

# On the Expressivity Role of LayerNorm in Transformers' Attention



**Shaked Brody**  
Technion



Uri Alon  
Language Technologies Institute  
Carnegie Mellon University

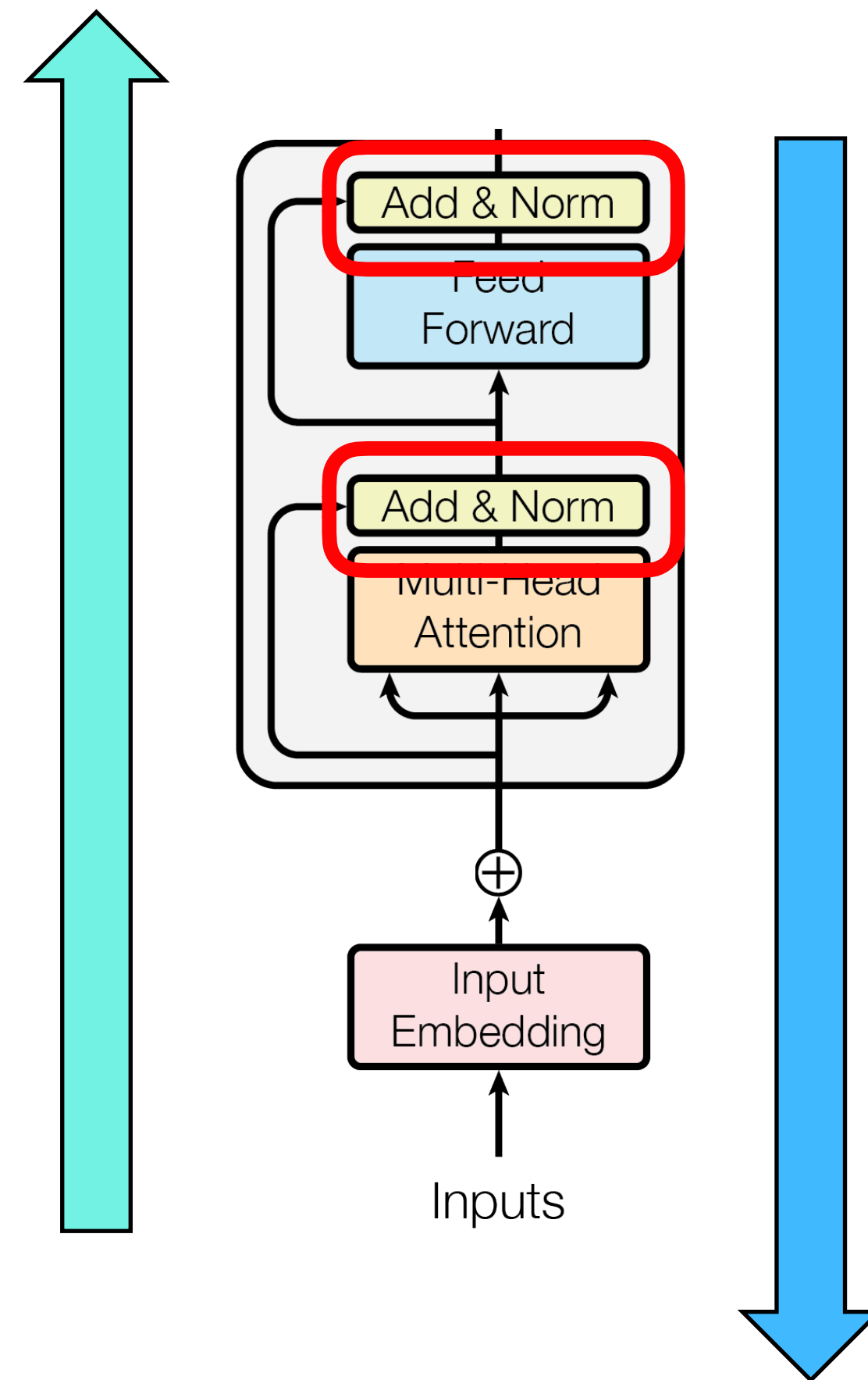


Eran Yahav  
Technion

# LayerNorm - Known so far

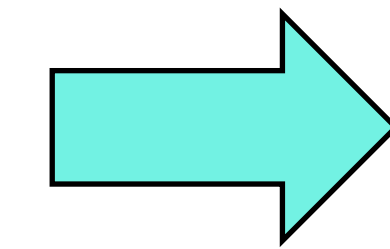
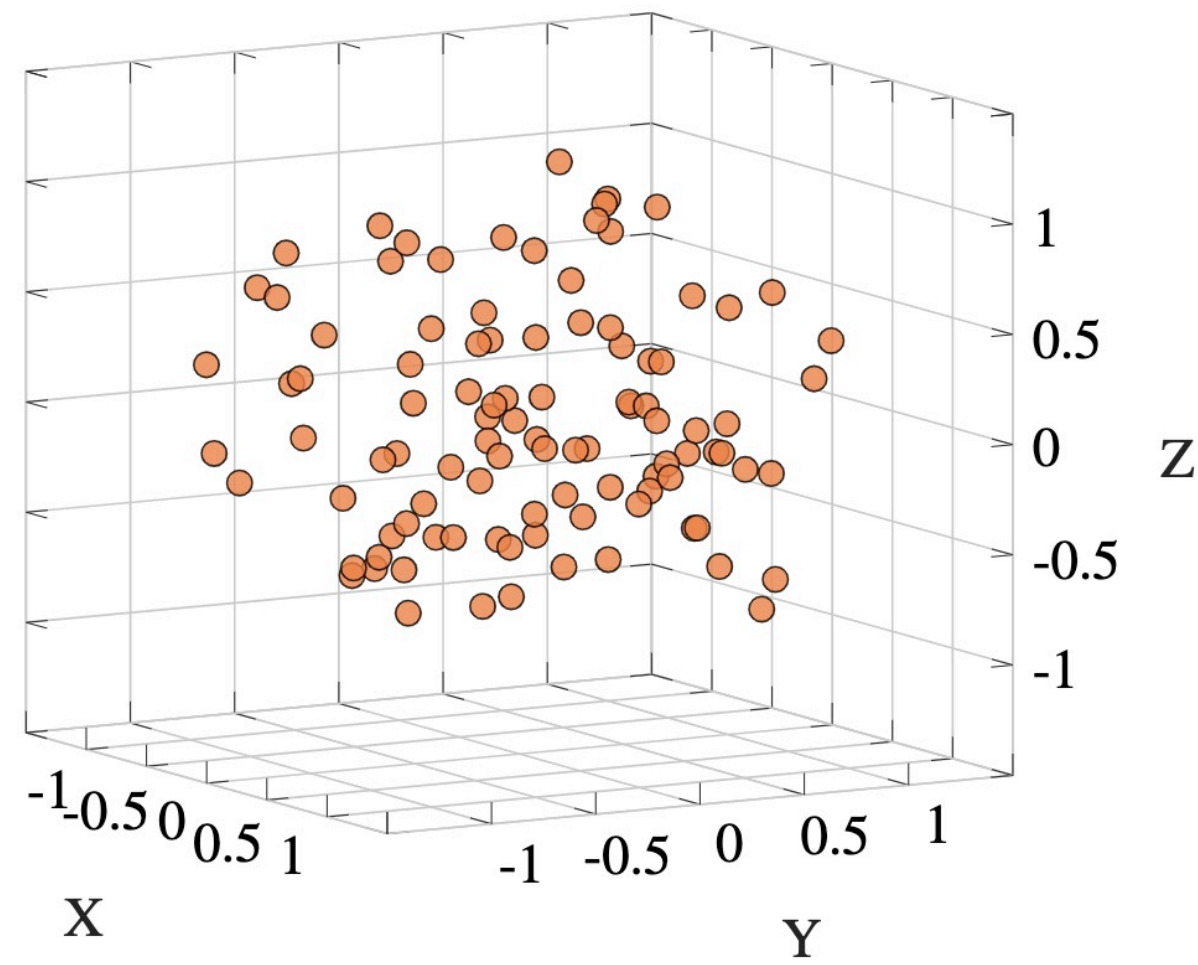
LayerNorm is the least studied component of Transformers

