An optimal inverse theorem for tensors over large fields

Alex Cohen (MIT)

February 23, 2022

Joint work with Guy Moshkovitz (CUNY)

Table of Contents

Overview

- A tensor is a multilinear form $T : \mathbb{F}_q^n \times \cdots \times \mathbb{F}_q^n \to \mathbb{F}_q$.
- The *analytic rank* (AR) measures randomness
- The *partition rank* (PR) measures structure
- Main result:

  $$\mathrm{PR}(T) \leq C_k \, \mathrm{AR}(T)$$

  if $|\mathbb{F}|$ is large enough.
- Main proof idea: work over the algebraic closure $\overline{\mathbb{F}}_q$, use tangent spaces / derivatives to construct small PR decomposition

The bias of polynomials

- For $f(x_1, \ldots, x_n) : \mathbb{F}_p^n \to \mathbb{F}_p$

$$\text{bias}(f) = \mathbb{E}_{x \in \mathbb{F}_q^n} \chi(f(x)), \quad \text{if } q = p \text{ is prime, } \chi(x) = e^{2\pi i x/p}$$

- For f a polynomial, $0 \leq \text{bias}(f) \leq 1$
- Bias is a measure of randomness / correlation
- **Goal:** when f is a polynomial function, explain large bias by the presence of structure

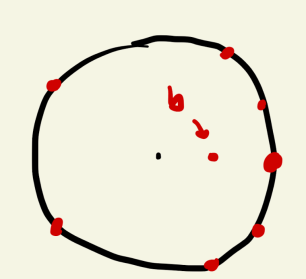


Table of Contents

Analytic rank

We focus on **multilinear forms** $T(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}) \in \mathbb{F}_q$

- A 1-tensor is a linear form $T(\mathbf{x}) = a_1 x_1 + \cdots + a_n x_n$
- A 2-tensor is a bilinear form

$$T(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t A \mathbf{y} = \sum_{ij} x_i y_j A_{ij}, \quad A \text{ a matrix}$$

- A 3-tensor is a trilinear form

$$T(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum_{ijk} T_{ijk} x_i y_j z_k$$

- k-tensors are a useful class of degree k polynomials

Definition (Analytic rank)

$$\mathrm{AR}(T) = -\log_{|\mathbb{F}|} \mathrm{bias}(T) = -\log_{|\mathbb{F}|} \mathbb{E}_{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)} \in \mathbb{F}_q^n} \chi(T(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}))$$

$$0 \le \mathrm{AR}(T) \le n, \quad \mathrm{AR}(T + L) \le \mathrm{AR}(T) + \mathrm{AR}(L), \quad \text{generically } \mathrm{AR}(T) \sim n$$

Examples

Example (2-tensors)

Let $T(\mathbf{x},\mathbf{y}) = \mathbf{x}^t A\mathbf{y}$.

$$\begin{aligned}\text{bias}(T) &= \mathbb{E}_{\mathbf{y}}\mathbb{E}_{\mathbf{x}}\chi(\mathbf{x}^t A\mathbf{y}) \\ &= \mathbb{E}_{\mathbf{y}} 1_{A\mathbf{y}=0} = \Pr_{\mathbf{y}}[A\mathbf{y}=0] = \frac{|\mathbb{F}|^{\dim\ker A}}{|\mathbb{F}|^n} = |\mathbb{F}|^{-\operatorname{rank} A} \\ \text{AR}(T) &= \operatorname{rank} A\end{aligned}$$

Example (Singleton tensor)

$$\begin{aligned}T(\mathbf{x},\mathbf{y},\mathbf{z}) &= x_1y_1z_1 \\ \text{bias}(T) &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\mathbb{E}_{\mathbf{z}}\chi(x_1y_1z_1) \\ &= \Pr_{\mathbf{x},\mathbf{y}}[x_1y_1=0] \\ &= \frac{2|\mathbb{F}|-1}{|\mathbb{F}|^2} \\ \text{AR}(T) &\approx 1\end{aligned}$$

Example (Diagonal tensor)

$$\begin{aligned}T(\mathbf{x},\mathbf{y},\mathbf{z}) &= x_1y_1z_1 + \cdots + x_ny_nz_n \\ \text{bias}(T) &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\mathbb{E}_{\mathbf{z}}\chi(x_1y_1z_1 + \cdots + x_ny_nz_n) \\ &= \Pr_{\mathbf{x},\mathbf{y}}[x_1y_1 = x_2y_2 = \cdots = x_ny_n = 0] \\ &= \left(\frac{2|\mathbb{F}|-1}{|\mathbb{F}|^2}\right)^n \\ \text{AR}(T) &= -\log_{|\mathbb{F}|}\left(\frac{2|\mathbb{F}|-1}{|\mathbb{F}|^2}\right)^n \approx n\end{aligned}$$

