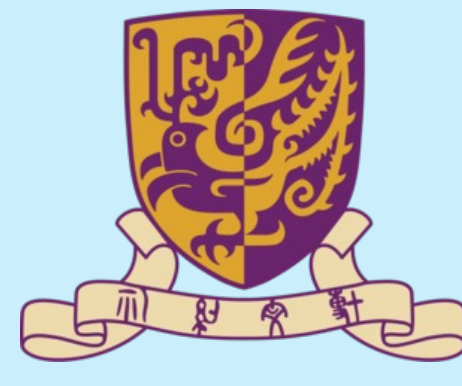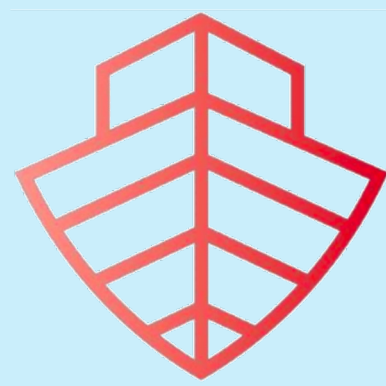# IDInit: A Universal and Stable Initialization Method for Neural Network Training

Yu Pan, Chaozheng Wang, Zekai Wu, Qifan Wang, Min Zhang, Zenglin Xu
iperryuu@gmail.com, czwang23@cse.cuhk.edu.hk, zenglin@gmail.com

## ➤ Motivation

At present, Xavier initialization is widely used, however:

- Insufficient stability:
  - cannot train deep models;
  - cannot support BN-free networks.

```
[[-0.6753,  0.7071, -0.9315],
 [ 0.5739,  0.2670, -0.1037],
 [-0.6213,  0.0926, -0.3897]]
```
A sample of Xavier

- Slow convergence:
  - cannot satisfy dynamical isometry.

- Random values:
  - hard to perform stably.

Solution: Identity Matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
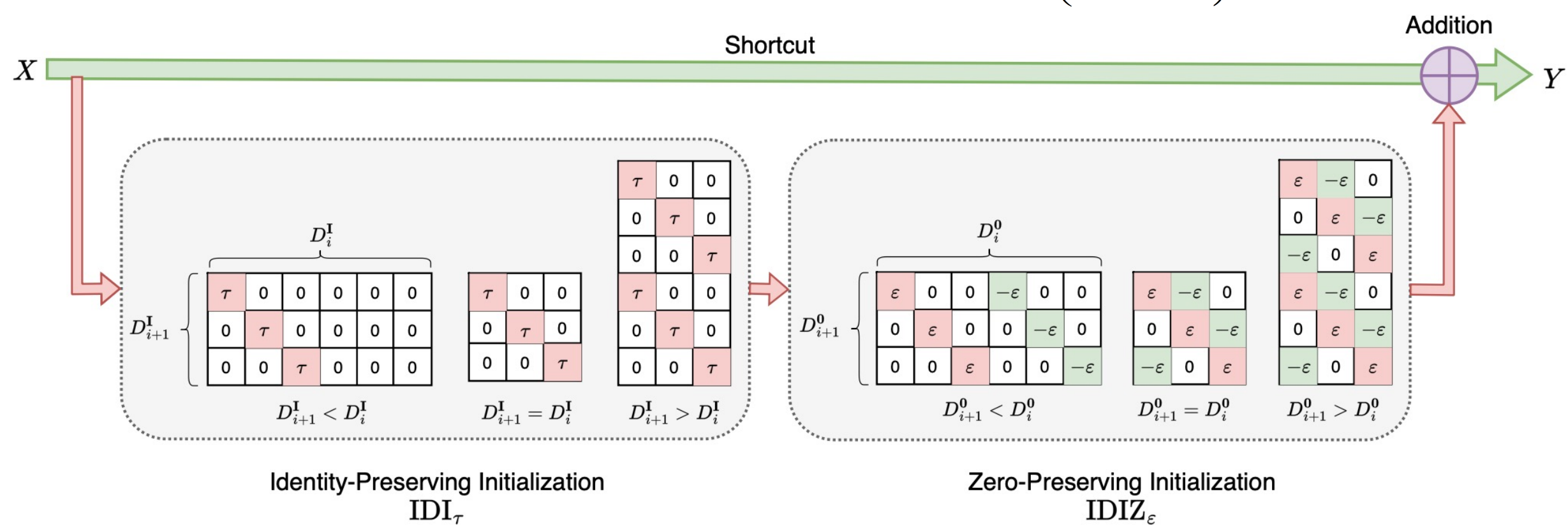

