The analytic rank of a tensor

by

Marcel K. Goh

22 April 2022

Note. These notes are more or less a retelling of some results in a 2019 paper entitled "The analytic rank of tensors and its applications" by S. Lovett. I rearranged results in an order that makes more sense to me. I also go more deeply into definitions (for my own sake) and skip fewer steps, so for other students this may be easier to follow than the original paper.

1. Introduction

There are various compatible definitions of the *rank* of a matrix. The one that extends most easily to the context of tensors, which we define later, is the following. An $m \times n$ matrix A is said to be *rank one* if there exist vectors $u \in \mathbf{F}^m$ and $v \in \mathbf{F}^n$ such that $A = uv^{\mathrm{T}}$. The *rank* of a general matrix A is the minimum number k such that we can write $A = A_1 + \cdots + A_k$, where A_i is a rank one matrix for all $1 \le i \le k$.

In a first course on linear algebra, one usually learns that the rank of a matrix is the dimension of its column (or row) space. To see that the above definition of rank is equivalent, let A be a rank-k matrix and let $B = \{b_1, \ldots, b_k\}$ be a basis of its column space. Since every column of A can be written as a linear combination of vectors in B, so there is a $k \times n$ matrix C such that $A = BC$. Now letting $c_1, \ldots, c_k$ be the rows of C, we have

$$A = b_1 {c_1}^{\mathrm{T}} + \cdots + b_k {c_k}^{\mathrm{T}},$$

so the rank of A is at most k. On the other hand, if

$$A = u_1 {v_1}^{\mathrm{T}} + \cdots + u_{k'} {v_{k'}}^{\mathrm{T}},$$

for some $k' < k$, then for all $x \in \mathbf{F}^n$, we have

$$Ax = u_1 {v_1}^{\mathrm{T}} x + \cdots + u_{k'} {v_{k'}}^{\mathrm{T}} x.$$

Since ${v_i}^{\mathrm{T}} x$ is a scalar for all $1 \le i \le k'$, we conclude that $u_1, \ldots, u_{k'}$ span the image of A and the column space of A is at most $k' < k$.

Any $m \times n$ matrix A gives a bilinear map from $\mathbf{F}^m \times \mathbf{F}^n$ to $\mathbf{F}$ by taking $(x, y) \mapsto xAy$. We extend this to more than two vector spaces by defining an *order-d tensor* to be a multilinear map $T : V_1 \times \cdots \times V_d \to \mathbf{F}$, where V_i is a vector space over $\mathbf{F}$ for all $1 \leq i \leq d$. From here on out, we restrict ourselves to the case where each V_i has the same dimension n and can thus be identified with $\mathbf{F}^n$. Then there exist n^d scalars $\{T_{j_1,\ldots,j_d}\}_{j_1,\ldots,j_d \in [n]}$ such that for all $x_1, \ldots, x_d \in \mathbf{F}^n$,

$$T(x_1, \ldots, x_d) = \sum_{j_1,\ldots,j_d \in [n]} T_{j_1,\ldots,j_d} x_{1,j_1} \cdots x_{d,j_d},$$

where $x_{i,k}$ denotes the kth component of the vector x_i. There is thus a one-to-one correspondence between order-d tensors and d-dimensional arrays of scalars (in our setting each dimension has size n). If T is an order-d tensor and T' is an order-d' tensor, then we can form a tensor of order $d + d'$ from the $(d + d')$-dimensional array of scalars

$$\{T_{i_1,\ldots,i_d} T'_{j_1,\ldots,j_{d'}}\}_{i_1,\ldots,i_d,j_1,\ldots,j_{d'} \in [n]}.$$

This tensor is denoted $T \otimes T'$ and is called the *tensor product* of T and T'.

Now we say that an order-d tensor T is *partition rank one* if there exists $A \subseteq [d]$ with $0 < |A| < d$, as well as an order-$|A|$ tensor T_1 and an order-$(d - |A|)$ tensor T_2 such that T can be written as

$$T(x_1, \ldots, x_d) = T_1(x_i : i \in A) T_2(x_i : i \notin A).$$

The *partition rank* $\mathrm{prk}(T)$ of a general tensor T is the minimum k such that T can be written as a sum of k partition rank one tensors. Note that in the case $d = 2$ this reduces to the ordinary matrix rank.

