NEURAL INFORMATION PROCESSING SYSTEMS

Accepted at NeurIPS 2025

Sinusoidal Initialization: Time for a New Start

Alberto Fernández-Hernández[1,*], Jose I. Mestre[2,*], Manuel F. Dolz[2], Jose Duato[3], Enrique S. Quintana-Ortí[1]

[1] Universitat Politècnica de València (UPV), Spain

[2] Universitat Jaume I (UJI), Spain

[3] Openchip Software & Technologies S.L.

* Equal contribution

UPV UJI

Do we need randomness in DNN initializations?

- Randomness is usually assumed essential for training
- Glorot & He initializations as key milestones

We propose Sinusoidal Initialization

- Fully deterministic
- Maximizes expressivity from layer one
- Boosts convergence and accuracy

Sinusoidal Initialization

Let $W \in \mathbb{R}^{m \times n}$ be the layer weight matrix

$$W[i,j] \; = \; a \cdot \sin(k_i \cdot x_j + \phi_i),$$

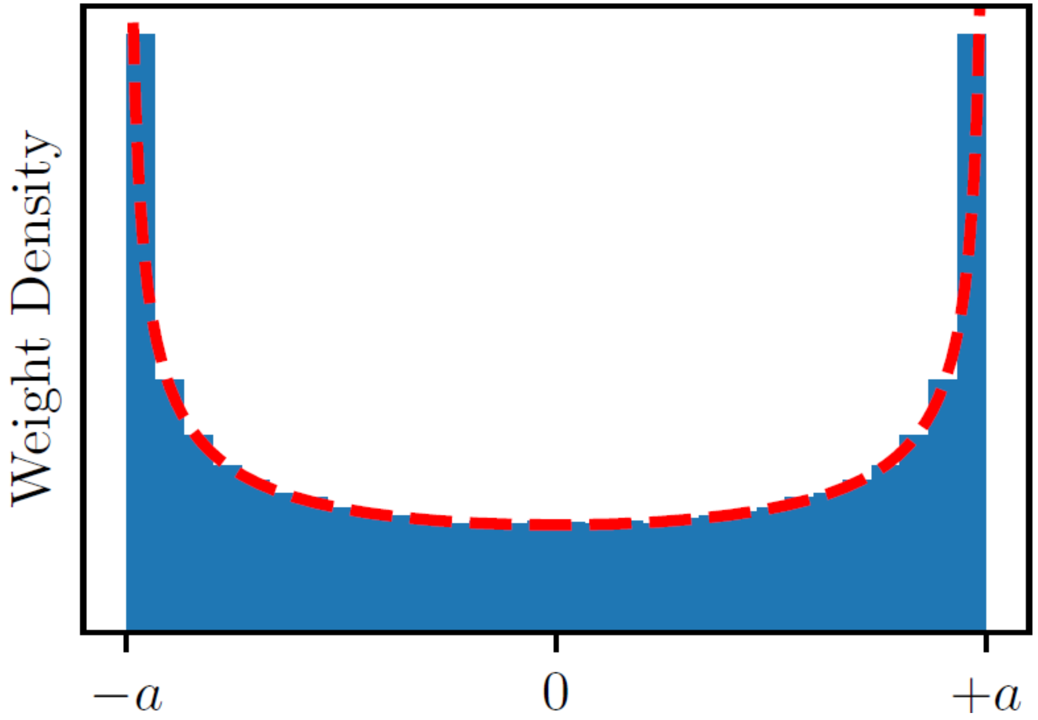
with

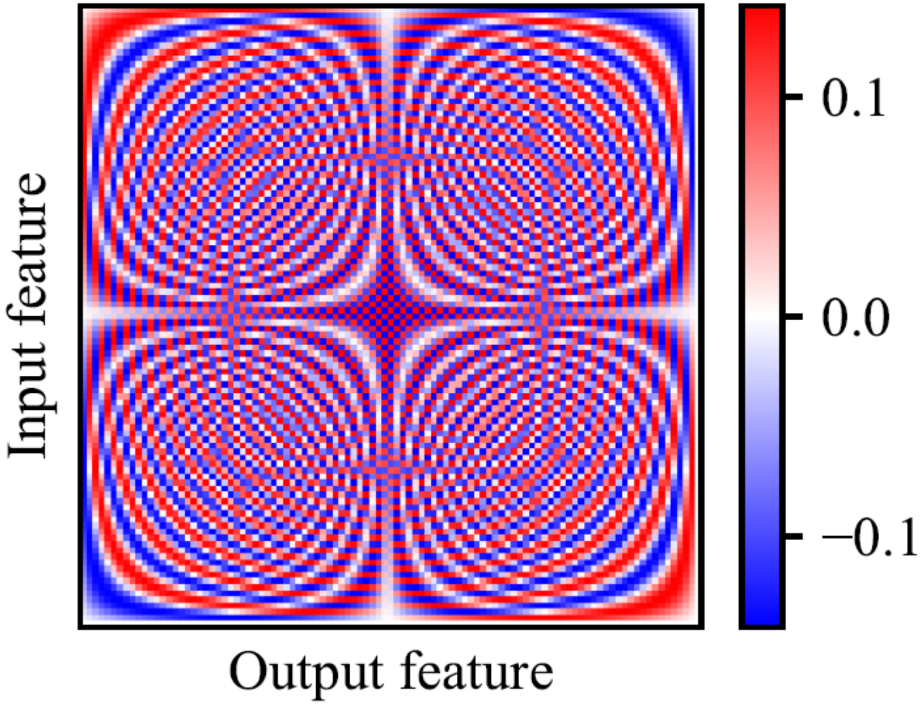
$k_i = 2\pi i, x_j = j/n,\ \phi_i = 2\pi i/m,$

