COMS 4774 Spring 2021 Scribe: Sean Harris (sh3715)
April 1, 2021 Editor(s): Karen Jan (kj2546), Jay Pakala (jp4102)

Tensor Rank and Decomposition

1 Tensor Rank

1.1 Generic Rank

A rank-1, order-p tensor is a p-linear form composed of the tensor product of at least p vectors. In the case where $p = 3$:

$$T = u^{(1)} \otimes u^{(2)} \otimes u^{(3)} \in (\mathbb{R}^n)^{\otimes 3}$$

where $u^{(i)} \in \mathbb{R}^n \neq \vec{0}$ for each $i = 1, 2, 3$. T is a trilinear form, so on vector inputs $x, y, z \in \mathbb{R}^n$ we have:

$$T(x, y, z) = \langle u^{(1)}, x\rangle \cdot \langle u^{(2)}, y\rangle \cdot \langle u^{(3)}, z\rangle$$

Like matrices, rank r tensors can be expressed as a sum of r rank-1 tensors. Thus the rank of $T \in (\mathbb{R}^n)^{\otimes p}$ is the smallest r such that there exists $T_1, ...T_r$ where $\text{rank}(T_i) = 1$ and $T = \sum\limits_{i=1}^{r} T_i$.

1.2 Border Rank

Border rank is another way to define the rank of a tensor. The border rank of $T \in (\mathbb{R}^n)^{\otimes p}$ is the smallest r such that there exists a sequence of rank r tensors $(T_k)_{k\in\mathbb{N}} = \sum\limits_{i=1}^{r} T_{k,i}$ where $\lim_{k\to\infty} T_k = T$. If we get better and better approximations of the tensor, we hope to eventually converge to the actual tensor in the space of the generic rank of the tensor. However, this is not guaranteed to happen. Unlike matrices, order $p \geq 3$ tensors do not necessarily have a border rank equal to their generic rank.

To illustrate: take any distinct $u, v \in S^{n-1}$ and define the following rank-3 tensor:

$$T := u \otimes u \otimes v + u \otimes v \otimes u + v \otimes u \otimes u$$

and the rank-2 tensor:

$$T_k := k(u + \frac{1}{k}v) \otimes (u + \frac{1}{k}v) \otimes (u + \frac{1}{k}v) - k \cdot u \otimes u \otimes v$$

To confirm the ranks of these tensors, note that $k(u+\frac{1}{k}v)\otimes(u+\frac{1}{k}v)\otimes(u+\frac{1}{k}v)$, $k\cdot u\otimes u\otimes v$, and $u\otimes u\otimes v$ are all tensor products of three vectors, and thus rank-1 tensors.

We claim that $\lim_{k\to\infty} T_k = T$:.

$$\begin{aligned}
T_k &= k(u+\frac{1}{k}v)\otimes(u+\frac{1}{k}v)\otimes(u+\frac{1}{k}v) - k\cdot u\otimes u\otimes v \\
&= (k\cdot u\otimes u\otimes u) + (u\otimes u\otimes v + u\otimes v\otimes u + v\otimes u\otimes u) + (k^{-2}\cdot v\otimes v\otimes v) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - (k\cdot u\otimes u\otimes u) \\
&= (u\otimes u\otimes v + u\otimes v\otimes u + v\otimes u\otimes u) + (k^{-2}\cdot v\otimes v\otimes v) \\
&= T + (k^{-2}\cdot v\otimes v\otimes v)
\end{aligned}$$

It is then easy to see that this sequence of rank-2 tensors converges on the rank-3 tensor T:

$$\lim_{k\to\infty} T_k = \lim_{k\to\infty}[T + (k^{-2}\cdot v\otimes v\otimes v)] = T + \lim_{k\to\infty}(k^{-2}\cdot v\otimes v\otimes v) = T$$

Since there is no unifying notion of rank for tensors, it is important to keep in mind which variety is being used in an application.

