Ablation Study on the Hyperparameter n

For NTKMTL-SR, the computational cost of calculating the NTK is minimal, allowing us to further investigate the impact of varying mini-batch sizes n on the results. Therefore, we set n to $[1, 2, 3, 4, 6]$ and conduct an ablation study on the QM9 (11-task) benchmark. For each value of n, we conduct 3 repeated experiments with different random seeds and calculate the mean and variance for $\Delta m\%$. The results are shown in Fig. 2. When $n = 1$, the performance of NTKMTL-SR is comparable to that of NTKMTL. However, when increasing n from 1 to 2 or more, NTKMTL-SR shows a noticeable improvement in performance, accompanied by a reduction in the variance of performance across repeated experiments. We attribute this to the fact that increasing the number of mini-batches leads to a larger NTK matrix dimension, which in turn reduces stochastic error and allows our method to more accurately characterize the convergence speed of the tasks.

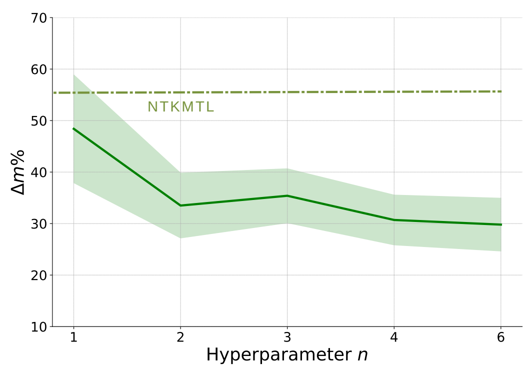

Figure 2: Ablation study on hyperparameter n on QM9. Each experiment is repeated over 3 random seeds, and the mean and stderr are reported.

However, we also find that the results for $n = 4$ and $n = 6$ are almost identical, and the performance differences observed were potentially weaker than the inherent variability stemming from different random seeds. Concurrently, Fig. 3 visualizes the training time per epoch for various n values on QM9. Despite only requiring the computation of the maximum eigenvalue of the NTK matrix with respect to z, training a single epoch when $n = 6$ already approached 1.7 times the duration of the LS method. Overall, we posit that the selection of hyperparameter n is a trade-off between performance and training speed, and $n = 4$ generally presents a favorable compromise.

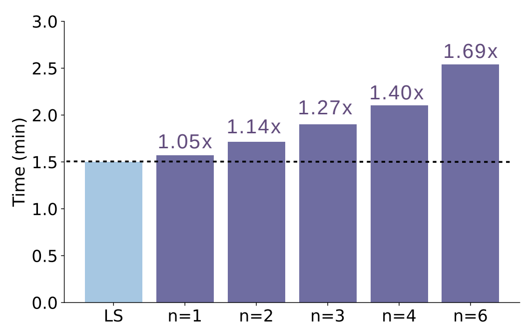

Figure 3: Training time per epoch for LS and NTKMTL-SR (with different hyperparameter n) on QM9.