Overview of IDInit
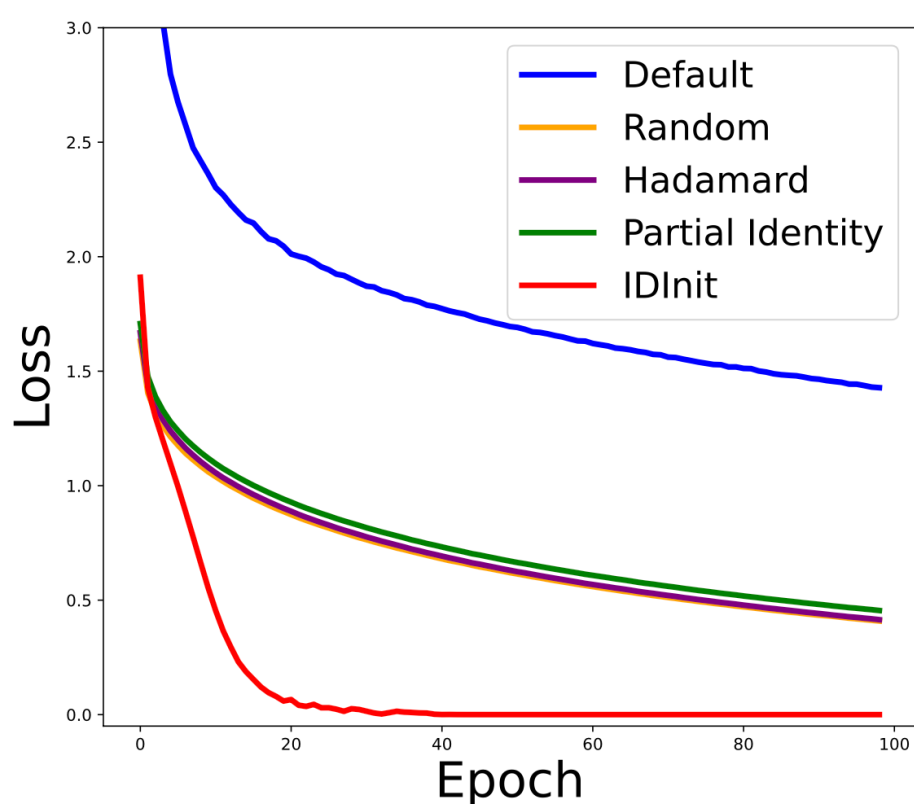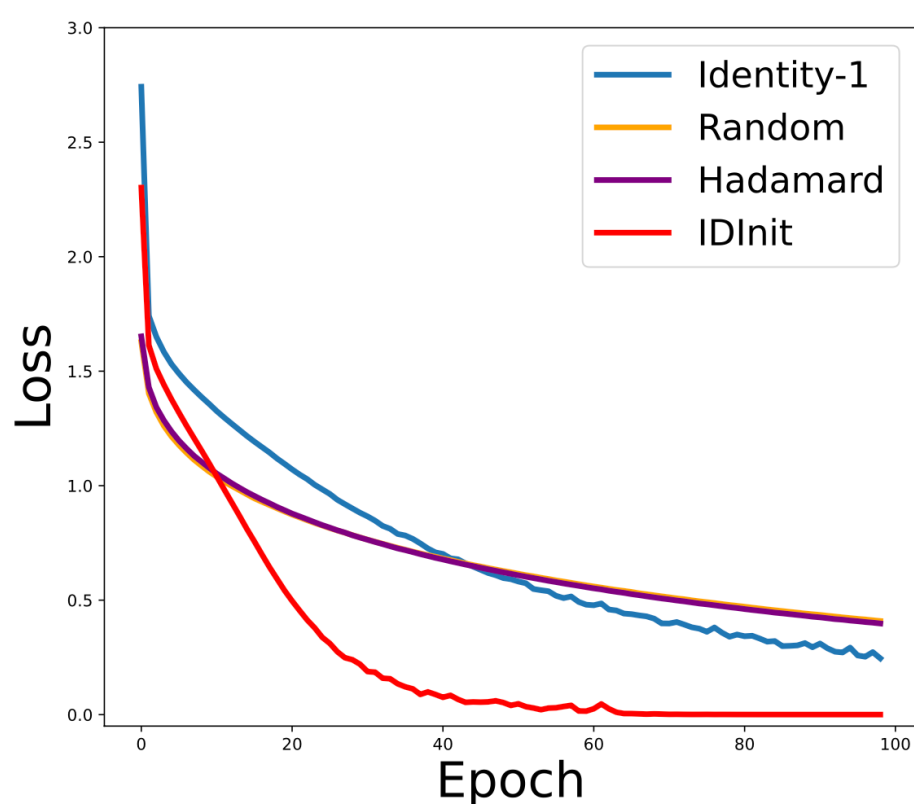

