# SQ Learning for Tensor PCA

Arvind Ramaswami

December 5, 2021

## 1 Abstract

Tensor PCA was introduced by Richard and Montanari in 2015 [1]. We look at a statistical query approach for proving lower bounds in terms of the number of samples for polynomial-time statistical query estimators to this problem from Dudeja, Hsu 2021 [2]. The SQ approach has gaps compared to the state of the art for Tensor PCA.

## 2 Notations

Here are some notations that are used in later parts of this report.

For two vectors $v, w \in RR^d$, define $v \otimes w$ to be a tensor $T$ where $T_{ij} = v_i w_j$.

This can be generalized to $T = v_1 \otimes v_2 \otimes ... \otimes v_k$ (i.e. $\bigotimes^k v_i$) where $T_{i_1 i_2 ... i_k} = v_{i_1} v_{i_2} ... v_{i_k}$

Let $\bigotimes^k \mathbb{R}^d$ be the space of tensors in $\mathbb{R}^d$ with $k$ indices and dimension $d$.

Given vectors $v_1, ..., v_k$, consider a mapping $\pi : [k] \to [K]$ (where $K < n$). In the lower bounds later in this paper, it is assumed that the vectors are of the form $v_{\pi(1)}, v_{\pi(2)}, ..., v_{\pi(k)}$, and we assume that this mapping $\pi$ is known beforehand (while it is not necessarily known by the algorithm, it can still imply lower bounds). We denote $o$ as $\{n \in [K] : |\pi^{-1}(n)| \text{ is odd } \}$ – the theorem that will be presented has a dependence on $o$, and $o$ can intuitively be thought of as a parameter that controls the asymmetry of $v_{\pi(1)} \otimes v_{\pi(2)} \otimes ... \otimes v_{\pi(k)}$.

Define $PARITY(.,.) : \bigotimes^k \mathbb{R}^d \times [k] \to \{0,1\}^d$ as follows:

$$PARITY(T, l)_i = \bigoplus_{j_1, ..., j_{l-1}, j_{l+1}, ..., j_k \in [d]} T_{j_1 j_2 \cdots j_{l-1} i j_{l+1} \cdots j_k}$$

For a $\pi$ as was defined above, denote $PARITY_\pi(T, i) := \bigoplus_{l:\pi(l)=i} PARITY(T, l)$

## 3 Tensor PCA Motivation and Problem Statement

We first look at the standard version of PCA. Oftentime, one wants to compute a rank-$k$ approximation of a data matrix $A$, i.e. $\sum_{i=1}^{k} \sigma_i u_i v_i^T$. (where $|\sigma_1| \geq ... \geq |\sigma_k|$ are the $k$ largest singular values of $A$, and $u_i, v_i$ are the corresponding left and right singular vectors for each $\sigma_i$). This is often useful for finding principal vectors to make analyzing data more tractable. We will just focus on rank 1 approximations here (which in the case of standard PCA is just $\sigma_1 u_1 v_1^T$).

There has been much interest in tensors for reasons such as data that requires indexing multiple times, and tensors also allow one to study behavior of higher moments of a random vector. PCA has a generalization to tensors.

We now move on to the formal problem statement of Tensor PCA. Dudeja 2021 [2] mentions two Tensor PCA problems, a testing problem and an estimation problem. The paper defines them as follows:

**Definition** *(Tensor PCA Testing Problem)*: One is given samples $T_{1:n} \in \bigotimes^k \mathbb{R}^n$. We want to determine whether $T \sim D_0$, or $T \sim D_V$ for some $D \in \mathcal{D}$. Here, $D_0$ is a distribution over tensors $T$ such that $T_{i_1 \ldots i_k} \sim \mathcal{N}(0, 1)$. And each $D_V \in \mathcal{D}$ is a distribution over tensors $T$, where $V$ is a rank 1 tensor with $||V|| = \sqrt{d^k}$, we have $T_{i_1 \ldots i_k} \sim \mathcal{N}(\frac{V_{i_1 \ldots i_k}}{\sqrt{d^k}}, 1)$

**Definition** *(Tensor PCA Estimation problem Problem)* Let $D_V$ be as defined in the previous definition. Given samples $T_{1:n} \in D_V$, one wants to find a $\hat{V}$ such that $||\hat{V} - V|| \leq \frac{1}{4}$

*Remark:* Many of the older works like [1] define the problem as finding a rank-1 approximation for a single tensor $T$, rather than a collection of tensors $T_{1:n}$. This tensor $T$ is assumed to be from a distribution where elements are of the form $\beta V + G$, where $V$ is a rank-1 tensor and $G$ is Gaussian noise, which is more specifically defined in that paper. $\beta$ is defined as the "signal-to-noise ratio," and it can be shown that $n$ samples in the problem above corresponds to a signal-to-noise ratio on the order of $\sqrt{n}$.

