

Figure 7: Autoencoder - Pytorch (NEW). We perform tridiag-SONew comparison with KFAC [1], Eva [2], FishLeg [3] on the Autoencoder benchmark. We keep the batch size=1000 for all experiments. Eva is trained for 100 epochs, and for others we change number of epochs to train such that each experiment takes same amount of time. Each optimizer is tuned using 600 hyperparameters using NNI [4]. KFAC and FishLeg optimizers' preconditioner is updated once every 10 steps. We observe that tridiag-SONew beats all the baselines by large margin.

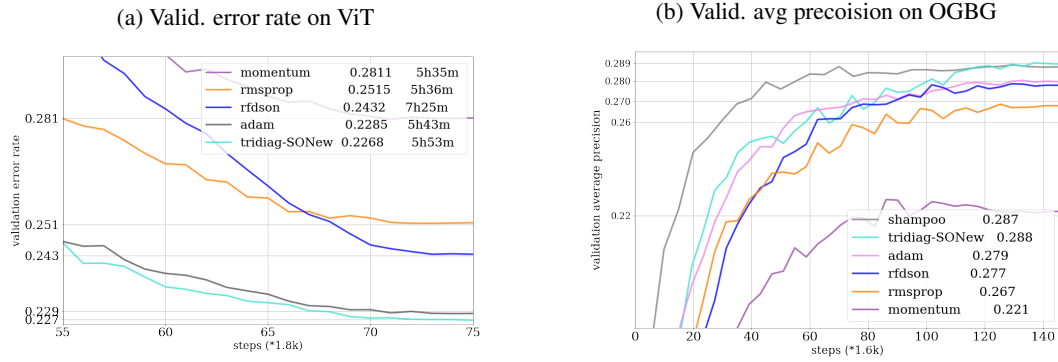


Figure 8: Large Scale Benchmarks with Momentum and RmsProp baselines added. (a) ViT benchmark - Since all baselines take similar time, we plot performance wrt steps. (b) GraphNetwork benchmark - in the Figure 1 we plotted performance wrt steps. Since Shampoo takes more time per step, we ran Shampoo on 80% fewer steps, so that end-to-end time taken is same by all the optimizers. We observe that tridiag-SONew outperforms even Shampoo in wall clock time, while simultaneously taking much less memory.

- 770 [1] Martens et. al, 2015 Optimizing Neural Networks with Kronecker-factored Approximate Curvature  
771 [2] Eva: Practical Second-order Optimization with Kronecker-vectorized Approximation, ICLR 2023  
772 [3] Gracia et al., 2023 Fisher-Legendre (FishLeg) optimization of deep neural networks.  
773 [4] <https://nni.readthedocs.io/en/v2.3/Tuner/BuiltInTuner.html#Anneal>