Assignment 2: Expressiveness

Dana Arad 311183321
Amit Elyasi 316291434

Part 1: Boolean AND-OR Networks

1. We will show that the lower bound on B cannot be larger than 2^{d-1}, since XOR_d is realizable by a shallow network of width $B = 2^{d-1}$.

   Denote the truth table of XOR_d:

x_1	x_2	....	x_d
+1	-1	....	-1
...	...	...	...
+1	+1	.....	+1

   Denote each row $i = (x_1, \dots, x_d)$ where $XOR_d(x_1, \dots, x_d) = +1$ as w_i.

   Notice that:

   a. $\forall v \in \{+1, -1\}^d.\ v * v = (+1)^d$
   b. $\forall u, v \in \{+1, -1\}^d \exists k. (u * v)_k = -1$

   We define the following shallow network, mark as N:

   - $B = 2^{d-1}$
   - $U = I$
   - $W = \begin{pmatrix} -w_1- \\ \cdots \\ -w_{2^{d-1}}- \end{pmatrix}$

   We will show that:

   1. $\forall (x_1, \dots, x_d)\ s.t. XOR_d(x_1, \dots, x_d) = +1.\ N(x_1, \dots, x_d) = +1$
   2. $\forall (x_1, \dots, x_d)\ s.t. XOR_d(x_1, \dots, x_d) = -1.\ N(x_1, \dots, x_d) = -1$

   1. Let $x_1, \dots, x_d$ be inputs $\in \{+1, -1\}^d$ s.t $XOR_d(x_1, \dots, x_d) = +1, so\ (x_1, \dots, x_d) = w_i$ for some $1 \le i \le 2^{d-1}$. Well show that $N(x_1, \dots, x_d) = +1$.

$$N(x) = OR\big(U * AND(W * x)\big)$$

$$AND(W * x) = AND\left(\begin{pmatrix} -w_1- \\ \cdots \\ -w_{2^{d-1}}- \end{pmatrix}\begin{pmatrix} | \\ w_i \\ | \end{pmatrix}\right) = (-1, -1, \dots + 1, . . -1)$$

      Where the only +1 is in index i (by claim a) .

$$OR\big(U * (-1,-1,\ldots+1,..-1)\big) = OR\big((-1,-1,\ldots+1,..-1)\big) = +1 \blacksquare$$

2. Similarly, let $x_1,\ldots,x_d$ be inputs $\in \{+1,-1\}^d$ s.t $XOR_d(x_1,\ldots,x_d) = -1.$ So, $\forall i. x \neq w_i$.

Therefore, $AND(W * x) = AND\left(\begin{pmatrix} -w_1 - \\ \cdots \\ -w_{2^{d-1}} - \end{pmatrix}\begin{pmatrix} | \\ x \\ | \end{pmatrix}\right) = (-1)^d \rightarrow OR(U * (-1)^d) = -1 \blacksquare$

2. Let us look at the number of functions a deep network can express:
   - $2\log_2(d)$ Layers:
     - Odd layers:
       - $W \in \mathbb{R}^{\bar{B}*d}$
       - $W^{(2)},\ldots,W^{(2\log_2(d))} \in \mathbb{R}^{\bar{B}*\bar{B}}$
       - $\forall k. W_{ij}^{(k)} \in \{-1,0,+1\}$
     - Even layers:
       - $U^{(1)},\ldots,U^{(2\log_2(d)-1)} \in \mathbb{R}^{\bar{B}*\bar{B}}$
       - $U^{(2\log_2(d))} \in \mathbb{R}^{\bar{B}*1}$
       - $\forall k. U_{ij}^{(k)} \in \{-1,+1\}$

Mark T as the total number of parameters in the network:

$$T = (2\log_2(d) - 2) * (\bar{B} * \bar{B}) + (\bar{B} * d) + (\bar{B} * 1)$$

The number of possible functions is therefore $2^{T+C} = 2^{poly(\bar{B})}$ (where C is some constant to change the exponent base from 3 to 2).

$|Y^X| = |Y|^{|X|} = 2^{2^d} \rightarrow$ unless $\bar{B} \in \exp(d)$, $2^{poly(\bar{B})} < 2^{2^d}$, so there exists a function that the network will not be able to express.

Part 2: Fully Connected ReLU Networks with 1D Input

1. First, we note that $B \geq 2$ is required for the condition to hold: when $B = 1$, $h \in H_1$ is equal to $h(x) = w_2[w_1 x + b_1]_+ + b_2$, this is a PWL function with 2 pieces- a constant part (negative values after Relu) and a linear part. Therefore, the number of pieces $= 2 \nleq 1 = B$.

