Tensor Diagrams Problem Set

Jacob Cohen

December 13, 2025

These problems have been carefully selected by me over a long and arduous selection process, which consisted of "omg it's the morning right before the event and Derik told me 'you have to actually do it, or I will be sad' so I should probably come up with some problems."

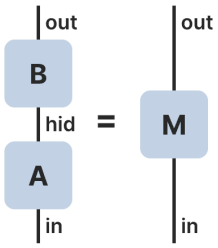
I am indebted to Jordan Taylor's "Graphical tensor notation for interpretability" (https://arxiv.org/abs/2402.01790) and Thomas Dooms' "Compositional Interpretability: Understanding neural networks from their weights." (localhost:4321), and stole their figures for this problem set.

EACH PAGE IS INDEPENDENT! You should probably start with the first page to check that you understand things, but then you can pick which page you like best based on the title.

1 Listening Comprehension

1.1

Interpret this diagram. "BAM" does not constitute a valid interpretation.



1.2

Draw a tensor diagram of some product that takes a 5-dimensional tensor and a 3-dimensional tensor and returns an 6-dimensional tensor. (I don't care about what the computation is, just that the result has the right dimension!)

1.3

Draw this as a tensor diagram. If it doesn't look elegant and symmetric, that's skill issue. What type is the result? (Is it a scalar, a vector, a matrix, etc.)

$$\sum_{ijklmnopqrstu} A_{ij} V_{ir} B_{jkl} W_{rks} C_{lmn} X_{smt} D_{nop} Y_{tou} E_{pq} Z_{uq}$$

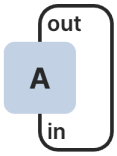
2 Creative Drawings

2.1

What operation does this represent?

2.2

What operation does this represent?



2.3

What operation does this represent?

2.4

Think about linear algebra operations that you like. Do they have natural representations as tensor networks? Do they require drawing new symbols?

3 Identities

3.1

Construct a "proof without words" that

$$(A \cdot B) \otimes (F \cdot G) = (A \otimes F) \cdot (B \otimes G).$$

3.2

How could you draw the identity matrix in tensor notation?

3.3

Anthropic asserts this in their paper "A Mathematical Framework for Transformer Circuits" (2021):

> Using tensor products, we can describe the process of applying attention as:
>
> $$h(x) = \underbrace{(\mathrm{Id} \otimes W_O)}_{\substack{\text{Project result vectors out for each token} \\ (h(x)_i = W_O r_i)}} \cdot \underbrace{(A \otimes \mathrm{Id})}_{\substack{\text{Mix value vectors } \textit{across} \text{ tokens to compute result vectors} \\ (r_i = \sum_j A_{i,j} v_j)}} \cdot \underbrace{(\mathrm{Id} \otimes W_V)}_{\substack{\text{Compute value vector for each token} \\ (v_i = W_V x_i)}} \cdot \; x$$
>
> Applying the mixed product property and collapsing identities yields:
>
> $$h(x) = \underbrace{(A \otimes W_O W_V)}_{\substack{A \text{ mixes across tokens while } W_O W_V \text{ acts on each vector independently.}}} \cdot \; x$$

Prove it with tensor networks. (There's probably not enough space on this page. Oops.)

4 Bilinear Magic

In this center of this diagram is a one-layer bilinear neural network that takes a vector x (twice), and returns some vector output.

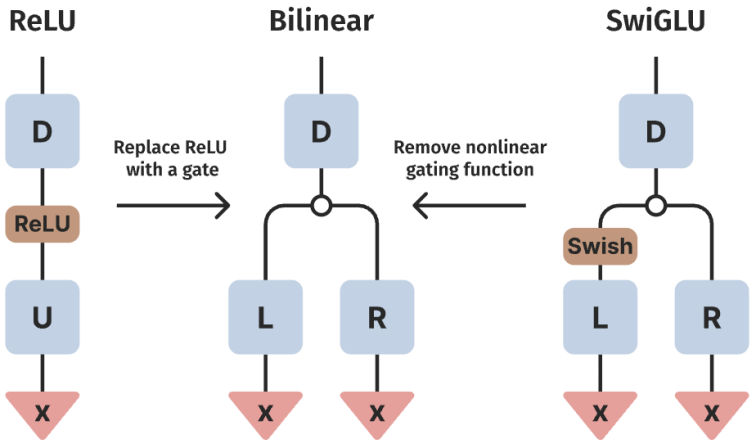


Figure 1: Different neural network layer architectures

Suppose x is a vector with two Booleans, i.e. $(0, 0)$ or $(0, 1)$ or $(1, 0)$ or $(1, 1)$.

4.1

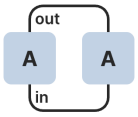
Determine matrices L, R, D such that the network returns a 1-dimensional vector with the product of the two components of x. (In other words, it returns (1) if $x = (1, 1)$, and otherwise returns (0).)

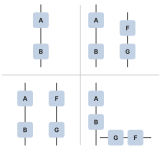
4.2

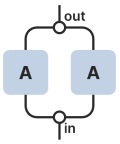
If you want to see something cool, replace both L and R with stack$(L+R, L-R)$. Does your network still work?

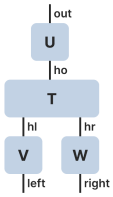
5 For Linear Algebra Knowers

Here are some more diagrams. What advanced linear algebra operations are they?





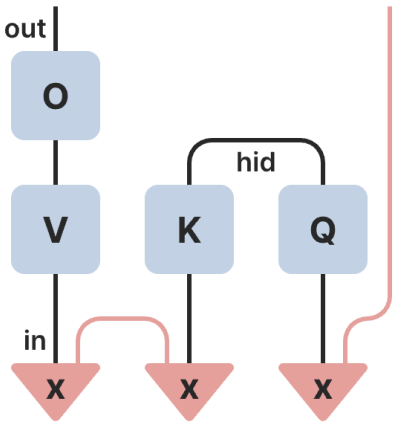




6 I Hope You Paid Attention

6.1

For machine learning knowers: what does this diagram depict? What insights can you learn from it?



6.2

I'm doing a research project into interpretability with tensor networks. What can we say about networks that use the attention mechanism above and a stack of bilinear layers? What results seem tractable?