a: scaling factor preserving variance

$\rightarrow$ Fully deterministic, variance-preserving initialization

Sinusoidal Initialization





Left: weight distribution. Right: visualization of the weight matrix w.

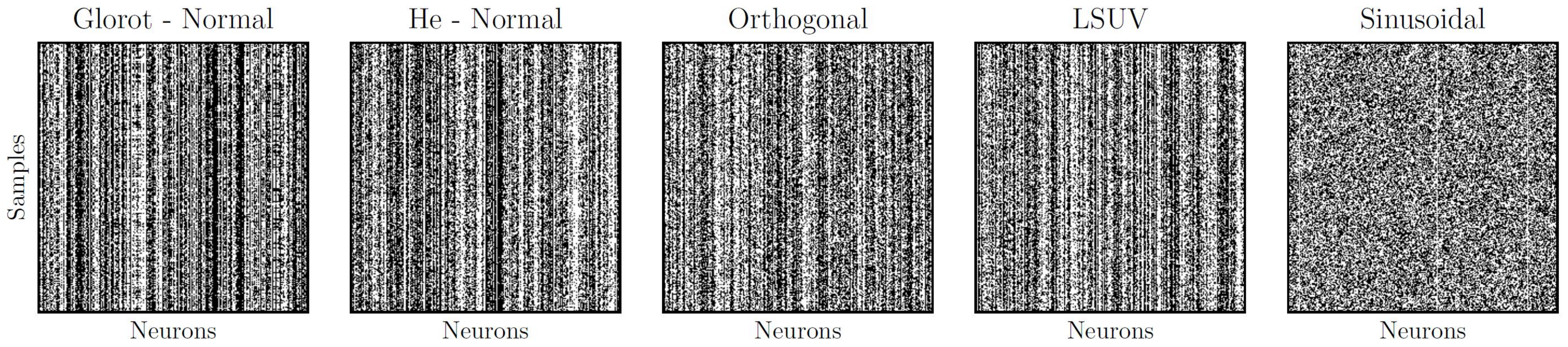
Why is this better than just random initializations?

Definition 1. A neuron Z is *skewed with degree* α if

$$|\, P(Z > 0\, | W_1, \ldots, W_n) - 1/2\, | > \alpha.$$

→ Skewed neurons activate unevenly (not 50%-50%)

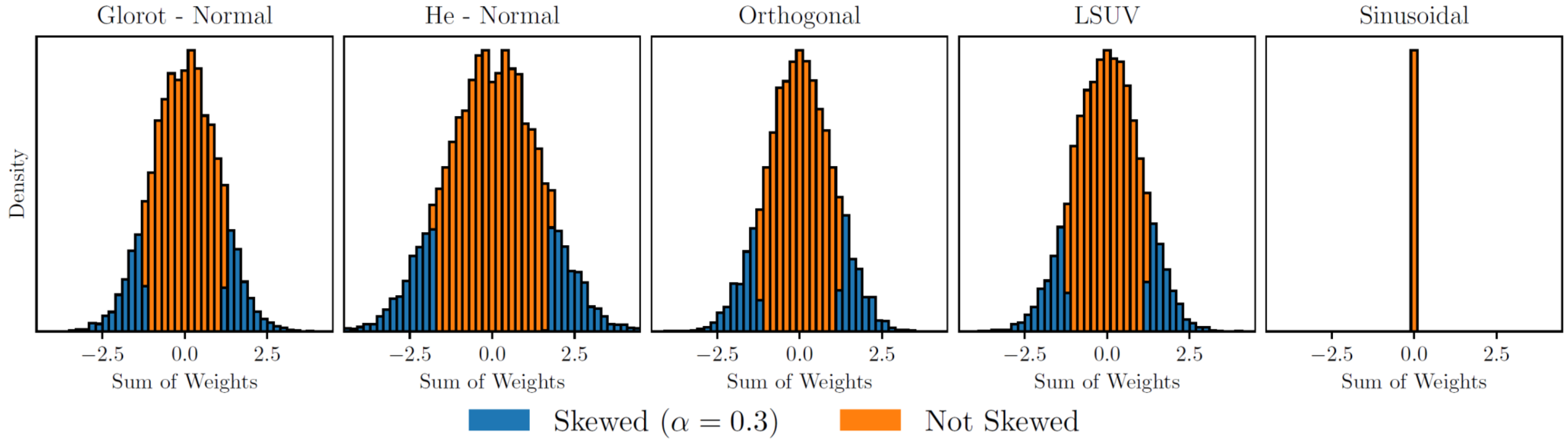
Why is this better than just random initializations?