   A shallow network of width $\boldsymbol{B} \geq \boldsymbol{2}$ can realize any PWL mapping with $\leq \boldsymbol{B}$ pieces:
   For $B \geq 2$ we will prove by induction by explicitly building such h.
$$h(x) = \sum_{i=1}^{B} w_2^i \left[w_1^i x^i + b_1^i\right]_+ + b_2$$
   with $g_1(x) = w_2^1[w_1^1 x + b_1^1]_+ + b_2 \; and \; g_i(x) = w_2^i\left[w_1^i x + b_1^i\right]_+$
   So $h(x) = \sum_{i=1}^{B} g_i(x)$

   Base case B=2:
   Let f be a PWL function with B=2 pieces, denote $f(x) = \begin{cases} m_1 x + n_1 & x < p \\ m_2 x + n_2 & x \geq p \end{cases}$
   We will define two functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$, and show that $h(x) \coloneqq (g_1 + g_2)(x) \equiv f(x)$:
   Set $w_1^1 = -1, w_2^1 = -m_1, b_1^1 = p, b_2 = n_1 + m_1 p$ and get:
$$g_1(x) = -m_1[p - x]_+ + n_1 + m_1 p = \begin{cases} m_1 x + n_1 & x < p \\ m_1 p + n_1 & x \geq p \end{cases}$$
   Now set $w_1^2 = 1, w_2^2 = m_2, b_1^2 = -p$ and get:
$$g_2(x) = m_2[x - p_1]_+ = \begin{cases} 0 & x < p \\ m_2 x - m_2 p & x \geq p \end{cases}$$
   Overall, we get:
$$(g_1 + g_2)(x) = \begin{cases} m_1 x + n_1 & x < p \\ m_1 p + n_1 + m_2 x - m_2 p & x \geq p \end{cases}$$
$$= \begin{cases} m_1 x + n_1 & x < p \\ m_1 p + n_1 + m_2 x - m_2 p & x \geq p \end{cases} \overset{(*)}{=} \begin{cases} m_1 x + n_1 & x < p \\ m_2 x + n_2 & x \geq p \end{cases} = f(x)$$
   $(*)\; f \; is \; continuos \; so \; m_1 p + n_1 = m_2 p + n_2$

   Induction Step:
   Let's assume there exist some B so that for any PWL f $\exists h \in H_B \; s.t. h(x) = f(x)$
   Let f be a PWL function with B+1 pieces $\{(-\infty, p_1), [p_1, p_2), \dots, [p_B, \infty)\}$
$$f(x) = m_i x + n_i \quad if \; x \in [p_{i-1}, p_i) \; for \; i \in [B+1] \quad (denote \; p_0 = -\infty, p_{B+1} = \infty)$$
   Let $f^*(x) = m_i x + n_i \quad if \; x \in [p_{i-1}, p_i) \; for \; i \in [B]$
   By induction we know that there exits $h^* \in H_B$ s.t. $h^*(x) = f^*(x)$, with $h^*(x) = \sum_{i=1}^{B} g_i(x)$
   And define $g_{B+1}(x) = w_2^{B+1}[w_1^{B+1} x + b_1^{B+1}]_+$
   with $w_1^{B+1} = 1, w_2^{B+1} = m_{B+1} - m_B, b_1^B = -p_B$
   So overall we will get:

$$h^*(x)+g_{B+1}(x)=\begin{cases}h^*(x) & x<p_B\\ m_Bx+n_B+(m_{B+1}-m_B)(x-p_B) & x\ge p_B\end{cases}$$
$$=\begin{cases}h^*(x) & x<p_B\\ m_Bx+n_B+m_{B+1}x-m_Bx+m_{B+1}p_B-m_Bp_B & x\ge p_B\end{cases}\overset{(*)}{\cong}$$
$$\begin{cases}h^*(x) & x<p\\ m_{B+1}x+m_{B+1}p_B+n_{B+1}-m_{B+1}p_B & x\ge p\_B\end{cases}=\begin{cases}h^*(x) & x<p\\ m_{B+1}x+n_{B+1} & x\ge p_B\end{cases}=f(x)$$

$(*)\ again, f\ is\ continuos\ so\ m_ip_i+n_i=m_{i+1}p_i+n_i$ ■

Any mapping realizable by such network is PWL with $\le \boldsymbol{B+1}$ pieces:

Let $h\in H_B$, so h is PWL with $\le B+1$ pieces because h can be written as a sum of PWL functions (like in the first direction):

Denote $p=\arg_x[w_1\cdot x+b_1=0]$, let $argsort(p)=(i_1,i_2,\dots,i_B)\ so\ that\ \ p_{i_1}\le p_{i_2}\le \cdots\le p_{i_B}$

Let $g_{i_1}(x)=w_2^{i_1}\left[w_1^{i_1}x+b_1^{i_1}\right]_+ +b_2$ and for $i_1\ne i\in argsort(p), g_i(x)=w_2^i\left[w_1^ix+b_1^i\right]_+$

$$h(x)=\sum_{j\in argsort(p)}g_j(x)=\begin{cases}g_{i_1}(x) & x<p_{i_1}\\ g_{i_1}(x)+g_{i_2}(x) & p_{i_1}\le x<p_{i_2}\\ \vdots & \\ \displaystyle\sum_{j\in argsort(p)}g_j(x) & x\ge p_B\end{cases}$$

all pieces of h are linear functions as sum of linear functions, and so by the equation above we have $\le B+1$ pieces of different linear functions. ■

2. As shown in class,

$$|S_<|=\#\ of\ sawteeth=2^{L-2},\qquad |S_>|=2^{L-2}+1\Rightarrow |S_<\cup S_>|=2^{L-1}+1$$

We'll show that a PWL function h with $\le B+1$ pieces could avoid missing $\le \left\lceil\frac{1}{2}(2^{L-1}+1)+\frac{1}{2}(B+1)\right\rceil$ out of $2^{L-1}+1$ intervals in $S_<, S_>$

Claim: linear h (B=1) can avoid missing $\le \left\lceil\frac{1}{2}(2^{L-1}+1)+\frac{1}{2}\right\rceil$

h is linear (assume WLOG h increasing) $\Rightarrow$ h intersect with $y=\frac{1}{2}$ once , denote the intersection point by x^* (assume WLOG x^* is a shared point of $S_>\ and\ S_<$, that is there exists $[x_1,x^*]\in S_<\ and\ [x^*,x_2]\in S_>$) $\Rightarrow$ h is below $y=\frac{1}{2}$ for all $x<x^*$, and above it for all $x>x^*$. Denote the number of $intervals\in S_<\cup S_>$ in $[0,x^*]$ by k_1 and the number of $intervals\in S_<\cup S_>$ in $[x^*,1]$ by k_2, note that $k_1+k_2=|S_<\cup S_>|=2^{L-1}+1$,

so h avoids missing no more than half of k_1 and no more than half of k_2 , that is $\le \frac{1}{2}k_2+\frac{1}{2}k_1=\frac{1}{2}(2^{L-1}+1)\le\left\lceil\frac{1}{2}(2^{L-1}+1)+\frac{1}{2}\right\rceil$

back to a general PWL h with $\leq \boldsymbol{B+1}$ pieces:

Let's divide $[0,1]$ into the intervals $[0 = C_0, C_1], [C_1, C_2], \dots, [C_B, 1 = C_{B+1}]$ where on each interval h is linear. Denote the number of $intervals \in S_< \cup S_>$ in (C_i, C_{i+1}) by k_{i+1} so we get $\sum_i k_i = |S_< \cup S_>| = 2^{L-1} + 1$.
Now for each $[C_i, C_{i+1}]$ h is linear and the claim above applies, so over all we get that the number of intervals h can avoid missing is $\leq \sum_i \left\lceil \frac{1}{2}(k_i) + \frac{1}{2} \right\rceil = \left\lceil \frac{1}{2}(2^{L-1}+1) + \frac{1}{2}(B+1) \right\rceil$ ■

3. **The universality and expressive efficiency analyses given in class, with the following modifications applies to leaky ReLU activation $\boldsymbol{\sigma(z) = \max\{az, z\}}$, $\boldsymbol{a \in (0,1)}$:**

Universality: for any $f \in \mathcal{F}$, $\epsilon > 0$ there exist $h \in \mathcal{H}_B$ that is built with leaky ReLU such that $d(f,h) < \epsilon$.
Note that this will also stand for $h \in \overline{\mathcal{H}}_B$ as $\mathcal{H}_B \subseteq \overline{\mathcal{H}}_B$.
We have seen in class that we can approximate any continuous function with PWL one, and that any PWL function can be realized by a ReLU network. In addition, we have seen that any ReLU neuron can be implemented by 2 leaky ReLU neurons, and vice versa.
So overall, any f can be approximated by PWL h with B pieces, that can be realized by a shallow ReLU network of width B, that can be modified to a leaky ReLU network of width 2B.