Normalizing activations in the **forward** pass

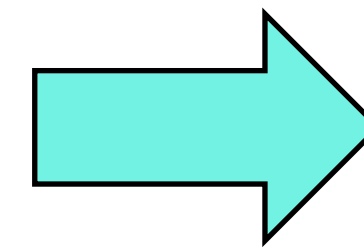
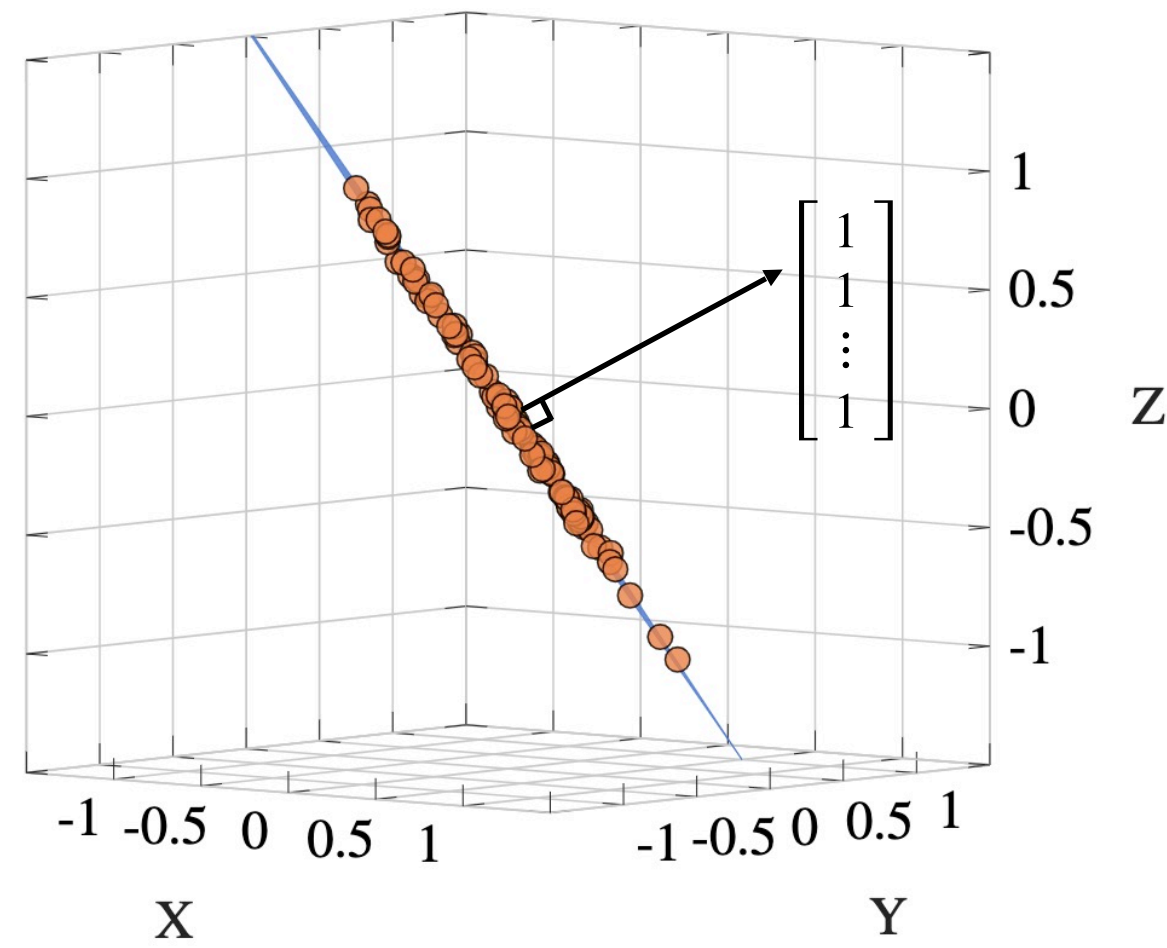


Normalizing gradients in the **backward** pass

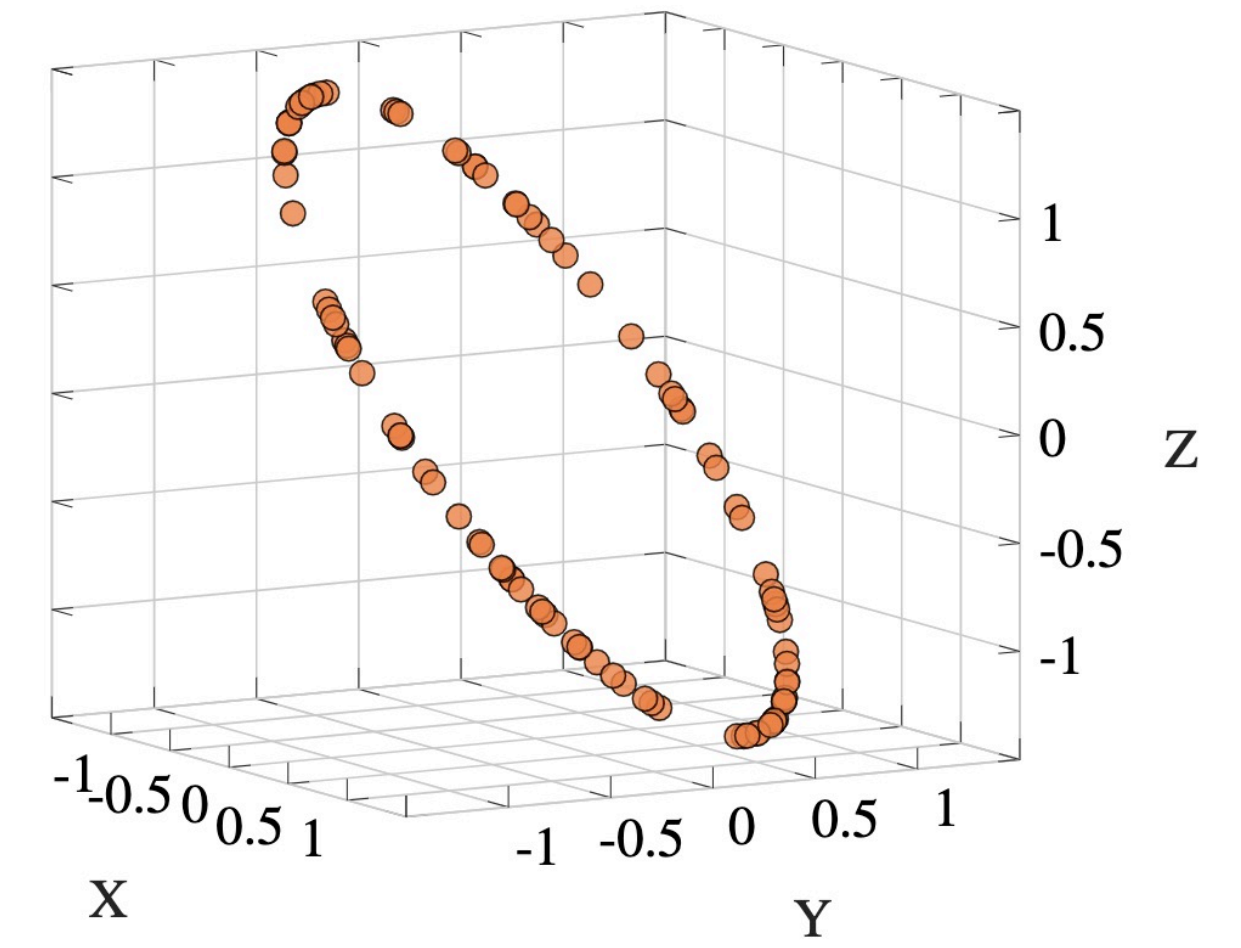
# This work: A Geometric Interpretation of LayerNorm