2 Tensor Decomposition

2.1 Comparison to Matrices

Consider a set of orthonormal basis vectors $v_1 ... v_n \in \mathbb{R}^n$ which compose some matrix M:

$$M = \sum_{i=1}^{n} v_i v_i^\top = \sum_{i=1}^{n} v_i \otimes v_i = \sum_{i=1}^{n} v_i^{\otimes 2} = I_n$$

If we are given such a matrix, clearly it is not possible to determine the values of $v_1 ... v_n$, as any orthonormal basis could produce M.

It turns out this is a property specifically of matrices, and tensors of order $p \geq 3$ behave differently.

$$T := \sum_{i=1}^{n} v_i \otimes v_i \otimes v_i = \sum_{i=1}^{n} v_i^{\otimes 3}$$

You *can* uniquely identify $v_1, ..., v_n$ from T, which need not be orthonormal, the only assumption here is linearly independent v_i.

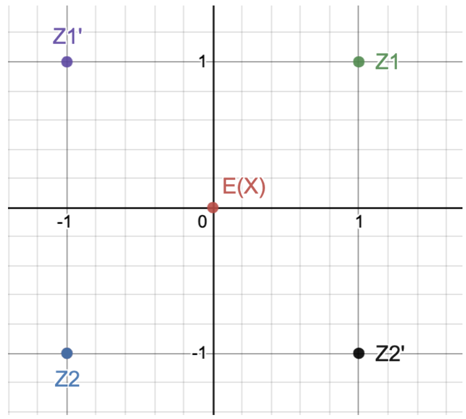

Figure 1: Points Z1 and Z2 have the same expected value and variance as points Z1' and Z2'

2.2 High-dimensional Support Recovery

We return to the support recovery problem, but now in higher dimensions. We have a random vector $\vec{X} \in \mathbb{R}^d$ supported on k distinct points $\vec{z}_1, ..., \vec{z}_k \in \mathbb{R}^d$ with distribution $\mathbb{P}[\vec{X} = \vec{z}_i] = w_i > 0$. The goal is to recover the k unknown parameters $\{w_1, \vec{z}_1, ..., w_k, \vec{z}_k\}$ using only the higher-order moments of $\vec{X}$. How many higher-order moments suffice?

- **First moment:** $\mathbb{E}(\vec{X}) = \sum_{i=1}^{k} w_i \vec{z}_i$. This does not suffice as multiple pairs of points can have the same expected value, as highlighted in Figure 1.
- **Second moment:** $\mathbb{E}(\vec{X} \otimes \vec{X}) = \sum_{i=1}^{k} w_i \vec{z}_i \otimes \vec{z}_i$. Generally this does not suffice.[1] Consider the case where $w_1, ..., w_i = \frac{1}{k}$. Any orthonormal $z_1, ..., z_i$ will have the same second moment matrix.

Since a one dimensional support recovery problem required on the order of $2k-1$ moments of the random variable to solve all unknowns, we might expect that expanding this problem into d dimensions would only make it harder. In fact, by taking advantage of those extra dimensions, and assuming $\vec{z}_i$ to be linearly independent (i.e. not on the same line), the first three moments suffice to recover the parameters.

[1]In the case where $\vec{z}_1, ..., \vec{z}_k$ are orthonormal and $w_1, ..., w_k$ are distinct, the first two moments suffice to recover the parameters, via eigendecomposition.

The intuition here is that the third-order moment tensor $\mathbb{E}(\vec{X} \otimes \vec{X} \otimes \vec{X}) = \sum_{i=1}^{k} w_i \vec{z}_i \otimes \vec{z}_i \otimes \vec{z}_i$ provides on the order of d^3 constraints on the unknown parameters, compared to the one constraint in the $d = 1$ case.

2.3 Jennrich's algorithm

Jennrich's algorithm retrieves $\{w_1, \vec{z}_1, ..., w_k, \vec{z}_k\}$ using only the first three moments of $\vec{X}$.

$$S := \mathbb{E}[\vec{X} \otimes \vec{X}] = \sum_{i=1}^{k} w_i \vec{z}_i \otimes \vec{z}_i$$

$$T := \mathbb{E}[\vec{X} \otimes \vec{X} \otimes \vec{X}] = \sum_{i=1}^{k} w_i \vec{z}_i \otimes \vec{z}_i \otimes \vec{z}_i$$

Here S is the second-order moment and T is the third-order moment. It is assumed $\vec{z}_1, ..., \vec{z}_i$ are linearly independent.