Expressive efficiency:
$\forall \overline{B} \in O(B)$ s.t. $\mathcal{H}_B \subseteq \overline{\mathcal{H}}_{\overline{B}}$ where here $\mathcal{H}_B$ is the set of functions that can be realized by leaky ReLU neural network.
$\exists \overline{h} \in \overline{\mathcal{H}}_{\overline{B}}$ for $\overline{B} \in O(1)$ s.t. $\overline{h} \notin \mathcal{H}_B$ unless $B \in \exp(L)$
As seen in class, condition (i) is an immediate consequence of the fact that a deep network can realize any mapping realizable by a shallow one of the same width.
Now for condition (ii) we'll construct a sawteeth function $g \coloneqq g^{\circ L}$ similar to the one from class, now g is PWL with 2^{L-1} pieces and we've seen that it can be realized by an L layer ReLU NN of width 3, which means the same function can be realized by an L layer leaky ReLU NN of width 6 ($(*)$ as any ReLU neuron can be implemented by 2 leaky ones and vice versa).

Next, let $h \in \mathcal{H}_B$ be a function that can be realized by a shallow leaky ReLU NN of width B $\Rightarrow$ h is PWL with $\leq 2B+2$ (again because $(*)$) pieces and like the proposition in class the most intervals in $S_< \cup S_>$ h can avoid missing is $\leq \left\lceil \frac{1}{2}(2^{L-1}+1) + \frac{1}{2}(2B+2) \right\rceil$ and like we've seen in class this implies that the closest a shallow leaky ReLU NN of width 2B+2 can get to g is $\geq \frac{1}{4} - (2B)2^{-L-1} - 3 \cdot 2^{-L}$ requiring $B > 2^L\left(\frac{1}{4} - \epsilon\right) - 6 \in \exp(L)$ ■

4. let f be a Reimann integrable function in $[0,1]$, $\epsilon^* > 0$, $B \in \mathbb{N}$, denote $\epsilon = \frac{\epsilon^*}{B}$.
   f is integrable so there exist $T, S: \mathbb{R} \to \mathbb{R}$ and intervals $(x_i, x_{i+1}]\ \forall i \in [B]$ in which S,T are constant functions (step functions) s.t. $\forall x \in [0,1],\ S(x) \le f(x) \le T(x)$ and $T(x) - S(x) < \epsilon$
   define $h: \mathbb{R} \to \mathbb{R}\ \ \forall x \in (x_i, x_{i+1}]\ by\ h(x) = x\frac{f(x_{i+1})-f(x_i)}{x_{i+1}-x_i} + x_i$ that is the line connecting the points $\left(x_i, f(x_i)\right), \left(x_{i+1}, f(x_{i+1})\right)$.
   Note that by definition h is PWL with B pieces and so $h \in \mathcal{H}_B \subseteq \overline{\mathcal{H}}_B$.
   Now, we know that $\forall x \in (x_i, x_{i+1}]\ \max_x |f(x)| \le T(x_i)$ and $\min_x |h(x)| \ge S(x_i)$ so we get

$$d(f,h) = \int_0^1 |f(x) - h(x)|dx \le \int_0^1 |f(x)|dx - \int_0^1 |h(x)|dx \le \int_0^1 |T(x)|dx - \int_0^1 |S(x)|dx$$
$$= \sum_i T(x_i)(x_{i+1} - x_i) - \sum_i S(x_i)(x_{i+1} - x_i)$$
$$= \sum_i \left(T(x_i) - S(x_i)\right)(x_{i+1} - x_i)$$
$$\overset{\forall x,\ T(x)-S(x)<\epsilon}{\lesssim} \sum_i \epsilon(x_{i+1} - x_i) \overset{\sum_i (x_{i+1}-x_i)=1}{\cong} B\epsilon \sum_i (x_{i+1} - x_i) = B\epsilon = \epsilon^* \quad \blacksquare$$

Part 3: Convolutional Arithmetic Circuits

1. Let T and $\bar{T}$ be tensors of orders n and $\bar{n}$ respectively. Let $I \subset [n+\bar{n}]$ and denote by $I-n$ the set obtained by subtracting n from each element of I, Then: $[\![T\otimes\bar{T}]\!]_I = [\![T]\!]_{I\cap[n]} \odot [\![\bar{T}]\!]_{I-n\cap[\bar{n}]}$.