The cap set problem. The partition rank was introduced to study the cap-set problem, and here we shall sketch how it applies. A *cap set* is a subset $A \subseteq \mathbf{F}_3^n$ such that for every triple $(x, y, z) \in A$ of pairwise distinct elements, $x + y + z \neq 0$. It was shown by T. C. Brown and J. P. Buhler that, loosely speaking, cap sets have zero density.

Theorem C (*Brown–Buhler*, 1986). *For every $\delta > 0$ there exists n such that every subset $A \subseteq \mathbf{F}_3^n$ with $|A| \geq \delta 3^n$ contains three pairwise distinct elements x, y, and z with $x + y + z = 0$.*

A later paper by R. Meshulam gave better quantitative bounds on n with respect to δ; namely, it was shown that we need only take $n > 2/\delta$. This means that if A is a cap set in $\mathbf{F}_3^n$, then $|A| \leq 2 \cdot 3^n/n$. However, it was long suspected that this bound could be improved to $|A| \leq O(c^n)$ for some $c < 3$. This was finally proved in 2017 by J. S. Ellenberg and D. Gijswijt, and T. Tao showed in a blog post (dated 18 May 2016) that thir proof can be restated in terms of the partition rank of a function in 3 variables. This can actually be modified so the function is a 3-tensor, but to just get the general idea, let us extend our definition of partition rank to general functions of three variables temporarily.

Given $A \subseteq \mathbf{F}_3^n$, let $T : V^3 \to \mathbf{F}_3$, where $V = \mathbf{F}_3^{\mathbf{F}_3^n}$, be given by

$$T(e_a, e_b, e_c) = \begin{cases} 1, & \text{if } a + b + c = 0; \\ 0, & \text{otherwise,} \end{cases}$$

for basis vectors e_v and extended to all other vectors by linearity. (The function e_v has $e_v(v) = 1$ and $e_v(x) = 0$ for all $x \neq v$.) Now for a tensor $T : (\mathbf{F}^X)^d \to \mathbf{F}$, we say that a set $A \subseteq X$ is an *independent set in* T if for all $i_1, \ldots, i_d \in A$, the condition that the coefficient $T_{i_1,\ldots,i_d}$ be nonzero is equivalent to $i_1 = \cdots = i_d$. We then give an upper bound on the size of a cap set by proving that

i) if A contains no nontrivial solutions to $x+y+z = 0$, then A is an independent set in T;
ii) if A is an independent set in T then $\operatorname{prk}(T) \geq |A|$; and
iii) the partition rank of T is low.

In these notes, we aim to show that this general strategy may be performed with the partition rank replaced by something called the analytic rank.

The analytic rank. In a 2011 paper, W. T. Gowers and J. Wolf introduced another definition of rank that is Fourier-analytic in nature. Now we require the field $\mathbf{F}$ to be finite, and let $\chi : \mathbf{F} \to \mathbf{C}$ be any nontrivial additive character. Recall that for such a character, $\mathbf{E}_{a\in\mathbf{F}}\, \chi(a) = 0$. The *bias* of a tensor $T : V^d \to \mathbf{F}$ is the average

$$\operatorname{bias}(T) = \mathbf{E}_{x\in V^d}\, \chi\big(T(x)\big).$$

Note that if T is a linear form (i.e., an order-1 tensor) that is not identically zero, then $\operatorname{bias}(T) = 0$, since we can bring the sum inside all three functions and the sum over all elements a vector space over a finite field is zero. If T is identically zero then $\operatorname{bias}(T) = 1$. Now to see that the bias of a tensor is always in $(0, 1]$, note that if we fix any $(x_2, \ldots, x_d) \in V^{d-1}$, then $T(x_1, x_2, \ldots, x_d)$ becomes a linear form (order-1 tensor) in x_1 and

$$\begin{aligned}\operatorname{bias}(T) &= \mathbf{E}_{x_2,\ldots,x_d\in V}\, \mathbf{E}_{x_1\in V}\, \chi\big(T(x_1, \ldots, x_d)\big)\\ &= \mathbf{P}_{x_2,\ldots,x_d\in V}\big\{T(x_1, \ldots, x_d) \equiv 0\big\},\end{aligned}$$

from our earlier observation about order-1 tensors.

