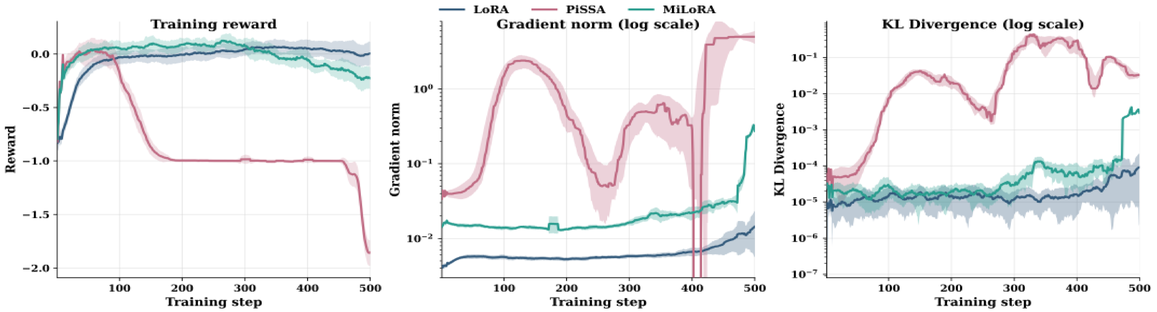
Geometry-Preserving Orthonormal Initialization for Low-Rank Adaptation in RLVR

Ruijia Zhang[1], Jiacheng Zhu[2], Hanqing Zhu[3], Laixi Shi[1]
Johns Hopkins University[1] | Meta Super Intelligence Labs[2]| Universitity of Texas, Austin[3]

ICML International Conference On Machine Learning

PiSSA/MiLoRA collapses under RLVR fine-tuning



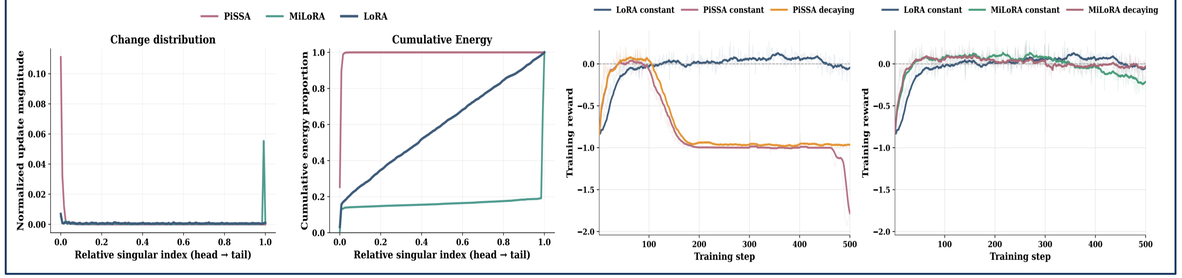
1. LoRA SVD variants (PiSSA and MiLoRA) designed for SFT do not transfer for RLVR.
2. Training reward collapses with the explosion of the gradient norm and KL divergence.

Stable RLVR training demands conservative updates

SFT objectives: $\mathcal{L}_{\mathrm{SFT}}(\theta) := -\mathbb{E}_{(q,a^*)\sim\mathcal{D}}\left[\log \pi_\theta(a^*|q)\right].$

RLVR objectives: $\mathcal{L}_{\mathrm{RLVR}}(\theta) = \mathbb{E}_{x\sim\mathcal{D},\, y\sim\pi_\theta(\cdot|x)}\left[R(x,y)\right] - \beta\,\mathrm{KL}(\pi_\theta\|\pi_{\mathrm{ref}})$

1. Initialization of LoRA largely governs the optimization geometry and the outcome.
2. SVD variants accelerates the optimization and destabilize training. Learning rate decay can mitigate.



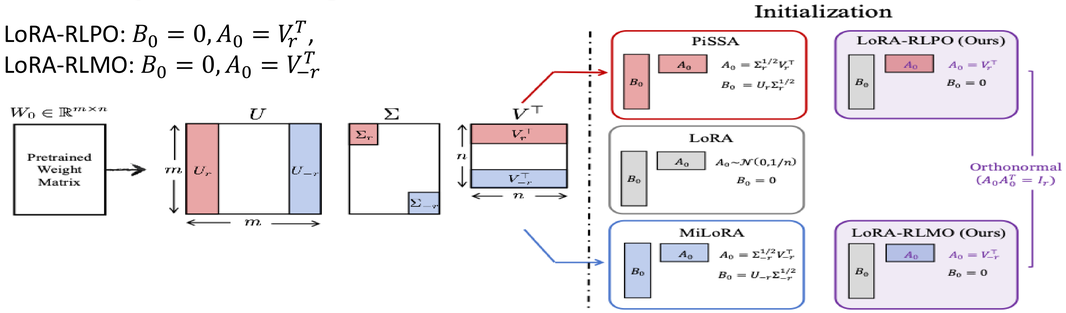
Orthonormal initialization in LoRA leads closer to full fine-tuning.