   WLOG we can assume the partition is over the first indices, otherwise we can apply a permutation for each of the tensors. By definition we get:

   $$\begin{aligned}\left[[\![T]\!]_{I\cap[n]} \odot [\![\bar{T}]\!]_{I-n\cap[\bar{n}]}\right]_{|I\cap[n]|*(i_1+j_1),|I-n\cap[\bar{n}]|*(i_2+j_2)} &= [[\![T]\!]_{I\cap[n]}]_{i_1,j_1} * \left[[\![\bar{T}]\!]_{I-n\cap[\bar{n}]}\right]_{i_2,j_2} \\ &= [T]_{i_1,j_1} * [\bar{T}]_{i_2,j_2} = [T\otimes\bar{T}]_{i_1,j_1,i_2,j_2} \\ &= [[\![T\otimes\bar{T}]\!]_I]_{|I\cap[n]|*(i_1+j_1),|I-n\cap[\bar{n}]|*(i_2+j_2)}\end{aligned}$$

   This holds for every i_1, j_1, i_2, j_2 where $i_1 \in I\cap[n], j_1 \in I^c \cap [n], i_2 \in I-n\cap[\bar{n}], j_2 \in I^c - n\cap[\bar{n}]$, and thus $[\![T]\!]_{I\cap[n]} \odot [\![\bar{T}]\!]_{I-n\cap[\bar{n}]} = [\![T\otimes\bar{T}]\!]_I$ ■

2. Modifications are only needed for part (ii) of the proof shown in class:
   We apply canonical matricization to the HT decomposition (corresponding to the deep network), while using the linearity of matricization and its relationship with outer/Kronecker product:

   $[\![\phi^{1,j,\gamma}]\!] = \sum_{\alpha=1}^{r_0} a_\alpha^{1,j,\gamma} [\![\underline{a}^{0,4j-1,\alpha} \otimes \underline{a}^{0,4j,\alpha}]\!]$, $\quad for\ j = 1,2,\dots,\frac{N}{4},\ \gamma = 1,2,\dots,r_1$

   ...

   $[\![\phi^{l,j,\gamma}]\!] = \sum_{\alpha=1}^{r_{l-1}} a_\alpha^{l,j,\gamma} [\![\underline{\phi^{l-1,4j-1,\alpha}}]\!] \odot [\![\underline{\phi^{l-1,4j,\alpha}}]\!]$, $\quad for\ j = 1,2,\dots,\frac{N}{4^l},\ \gamma = 1,2,\dots,r_l$

   ....

   $[\![\phi^{L-1,j,\gamma}]\!] = \sum_{\alpha=1}^{r_{L-2}} a_\alpha^{L-1,j,\gamma} [\![\underline{\phi^{L-2,4j-1,\alpha}}]\!] \odot [\![\underline{\phi^{L-2,4j,\alpha}}]\!]$, $\quad for\ j = 1,2,3,4,\ \gamma = 1,2,\dots,r_{L-1}$

   $$[\![A^{HT}]\!] = \sum_{\alpha=1}^{r_{L-1}} a_\alpha^L [\![\underline{\phi^{L-1,1,\alpha}}]\!] \odot [\![\underline{\phi^{L-1,2,\alpha}}]\!] \odot [\![\underline{\phi^{L-1,3,\alpha}}]\!] \odot [\![\underline{\phi^{L-1,4,\alpha}}]\!]$$

   Let r_0 be $\geq M$, and assign its filter sets - $\left\{\left(\underline{a}^{0,1,\gamma},\dots,\underline{a}^{0,N,\gamma}\right)\right\}_{\gamma=1}^{r_0}$ by $\underline{a}^{0,j,\gamma} = e^\gamma$ for $\gamma \in [M]$ and $\underline{a}^{0,j,\gamma} = \underline{0}$ for $\gamma > M$.

   Assign the filter sets in hidden layer 1 - $\left\{\left(\underline{a}^{1,1,\gamma},\dots,\underline{a}^{1,N/4,\gamma}\right)\right\}_{\gamma-1}^{r_1}$- by $\underline{a}^{1,j,1} = \underline{1}\ (all\ ones)$ and $\underline{a}^{1,j,\gamma} = \underline{0}\ for\ \gamma > 1$.

