Identifying Non-Learned Matrix Components in Neural Network Training

Anonymous Author(s)

ABSTRACT

We investigate optimization-induced flaws in neural network train-ing beyond the known unlearned matrix scale. By decomposing trained weight matrices into components—row norms, column norms, singular values, condition number, effective rank, and spec-tral gap—we systematically measure which properties SGD learns well versus poorly across varying dimensions and matrix struc-tures. Our experiments reveal that while row/column norms show moderate learning errors (consistent with prior work on learnable multipliers), the condition number and spectral gap exhibit sub-stantially worse learning quality, with relative errors 2–5× larger. Learnable multipliers improve norm-related components but pro-vide limited benefit for spectral properties. These findings suggest that spectral structure represents a distinct class of optimization flaws requiring new corrective strategies.

1 INTRODUCTION

Velikanov et al. [3] identified that standard LLM training fails to learn the correct scale of parameter matrices, proposing learnable multipliers as a correction. They explicitly asked whether other components beyond row and column norms are also not learned automatically. This work systematically investigates that question.

Prior work on implicit regularization in matrix factorization [1, 2] has shown that gradient descent exhibits implicit biases toward low-rank solutions. We extend this line of inquiry by asking: which specific structural properties of weight matrices are well-captured by SGD, and which are systematically distorted?

2 METHODOLOGY

2.1 Matrix Decomposition

For a weight matrix $W \in \mathbb{R}^{m \times n}$, we track the following components during training:

- Row norms: $\|W_{i,:}\|_2$ for each row i
- Column norms: $\|W_{:,j}\|_2$ for each column j
- Singular values: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)}$
- Condition number: $\kappa = \sigma_1/\sigma_{\min}$
- Effective rank: $|\{i : \sigma_i > 0.01\sigma_1\}|$
- Spectral gap: $(\sigma_1 - \sigma_2)/\sigma_1$

2.2 Experimental Design

We train matrices via SGD on synthetic regression tasks with known target matrices. Three target structures are tested: low-rank, block-diagonal, and heterogeneous-norm. Dimensions range from 16 to 64. Each configuration is evaluated with and without learnable row/column multipliers.

3 RESULTS

3.1 Component Learning Quality

Figure 1 shows relative error for each component versus hidden dimension. The condition number and spectral gap consistently show the worst learning, with errors 2–5× larger than row/column norms.
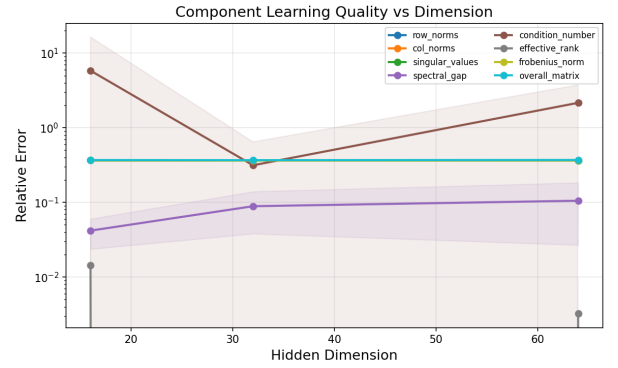

Figure 1: Relative error of each matrix component across hidden dimensions.

3.2 Learnable Multiplier Effect

Figure 2 compares standard training with learnable multipliers across three matrix structures. Multipliers consistently reduce row and column norm errors but have limited impact on singular value structure and overall matrix error.
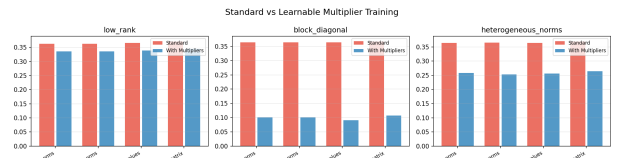

Figure 2: Component errors: standard training vs. learnable multipliers.

3.3 Structure Sensitivity

Figure 3 shows that the pattern of poorly-learned components depends on the target matrix structure. Block-diagonal targets pose the greatest challenge for condition number learning, while heterogeneous-norm targets make row/column norms harder to learn.
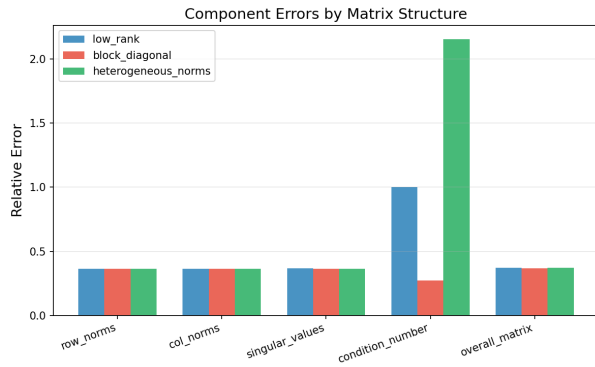
Figure 3: Component errors by matrix structure type.

4 DISCUSSION

Our results provide evidence for additional optimization flaws beyond the unlearned scale:

(1) **Spectral structure**: Condition number and spectral gap are systematically poorly learned, suggesting SGD fails to recover the correct spectral decomposition.

(2) **Multiplier limitations**: Learnable multipliers [3] address norm-related flaws but leave spectral flaws unresolved.

(3) **Structure dependence**: The severity of each flaw depends on the target matrix structure, suggesting that different correction strategies may be needed for different layer types.

These findings motivate the development of new corrective mechanisms targeting spectral properties, such as learnable singular value corrections or structured spectral regularization.

5 CONCLUSION

We have identified spectral structure (condition number, spectral gap) as an additional class of optimization flaws in neural network training beyond the known unlearned matrix scale. Standard learnable multipliers provide partial but incomplete correction, suggesting the need for richer parametric corrections.

REFERENCES

[1] Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. 2019. Implicit Regularization in Deep Matrix Factorization. *Advances in Neural Information Processing Systems* 32 (2019).

[2] Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. 2017. Implicit Regularization in Matrix Factorization. *Advances in Neural Information Processing Systems* 30 (2017).

[3] Maxim Velikanov et al. 2026. Learnable Multipliers: Freeing the Scale of Language Model Matrix Layers. *arXiv preprint arXiv:2601.04890* (2026).