The *analytic rank* is defined to be the quantity

$$\operatorname{ark}(T) = -\log_{|\mathbf{F}|} \operatorname{bias}(T);$$

since $\operatorname{bias}(T) \in (0, 1]$ we have $\operatorname{ark}(T) \geq 0$. In the case of order-2 tensors, the analytic rank is once again equivalent to ordinary matrix rank. To see this, suppose that $T : (\mathbf{F}^n)^2 \to \mathbf{F}$ is defined as $T(x, y) = \sum_{i=1}^{r} x_i y_i$. Then $\operatorname{bias}(T)$ is the probability that, fixing y, the linear form $T(x, y)$ is identically zero. This is equivalent to every coordinate of y being zero, which happens with probability $1/|\mathbf{F}|^r$, and hence we see that $\operatorname{ark}(T) = r$.

2. Subadditivity of analytic rank

The goal of this section is to prove that if T and S are tensors, then $\operatorname{ark}(T+S) \leq \operatorname{ark}(T) + \operatorname{ark}(S)$. Our first small lemma is the following.

Lemma 1. *Let $W_0, W_1, \ldots, W_n : \mathbf{F}^m \to \mathbf{F}$ be functions. Let functions $A, B : \mathbf{F}^n \times \mathbf{F}^m \to \mathbf{F}$ be given by*

$$A(x,y) = \sum_{i=1}^{n} x_i W_i(y) \qquad \text{and} \qquad B(x,y) = A(x,y) + W_0(y).$$

Then

$$\bigl|\mathrm{bias}(B)\bigr| \le \mathrm{bias}(A).$$

Proof. We expand

$$\mathrm{bias}(B) = \mathbf{E}_{x\in\mathbf{F}^n, y\in\mathbf{F}^m}\, \chi\bigl(B(x,y)\bigr) = \mathbf{E}_{y\in\mathbf{F}^m}\, \mathbf{1}_{[W_1(y)=\cdots=W_n(y)=0]} \cdot \chi\bigl(W_0(y)\bigr)$$

and by the triangle inequality,

$$\bigl|\mathrm{bias}(B)\bigr| \le \mathbf{E}_y\, \mathbf{1}_{[W_1(y)=\cdots=W_n(y)=0]} = \mathrm{bias}(A). \qquad \blacksquare$$

This lemma is used to prove a bound on the bias of a certain sum of tensors. First, we introduce some notation. With some d fixed, we let $\mathbf{x} = (x_1, \ldots, x_d)$ and similarly for $\mathbf{y} = (y_1, \ldots, y_d)$. Then for $I \subseteq [d]$, define $I^c = [d] \setminus I$ and let $\mathbf{x}_I = (x_i : i \in I)$.

Lemma 2. *Let $d \ge 1$ and for each $I \subseteq [d]$, let $R_I : V^I \to \mathbf{F}$ be an order-$|I|$ tensor. Consider the function*

$$R(\mathbf{x}) = \sum_{I\subseteq[d]} R_I(x_I).$$

Then

$$\bigl|\mathrm{bias}(R)\bigr| \le \mathrm{bias}(R_{[d]}).$$

Proof. Fix some $i \in [d]$ and write $R(\mathbf{x})$ as

$$R(\mathbf{x}) = \sum_{I\ni i} R_I(x_I) + \sum_{I\not\ni i} R_I(x_I).$$

Setting $x = x_i$ and $y = \mathbf{x}_{[d]\setminus\{i\}}$, the first sum has the form

$$\sum_{i=1}^{n} x_i W_i(y)$$

and the second sum does not depend on x at all, so we can set it to be $W_0(y)$. The previous lemma now tells us that

$$\bigl|\mathrm{bias}(R)\bigr| \le \mathrm{bias}\Bigl(\sum_{I\ni i} R_I(\mathbf{x}_I)\Bigr).$$

Now iterate this with $i = d$ all the way down to $i = 1$ (replacing d by $d-1$ each time) to get the statement of the lemma. ▮