   Assign the filter sets in hidden layers 2 through L-1 by:

$$\underline{a}^{l,j,\gamma} = \begin{cases} \underline{e}^1 \; for \; \gamma = 1 \\ \underline{0} \; otherwise \end{cases}$$

$$where \; l = 2,3,\dots,L-1, \quad j = 1,2,\dots,\frac{N}{4^l}, \quad \gamma = 1,2,\dots,r_l.$$

And the output weights by $\underline{a}^L = \underline{e}^1$. Under our assignments:

$$[\![A^{HT}]\!] = \underbrace{Id_m \odot \dots \odot Id_m}_{\frac{N}{4} times} = Id_{M^{\frac{N}{4}}}$$

In particular, $rank([\![A^{HT}]\!]) = M^{\frac{N}{4}}$. Our results on $rank([\![A^{HT}]\!])$ implies that for $\bar{B} \geq M$, $\bar{H}_{\bar{B}}$ includes a function $\bar{h}$ whose corresponding tensor has (canonical) matricization rank $M^{N/4}$. On the other hand, our upper bound on $rank([\![A^{CP}]\!])$ implies that H_B includes only functions whose corresponding tensors have (canonical) matricization ranks $\leq B$. This means that $\bar{h} \notin H_B$ unless $B \geq M^{\frac{N}{4}}$, establishing condition (ii).

3. To prove $Sep(f_{deep}; I_{quad}) \leq r_{L-1} r_{L-2}^2$, we can use the following lemma from class:

$$Sep(f; I) = rank(A)_I$$

Where A is the tensor that identifies with the function f.

In the case of a deep network $A = A^{HT} = \sum_{\alpha=1}^{r_{L-1}} a_\alpha^L * (\phi^{L-1,1,\alpha} \otimes \phi^{L-1,2,\alpha})$ and

$\phi^{L-1,j,\gamma} = \sum_{\alpha=1}^{r_{L-2}} a_\alpha^{L-1,j,\gamma} \phi^{L-2,2j-1,\alpha} \otimes \phi^{L-2,2j,\alpha}, \quad for \; j = 1,2,3,4, \; \gamma = 1,2,\dots,r_{L-1}$

Applying the matricization we get:

$$[\![A^{HT}]\!]_{I_{quad}} = \sum_{\alpha=1}^{r_{L-1}} a_\alpha^L * [\![\phi^{L-1,1,\alpha}]\!]_{I_{quad}} \odot [\![\phi^{L-1,2,\alpha}]\!]_{I_{quad}}$$

$$= \sum_{\alpha=1}^{r_{L-1}} a_\alpha^L \left[\!\!\left[\sum_{\alpha=1}^{r_{L-2}} a_\alpha^{L-1,1,\gamma} \phi^{L-2,1,\alpha} \bigotimes \phi^{L-2,2,\alpha}, \right]\!\!\right]_{I_{quad \cap \left[\frac{N}{2}\right]}} \odot \left[\!\!\left[\sum_{\alpha=1}^{r_{L-2}} a_\alpha^{L-1,2,\gamma} \phi^{L-2,3,\alpha} \bigotimes \phi^{L-2,4,\alpha}, \right]\!\!\right]_{I_{quad} - \frac{N}{2} \cap \left[\frac{N}{2}\right]}$$

$$= \sum_{\alpha=1}^{r_{L-1}} a_\alpha^L \sum_{\alpha=1}^{r_{L-2}} a_\alpha^{L-1,1,\gamma} \overbrace{[\![\phi^{L-2,1,\alpha}]\!]_{I_{quad} \cap \left[\frac{N}{4}\right]}}^{rank=1} \odot [\![\phi^{L-2,2,\alpha}]\!]_{\underbrace{(I_{quad} \cap \left[\frac{N}{2}\right]) - \frac{N}{4} \cap \left[\frac{N}{4}\right]}_{=\emptyset}}$$

$$\odot \sum_{\alpha=1}^{r_{L-2}} a_\alpha^{L-1,2,\gamma} \underbrace{[\![\phi^{L-2,3,\alpha}]\!]_{(I_{quad} - \frac{N}{2} \cap \left[\frac{N}{2}\right]) \cap \left[\frac{N}{4}\right]}}_{rank=1} \odot [\![\phi^{L-2,4,\alpha},]\!]_{\underbrace{(I_{quad} - \frac{N}{2} \cap \left[\frac{N}{2}\right]) - \frac{N}{4} \cap \left[\frac{N}{4}\right]}_{=\emptyset}}$$

Using $rank(A \odot B) = rank(A) * rank(B)$ we get exactly $rank\left(\llbracket A^{HT} \rrbracket_{I_{quad}}\right) \leq r_{L-1} r_{L-2}^2$