Kernel definition of AR

Definition (Kernel of a tensor)

Let T be a 3-tensor.

$$\begin{aligned}\ker T &= \{(\mathbf{x}, \mathbf{y}) \in \mathbb{F}^n \times \mathbb{F}^n \ : \ T(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0 \text{ for all } \mathbf{z}\} \\ &= \{(\mathbf{x}, \mathbf{y}) \in \mathbb{F}^n \times \mathbb{F}^n \ : \ \mathbf{x}^t T_1 \mathbf{y} = \cdots = \mathbf{x}^t T_n \mathbf{y} = 0\}\end{aligned}$$

For T a k-tensor,

$$\ker T = \{(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}) \in \mathbb{F}^n \times \cdots \times \mathbb{F}^n \ : \ T(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}, \cdot) = 0\}$$

We have $\mathrm{AR}(T) = (k-1)n - \log_{|\mathbb{F}|} |\ker T|$.

$$\mathbb{E}\chi(T(\mathbf{x}, \mathbf{y}, \mathbf{z})) = \mathbb{E}_{x,y}\mathbb{E}_z\chi(T(\mathbf{x}, \mathbf{y}, \mathbf{z})) = \Pr[T(\mathbf{x}, \mathbf{y}, \cdot) = 0] = \frac{|\ker T|}{|\mathbb{F}|^{2n}}.$$

The kernel is large if the analytic rank is small

Table of Contents

Partition rank

A matrix has rank one if it is an outer product of two vectors

Example

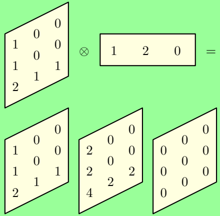
$$\begin{pmatrix} 1 & 2 & 0 \\ 3 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \end{pmatrix}$$

Definition

The rank of a matrix A is the minimal r so that A is a sum of r rank one matrices.

A tensor has partition rank one if it is an outer product of two smaller tensors

Example

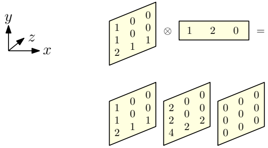


Definition

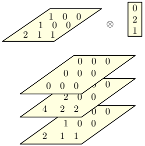
The *partition rank* of a tensor, $\mathrm{PR}(T)$, is the minimal r so that T is the sum of r partition rank one tensors.

Partition rank, multilinear form perspective

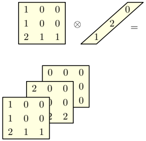
$T(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)})$ has partition rank one if it is a product of two tensors in disjoint subsets of the variables



$T(x, y, z) = \ (x_1 + 2x_2)(2y_1z_1 + y_1z_2 + y_1z_3 + y_2z_1 + y_3z_1)$



$T(x, y, z) =$
$(y_1 + 2y_2)(2x_1z_1 + x_1z_2 + x_1z_3 + x_2z_1 + x_3z_1)$



$T(x, y, z) =$
$(z_1 + 2z_2)(2x_1y_1 + x_1y_2 + x_1y_3 + x_2y_1 + x_3y_1)$

Definition

The partition rank of a tensor is the minimal r so that we can write

$$T(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}) = \sum_{j=1}^{r} f_j(\mathbf{x}^{I_j}) g_j(\mathbf{x}^{I_j^c}), \quad I_j \subsetneq [k]$$

More on partition rank

Example

The tensor $T(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}) = (x_1 y_1 + \cdots + x_n y_n)(z_1 w_1 + \cdots + z_n w_n)$ has PR $= 1$ but has no linear form factor.