Before proving subadditivity, we introduce another bit of notation. For $I \subseteq [d]$, let

$$T_I(\mathbf{x}, \mathbf{y}) = T(\mathbf{x}_I, \mathbf{y}_{I^c}) = T(z_1, \ldots, z_d),$$

where $z_i = x_i$ if $i \in I$ and $z_i = y_i$ if $i \notin I$. After expanding out the definition of multilinearity, we see that $T(\mathbf{x} + \mathbf{y})$ decomposes as

$$T(\mathbf{x} + \mathbf{y}) = \sum_{I \subseteq [d]} T_I(\mathbf{x}, \mathbf{y}).$$

Theorem 3. *Let $T, S : V^d \to \mathbf{F}$ be order-d tensors. Then*

$$\operatorname{ark}(T + S) \le \operatorname{ark}(T) + \operatorname{ark}(S).$$

Proof. It is enough to show that

$$\operatorname{bias}(T + S) \ge \operatorname{bias}(T) \operatorname{bias}(S).$$

We express

$$\begin{aligned} \operatorname{bias}(T) \operatorname{bias}(S) &= \Bigl(\mathbf{E}_{\mathbf{x} \in V^d}\, \chi\bigl(T(\mathbf{x})\bigr)\Bigr)\Bigl(\mathbf{E}_{\mathbf{y} \in V^d}\, \chi\bigl(T(\mathbf{y})\bigr)\Bigr) \\ &= \mathbf{E}_{\mathbf{x} \in V^d}\, \mathbf{E}_{\mathbf{y} \in V^d}\, \chi\bigl(T(\mathbf{x}) + S(\mathbf{y})\bigr) \\ &= \mathbf{E}_{\mathbf{x} \in V^d}\, \mathbf{E}_{\mathbf{y} \in V^d}\, \chi\bigl(T(\mathbf{x}) + S(\mathbf{x} + \mathbf{y})\bigr) \\ &= \operatorname{bias}\bigl(T(\mathbf{x}) + S(\mathbf{x} + \mathbf{y})\bigr), \end{aligned}$$

and by our earlier decomposition the tensor product of a sum, this gives us

$$\operatorname{bias}(T) \operatorname{bias}(S) = \operatorname{bias}\Bigl(T(\mathbf{x}) + \sum_{I \subseteq [d]} S_I(\mathbf{x}, \mathbf{y})\Bigr) = \operatorname{bias}\Bigl((T+S)(\mathbf{x}) + \sum_{I \ne [d]} S_I(\mathbf{x}, \mathbf{y})\Bigr),$$

Setting $\mathbf{y} = \mathbf{b} \in V^d$ to maximise the norm of this right-hand side, we have

$$\operatorname{bias}(T) \operatorname{bias}(S) \le \Bigl|\operatorname{bias}\Bigl((T + S)(\mathbf{x}) + \sum_{I \ne [d]} S_I(\mathbf{x}, \mathbf{b})\Bigr)\Bigr|.$$

Here $S_I(\mathbf{x}, \mathbf{b})$ is an order-$|I|$ tensor in the variables $\mathbf{x}_I$. Thus we may apply the previous lemma with

$$R_{[d]}(\mathbf{x}) = (T + S)(\mathbf{x}) \qquad \text{and} \qquad R_I(\mathbf{x}_I) = S_I(\mathbf{x}, \mathbf{b}) \text{ for } I \ne [d]$$

to get $\operatorname{bias}(T) \operatorname{bias}(S) \le \operatorname{bias}(T + S)$. ▮

As an application of this theorem, we show that common roots of tensors on a common input are positively correlated.

Corollary 4. *Let $T_1, \ldots, T_m, S_1, \ldots, S_n : V^d \to \mathbf{F}$ be order-d tensors. Then*

$$\mathbf{P}_{x \in V^d}\big\{T_1(x) = \cdots = T_m(x) = S_1(x) = \cdots = S_n(x) = 0\big\} \geq \mathbf{P}_{x \in V^d}\big\{T_1(x) = \cdots = T_m(x) = 0\big\} \cdot \mathbf{P}_{x \in V^d}\big\{S_1(x) = \cdots = S_n(x) = 0\big\}$$

