young, which does not get us smaller than $\sqrt(3)$. Any other entries being 0 yields the same result. If none of them are 0, let $a = x_2/x_1$, $b = y_2/y_1$, $c = z_2/z_1$ and $x = x_1x_2x_3$ (the entries need to be non-zero for the divisions), then:

$$\mathcal{Y} = x \begin{bmatrix} 1 \\ a \end{bmatrix} \otimes \begin{bmatrix} 1 \\ b \end{bmatrix} \otimes \begin{bmatrix} 1 \\ c \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} x & xa \\ xb & xab \end{bmatrix} & \begin{bmatrix} xc & xac \\ xbc & xabc \end{bmatrix} \end{bmatrix}$$

Our task now becomes minimizing

$$\|\mathcal{X} - \mathcal{Y}\|^2 = (2-x)^2 + (xa)^2 + (xb)^2 + (xc)^2 + (1-xab)^2 + (1-xbc)^2 + (1-xca)^2 + (xabc)^2.$$

Rearranging the RHS into a quadratic equation in x, we get:

$$f(x) = 7 - 2x(2 + ab + bc + ca) + x^2(1 + a^2)(1 + b^2)(1 + c^2)$$

To prove that $f(x) \geq 3$ for all x, we can show that the discriminant of $f(x) - 3$ is non-positive i.e.

$$(2 + ab + bc + ca)^2 - 4(1 + a^2)(1 + b^2)(1 + c^2) \leq 0$$

Expanding the equation, we find it equivalent to:

$$4(a^2+b^2+c^2)+3(a^2b^2+b^2c^2+c^2a^2)+4a^2b^2c^2 \geq 4(ab+bc+ca)+2abc(a+b+c). \tag{5.8}$$

This is always true because

$$(a-b)^2 + (b-c)^2 + (c-a)^2 \geq 0 \Leftrightarrow 2(a^2 + b^2 + c^2) \geq 2(ab + bc + ca) \tag{5.9}$$

and

$$(ab - bc)^2 + (bc - ca)^2 + (ca - ab)^2 \geq 0 \Leftrightarrow 2(a^2b^2 + b^2c^2 + c^2a^2) \geq 2abc(a + b + c)$$

This means we have $\|\mathcal{X} - \mathcal{Y}\|^2 \geq 3$.

Moreover, equality can only happen if $a = b = c = 0$ (because of (5.9) and the leftover $4a^2b^2c^2 \geq 0$ in (5.8)). Solving for x then yields $x = 2$, giving us the tensor $\mathcal{S}$. □

We exploited the small size of the tensor ($2 \times 2 \times 2$), along with it's symmetry in order to make it simpler.

The fact that rank-1 approximation is easy, while tensor rank is hard, highlights another complication with tensor rank: a computing low-rank approximation does not get us a lower rank tensor. This is completely different from matrices, where the Eckart-Young theorem implies that subtracting away the best rank-1 approximation from a rank r matrix will result in a rank $r - 1$ matrix.

Theorem 5.10. *Subtracting a best rank-1 approximation can increase the rank of a tensor.*

Proof. We first show that the following tensor is rank-2

$$\mathcal{X} = \left[\begin{bmatrix} a^2 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right].$$

Indeed, lets look at the following 2 matrices:

$$A = \begin{bmatrix} a^2 & a \\ a & 1 \end{bmatrix}$$

and

$$B = \begin{bmatrix} a^2 & -a \\ -a & 1 \end{bmatrix}$$

which are both singular. The slices of $\mathcal{X}$ can be written as $(A+B)/2$ and $(A-B)/2a$, respectively. Thus, by Lemma 2.2, $\mathcal{X}$ is rank-2.

Choosing $a = \sqrt{2}$, we can show (and will in the lemma below) that APPROX($\mathcal{X}$, 1) has only one non-zero entry (which is 2, corresponding with $\mathcal{X}_{122}$).

Subtracting APPROX($\mathcal{X}$, 1) from $\mathcal{X}$ gives us the tensor in (2.3), which as we have shown is rank-3. □

The phenomena shown in Theorem 5.10 is not rare, surprisingly. For almost all $2 \times 2 \times 2$ tensors, subtracting away the best rank 1 approximation will either increase the rank or keep it the same; the proof can be found in [5].

