Why Deep Jacobian Spectra Separate

Depth-Induced Scaling and Singular-Vector Alignment

Nathanaël Haas François Gatine Augustin Cosse Zied Bouraoui

ICML 2026

CRIL (Univ. Artois / CNRS) LMPA (ULCO) IMJ-PRG (Sorbonne)

Motivation

Observation

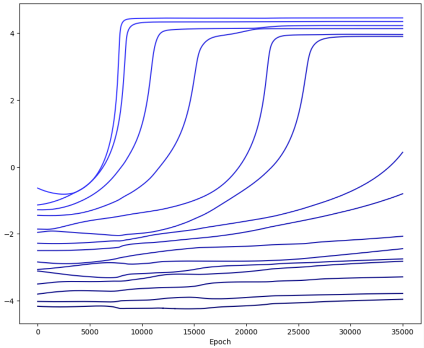
During training, the input–output Jacobian J of factored models like MLPs is observed to develops a *few large singular values* while the rest stay small. This phenomenon called **low rank bias** is usually linked to questions of generalization.

Key question

What mechanism creates and maintains this spectral separation?

Our thesis

Factored models are inherently drawn toward spectral separation, independently of the training data.



Top-15 log-singular values during training
(fixed-gates net, depth 10, width 64)

Setup: Spectral separation and Vector alignment

Provided a Jacobian $J = \underbrace{M_N \cdots M_\ell}_{A_\ell} M_{\ell-1} \cdots M_1$ in some factored model, we let s_k be the k-th largest singular value of J.

We will focus on the following phenomena:

- **Spectral Separation**:

  Every s_k dominates s_{k+1}, i.e. $\frac{s_{k+1}}{s_k} \ll 1$.

- **Vector Alignment**:

  If $\ell \ll L$, *the left singular vectors of A_ℓ approximate those of J.*

  In other words, letting

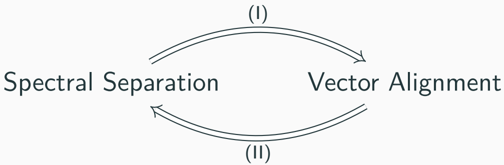
  $$J = U_J S V_J^\top, \qquad A_\ell = U_{A_\ell} S_{A_\ell} V_{A_\ell}^\top$$

  be the respective SVDs, then

  $$U_{A_\ell}^\top U_J \approx \text{diagonal}.$$

Setup: Spectral Separation and Vector alignment

We hope to describe singular value dynamics as emerging from a feedback loop:



In our framework:

- We show **Spectral Separation holds at random initialization.**
- We **prove implication (I).**
- We **prove implication (II) under stronger assumptions**.

Spectral Separation implies Vector Alignment

Theorem

Fix ℓ. Let $J_N = \underbrace{M_N \cdots M_\ell}_{A_\ell} M_{\ell-1} \cdots M_1$ and assume A_ℓ satisfies **spectral separation**, i.e. $s_{i+1}^{A_\ell}/s_i^{A_\ell} \xrightarrow{N\to\infty} 0$.

Then **the left singular vectors of A_ℓ and J_N align as $L \to \infty$.**

Equivalently, up to signs,

$$U_{A_\ell}^\top U_{J_N} \xrightarrow{N\to\infty} I_r.$$

The rate of convergence is controlled by the ratios $s_j^{A_\ell}/s_i^{A_\ell}$, $i \neq j$.

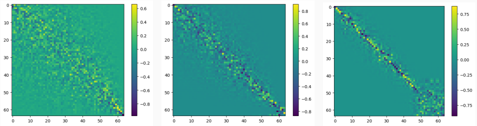


Figure 1: The matrix $U_{J_N}^\top U_{A_\ell}$ at initialization with $p = 1$. From left to right: $N = 2$, $\ell = 1$, diagonal correlation of 0.26; $N = 10$, $\ell = 5$, diagonal correlation of 0.49 and finally $N = 20$, $\ell = 10$, diagonal correlation of 0.84. Note that only the upper left 64×64 sub-matrix is displayed from the original 128×128.

Singular Value Dynamics 1/2

Provided a Jacobian $J = \underbrace{M_N \cdots M_\ell}_{A_\ell} \underbrace{M_{\ell-1} \cdots M_1}_{B_\ell}$ in some factored model, we assume:

1. **Depth scaling:** for some $\gamma_k(t)$ and $\delta_k(t)$,

$$\forall \ell,\ s_{k,A_\ell}(t) = e^{(N-\ell+1)\gamma_k(t)+\delta_k(t)}, \qquad s_{k,B_\ell}(t) = e^{(\ell-1)\gamma_k(t)+\delta_k(t)}.$$

2. **Strong vector alignment:** for some $\epsilon \ll 1$,

$$\forall \ell,\ \|U_{A_\ell}^\top U_J - I\|_\infty < \varepsilon, \qquad \|V_J^\top V_{B_\ell} - I\|_\infty < \varepsilon.$$

Singular Value Dynamics 2/2

Under these assumptions, **we derive general singular value dynamics**:

Proposition (general dynamics)

$$\dot{s}_k \overset{\varepsilon \to 0}{\sim} e^{(1+\frac{1}{N})\delta_k} s_k^{1-\frac{1}{N}} K(t)$$

For particular architectures (FGLNs) we approximately recover previously known results:

Proposition (deep-linear, Arora & all. "Implicit regularization in deep matrix factorization" 2019)

$$\dot{s}_k = -N\, s_k^{2-2/N} \left\langle \nabla_J \mathcal{L},\ u_k v_k^\top \right\rangle.$$

Dynamics of this form are expected to **induce spectral separation**.

Conclusion

Interpretation

Our work suggests that the feedback loop inducing low rank bias is present in *any* factored network. It would in particular be independent of the training data.

Thank you! [1]

[1] Code: `https://github.com/Naloween/separation_deep_jacobian-spectra`