Inicialización Sinusoidal: una alternativa determinista para el entrenamiento de DNNs

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

[1] Universitat Politècnica de València, Valencia, Spain
a.fernandez@upv.es, cpercor@upv.es, quintana@disca.upv.es
[2] Universitat Jaume I, Castellón, Spain
jmiravet@uji.es, dolzm@uji.es
[3] Openchip & Software Technologies S.L., Spain
jose.duato@openchip.com

Basado en la referencia [1] titulada *Sinusoidal Initialization, Time for a New Start*, publicada en NeurIPS 2025 por los autores.

Abstract. La inicialización de pesos condiciona la estabilidad de la propagación, la velocidad de convergencia y el aprovechamiento de la capacidad expresiva de una red profunda. Los esquemas dominantes, como Glorot y He, preservan la varianza en esperanza, pero dependen de una realización aleatoria concreta y pueden inducir configuraciones iniciales poco favorables. En este trabajo se resume una alternativa determinista, la Inicialización Sinusoidal, que construye las matrices de pesos mediante funciones sinusoidales y conserva el escalado de varianza necesario para una propagación estable. La propuesta parte de dos ideas: que una fracción relevante de la ineficiencia temprana proviene de neuronas sesgadas, y que una estructura determinista con simetría y diversidad neuronal puede reducir ese fenómeno sin perder expresividad. La construcción sinusoidal impone cancelación exacta por filas y favorece un régimen inicial más equilibrado. Experimentalmente, mejora la convergencia y, en promedio, también la precisión final frente a inicializaciones aleatorias y semiestructuradas en CNN, Vision Transformers y modelos de lenguaje.

Keywords: inicialización de pesos · aprendizaje profundo · redes neuronales profundas · Transformers · optimización

1 Introducción

La inicialización sigue siendo un componente decisivo en el entrenamiento de redes neuronales profundas, aunque a menudo quede relegada frente al diseño arquitectónico o la elección del optimizador. Los métodos clásicos de Glorot [2] y He [3] se apoyan en el principio de preservación de varianza para evitar la degradación de activaciones y gradientes al aumentar la profundidad. Sobre esa base se han propuesto variantes como la inicialización LSUV [4] o la ortogonal

[5], que buscan mejorar el acondicionamiento inicial sin abandonar por completo la aleatoriedad.

El punto de partida aquí es simple: incluso cuando la varianza está bien ajustada en esperanza, cada entrenamiento arranca desde una única realización aleatoria. Esto invita a preguntarse si esa dependencia del azar es realmente necesaria. La alternativa considerada sustituye el muestreo i.i.d. por una construcción determinista basada en funciones sinusoidales, diseñada para conservar el control sobre la varianza e imponer una organización más simétrica en las contribuciones entrantes a cada neurona.

La motivación no es meramente formal. Las inicializaciones aleatorias tienden a producir neuronas sesgadas, esto es, unidades que nacen mayoritariamente activas o inactivas desde la primera propagación. Cuando esto ocurre, la red aprovecha peor sus no linealidades y arranca desde una configuración menos eficiente. Frente a ello, la Inicialización Sinusoidal persigue dos objetivos: equilibrio entre activación e inactivación, e independencia funcional entre neuronas.

2 Inicialización Sinusoidal

Sea $W \in \mathbb{R}^{m \times n}$ la matriz de pesos de una capa con n entradas y m salidas. La construcción propuesta define

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

con $k_i = 2\pi i$, $x_j = j/n$, $\phi_i = 2\pi i/m$, para $i \in \{1, \ldots, m\}$ y $j \in \{1, \ldots, n\}$. Cada fila de W queda asociada a una onda sinusoidal muestreada uniformemente, con frecuencia y fase propias para cada neurona. La amplitud a se ajusta para que la varianza coincida con el escalado deseado, en particular el de Glorot, $2/(m + n)$, y los sesgos se inicializan a cero.

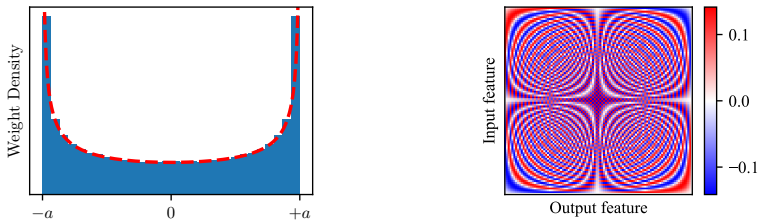


Fig. 1. Visualización de la Inicialización Sinusoidal. Izquierda: distribución marginal de pesos inducida por la construcción. Derecha: mapa de calor de la matriz inicializada, que muestra la estructura armónica por filas.

Como muestra la Figura 1, esta construcción induce una distribución marginal no gaussiana, de tipo arcoseno, y genera una matriz con un patrón oscilatorio bien definido. En ese sentido, existe una afinidad formal con las codificaciones sinusoidales del transformer, aunque aquí el objetivo es construir una inicialización más equilibrada.

3 Motivación teórica

La motivación teórica parte de una idea simple: una neurona es poco útil si nace casi siempre activa o casi siempre inactiva. Para activaciones tipo ReLU y afines, ese comportamiento puede formalizarse mediante la noción de neurona sesgada. Si Z representa la salida de una neurona y $\alpha \in (0, 1/2)$, diremos que la neurona es sesgada con grado α si $|\mathbb{P}(Z > 0 \mid W_1, \ldots, W_n) - 1/2| > \alpha$.

Esta noción está estrechamente ligada a una estadística muy simple: la suma de los pesos entrantes, $S_n = \sum_{i=1}^{n} W_i$. Si S_n se desvía de cero, la neurona arranca con una preferencia direccional agregada antes incluso de aprender.