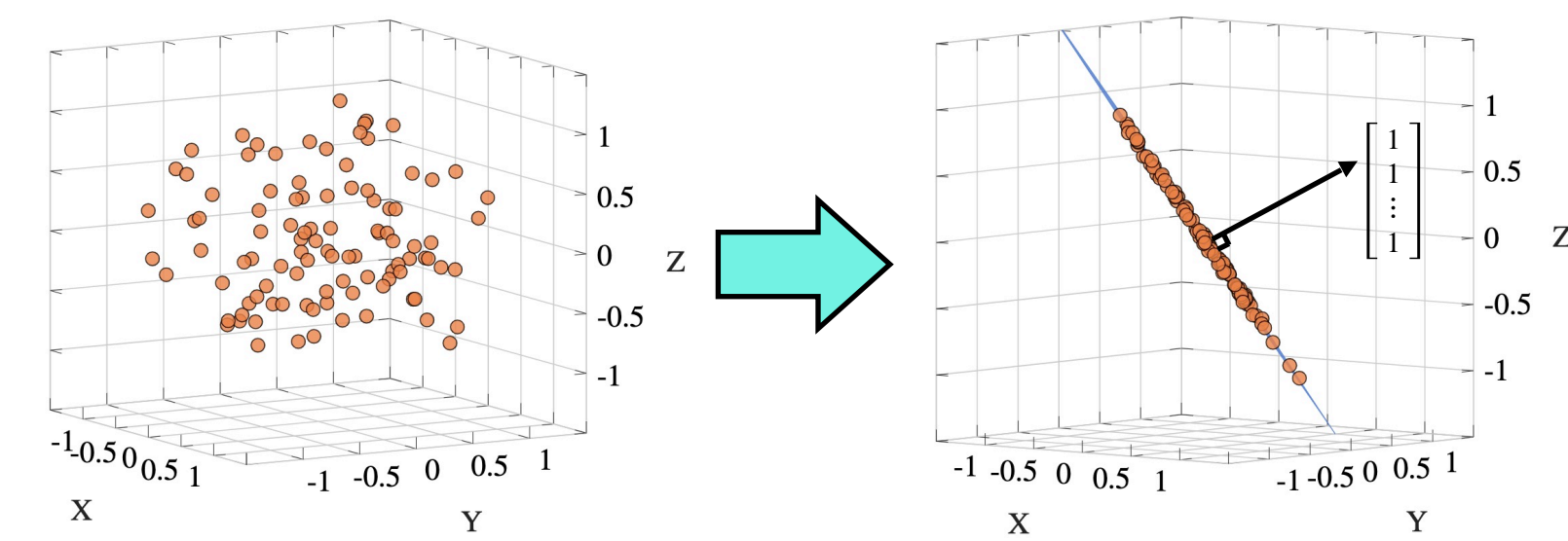
*projection*



*scaling*

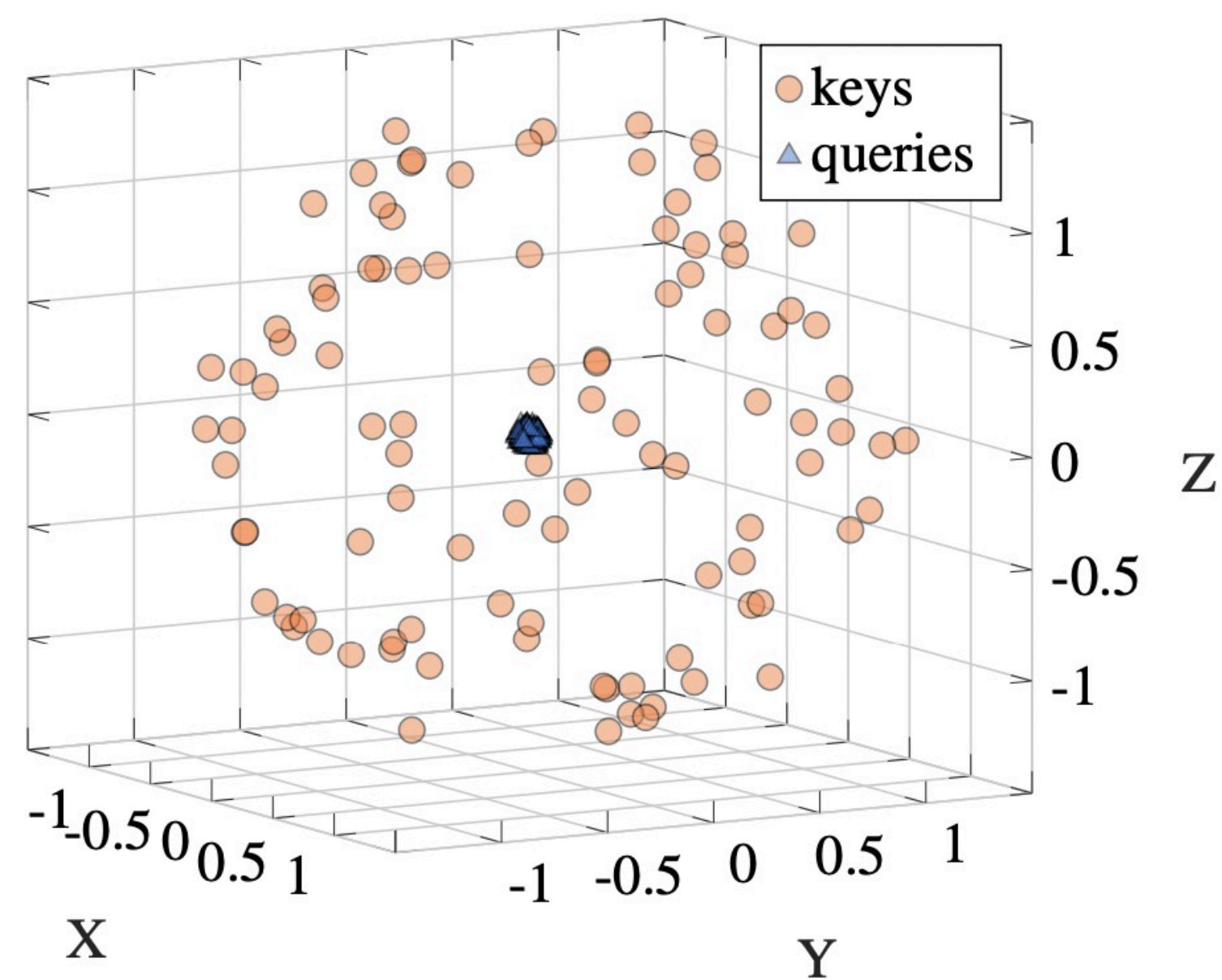


# LayerNorm – Projection

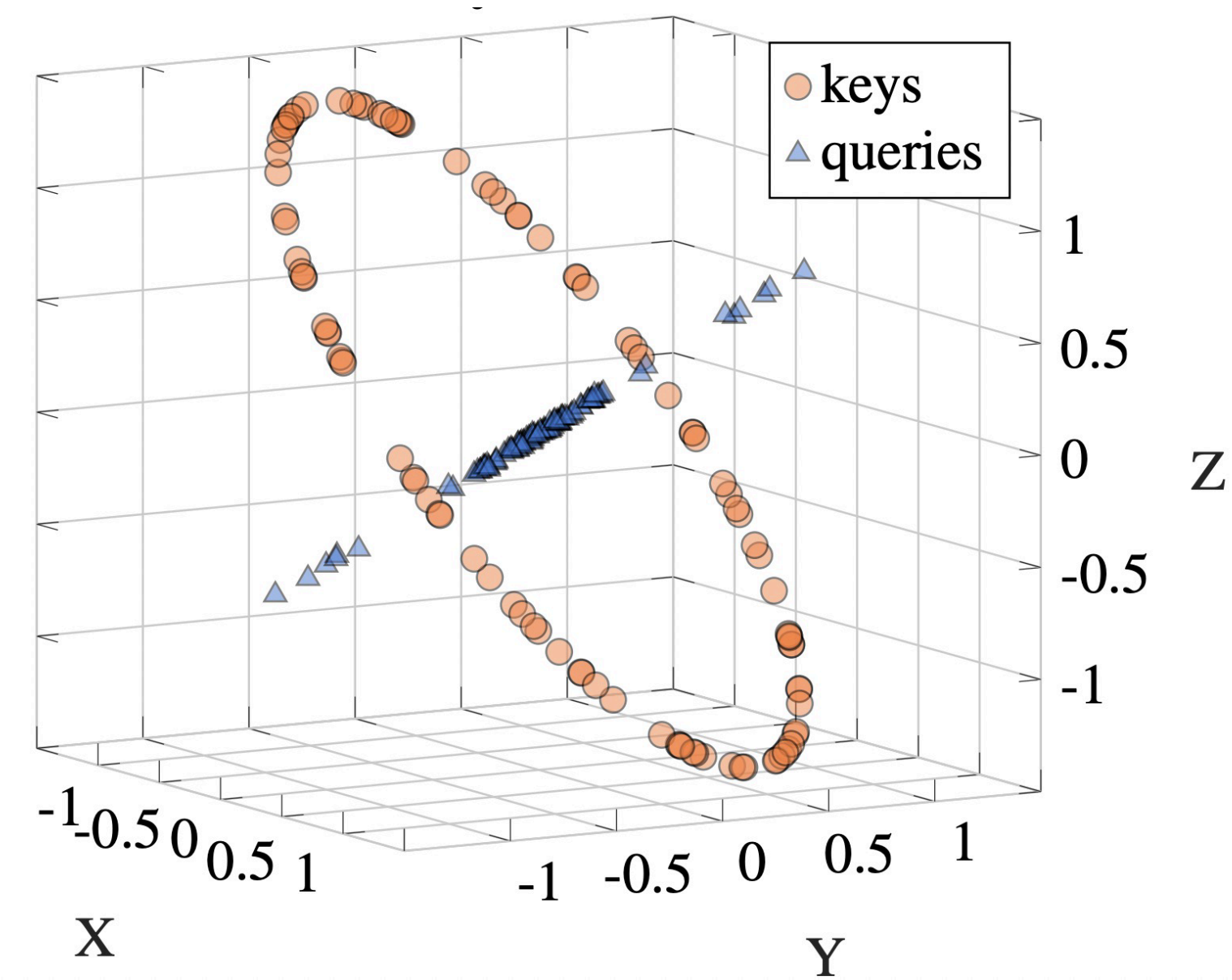