- $0 \leq \mathrm{PR}(T) \leq n$
- $\mathrm{AR}(T) \leq \mathrm{PR}(T)$.

$$T(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x})g(\mathbf{y}, \mathbf{z}) \Rightarrow \{(\mathbf{x}, \mathbf{y}) : f(\mathbf{x}) = 0\} \subset \ker T$$

$$T(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f_1(\mathbf{x})g_1(\mathbf{y}, \mathbf{z}) + f_2(\mathbf{y})g_2(\mathbf{x}, \mathbf{z}) + f_3(\mathbf{x}, \mathbf{y})g_3(\mathbf{z}) \Rightarrow$$
$$\{(\mathbf{x}, \mathbf{y}) : f_1(\mathbf{x}) = f_2(\mathbf{y}) = f_3(\mathbf{x}, \mathbf{y}) = 0\} \subset \ker T$$

- When $k = 3$, partition rank is the *slice rank* of Tao, which was used in the cap-set problem breakthrough

Table of Contents

Prior work

- A bound $\mathrm{PR}(T) \leq f(\mathrm{AR}(T))$ is an *inverse theorem*: given that T is *biased*, can we conclude that T is *structured*?
- This is the simplest case of inverse conjectures for the Gowers norm, the functions in consideration are themselves polynomials

- Green & Tao (2009), Kaufman & Lovett (2008), Lovett & Bhowmick (2015): $\mathrm{PR}(T) \leq f_k(\mathrm{AR}(T))$
  PR the partition rank, f_k an Ackermann type function
- Haramaty & Shpilka (2010): for 3-tensors, $\mathrm{PR}(T) \leq C\,\mathrm{AR}(T)^4$
- Milićević (2019), Janzer (2020): $\mathrm{PR}(T) \leq \mathrm{AR}(T)^c$, $c = c(k)$
- It has been conjectured by several authors that $\mathrm{PR}(T) \leq C_k\,\mathrm{AR}(T)$

Main theorem

Theorem

For every $k \geq 2$ *and* $r \geq 0$ *there is a* $F = F(r, k)$ *such that for every finite field* $\mathbb{F}$ *with* $|\mathbb{F}| \geq F$*, every* k*-tensor* T *over* $\mathbb{F}$ *with* $\mathrm{AR}(T) \leq r$*,*

$$\mathrm{PR}(T) \leq (2^{k-1} - 1)\,\mathrm{AR}(T) + 1$$

- 3-tensors came first, works over all fields but $\mathbb{F}_2$, and was independently discovered by Adiprasito, Kazhdan, and Ziegler
- We go through a smooth analogue of AR—the *geometric rank* (GR)

Geometric rank

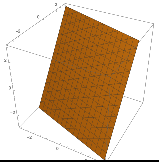
View T over $\overline{\mathbb{F}}$.

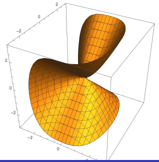
- $(k = 2)$ $T(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t A\mathbf{y}$, $\ker T = \{\mathbf{x} \in \overline{\mathbb{F}}_q^n : \mathbf{x}^t T = 0\} = \ker A^t$
- $(k = 3)$ $\ker T = \{(\mathbf{x}, \mathbf{y}) \in \overline{\mathbb{F}}_q^n \times \overline{\mathbb{F}}_q^n : T(\mathbf{x}, \mathbf{y}, \cdot) = 0\}$
  $\{(\mathbf{x}, \mathbf{y}) \in \overline{\mathbb{F}}_q^n \times \cdots \times \overline{\mathbb{F}}_q^n : \mathbf{x}^t T_1 \mathbf{y} = \cdots = \mathbf{x}^t T_n \mathbf{y} = 0\}$
- $(k > 3)$ $\ker T = \{(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}) : T(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}, \cdot) = 0\}$

Definition

$\mathrm{GR}(T) = \operatorname{codim} \ker_{\overline{\mathbb{F}}} T$

- Doesn't depend on how we "slice" T
- Only makes sense over algebraically closed fields
- $0 \leq \mathrm{GR}(T) \leq \mathrm{PR}(T) \leq n$



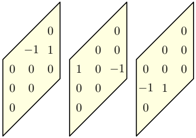


Geometric rank example

Let T be the *determinant tensor*, given by

$$T(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \det \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix}$$

$$(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} \times \mathbf{y}$$



$$\begin{aligned} \ker T &= \{(\mathbf{x}, \mathbf{y}) \in \overline{\mathbb{F}}^6 \text{ such that } \mathbf{x} \times \mathbf{y} = 0\} \\ &= \{(\mathbf{x}, \mathbf{y}) \text{ scalar multiples of each other}\} \end{aligned}$$

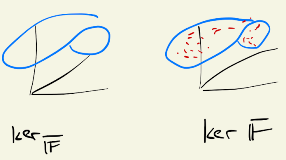
- This is a four dimensional space, so $\mathrm{GR}(T) = 6 - \dim \ker T = 2$.
- On the other hand, $\mathrm{PR}(T) = 3$.