## 4 Standard Tensor PCA Results

Montanari 2014 [1] study the symmetric version of this problem, i.e. where $V = \bigoplus^k v_0$ for some $v_0$, and some others study the asymmetric case, where $V$ is any rank 1 tensor $v_1 \oplus ... \oplus v_k$ where $||v_i|| \leq \sqrt{d}$.

There are methods in [1] like tensor unfolding (decomposing a tensor into a $d^{\lfloor \frac{k}{2} \rfloor} \times d^{\lfloor \frac{k}{2} \rfloor}$ matrix, reducing this to a problem where $k = 2$, which produces polynomial time algorithms for $n = \Omega(d^{\lceil \frac{k}{2} \rceil})$ and runs exponentially for larger $n$. There are also other proposed methods like tensor power iteration, but have a lower complexity.

# 5 Statistical Query Learning

Statistical query learning is learning in a setting with some certain restrictions; one does not have direct access to samples from a distribution, but can use oracles to query the expectation of an arbitrary function within a certain tolerance. This report will discuss about Tensor PCA from a statistical query learning perspective.

**Definition** *(VSTAT oracle)* Given a distribution $D$ with sample space $X$ and an integer $t$, denote $VSTAT(t)$ as an oracle that given a query $q : X \to [0, 1]$, returns a real number in the interval $[p - \tau, p + \tau]$ where $\tau \leq \max(\frac{1}{t}, \sqrt{\frac{p(1-p)}{2}})$, and $p = \mathbb{E}_{X \sim D} q(X)$

The above tolerance $\tau$ can be intuitively be thought of as the high-probability regions of $n$ Bernoulli variables.

# 6 SQ results for Tensor PCA

We will discuss the results from Dudeja 2021 [2]. The following lower bound is presented in that paper (this bound gives conditions that require an exponential number of queries). Since this is the main (and only) theorem that will be described in depth in this paper, we will refer to it as "Theorem 4" to be consistent with Dudeja 2021 [2]

**Theorem** *(Theorem 4 from Dudeja 2021 [2])* Let $C_0 \geq 0$, and let $\epsilon \in (0, 1)$. For the following problems:

- Tensor PCA with $o = 0$ and sample size $n \leq C_0 d^{k/2-t}$
- Tensor PCA estimation with $o = 0$ and sample size $n \leq C_0 d^{\frac{k+2}{2}-\epsilon}$
- Tensor PCA testing or estimation with sample size $n \leq C_0 d^{\frac{k+2}{2}-\epsilon}$

For any $L \in \mathbb{N}$, there exists constants $c_1(k, L, \epsilon, C_0) > 0$ and $C_1(k, L, \epsilon) < \infty$ such that for all $d \geq c_1(k, L, \epsilon)$, any SQ algorithm that solves any of the above must make at least $c_1(k, L, \epsilon, c_0) d^L$ queries.

The paper also shows that there are statistical query algorithms with matching upper bounds (i.e. polynomial time algorithms for $n$ above the threshold in the above theorem); more details can be seen there [2].

The result for $o \geq 1$ is pessimistic in the sense that it doesn't match state of the art bounds. When the tensor is asymmetric, $o = k$, giving that $n \leq C_0 d^{k-\epsilon}$, which is worse than the well-known bound $d^{\frac{k}{2}}$. The implications will be discussed in the final section of this report.

### 6.1 Proof ideas

This section will focus on the high level proof ideas for the above theorem (i.e. exponential lower bounds for query complexity given a limited number of samples).

One way such lower bounds are proven in general is to show that no matter what sequence of queries is used, one can only eliminate a small fraction of distributions at each step.

Past works (including Kearns 98 [3], and more recently Feldman 2015 and 2018 [4] [5]), use some notion of *statistical dimension* to prove lower bounds. On a high level, statistical dimension is a measure of the complexity of the candidate distributions, for example, the maximum number of functions that have low pairwise correlation (defined in [3]). There have been many notions of varying strenghts of statistical dimension, such as SQ-DIM [3] and SDA (Statistical Dimension with average correlation) that have been used to improve lower bounds for problems like planted clique [4]. Dudeja 2021 focuses on using SDN (statistical dimension with discrimination norm), taken from Feldman et. 2018 (which has been used to show bounds for problems like $k - SAT$), defined as follows, to prove the TPCA lower bounds:

**Definition** *(statistical dimension with discrimination norm)*Let $\epsilon > 0$, let $D_*$ be a reference distribution, and $\mathcal{H}$ be a finite collection of distributions such that $D_* \in \mathcal{H}$. We define the *statistical dimension with discrimination norm* $\epsilon$ for $D_*$ versus $\mathcal{H}$ ($SDN(D_*, \mathcal{H}, \epsilon)$) to be the largest $m$ such that for any $S \subset H$ with $|S| \geq \frac{|\mathcal{H}|}{m}$,

$$\frac{1}{|S|} \sum_{D \in S} |\mathbb{E}_D q(T) - \mathbb{E}_{D_*} q(T)| \leq \epsilon \forall q : \bigotimes^k \mathbb{R}^d \to R, \mathbb{E}_{D_*} q^2(T) = 1$$