LayerNorm projects the key vectors onto the same hyperplane

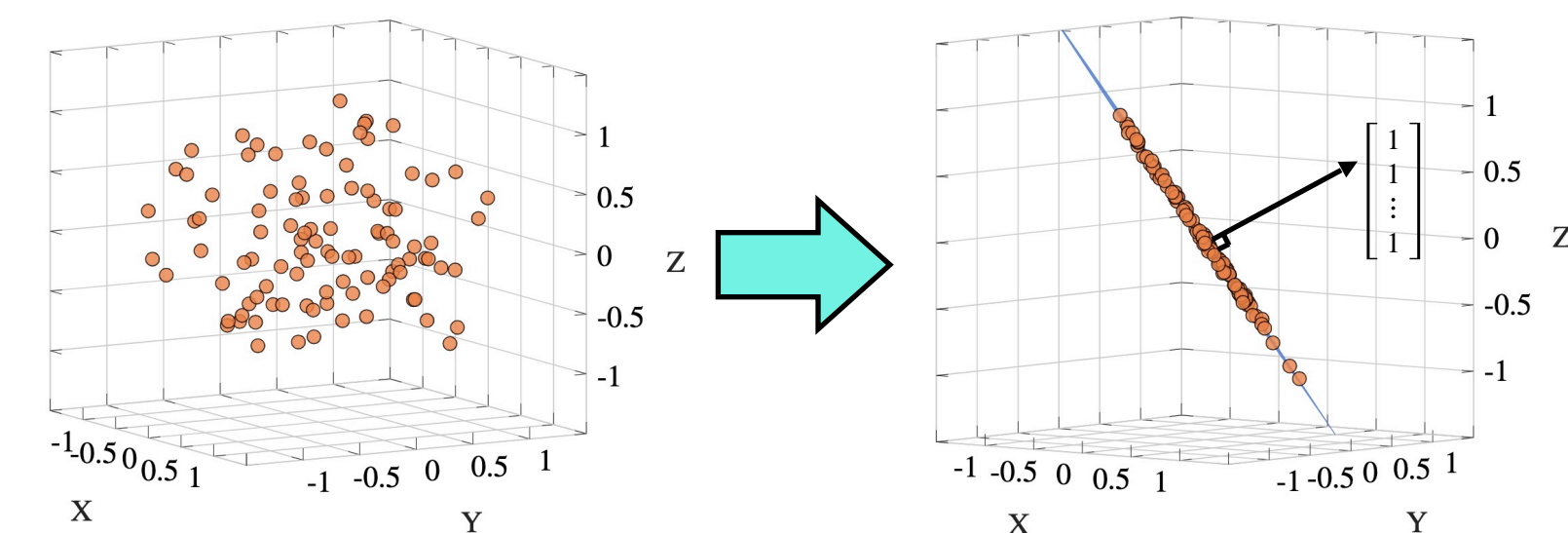
*w/o projection*



*w/ projection*

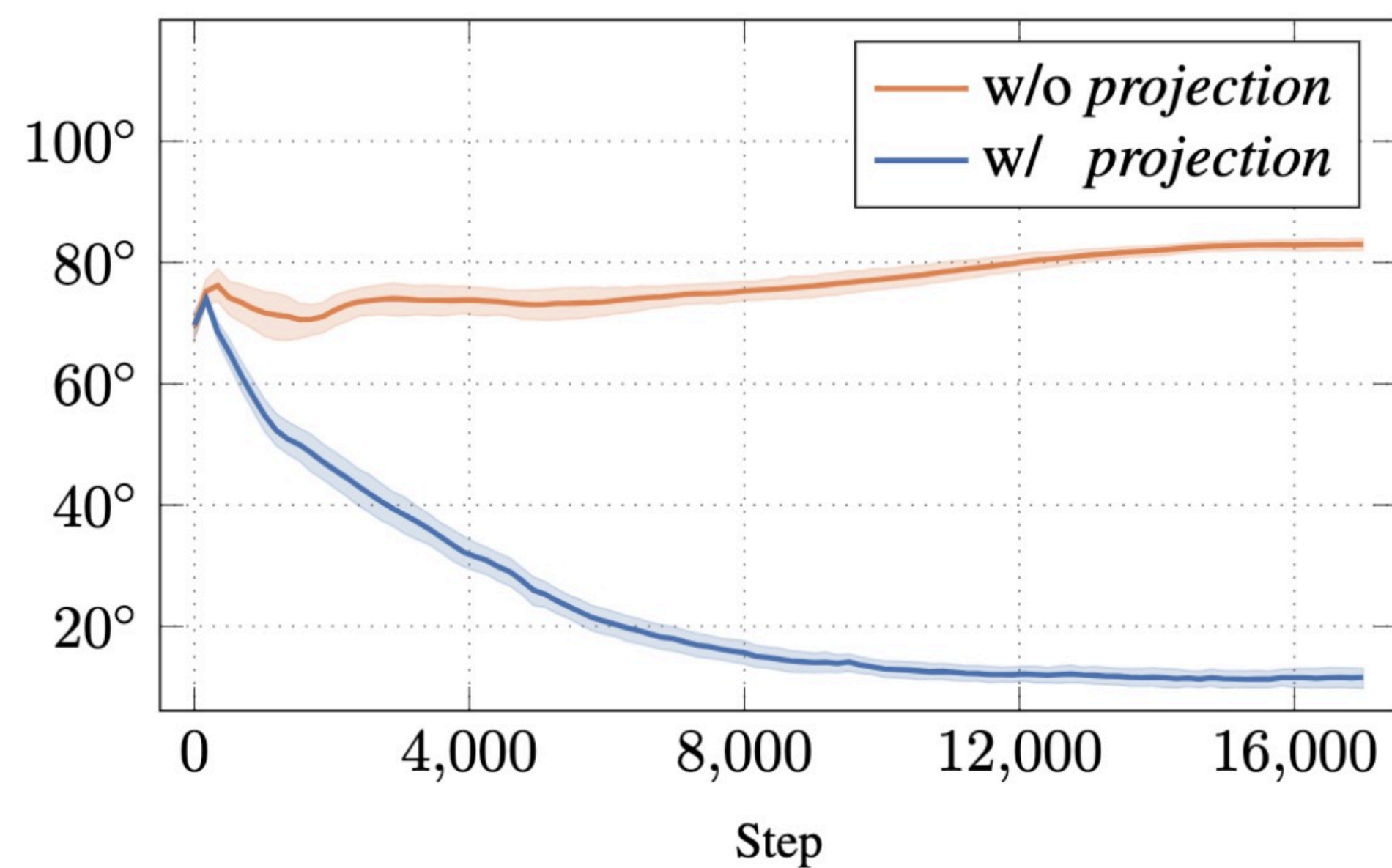


# LayerNorm – Projection

