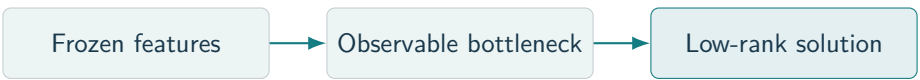
NeurIPS 2026

How Finite-Rank Bottlenecks Shape the Low-Rank Adaptation Landscape

Long Nguyen-Chi Quynh Nguyen Thanh Nguyen-Cung Binh T. Nguyen

College of Engineering & Computer Science, VinUniversity



Why can low-rank updates suffice without NTK linearization?

Frozen features hide directions of the update



$\mathrm{rank}(T_1) = 2, \ \|T_1\|_* = 5$ $\mathrm{rank}(T_2) = 1, \ \|T_2\|_* = 2$

For the frozen feature $H = (1, 0)^\top$, both updates give the same prediction.

$$T_1 H = T_2 H = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

The loss cannot distinguish these two updates, but trace-norm regularization can.

The bottleneck controls the rank

The loss only observes T through a finite-rank linear map $\Phi(T) = TH$.

Low-rank representer theorem

For a well-posed trace-norm-regularized objective, there exists a global minimizer $T^\star$ with

$$\boxed{\frac{r(r+1)}{2} \le d_\Phi, \qquad r = \operatorname{rank}(T^\star)}.$$

$d_\Phi = 1 \;\Rightarrow\; r \le 1$ $d_\Phi = 3 \;\Rightarrow\; r \le 2$ $d_\Phi = 6 \;\Rightarrow\; r \le 3$

This guarantees the existence of a low-rank optimum; it does not guarantee that gradient descent finds it.

Attention hides even more directions

Softmax ignores row shifts

$$\mathrm{softmax}([1, 2, 4]) = \mathrm{softmax}([11, 12, 14]).$$

Centering removes one unobservable direction from each row.

Value updates pass through two filters



Value updates are observed only after attention weighting and a fixed output projection.

Query, key and value ranks are governed by the subspaces the frozen computation actually exposes.

Stationary points in nonconvex LoRA

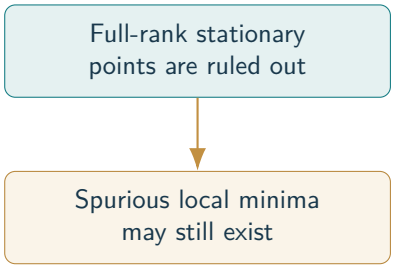
The factorized update has the form

$$T = UV^\top, \qquad R = \begin{bmatrix} U \\ V \end{bmatrix}.$$

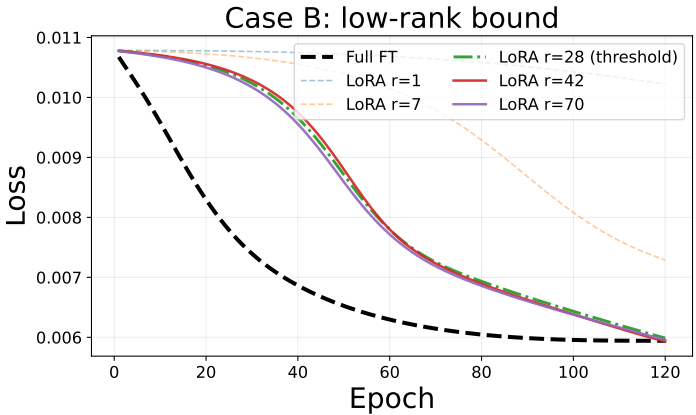
After an arbitrarily small generic PSD perturbation, the condition

$$\boxed{\frac{r(r+1)}{2} > p}$$

forces every first-order stationary point to be rank deficient, under the paper's assumptions.



MNLI experiment with a four-example training set



Estimated rank threshold

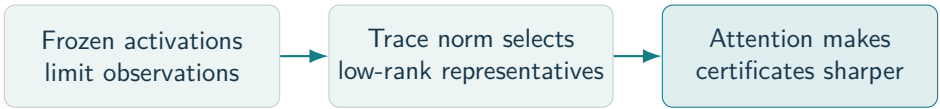
$r = 28$

Ranks 1 and 7 show worse training-loss trajectories.

Ranks 28, 42 and 70 cluster closely.

Full fine-tuning follows a different path.

A structural explanation for low rank



What the paper establishes

Low-rank global minimizers exist without NTK linearization. Under a separate condition, generic perturbations force rank-deficiency at first-order stationary points.