S can be rewritten as $S = ZWZ^\top$, where

$$Z = [\vec{z}_1 | \cdots | \vec{z}_k]$$

$$W = \operatorname{diag}(w_1, ..., w_k)$$

Note that while all $\vec{z}_i$ are linearly independent, they are not orthonormal, and this is not necessarily the eigendecomposition.

Since T is a trilinear form $T : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, meaning it is linear with respect to all three inputs, we can hold one input constant and T becomes a bilinear form, or matrix:

$$T_u := T(u, \cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

$$T_u(e_i, e_j) = \sum_{i=1}^{k} w_t \langle \vec{z}_t, u \rangle \langle \vec{z}_t, e_i \rangle \langle \vec{z}_t, e_j \rangle$$

$$\begin{aligned} T_u &= \sum_{i=1}^{k} w_t \langle \vec{z}_t, u \rangle \vec{z}_t \otimes \vec{z}_t \\ &= Z W_u Z^\top \end{aligned}$$

where $W_u = \text{diag}(w_1\langle \vec{z}_1, u\rangle, ..., w_k\langle \vec{z}_k, u\rangle)$. This is also called flattening a tensor. One can imagine the three-dimensional array of numbers T being flattened into a two-dimensional matrix T_u. Since all $\vec{z}_i$ are linearly independent, Z is full rank, and we can exploit the properties of the pseudoinverse of a full rank matrix:

$$\begin{aligned}
T_u S^\dagger &= (ZW_u Z^\top)(ZWZ^\top)^\dagger && (1)\\
&= (ZW_u Z^\top)(Z^{\top\dagger} W^{-1} Z^\dagger) && (2)\\
&= ZW_u W^{-1} Z^\dagger && (3)\\
&= Z \ \text{diag}(\langle \vec{z}_1, u\rangle, ..., \langle \vec{z}_k, u\rangle) Z^\dagger && (4)\\
&= ZD_u Z^\dagger && (5)
\end{aligned}$$

Where $D_u = \ \text{diag}(\langle \vec{z}_1, u\rangle, ..., \langle \vec{z}_k, u\rangle)$.

Note that in (2), $W^\dagger = W^{-1}$ because W is a square diagonal matrix with a positive diagonal. In (3), since Z is full rank, $Z^\top Z^\dagger = I$. In (4), W^{-1} cancels out all the w_i in W_u.

Observe that $T_u S^\dagger$ has eigenvectors of $\vec{z}_i$:

$$\begin{aligned}
T_u S^\dagger \vec{z}_i &= T_u S^\dagger Z \vec{e}_i && (6)\\
&= ZD_u Z^\dagger Z \vec{e}_i && (7)\\
&= ZD_u \vec{e}_i && (8)\\
&= \langle \vec{z}_i, \vec{u}\rangle Z \vec{e}_i && (9)\\
&= \langle \vec{z}_i, \vec{u}\rangle \vec{z}_i && (10)
\end{aligned}$$

Note that $M\vec{e}_i$ selects the ith column of M.

There is the possibility that $\langle \vec{z}_i, \vec{u}\rangle$ is not distinct, meaning not all eigenvectors can be uniquely determined. This can be handled by selecting $\vec{u}$ randomly from a continuous distribution, i.e. $\vec{u} \sim \text{Unif}(S^{d-1})$. Then, with probability 0, $\vec{u}$ will have the exact same inner product with any two distinct $\vec{z}_i$.

This means we can recover a scaling of all $z_1, ..., z_i$ with a flattened third-order moment tensor and the pseudoinverse of the second-order moment matrix. A z_i might be scaled by $\sigma \in \mathbb{R} \neq 0$ because any vector in the span of an eigenvector is also an eigenvector. These unwanted scaling factors can be found by calculating:

$$\frac{\langle \sigma_i z_i, \vec{u}\rangle}{\langle z_i, \vec{u}\rangle} = \sigma_i$$