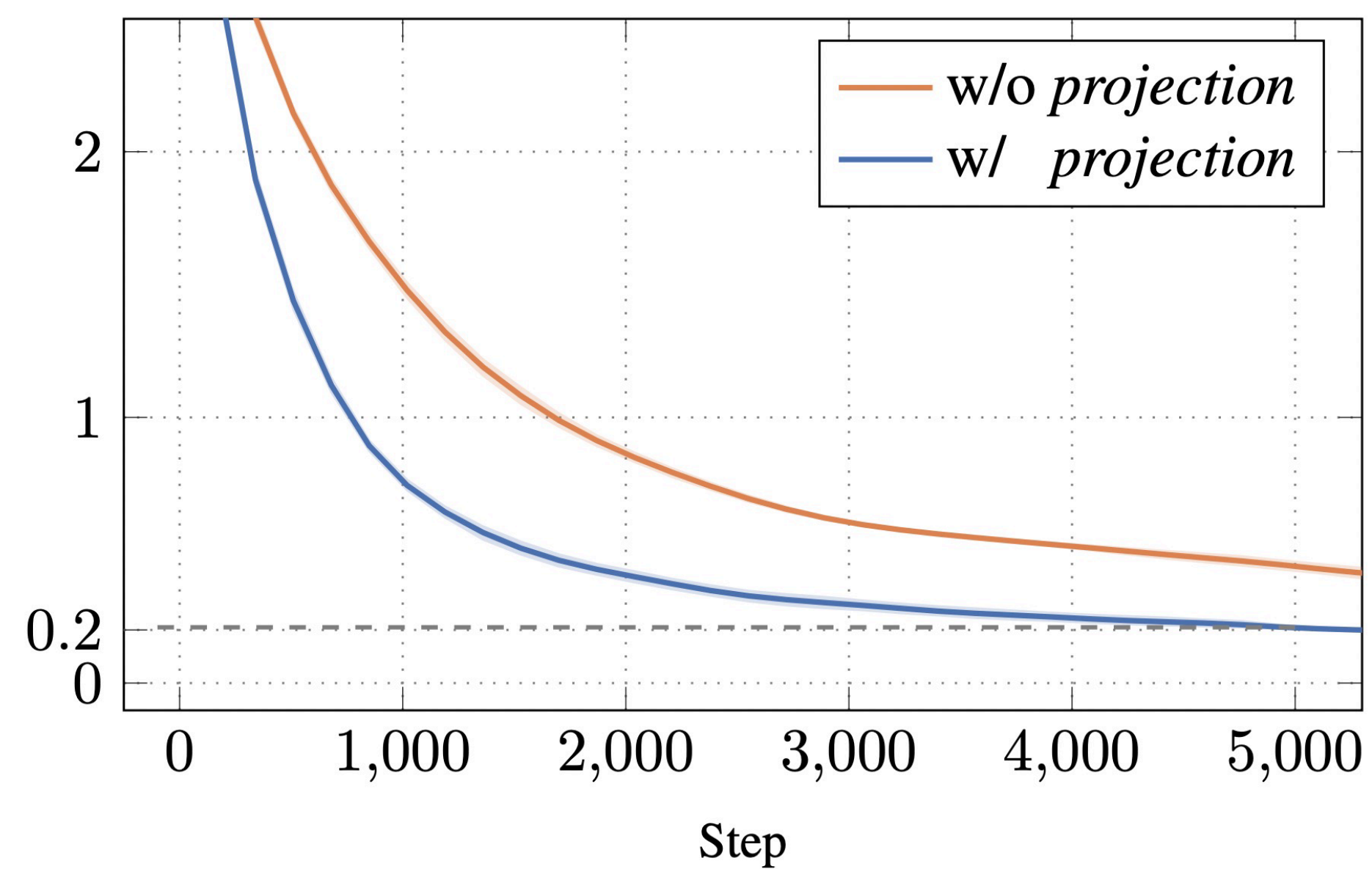


“Majority” task: Predict the most frequent token

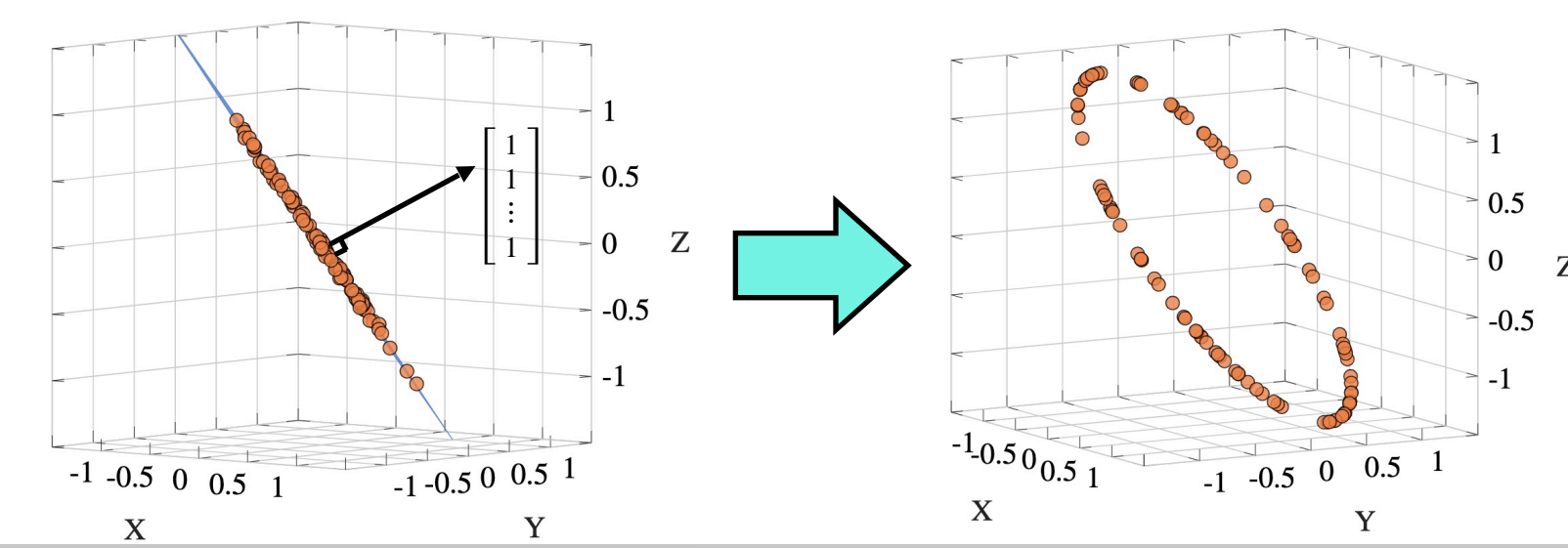
Angle to  $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$