Why are some neurons skewed? Explained by $S = W_1 + W_2 + \cdots + W_n$.

Theorem 1 (simplified version). A neuron Z is skewed with degree α **if and only if** $|S| > \lambda$, where λ is a unique constant depending on α.

Why is this better than just random initializations?



→ Skewed neurons = tails of the s distribution

→ Random inits → stochastic s → skewed neurons

→ **Sinusoidal Initialization:** sine symmetry ⇒ $s = 0$ ⇒ <u>no skewness</u>

Experimental results

- Several network-dataset-optimizer combinations
- Compared initialization schemes (Default, Orthogonal, LSUV, Sinusoidal).

Model / Dataset	Optim.	Maximum accuracy (%)				AUC			
		Def.	Orth.	LSUV	**Sin.**	Def.	Orth.	LSUV	**Sin.**
ResNet-50 CIFAR-100	SGD	37.3	46.5	44.8	**51.9**	25	35	31	**42**
	Adam	53.1	56.6	56.3	**61.5**	48	51	51	**57**
	AdamW	67.5	67.7	69.7	**71.0**	64	65	66	**68**
MobileNetV3 TinyImageNet	SGD	18.4	25.8	**28.0**	21.6	26	**38**	35	36
	Adam	32.8	34.4	**35.2**	34.8	62	66	**71**	65
	AdamW	40.9	**43.6**	40.1	42.6	79	**85**	76	82
EfficientNetV2 TinyImageNet	SGD	28.1	30.9	**32.1**	32.0	47	52	**56**	56
	Adam	27.7	29.8	32.7	**36.6**	53	56	62	**70**
	AdamW	50.0	50.2	49.3	**53.5**	100	100	100	**106**
ViT-16 ImageNet-1k	SGD	28.6	28.2	29.6	**31.5**	23	**25**	24	**25**
BERT-mini WikiText	AdamW	40.4	**42.2**	15.9	41.1	58	**72**	32	**72**

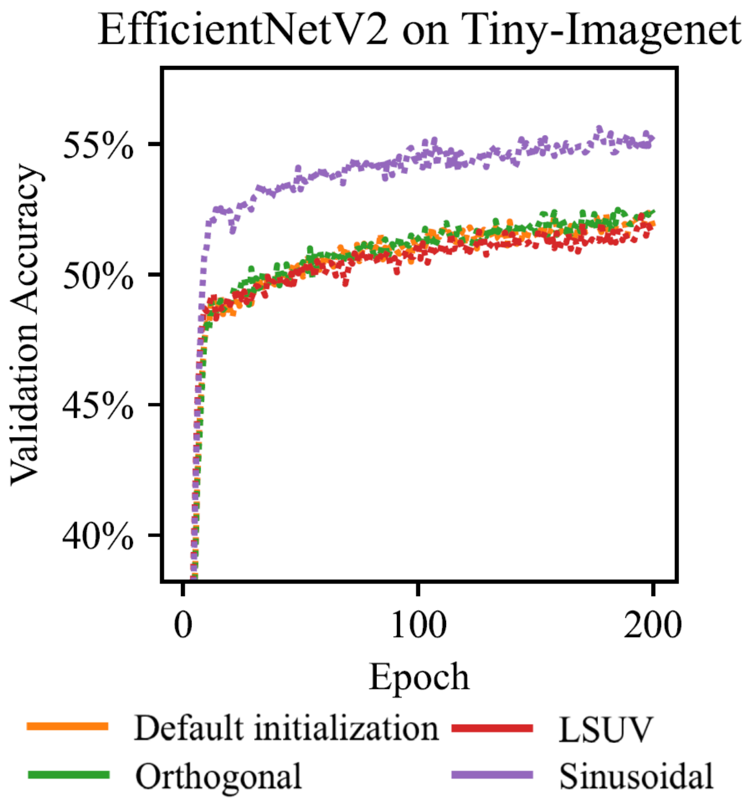
Experimental results

Performance summary

+4.9% average gain in final accuracy

+20.9% faster convergence (AUC)

Consistent across models and datasets



Conclusion

- Randomness **not essential** for effective initialization
- **Sinusoidal Initialization** → deterministic, expressive, stable
- Strong **theoretical & empirical** support
- A step forward in understanding DNN initialization

a.fernandez@upv.es