Convergence ability of IDInit

IDInit can achieve the fastest convergence among the baselines, which demonstrates its significance in accelerating training.
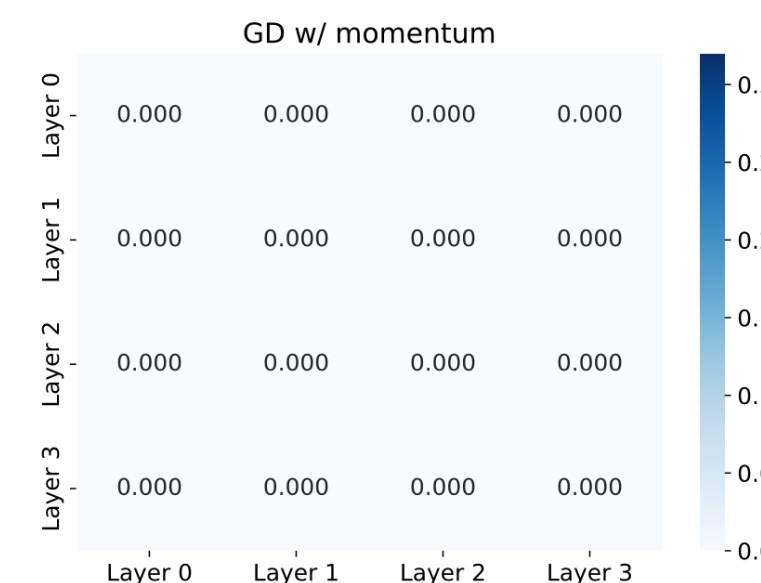
## ➤ Three Main Obstacles of Identical Initialization

