Titre: Learning Rate Transfer and Feature Learning Across Depth for Constrained Spectral Optimizers: Complete Scion.

Orateur : Eline Pot.

Encadrants de thèse : Laurent Lefebvre (Framatome), Mathieu Segond (Framatome), Antonio Silveti-Falls (CVN, CentraleSupélec), Jean-Christophe Pesquet (CVN, CentraleSupélec).

Date de commencement de thèse : 01/02/2025.

Abstract: Hyperparameter transfer across model depth is instrumental in training deep models, as it enables learning rates calibrated on shallow networks to be effectively applied to deeper architectures. The recent CompleteP framework establishes a $1/L$ residual scaling rule that demonstrates this for Adam-family optimizers. We study whether similar transfer extends to constrained optimization methods, specifically Frank-Wolfe and projected non-Euclidean normalized steepest descent, where updates are determined by linear minimization oracles. We derive scaling rules for these methods which confirm that the $1/L$ residual scaling applies, although the arguments differ from the Adam-family optimizers. Our analysis supports depth-wise transfer and width-wise transfer of the learning rate for the Scion algorithm while ensuring feature learning and avoiding the kernel-regime. In experiments on NLP transformers (nanoGPT on FineWeb), vision transformer (DeiT on ImageNet), and vision MLPs (MLP-Mixer on CIFAR-10), we observe learning rate transfer across depth and feature learning for appropriately chosen residual scaling.