To conclude, the problem APPROX is often ill-posed. Whether or not one can overcome this is still an open question, as noted in [12].

6. Approximation algorithm

How does one actually compute the best rank-r decomposition of a tensor? Recall that APPROX(A, r) is finding $3r$ vectors $u_1, v_1, w_1, \ldots, u_r, v_r, w_r$. An iterative approach might be appropriate here: selecting some vectors and optimize them while fixing the others. However, we can't select u_1, v_1, w_1 at the same time, as subtracting the best rank-1 approximation does not reduce the rank.

Instead, the state of the art algorithms select the u_i's, then v_i's, then w_i's, and repeat. To ease calculation, we combine the u_i into a **factor matrix** U (similarly for V, W) by way of :

$$U = \begin{bmatrix} u_1 & u_2 & \ldots & u_r \end{bmatrix}.$$

We can solve for factor matrices instead of each vector individually. $[\![U, V, W]\!]$ denotes the tensor reconstructed from U, V and W.

To compute APPROX, we will utilize the **Alternating Least Square** method:

(1) Choose the desired rank r.
(2) Initialize factor matrices U, V, W (usually as random matrices).
(3) Fix V and W, choose U to minimize $\|\mathcal{X} - [\![U, V, W]\!]\|$. (hence, Least Square)
(4) Choose V and W similarly as in step (3). (hence, Alternating)
(5) Repeat (3) and (4) until convergence.

Since steps (1) and (2) are only done once and steps (4), (5) is the repetition of (3), we turn our attention to optimizing these latter steps. Consider a rank-r decomposition of a $m \times n \times p$ tensor:

$$\mathcal{X} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

Let's take a closer look at an $m \times n$ frontal slice of $\mathcal{X}$. We can write

$$\mathcal{X}_k = \sum_{i=1}^{r} w_{i,k} \cdot u_i \otimes v_i = \sum_{i=1}^{r} u_i \otimes (w_{i,k} \cdot v_i) = UV_k^T$$

where V_k is

$$V_k = \begin{bmatrix} w_{1,k} \cdot v_1 & w_{2,k} \cdot v_2 & \dots & w_{r,k} \cdot v_r \end{bmatrix}.$$

If we line up the slices of $\mathcal{X}$ as follows, we get:

$$\begin{bmatrix} \mathcal{X}_1 & \mathcal{X}_2 & \dots & \mathcal{X}_p \end{bmatrix} = \mathcal{X}_{(1)} = U \begin{bmatrix} V_1^T & V_2^T & \dots & V_p^T \end{bmatrix},$$

where $\mathcal{X}_{(1)}$, which is a $m \times np$ matrix, is the **mode-1 matricization** of $\mathcal{X}$: turning the tensor into a matrix via reshaping, so that the first index is preserved, meaning:

$$\mathcal{X}_{j::} \mapsto \left(\mathcal{X}_{(1)}\right)_{j:}$$

Similarly, we have $\mathcal{X}_{(2)}$ and $\mathcal{X}_{(3)}$, which have shapes $n \times mp$ and $p \times mn$ and preserve the second and third index, respectively.

We also define the **Khatri-Rao product** $\odot$ of $W \in \mathbb{R}^{p \times r}$ and $V \in \mathbb{R}^{n \times r}$, which is the $np \times r$ matrix:

$$W \odot V = \begin{bmatrix} w_{1,1}v_1 & w_{1,2}v_2 & \dots & w_{1,r}v_r \\ w_{2,1}v_1 & w_{2,2}v_2 & \dots & w_{2,r}v_r \\ \dots & \dots & \dots & \dots \\ w_{p,1}v_1 & w_{p,2}v_2 & \dots & w_{p,r}v_r \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \\ \dots \\ V_p \end{bmatrix}$$

So that:

$$\begin{aligned} \mathcal{X}_{(1)} &= U(W \odot V)^T \\ \mathcal{X}_{(2)} &= V(W \odot U)^T \\ \mathcal{X}_{(3)} &= W(V \odot U)^T \end{aligned} \tag{6.1}$$

With this, for step (3) of our algorithm, we can choose U to minimize

$$\|\mathcal{X}_{(1)} - U(W \odot V)^T\|.$$

This is a linear least squares problem, which has an efficient solution.