❑ **Convergent Problem:** Difficulty on converging to matrix with negative eigenvalues.
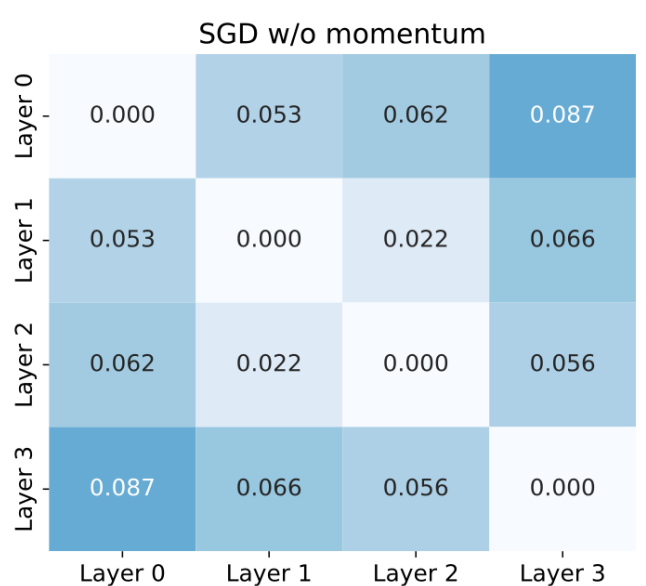
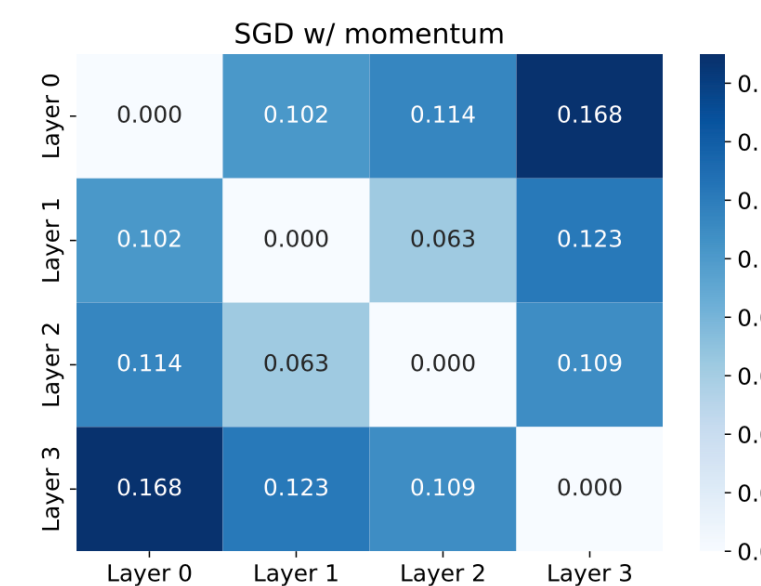Solution: The problem can be solved by SGD optimization.
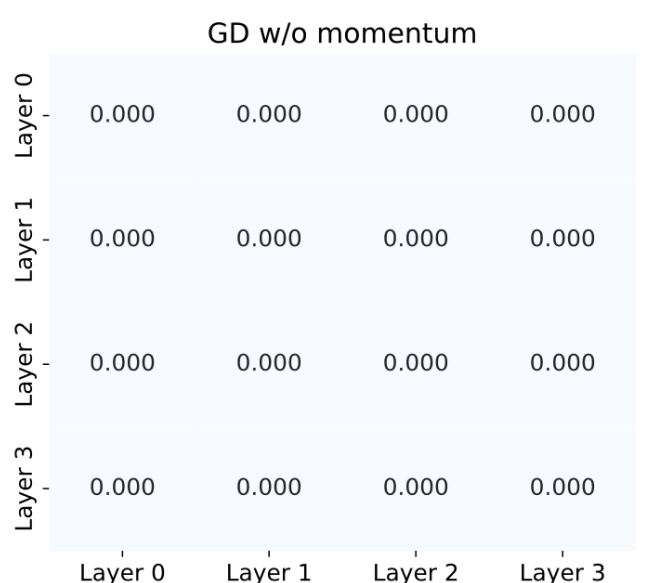