A partir de esta cantidad, el análisis establece dos hechos. El primero es que, asintóticamente, el sesgo de una neurona queda ligado a que $|S_n|$ supere un cierto umbral. El segundo es más fuerte: cuando los pesos se inicializan como variables i.i.d. centradas y con varianza finita, la probabilidad de obtener neuronas sesgadas tiende a uno al crecer la dimensión de entrada. Esto sugiere que el problema no es una rareza asociada a malas semillas, sino una consecuencia bastante natural del propio mecanismo de inicialización aleatoria.

La Inicialización Sinusoidal actúa precisamente sobre ese punto imponiendo una simetría fuerte: la suma de cada fila es exactamente cero. Concretamente, para todo $m < n$ y para todo $i \in \{1, \ldots, m\}$, se tiene que $\sum_{j=1}^{n} W[i, j] = 0$.

Esa cancelación evita que una neurona nazca con un desequilibrio agregado en sus entradas. La diferencia entre ambos regímenes se aprecia en la Figura 2: mientras que una inicialización aleatoria produce columnas dominadas por neuronas persistentemente activas o inactivas, la construcción sinusoidal da lugar a patrones mucho más homogéneos. A esto se suma un segundo rasgo importante: la combinación de frecuencias y fases distintas hace que las neuronas respondan de forma más diversa entre sí.

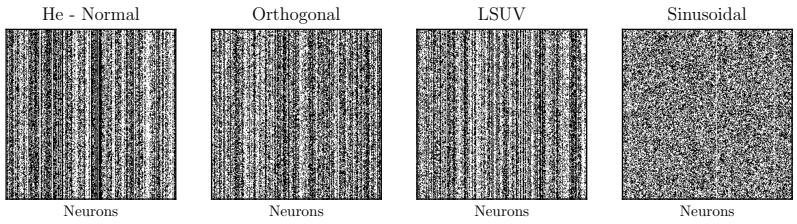


Fig. 2. Comparación cualitativa del estado inicial de las neuronas para distintas inicializaciones.

4 Resultados experimentales

La evaluación empírica considera cinco configuraciones: ResNet-50 sobre CIFAR-100, MobileNetV3 sobre Tiny-ImageNet, EfficientNetV2 sobre Tiny-ImageNet, ViT-B16 sobre ImageNet-1K y BERT-mini sobre WikiText. La comparación incluye la inicialización por defecto de cada modelo, la inicialización ortogonal y LSUV, usando SGD, Adam o AdamW según la estabilidad de cada caso. El

objetivo es responder a dos preguntas: si la Inicialización Sinusoidal mejora la precisión final y si, además, permite llegar antes a buenas soluciones.

La tendencia general es clara. En promedio, la Inicialización Sinusoidal mejora la precisión final en un 4.9% y acelera la convergencia, medida mediante AUC, en un 20.9% respecto a la referencia por defecto. Lo relevante no es solo la magnitud de la mejora, sino su regularidad: aparece en arquitecturas muy distintas entre sí y no queda ligada a un único optimizador. Como puede verse en la Fig. 3, la ventaja suele manifestarse desde las primeras épocas y, en varios casos, se conserva hasta el final del entrenamiento.

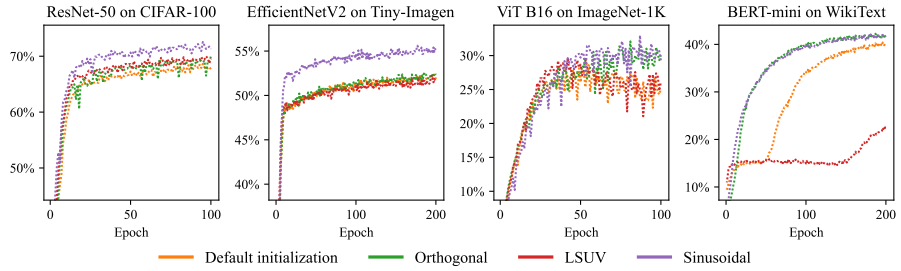


Fig. 3. Curvas de entrenamiento comparando la precisión en validación obtenida por la Inicialización Sinusoidal con las alternativas de referencia en varias arquitecturas.

5 Conclusión

La idea central puede expresarse de forma directa: la aleatoriedad no es un requisito imprescindible para construir una buena inicialización. La Inicialización Sinusoidal combina preservación de varianza, simetría estructural y diversidad funcional dentro de una construcción completamente determinista. El análisis teórico conecta las inicializaciones aleatorias con la aparición de neuronas sesgadas, y los resultados experimentales muestran que esta alternativa se traduce en mejoras consistentes de convergencia y de precisión final.

References

1. Fernández-Hernández, A., Mestre, J.I., Dolz, M.F., Duato, J., Quintana-Ortí, E.S.: Sinusoidal Initialization, Time for a New Start. In: Advances in Neural Information Processing Systems (2025), https://openreview.net/forum?id=FGliQVcrDZ
2. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. Proceedings of Machine Learning Research, vol. 9, pp. 249–256. PMLR (2010), https://proceedings.mlr.press/v9/glorot10a.html
3. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: IEEE International Conference on Computer Vision (ICCV) (2015). https://doi.org/10.1109/ICCV.2015.123
4. Mishkin, D., Matas, J.: All you need is a good init (2016), https://arxiv.org/abs/1511.06422
5. Saxe, A.M., McClelland, J.L., Ganguli, S.: Exact solutions to the nonlinear dynamics of learning in deep linear neural networks (2014), https://arxiv.org/abs/1312.6120