The algorithm above is called canonical polyadic (CP) decomposition. The implementation details, along with discussion of applications and variations, can be found in [1].

7. Applications

To conclude, here we present 2 applications of tensors in different fields. For an overview, see [1].

7.1. Chemometrics.

In chemometrics, data is often collected from experiments that measure various properties of chemical samples across multiple conditions. For example, fluorescence spectroscopy might measure the emission intensity of a sample across different excitation wavelengths and emission wavelengths, generating a 3-dimensional tensor with dimensions corresponding to intensity, excitation wavelength, and emission wavelength.

Then, using a tensor decomposition, we get

$$\mathcal{X} \approx \sum_{i=r}^{R} \lambda_r u_i \otimes v_i \otimes w_i$$

Where each rank-1 $\lambda_r u_i \otimes v_i \otimes wir$ correspond to a distinct chemical component. u, v, w represent the intensity, excitation wavelength and emission wavelength of the pure component. This makes them easily identifiable. A deeper analysis, along with comparisons to similar tools, can be found in [4].

Acknowledgments

I would like to express my gratitude toward my mentor, Phillip Lo, who both introduced me to this fascinating topic and provided meticulous feedback on the draft. This paper wouldn't be possible without his mentorship and guidance. I would also like to thank Professor Peter May for this exceptional REU experience.

References

[1] Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. SIAM Review, 51(3), 455–500. https://doi.org/10.1137/07070111x

[2] Howell, T. D. (1978). Global properties of tensor rank. Linear Algebra and Its Applications, 22, 9–23. https://doi.org/10.1016/0024-3795(78)90052-6

[3] Atkinson, M. D., & Stephens, N. M. (1979). On the maximal multiplicative complexity of a family of bilinear forms. Linear Algebra and Its Applications, 27, 1–8. https://doi.org/10.1016/0024-3795(79)90026-0

[4] Bro, R. (1997). PARAFAC. tutorial and applications. Chemometrics and Intelligent Laboratory Systems, 38(2), 149–171. https://doi.org/10.1016/s0169-7439(97)00032-4

[5] Stegeman, A., & Comon, P. (2010). Subtracting a best rank-1 approximation may increase tensor rank. Linear Algebra and Its Applications, 433(7), 1276–1300. https://doi.org/10.1016/j.laa.2010.06.027

[6] Lickteig, T. (1985). Typical tensorial rank. Linear Algebra and Its Applications, 69, 95–120. https://doi.org/10.1016/0024-3795(85)90070-9

[7] Strassen, V. Gaussian elimination is not optimal. Numer. Math. 13, 354–356 (1969). https://doi.org/10.1007/BF02165411

[8] Lim, L.-H. (2021). Tensors in computations. Acta Numerica, 30, 555–764. https://doi.org/10.1017/s0962492921000076

[9] Håstad, J. (1990). Tensor rank is NP-complete. Journal of Algorithms, 11(4), 644–654. https://doi.org/10.1016/0196-6774(90)90014-6

[10] Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. Psychometrika, 1(3), 211–218. https://doi.org/10.1007/bf02288367

[11] Zuiddam, J. (2017). A note on the gap between rank and Border Rank. Linear Algebra and Its Applications, 525, 33–44. https://doi.org/10.1016/j.laa.2017.03.015

[12] de Silva, V., & Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM Journal on Matrix Analysis and Applications, 30(3), 1084–1127. https://doi.org/10.1137/06066518x

[13] Rosowski, A. (2023). Fast commutative matrix algorithms. Journal of Symbolic Computation, 114, 302–321. https://doi.org/10.1016/j.jsc.2022.05.002

[14] Williams, V. V., Xu, Y., Xu, Z., & Zhou, R. (2024). New Bounds for matrix multiplication: From alpha to omega. Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 3792–3835. https://doi.org/10.1137/1.9781611977912.134