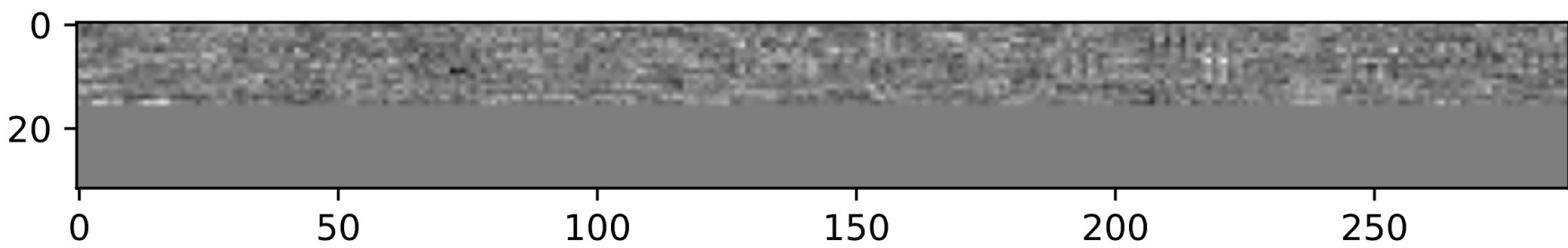(a) GD w/o momentum  (b) GD w/ momentum

(c) SGD w/o momentum  (d) SGD w/ momentum
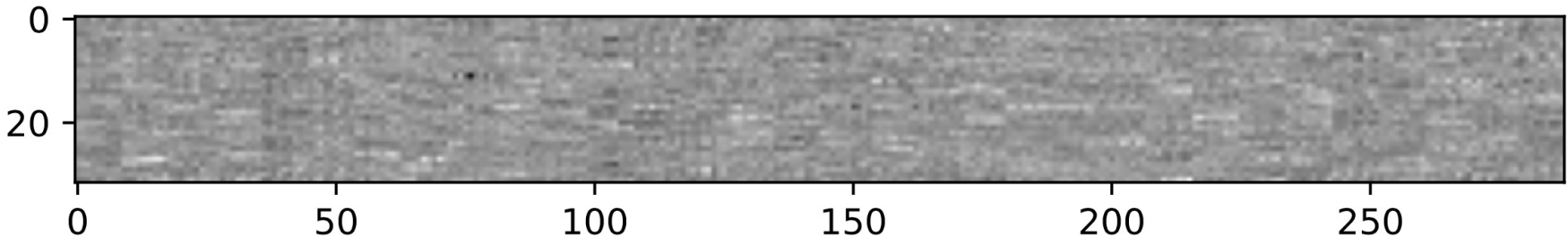
Solving Convergent Problem

❑ **Dead Neurons:** Direct using the zero matrix will cause dead neurons.

Solution: Using small values $\varepsilon$ and $-\varepsilon$: For a matrix production, results can be calculated as $v = \varepsilon(v1 - v2)$, thus

$$\begin{cases} \mu(v) = 0, \\ \sigma^2(v) = 2\varphi\varepsilon^2, \end{cases}$$
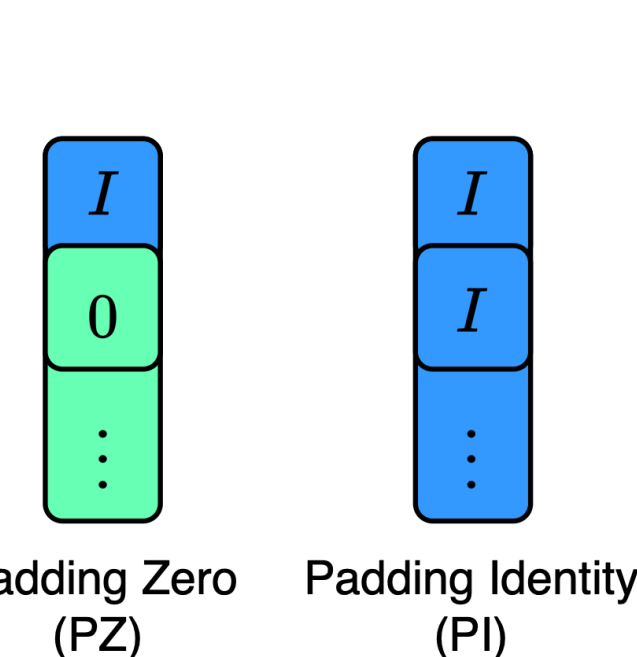

(a) Weight initialized with numerical value 0.

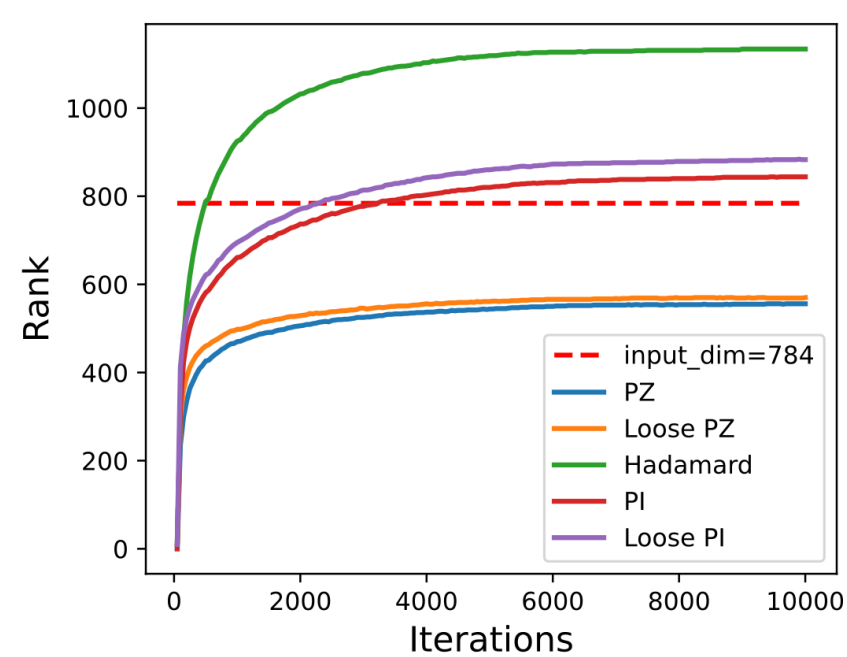
(b) Weight initialized with $\text{IDIZ}_{1e-6}$.