Feldman et al. 2018 [5] has done work on finding a lower bound for the number of queries as a function of the statistical dimension defined above. Using this, along with some other techniques (Fourier analysis and hypercontractivity), Dudeja 2021 [2] proves the following concentration bound:

**Proposition** *(Proposition 6 from Dudeja 2021)* Let $\epsilon > 0$, $u \in \mathbb{N}, u \geq 2$ be arbitrary. For the reference distribution $D_0$, and for every query $q$ with $\mathbb{E}_{D_0} q^2(T) = 1$, we have

$$\mathbb{P}\left[\mathbb{E}_D q(T) - \mathbb{E}_{D_0} q(T) \rangle \, \epsilon\right] \leq \frac{e}{\epsilon^2} \max_{\ell_1, \ldots, \ell_k \in \mathbb{N}_0} \left( (u-1)^{\ell_1 + \ldots + \ell_k} p_\pi(\ell_{1:k}) \right)$$

where

$$p_\pi(\ell_1, ..., \ell_k) := \sum_{c \in \bigoplus^k \mathbb{N}_0 : ||c||_1 \geq 1, PARTITY_\pi(c,i) = (1_{\ell_i}, 0) \forall i \in [k]} \frac{1}{d^{k||c||_1} c!}$$

Here, $c!$ denotes the product of the entrywise factorials of $c$. Setting $\epsilon = \frac{1}{\sqrt{n}}$ has the implication that if $|\mathbb{E}_{D_V} q(T) - \mathbb{E}_{D_0} q(T)| \geq n^{-\frac{1}{2}}$, then the difference in the expectation of $Q(T)$ on $D_0$ and $D$ is greater than that which is permitted by $VSTAT(n)$ (since $VSTAT(n)$ has tolerance at most $\frac{n}{\frac{1}{2}}$), so one can distinguish between $D$ and $D_0$. In order to prove an exponential lower bound on the number of queries, we want the number of candidate distributions $D_V$ from $\mathcal{D}$ that are eliminated after each query to be small, i.e. $o(d^{-t})$. Thus, it makes sense to try to find a tight upper bound for the right-hand side of the above proposition, i.e. for $p_\pi$.

Using some probabilistic arguments, Dudeja 2021 gives the following upper bounds for $p_\pi$ (for some constant $C_k$):

$$p_\pi(\ell_1, ..., \ell_k) \leq \begin{cases} C_k d^{\frac{-k}{2}}, & \ell_1 = ... = \ell_k = 0, o = 0 \\ C_k d^{-k}, & \ell_1 = ... = \ell_k, o \geq 1 \\ C_k d^{\frac{-}{k+\sum_{i=1}^k \ell_i}2}, & \ell_i \leq |\pi^{-1}(i)|, \ell_i + |\pi^{-1}(i)| \text{ even} \\ C_k d^{-\frac{\ell_1 \vee ... \vee \ell_k}{2}}, & \ell_1 \vee ... \vee \ell_k \geq 2k \\ C_k d^{-k}, & \text{otherwise} \end{cases}$$

Dudeja proves an analogous version of the proposition for the tensor estimation problem, and substituting $n = O(\epsilon^{\frac{1}{2}})$ gives the desired bound in Theorem 4.

Dudeja spends the rest of the paper giving matching upper bounds of SQ algorithms and analyzing what happens when the variance is modified. More information about that can be found in that paper [2]

# 7 Limitations and remaining challenges

In the asymmetric case (especially evident when setting $o = d$ in Theorem 4), we have that when using a VSTAT oracle, one must have must have $\Omega(d^k)$ samples in order to solve Tensor PCA efficiently. However, the state of the art gives that the bound is $\Omega(d^{\frac{k}{2}})$, so there is a gap. The guarantees for the VSTAT lower bounds in Dudeja 2021 [2] are definitely tight, given that there are matching upper bounds. This shows a weakness in the $VSTAT(n)$ oracle to capture the underlying structure in the problem.

Feldman et al. 2018 [5] use an oracle called $MVSTAT$, which includes an extra parameter $\ell$ that controls the strength of the oracle $VSTAT$ and can be used to prove tighter lower bounds for statistical query problems. This has been used for $k$-CSP problems. Analyzing the guarantees of $MVSTAT$ on the Tensor PCA problem is still open, and a potential future direction could be to find bounds for some SQ oracle stronger than $VSTAT$ on TPCA and see if it is able to match the state of the art.

## References

[1] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. *CoRR*, abs/1411.1076, 2014.

[2] Rishabh Dudeja and Daniel Hsu. Statistical query lower bounds for tensor pca, 2021.

[3] Michael Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, nov 1998.

[4] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, and Santosh S. Vempala. The complexity of statistical algorithms. *CoRR*, abs/1201.1214, 2012.

[5] Vitaly Feldman, Will Perkins, and Santosh S. Vempala. On the complexity of random satisfiability problems with planted solutions. *CoRR*, abs/1311.4821, 2013.