Proof. Define order-$(d+1)$ tensors T and S by setting

$$T(x_0, x_1, \ldots, x_d) = \sum_{i=1}^{m} x_{0,i} T_i(x_1, \ldots, x_d)$$

and

$$S(x_0, x_1, \ldots, x_d) = \sum_{i=m+1}^{m+n} x_{0,i} S_{i-m}(x_1, \ldots, x_d).$$

The left-hand side of the desired inequality is $\operatorname{bias}(T+S)$ and the right-hand side is $\operatorname{bias}(T)\operatorname{bias}(S)$. ∎

3. Analytic rank and partition rank

In this section, we shall discuss properties of the analytic rank that allow it to be used as the partition rank was used above in the cap-set problem. First, we show that the analytic rank is bounded above by the partition rank.

Theorem 5. *Let $T : V^d \to \mathbf{F}$ be an order-d tensor. Then* $\operatorname{ark}(T) \leq \operatorname{prk}(T)$.

Proof. By Theorem 3, we only need to consider the case where T has partition rank one. So we assume that $T : V^d \to \mathbf{F}$ factors as

$$T(\mathbf{x}) = T_1(\mathbf{x}_A) T_2(\mathbf{x}_B),$$

where $A \cup B = [d]$ is a nontrivial partition of $[d]$. We shall show that $\operatorname{ark}(T) \leq 1 = \operatorname{prk}(T)$ by showing that $\operatorname{bias}(T) \geq 1/|\mathbf{F}|$. For $a, b \in \mathbf{F}$ define the function

$$F_{a,b}(\mathbf{x}) = \big(T_1(\mathbf{x}_A) + a\big)\big(T_2(\mathbf{x}_B) + b\big).$$

Letting $R_{[d]} = T$, $R_\emptyset = ab$, $R_A = bT_1(\mathbf{x}_A)$, $R_B = aT_2(\mathbf{x}_B)$, and R_I be the zero tensor (of the corresponding order) for all other $I \subseteq [d]$, we see that $R = F_{a,b}$ and by Lemma 2, $\big|\operatorname{bias}(F_{a,b})\big| \leq \operatorname{bias}(T)$. On the other hand, we have

$$\begin{aligned}
\mathbf{E}_{a,b \in \mathbf{F}} \operatorname{bias}(F_{a,b}) &= \mathbf{E}_{a,b \in \mathbf{F}} \mathbf{E}_{\mathbf{x} \in V^d}\, \chi\big((T_1(\mathbf{x}_A) + a)(T_2(\mathbf{x}_B) + b)\big) \\
&= \mathbf{E}_{a,b \in F}\, \chi(ab) \\
&= \mathbf{P}_{b \in \mathbf{F}}\{b = 0\} \\
&= \frac{1}{|\mathbf{F}|}.
\end{aligned}$$

So we may find a, b such that $\mathrm{bias}(T) \geq \left|\mathrm{bias}(F_{a,b})\right| \geq 1/|\mathbf{F}|$, which is what we wanted. ∎

Since we are often interested in finding some upper bound on the partition rank in applications, this theorem suggests that it should actually be easier to upper bound the analytic rank, though *ad hoc* methods to do so have not currently been developed.

In applications, it is also important that the partition rank of a tensor T is bounded below by the size of an independent set in T. Towards showing a version of this for the analytic rank, we first show that the analytic rank does not increase when the tensor is restricted to a subspace.

Lemma 6. *Let $T : V^d \to \mathbf{F}$ be an order-d tensor and let $U \subseteq V$ be a subspace. If $T|_U : U^d \to \mathbf{F}$ is the restriction of T to U, then $\mathrm{ark}(T|_U) \leq \mathrm{ark}(T)$.*

Proof. Let $W \subseteq V$ be a subspace so that $V = U \oplus W$. Then any $v \in V$ has a unique expression as $v = u + w$ for $u \in U$ and $w \in W$, so

$$\mathrm{bias}(T) = \mathbf{E}_{u_1,\ldots,u_d \in U}\, \mathbf{E}_{w_1,\ldots,w_d \in W}\, \chi\big(T(u_1 + w_1, \ldots, u_d + w_d)\big).$$