Solving dead neurons

## ❑ Rank Constraints : Direct dimension-increasing matrix may face a rank constraint problem if padding zero values.

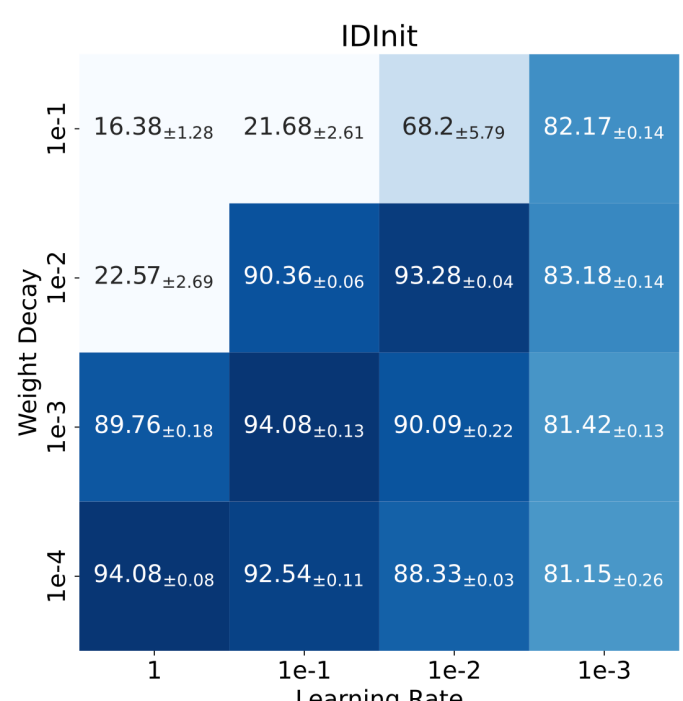Solution: Padding identity matrix can solve the problem.
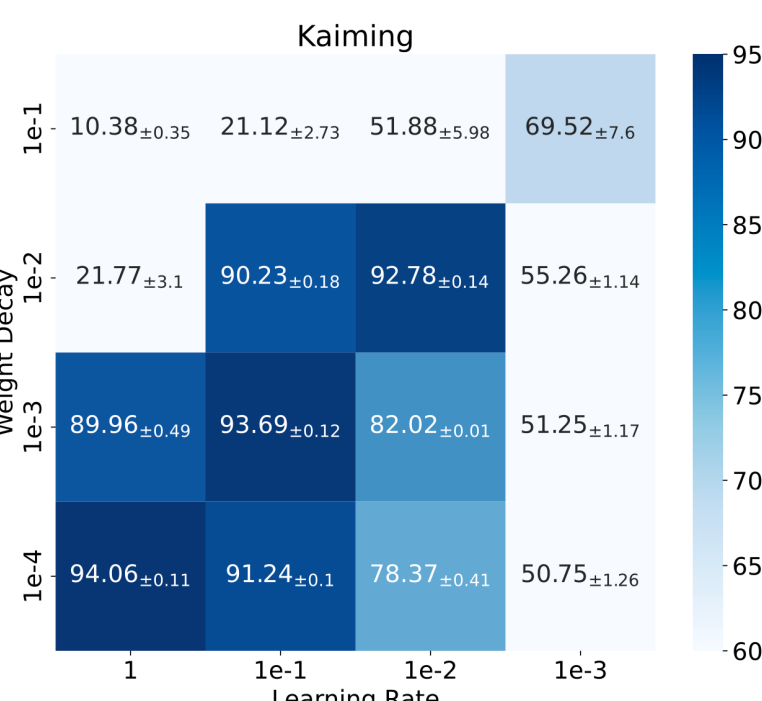


(a) Padding schemes.  (b) Rank plot.