Theorem: $\frac{1}{T}\|W_T^{\mathrm{LoRA}} - W_T^{\mathrm{full}}\|_F \le M\eta\,\Gamma_T(\eta)\,\|I_n - A_0^\top A_0\|_2 + \mathcal{O}(\eta^2).$

The equality holds if A_0 is orthornormal

Geometry-Preserving Orthonormal Initializations for RLVR LoRA

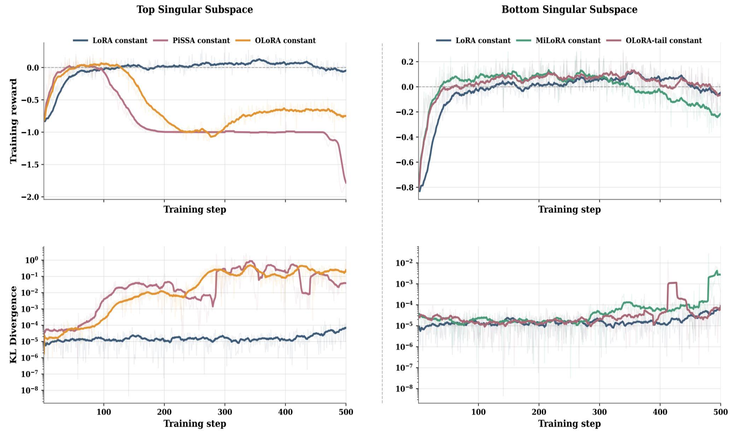
LoRA-RLPO: $B_0 = 0, A_0 = V_r^T$,
LoRA-RLMO: $B_0 = 0, A_0 = V_{-r}^T$



Results

	LoRA	MiLoRA	LoRA-RLMO (Ours)	PiSSA	LoRA-RLPO (Ours)
B_0	$\mathbf{0}$	$U_{-r}\Sigma_{-r}^{1/2}$	$\mathbf{0}$	$U_r\Sigma_r^{1/2}$	$\mathbf{0}$
A_0	$\mathcal{N}(0, \frac{1}{n})$	$\Sigma_{-r}^{1/2}V_{-r}^\top$	$V_{-r}^\top$	$\Sigma_r^{1/2}V_r^\top$	$V_r^\top$
GSM8K$^{@1}$	$75.64_{\pm 1.25}$	$75.41_{\pm 1.35}$	$\mathbf{76.42}_{\pm 0.76}$	$9.65_{\pm 8.20}$	$75.59_{\pm 1.14}$
MATH500$^{@4}$	$81.93_{\pm 7.56}$	$76.47_{\pm 6.79}$	$86.80_{\pm 2.31}$	$13.20_{\pm 9.85}$	$\mathbf{87.33}_{\pm 1.81}$
AIME22$^{@32}$	$43.33_{\pm 6.67}$	$28.89_{\pm 1.92}$	$42.22_{\pm 1.92}$	$0.00_{\pm 0.00}$	$\mathbf{46.67}_{\pm 3.33}$
AIME23$^{@32}$	$38.89_{\pm 5.09}$	$30.00_{\pm 3.33}$	$41.11_{\pm 5.09}$	$0.00_{\pm 0.00}$	$\mathbf{42.22}_{\pm 6.94}$
AIME24$^{@32}$	$72.22_{\pm 3.85}$	$47.78_{\pm 6.94}$	$72.22_{\pm 1.92}$	$0.00_{\pm 0.00}$	$\mathbf{73.33}_{\pm 0.00}$
Avg	$62.40_{\pm 2.96}$	$51.71_{\pm 0.98}$	$63.76_{\pm 1.64}$	$4.57_{\pm 3.26}$	$\mathbf{65.03}_{\pm 0.55}$

Ablations



OLoRA: $B_0 = U_r, A_0 = V_r^T$,
OLoRA-tail: $B_0 = U_{-r}, A_0 = V_{-r}^T$

Singular value scaling exacerbates instability beyond subspace selection.

Resources

Paper | Code | Contact