$\mathrm{GR} \leq \mathit{CAR}$

$$\mathrm{AR}(T) = n(k-1) - \log_{|\mathbb{F}|} |\ker_{\mathbb{F}} T|$$

$$\mathrm{GR}(T) = n(k-1) - \dim \ker_{\overline{\mathbb{F}}} T$$

- If $\ker T$ is linear, $|\ker_{\mathbb{F}} T| = |\mathbb{F}|^{\dim \ker T}$
- We expect $\mathrm{AR} \sim \mathrm{GR}$ if $\mathbb{F}$ is large (Lang-Weil)



Proposition

For a tensor T over a finite field $\mathbb{F}$,

$$\mathrm{GR}(T) \leq \left(1 - \frac{\log(k-1)}{\log|\mathbb{F}|}\right)^{-1} \mathrm{AR}(T)$$

Lemma (Generalized Schwartz-Zippel)

For $\mathbf{V} \subset \overline{\mathbb{F}}^n$ a variety cut out by degree $\leq d$ equations,

$$|V_{\mathbb{F}}| \leq d^{\operatorname{codim} \mathbf{V}} |\mathbb{F}|^{\dim \mathbf{V}}$$

PR vs GR

Theorem

$$\mathrm{PR}_{\overline{\mathbb{F}}}(T) \leq (2^{k-1} - 1)\,\mathrm{GR}(T)$$

- Proof works locally (pick a point on the kernel, take tangent spaces / derivatives)
- Same result appeared in work of Schmidt
- Originally, PR decomposition over $\overline{\mathbb{F}}$
- A key part of our work is adapting this strategy to work over the base field $\mathbb{F}$ (important to avoid dependence on n). Need large fields.

Theorem

If $\mathbf{X} \subset \ker T$ *is defined over* $\mathbb{F}$ *and has a nonsingular* $\mathbb{F}$*-point,*

$$\mathrm{PR}_{\mathbb{F}}(T) \leq (2^{k-1} - 1)\,\mathrm{codim}\,\mathbf{X}$$

Two parts to the argument

$\mathbf{f} : \mathbb{F}^n \to \mathbb{F}^m$, $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$
$D\mathbf{f} : \mathbb{F}^n \to \mathbb{F}^m$ (linear), $\quad D^k\mathbf{f}|_{\mathbf{x}} : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}^m$ (multilinear)

- $D\mathbf{f} : \mathbb{F}^n \to \mathbb{F}^n \otimes \mathbb{F}^m$
- $D^2\mathbf{f} : \mathbb{F}^n \to \mathbb{F}^n \otimes \mathbb{F}^n \otimes \mathbb{F}^m$
- $D^k\mathbf{f} : \mathbb{F}^n \to \mathbb{F}^n \otimes \cdots \otimes \mathbb{F}^n \otimes \mathbb{F}^m$

$D\mathbf{f}|_{\mathbf{x}} = (\partial_{x_k} f_j|_{\mathbf{x}})_{j=1,k=1}^{j=m,k=n}$

$D^2\mathbf{f}|_{\mathbf{x}} = (\partial_{x_k}\partial_{x_l} f_j|_{\mathbf{x}})_{j=1,k=1,l=1}^{j=m,k=n,l=n}$

Theorem A (PR decomposition)

Suppose $\mathbf{f} : \mathbb{F}^n \to \mathbb{F}^m$ *and* $\mathbf{X} \subset \mathbb{F}^n$ *is an irreducible variety with* $\mathbf{f}|_{\mathbf{X}} = 0$.
Then $\mathrm{PR}(D^k\mathbf{f}|_{\mathbf{x}}) \leq (2^k - 1)\,\mathrm{codim}\,\mathbf{X}$ *for all nonsingular* $\mathbf{x} \in \mathbf{X}$.
If everything is defined over $\mathbb{F}_q$, *so is the* PR *decomposition.*

Theorem B (Obtaining a nonsingular point, heuristic statement)

Let $\mathbf{X} \subset \overline{\mathbb{F}}_q^n$ *be a variety defined over* $\mathbb{F}_q$ *with* ***bounded complexity***.
There exists an irreducible variety $\mathbf{Z} \subset \mathbf{X}$ *which has controlled complexity, has a nonsingular* $\mathbb{F}_q$*-rational point, and contains all the* $\mathbb{F}_q$*-rational points of* $\mathbf{X}$.