Training Loss

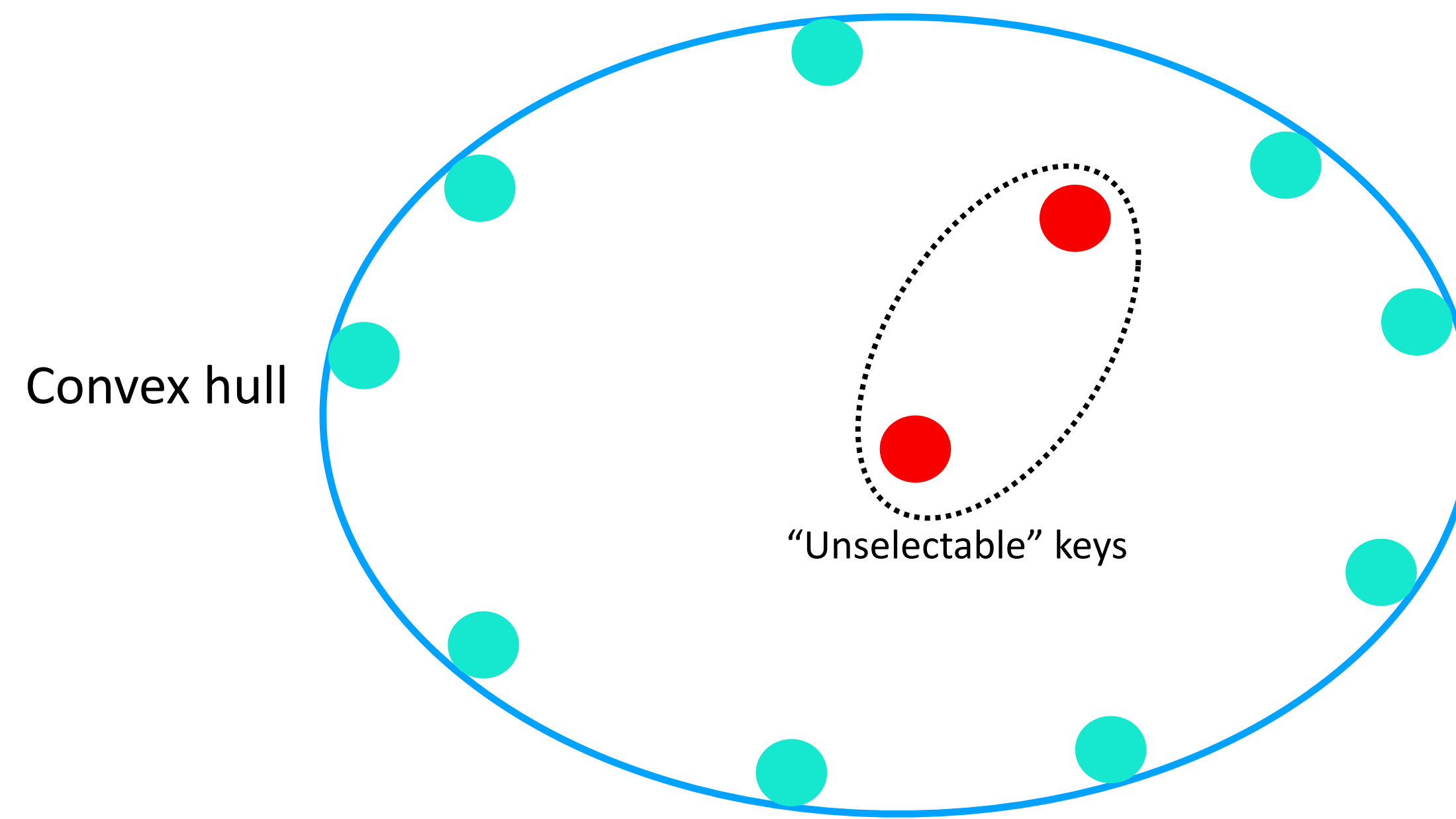


# LayerNorm – Scaling

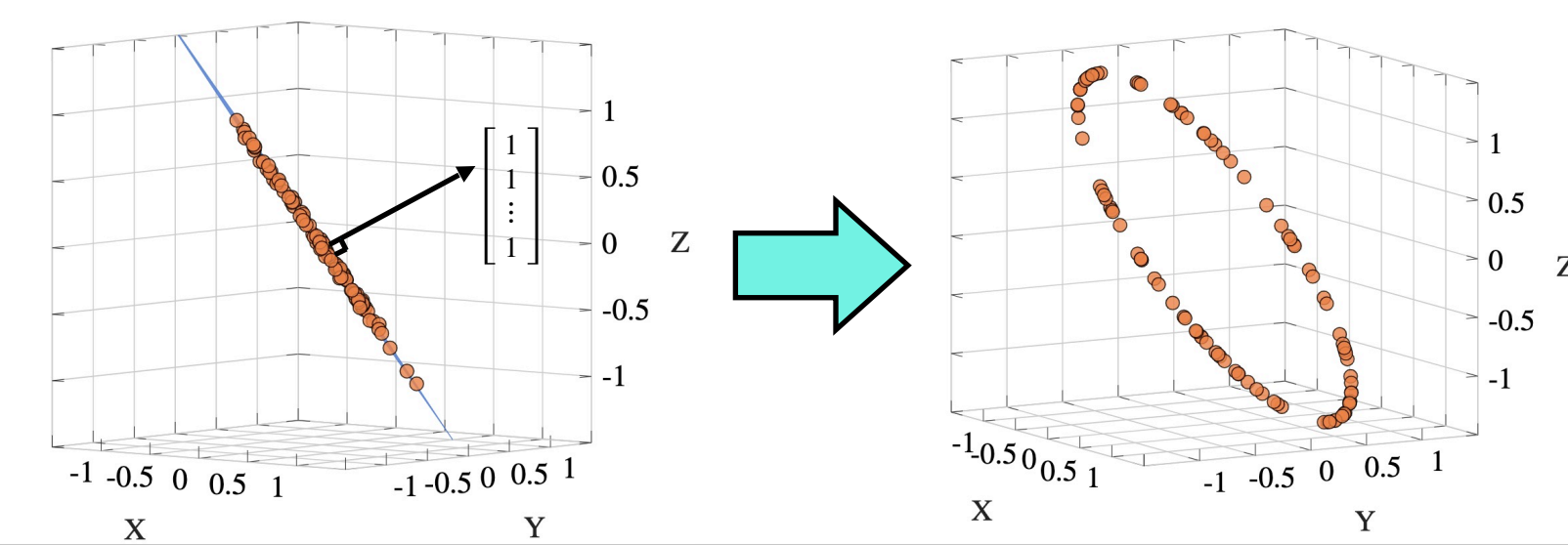


“Unselectable” keys problem [Demter et al., 2020; Grivas et al., 2022]