Fix an arbitrary choice of $w_1, \ldots, w_d \in W$. Then

$$T(u_1 + w_1, \ldots, u_d + w_d) = \sum_{I \subseteq [d]} T_I(u_I, w_{I^c}),$$

where $T_I(u_I, w_{I^c}) = T(z_1, \ldots, z_d)$ where $z_i = u_i$ if $i \in I$ and $z_i = w_i$ if $i \notin I$. By Lemma 2, we have

$$\begin{aligned}\left|\mathbf{E}_{u_1,\ldots,u_d \in U}\, \chi\big(T(u_1 + w_1, \ldots, u_d + w_d)\big)\right| &\leq \mathbf{E}_{u_1,\ldots,u_d \in U}\, \chi\big(T(u_1, \ldots, u_d)\big) \\ &= \mathrm{bias}(T|_U)\end{aligned}$$

Since this is for all $w_1, \ldots, w_d$, averaging the left-hand side over all choices and applying the triangle inequality gives us what we want. ∎

We are now able to show that for any tensor T and independent set A in T, we have $\mathrm{ark}(T) \geq c|A|$ for c depending on d and $|\mathbf{F}|$.

Theorem 7. *Let $T : (\mathbf{F}^n)^d \to \mathbf{F}$ be an order-d tensor. Assume that $A \subseteq [n]$ is an independent set in T. Then $\mathrm{ark}(T) \geq c|A|$ where*

$$c = -\log_{|\mathbf{F}|}\left(1 - \left(1 - \frac{1}{|\mathbf{F}|}\right)^{d-1}\right),$$

so that $c \geq 2^{-d}$ and $c \geq 1 - \log(d-1)/\log|\mathbf{F}|$.

Proof. Let $S : (F^A)^d \to \mathbf{F}$ be the restriction of T to $\mathbf{F}^A$. By the previous lemma, $\mathrm{ark}(S)$ is a lower bound on $\mathrm{ark}(T)$, so we shall prove the theorem by expressing $\mathrm{ark}(S)$ as a function of $|A|$. We have

$$\mathrm{bias}(S) = \mathbf{E}_{x_1,\ldots,x_d \in \mathbf{F}^A}\, \chi\Big(\sum_{i \in A} c_i \prod_{j \in [d]} x_{i,j}\Big),$$

where, since A is an independent set, $c_i \neq 0$ for all $i \in A$. But we also know that

$$\operatorname{bias}(S) = \mathbf{P}_{x_2,\ldots,x_d \in \mathbf{F}^A}\Big\{\bigcap_{i\in A}\{x_{i,2}\cdots x_{i,d} = 0\}\Big\},$$

and this is easily seen to be

$$\operatorname{bias}(S) = \left(1 - \left(1 - \frac{1}{|\mathbf{F}|}\right)^{d-1}\right)^{|A|}.$$

Taking the logarithm of both sides, we see that $\operatorname{ark}(S) = c|A|$, where c is as in the theorem statement.

By convexity,

$$c \geq -\log_2(1 - 2^{-(d-1)}) \geq 2^{-(d-1)}.$$

Also, as long as $|\mathbf{F}| \geq d$, we have

$$c \geq -\log_{|\mathbf{F}|}\big((d-1)/|\mathbf{F}|\big) = 1 - \log_{|\mathbf{F}|}(d-1). \quad \blacksquare$$

References

Thomas Craig Brown and Joe Peter Buhler, "Lines imply spaces in density Ramsey theory," *Journal of Combinatorial Theory, Series A* **36** (1984), 214–220.

Jordan Stuart Ellenberg and Dion Gijswijt, "On large subsets of $\mathbf{F}_q^n$ with no three-term arithmetic progression," *Annals of Mathematics* **185** (2017), 339–343.

William Timothy Gowers and Julia Wolf, "Linear forms and higher-degree uniformity for functions on $\mathbf{F}_p^n$," *Geometric and Functional Analysis* **21** (2011), 36–69.

Shachar Lovett, "The analytic rank of tensors and its applications," *Discrete Analysis* (2019), #7, 10 pp.

Roy Meshulam, "On subsets of finite abelian groups with no 3-term arithmetic progressions," *Journal of Combinatorial Theory, Series A* **71** (1995), 168–172.