Expressive Power of Recurrent Neural Networks, II

Emil Alkin

January 16, 2024

1 Introduction

Let us denote $\mathbf{n} = (n_1, n_2 \dots n_d)$. Set of all tensors $\mathcal{X}$ with mode sizes $\mathbf{n}$ representable in TT-format with

$$\operatorname{rank}_{TT} \mathcal{X} \leqslant \mathbf{r},$$

for some vector of positive integers $\mathbf{r}$ (inequality is understood entry-wise) forms an *irreducible algebraic variety* (Shafarevich, Hirsch (1994)), which we denote by $\mathcal{M}_{\mathbf{n},\mathbf{r}}$.

2 Main part

Lemma 1. *Let $\mathcal{X}^{i_1 i_2 \dots i_d}$ and $\operatorname{rank}_{CP} \mathcal{X} = r$. Then for any matricization $\mathcal{X}^{(s,t)}$ we have $\operatorname{rank} \mathcal{X}^{(s,t)} \leqslant r$, where the ordinary matrix rank is assumed.*

Theorem 2. *Suppose that $d = 2k$ is even. Define the following set*

$$B := \left\{ \mathcal{X} \in \mathcal{M}^{eq}_{\mathbf{n},\mathbf{r}} : \operatorname{rank}_{CP} \mathcal{X} < q^{\frac{d}{2}} \right\},$$

where $q = \min\{n, r-1\}$.

Then

$$\mu(B) = 0,$$

where μ is the standard Lebesgue measure on $\mathcal{M}^{eq}_{\mathbf{n},\mathbf{r}}$.

Proof. Our proof is based on applying Lemma 1 to a particular matricization of $\mathcal{X}$. Namely, we would like to show that for $s = \{1, 3, \dots, d-1\}$, $t = \{2, 4, \dots, d\}$ the following set

$$B^{(s,t)} := \left\{ \mathcal{X} \in \mathcal{M}^{eq}_{\mathbf{n},\mathbf{r}} : \operatorname{rank} \mathcal{X}^{(s,t)} \leqslant q^{\frac{d}{2}} - 1 \right\}$$

has measure 0. Indeed, by Lemma 1 we have

$$B \subset B^{(s,t)},$$

so if $\mu(B^{(s,t)}) = 0$ then $\mu(B) = 0$ as well. Note that $B^{(s,t)}$ is an algebraic subset of $\mathcal{M}^{eq}_{\mathbf{n},\mathbf{r}}$ given by the conditions that the determinants of all $q^{\frac{d}{2}} \times q^{\frac{d}{2}}$ submatrices of $\mathcal{X}^{(s,t)}$ are equal to 0. Thus to show that $\mu(B^{(s,t)}) = 0$ we need to find at least one $\mathcal{X}$ such that $\operatorname{rank} \mathcal{X}^{(s,t)} \geqslant q^{\frac{d}{2}}$. This follows from the fact that because $B^{(s,t)}$ is an algebraic subset of the irreducible algebraic variety $\mathcal{M}^{eq}_{\mathbf{n},\mathbf{r}}$, it is either equal to $\mathcal{M}^{eq}_{\mathbf{n},\mathbf{r}}$ or has measure 0, as was explained before.

One way to construct such tensor is as follows. Let us define the following tensors:

$$G_1^{i_1 \alpha_1} = [i_1 = \alpha_1], \quad G_1 \in \mathbb{R}^{1 \times n \times r}$$

$$G_k^{\alpha_{k-1} i_k \alpha_k} = \begin{cases} [\alpha_{k-1} = i_k] \cdot [\alpha_k = q+1], & \text{if } \alpha_{k-1} \leqslant q \\ [i_k = \alpha_k], & \text{if } \alpha_{k-1} = q+1 \\ 0, & \text{if } \alpha_{k-1} > q+1 \end{cases}, \quad G_k \in \mathbb{R}^{r \times n \times r}, \quad k = 2, 3, \dots, d-1$$

$$G_d^{\alpha_{d-1} i_d} = [\alpha_{d-1} = i_d], \quad G_d \in \mathbb{R}^{r \times n \times 1},$$

where $[\,\cdot\,]$ is the Iverson bracket notation.

The TT-ranks of the tensor $\mathcal{X}$ defined by the TT-cores are equal to $\operatorname{rank}_{TT} \mathcal{X} = (r, r, \dots, r)$.