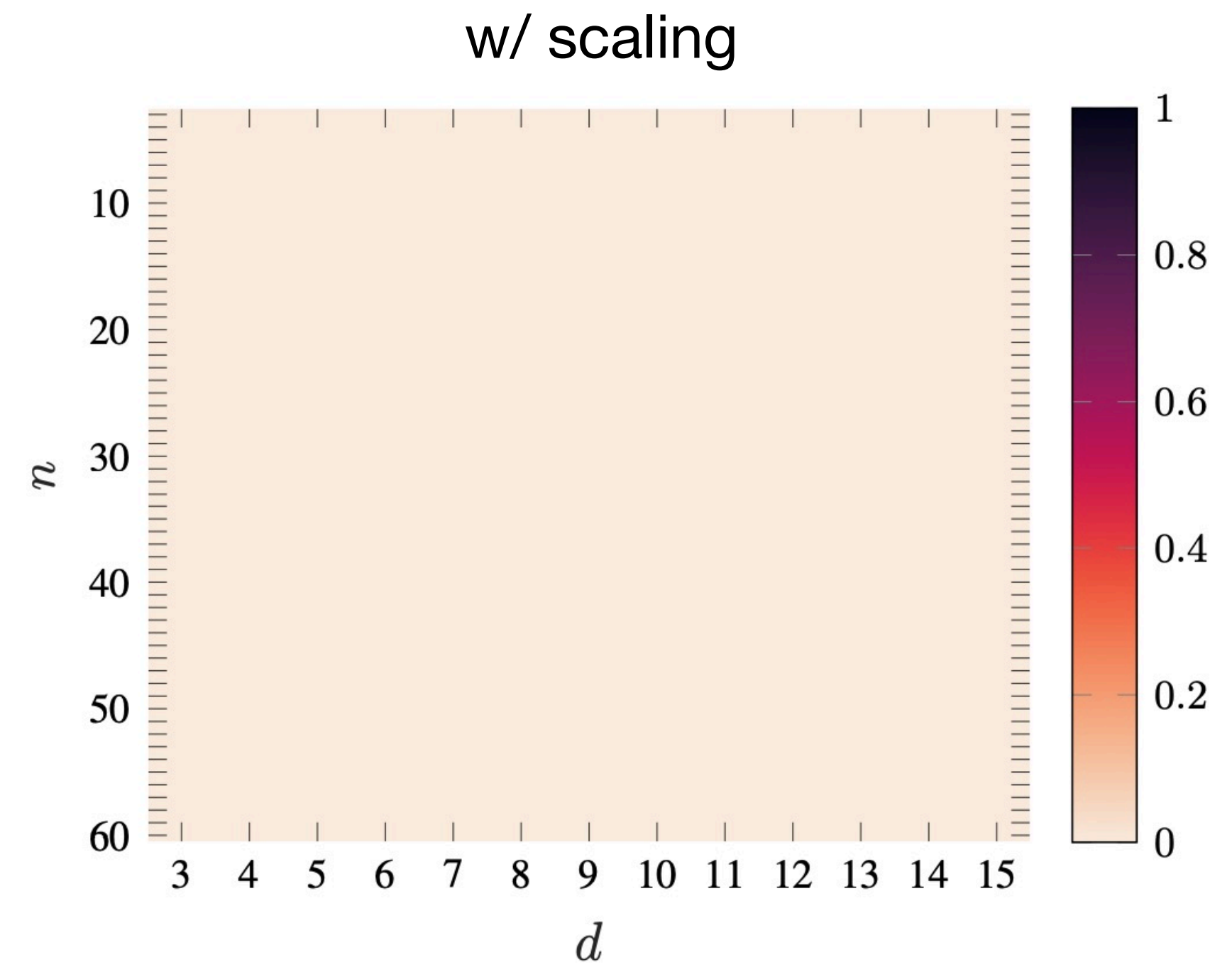
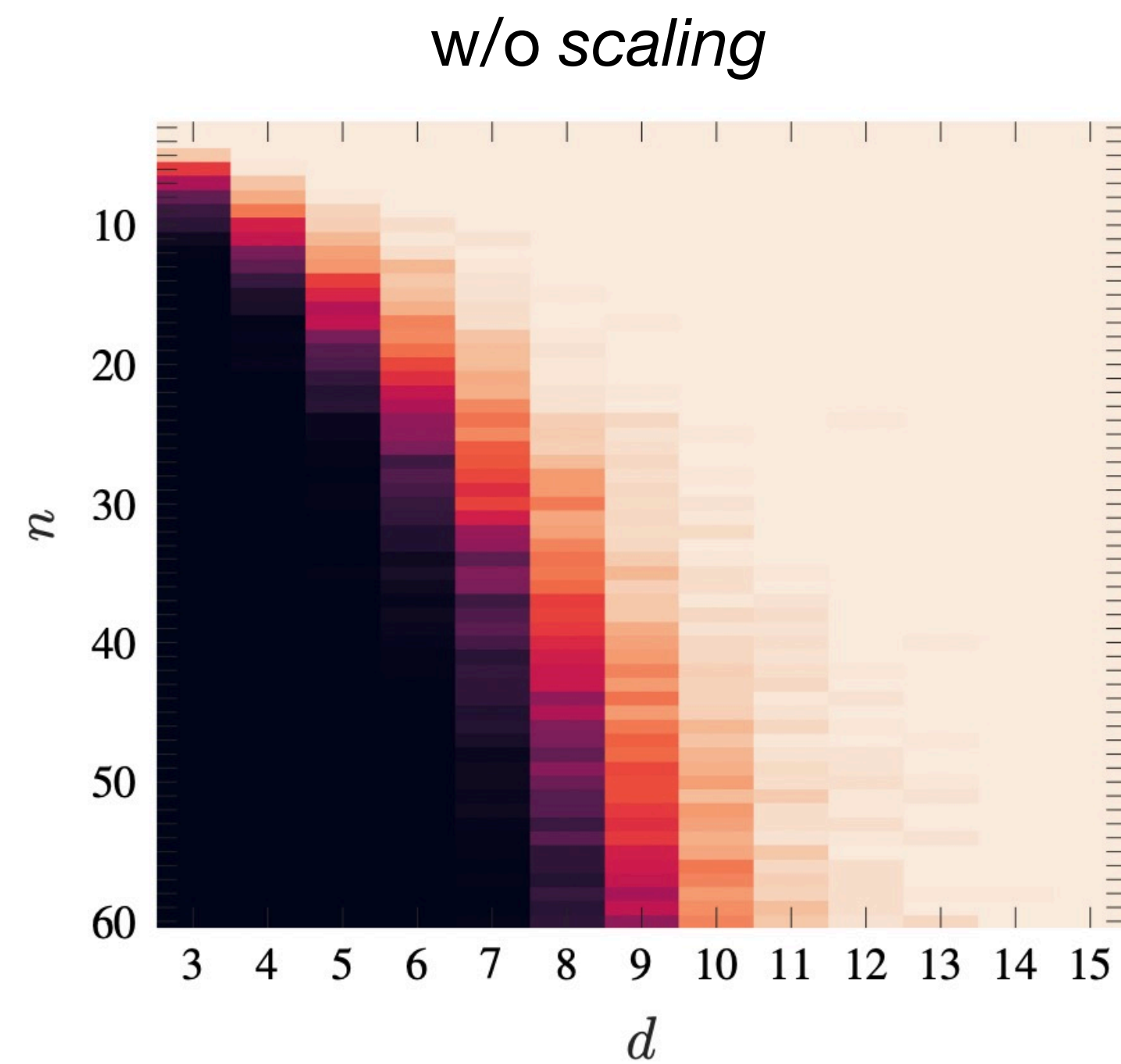
**Interior** key vectors cannot receive the highest attention score