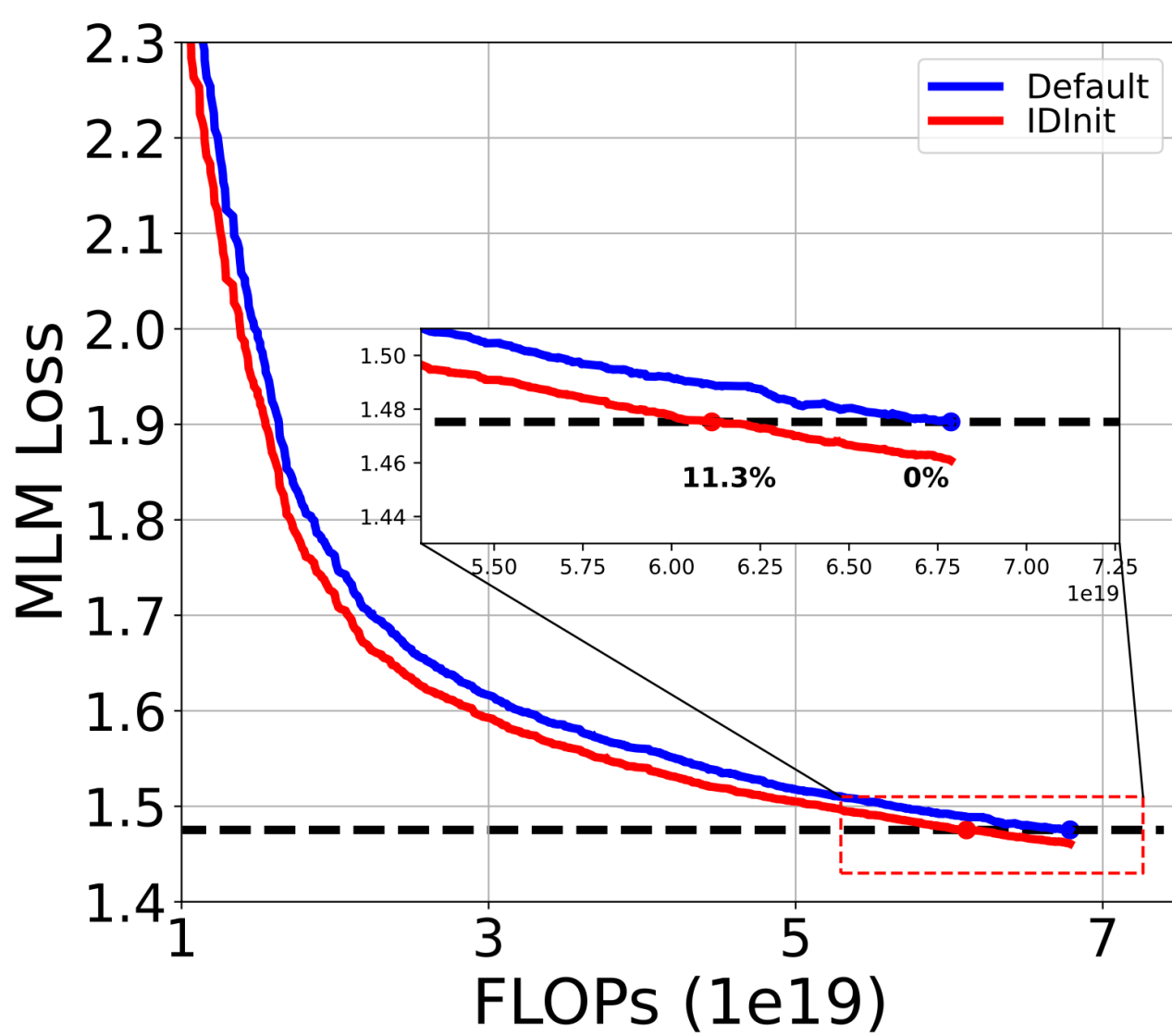Breaking rank constraints

## ➤ Empirical Results



(a) IDInit  (b) Kaiming

The hyperparameter experiment on Cifar10



Results of BERT-Base