Lemma 3. *Consider* $(i_1, i_2, \ldots, i_d) \in [q]^d$ *and* $(\alpha_1, \alpha_2, \ldots, \alpha_{d-1}) \in [r]^{d-1}$ *such that*

$$G_1^{i_1\alpha_1} \cdot \ldots \cdot G_d^{\alpha_{d-1}i_d} \neq 0.$$

Then

$$\alpha_k = \begin{cases} i_k, & \text{if } k \text{ is odd} \\ q+1, & \text{if } k \text{ is even} \end{cases} \qquad \text{for any } k \in [d-1].$$

Proof. Let us prove the lemma by induction over k.

- *Base of the induction:* $k = 1$.

  Since $G_1^{i_1\alpha_1} = [i_1 = \alpha_1] \neq 0$, then $\alpha_1 = i_1$.

- *The induction step:* $k \to k+1$, $k \in [d-2]$.

  If k is odd, then $\alpha_k = i_k \in [q]$. Hence $G_{k+1}^{\alpha_k i_{k+1}\alpha_{k+1}} = [\alpha_k = i_{k+1}] \cdot [\alpha_{k+1} = q+1]$. Since $G_{k+1}^{\alpha_k i_{k+1}\alpha_{k+1}} \neq 0$, then $[\alpha_{k+1} = q+1] \neq 0$, so $\alpha_{k+1} = q+1$.

  If k is even, then $\alpha_k = q+1$. Hence $G_{k+1}^{\alpha_k i_{k+1}\alpha_{k+1}} = [i_{k+1} = \alpha_{k+1}]$. Since $G_{k+1}^{\alpha_k i_{k+1}\alpha_{k+1}} \neq 0$, then $[i_{k+1} = \alpha_{k+1}] \neq 0$, so $\alpha_{k+1} = i_{k+1}$.

□

Lets consider the following matricization of the tensor $\mathcal{X}$

$$\mathcal{X}^{(i_1,i_3,\ldots,i_{d-1}),(i_2,i_4,\ldots,i_d)}$$

The following identity holds true for any values of indices such that $i_k = 1, \ldots, q$, $k = 1, \ldots, d$.

$$\mathcal{X}^{(i_1,i_3,\ldots,i_{d-1}),(i_2,i_4,\ldots,i_d)} = \sum_{\alpha_1,\ldots,\alpha_{d-1}} G_1^{i_1\alpha_1} \ldots G_d^{\alpha_{d-1}i_d} = (\text{by Lemma 3}) = G_1^{i_1 i_1} G_2^{i_1 i_2 (q+1)} G_3^{(q+1) i_3 i_3} \ldots G_d^{i_{d-1} i_d} =$$

$$\Big([i_1 = i_1]\Big) \cdot \Big([i_1 = i_2] \cdot [q+1 = q+1]\Big) \cdot \Big([i_3 = i_3]\Big) \cdot \ldots \cdot \Big([i_{d-1} = i_d]\Big) =$$
$$[i_1 = i_2] \cdot [i_3 = i_4] \cdot \ldots \cdot [i_{d-1} = i_d].$$

We obtain that

$$\mathcal{X}^{(i_1,i_3,\ldots,i_{d-1}),(i_2,i_4,\ldots,i_d)} = [i_1 = i_2] \cdot [i_3 = i_4] \cdot \ldots \cdot [i_{d-1} = i_d] = I^{(i_1,i_3,\ldots,i_{d-1}),(i_2,i_4,\ldots,i_d)},$$

where I is the identity matrix of size $q^{\frac{d}{2}} \times q^{\frac{d}{2}}$.

To summarize, we found an example of a tensor $\mathcal{X}$ such that $\operatorname{rank}_{TT} \mathcal{X} \leqslant \mathbf{r}$ and the matricization $\mathcal{X}^{(i_1,i_3,\ldots,i_{d-1}),(i_2,i_4,\ldots,i_d)}$ has a submatrix being equal to the identity matrix of size $q^{\frac{d}{2}} \times q^{\frac{d}{2}}$, and hence $\operatorname{rank} \mathcal{X}^{(i_1,i_3,\ldots,i_{d-1}),(i_2,i_4,\ldots,i_d)} \geqslant q^{\frac{d}{2}}$. This means that the canonical $\operatorname{rank}_{CP} \mathcal{X} \geqslant q^{\frac{d}{2}}$ which concludes the proof.

□

References

[1] Nadav Cohen, Or Sharir, and Amnon Shashua. On the expressive power of deep learning: A tensor analysis. In Conference on Learning Theory, pp. 698–728, 2016.

[2] Valentin Khrulkov, Alexander Novikov and Ivan Oseledets. Expressive power of recurrent neural networks, 2017, https://arxiv.org/abs/1711.00811.