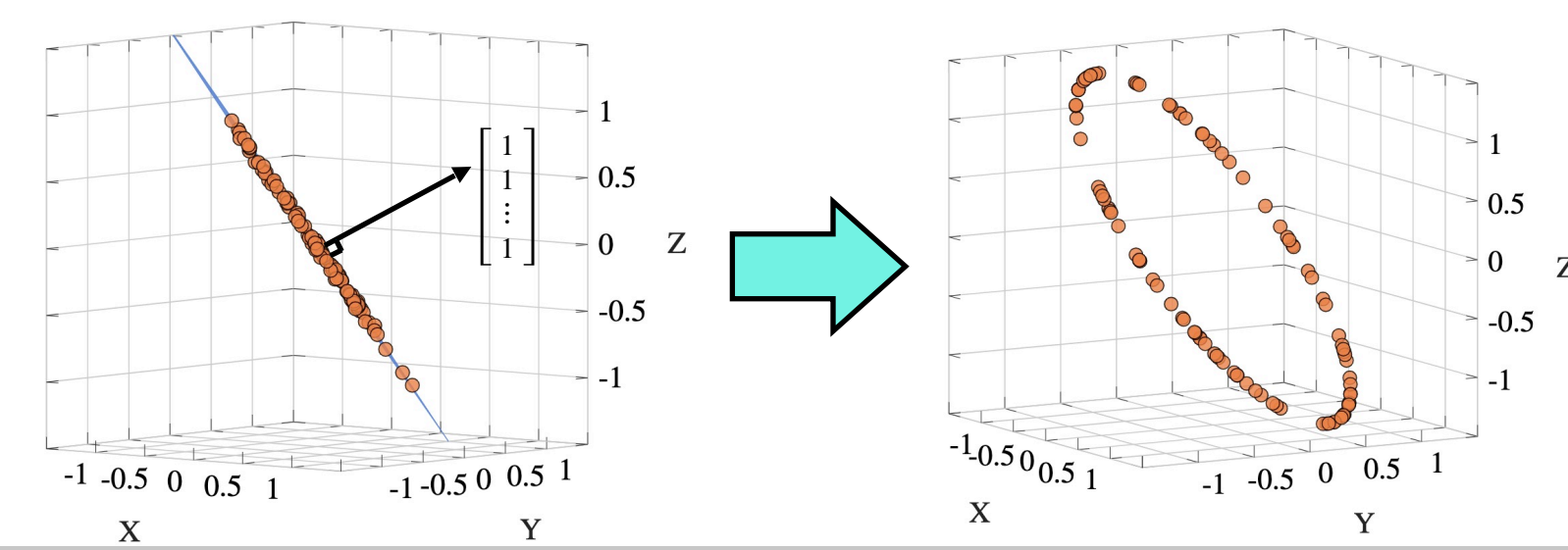
# LayerNorm – Scaling



The fraction of “unselectable”  $n$  vectors of dimension  $d$



# LayerNorm – Scaling



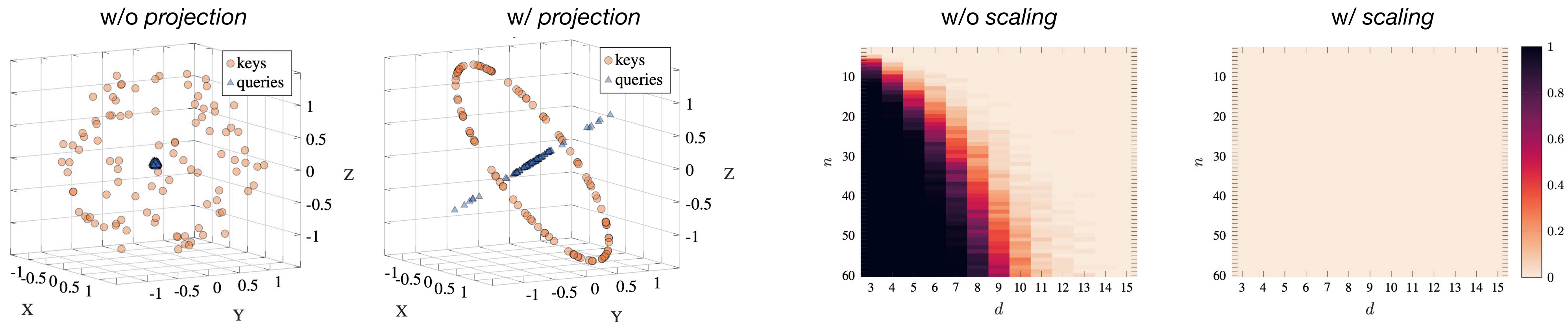
The fraction of “unselectable” key vectors in different layers of a language model

Model	$L_1$	$L_2$	$L_3$	$L_4$
w/o <i>scaling</i>	51.0	32.2	34.7	36.8
w/ <i>scaling</i>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>



# Summary

- LayerNorm is crucial for the expressivity of attention in Transformers
- LayerNorm can be seen as a **projection** followed by a **scaling** operation
- *Projection* helps the model to learn to attend equally to all keys
- *Scaling* eliminates the “unselectable” keys problem



[shakedbr@cs.technion.ac.il](mailto:shakedbr@cs.technion.ac.il)

[https://github.com/tech-srl/layer\\_norm\\_expressivity\\_role](https://github.com/tech-srl/layer_norm_expressivity_role)