Implicit Regularization with Polynomial Growth in Deep Tensor Factorization

Kais Hariz[1,2], Hachem Kadri[1], Stéphane Ayache[1], Maher Moakher[2], Thierry Artières[1,3]

[1]Aix-Marseille University, CNRS, LIS, Marseille, France

[2]LAMSIN, National Engineering School of Tunis, University of Tunis El Manar, Tunis, Tunisia

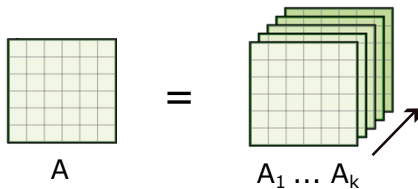[3]Ecole Centrale de Marseille, Marseille, France

Generalization in deep learning

How deep neural networks generalize well?

- # learnable parameters $\gg$ # training examples

- **Role of implicit regularization**

- Measure of complexity: matrix rank, tensor rank, ... ?

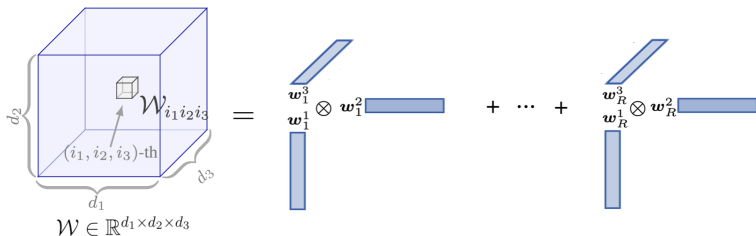# The role of implicit regularization in deep learning

- Deep Matrix Factorization

A = A₁ ... Aₖ

> ### Arora et al., 2019
>
> For $A \in \mathbb{R}^{d_1 \times d_2}$ and depth $k \geq 2$, for any $r = 1, \ldots, \min(d_1, d_2)$,
>
> $$\frac{d}{dt}\sigma_r(t) = k\alpha_r(t) \cdot (\sigma_r(t))^{2-\frac{2}{k}}$$

The role of implicit regularization in deep learning
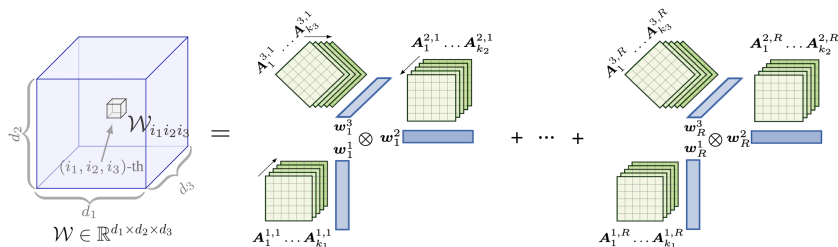
- (Shallow) Tensor CP Factorization

Razin et al., 2021

Under certain assumptions, for any $r = 1, \ldots, R$,

$$\frac{d}{dt} \left\| \bigotimes_{n=1}^{N} \boldsymbol{w}_r^n(t) \right\| = N \gamma_r(t) \cdot \left\| \bigotimes_{n=1}^{N} \boldsymbol{w}_r^n(t) \right\|^{2 - \frac{2}{N}}$$

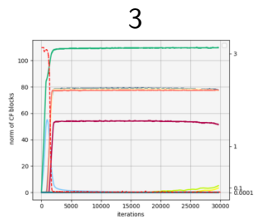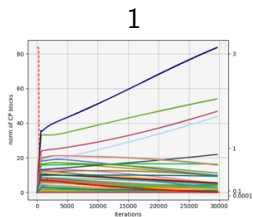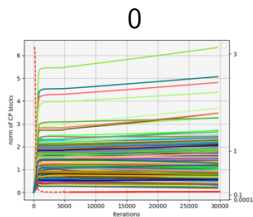# The role of implicit regularization in deep learning

- Deep Tensor CP Factorization

This work

Under certain assumptions, for any $r = 1, \ldots, R$,

$$\frac{d}{dt} \left\| \bigotimes_{n=1}^{N} \boldsymbol{w}_r^n(t) \right\| = N\delta_r(t) \left\| \bigotimes_{n=1}^{N} \boldsymbol{w}_r^n(t) \right\|^{2 - \frac{2}{N} + k}$$

Effect of the depth on tensor rank

Depth

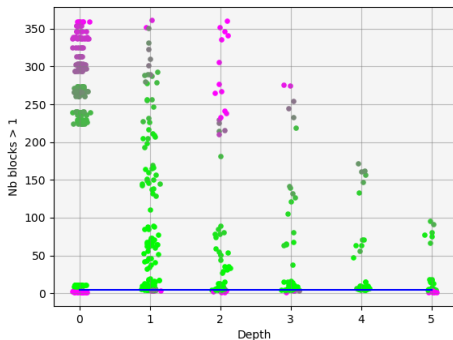0   1   3

How this impacts the implicit regularization on the learned tensor $\mathcal{W}$?

$$\left\| \bigotimes_{n=1}^{N} \prod_{i=1}^{k} \boldsymbol{A}_i^{n,r}(t) \boldsymbol{w}_r^n(t) \right\| \leq \left\| \bigotimes_{n=1}^{N} \boldsymbol{w}_r^n(t) \right\|^{1+k}$$

How depth yields low-rank solutions?

Rank

Generalization Loss