DP-MicroAdam: Private and Frugal Algorithm for Training and Fine-tuning

Machine learning models are often trained on sensitive data. To mitigate privacy risks, **Differential Privacy (DP)** has become the standard framework, providing mathematical guarantees that the output of an algorithm does not depend significantly on any single data point. The most popular differentially private algorithm is DP-SGD [1], which applies per-sample gradient clipping, adds calibrated Gaussian noise, and employs early stopping. Unfortunately, ensuring strong privacy guarantees significantly degrades the utility of the trained model, especially for **private fine-tuning**, where limited data amplifies the negative effects of noise and makes optimization more difficult. DP training is computationally demanding, sensitive to hyperparameter tuning, memory intensive, and does not scale well with gradient dimension, posing significant challenges when applied to large models.

A promising approach to improve the privacy-utility tradeoff is exploiting sparsity. Recently, DP-BiTFiT [3] and SPARTA [2] showed that updating only a small subset of parameters during fine-tuning can improve model accuracy. Another natural idea is to use Adam, which outperforms SGD in non-private training. However, adding DP to Adam is challenging, as naive noise injection and clipping create an additional bias that is hard to remove from Adam updates even with de-biasing techniques [5], limiting the accuracy of DP-Adam in practice.

MicroAdam [6] is a recent non-private optimizer that reduces memory usage while maintaining convergence guarantees. It incorporates four key mechanisms: top-k selection retaining only the largest 1% of gradient components; error feedback to recover lost information in the remaining gradients; quantization of the residuals; and a sliding window of sparse gradients to reconstruct Adam's moment estimates.

In this work, we propose **DP-MicroAdam**, motivated by the fact that MicroAdam fixes by design several issues of DP-Adam regarding bias and computational costs. We analyze the convergence of DP-MicroAdam for stochastic non-convex optimization. Our results show that it matches the optimal non-private convergence rate of $\mathcal{O}(1/\sqrt{T})$, up to a constant factor depending only on the gradient clipping threshold and noise variance, thus achieving **the same theoretical guarantees as DP-SGD**.

Beyond these theoretical guarantees, we empirically evaluate the performance of DP-MicroAdam. Interestingly enough, accuracy is largely unaffected by the choice of clipping threshold in comparison to DP-SGD, suggesting better robustness and reducing the need for extensive tuning. Moreover, Adam's adaptive update rule allows the learning rate to remain constant throughout training, further **simplifying hyperparameter tuning**. The top-k operator, which updates only the most significant entries, further improves robustness under gradient clipping and noise addition. In particular, the bias in DP-Adam primarily affects small-magnitude components [5], thus focusing on the largest components appears to mitigate this issue, **making de-biasing techniques unnecessary**.

We evaluate DP-MicroAdam on CIFAR-10 following the Wide-ResNet setup from [4], which integrates group normalization, large batch sizes, weight standardization, data augmentation, and parameter averaging. In this setting, DP-MicroAdam outperforms the previously reported state-of-the-art for private training on CIFAR-10, achieving **84.8%** test accuracy under $(8, 10^{-5})$-DP compared to the prior best of 81.4%. These preliminary results demonstrate that DP-MicroAdam combines strong theoretical guarantees with competitive empirical performance.

We are currently extending our evaluation of DP-MicroAdam to larger datasets such as ImageNet to test its scalability, where we expect even stronger gains. Ongoing work benchmarks the method on private fine-tuning tasks, where sparsity and adaptive updates are expected to provide the greatest benefits.

References

[1] Martin Abadi et al. "Deep Learning with Differential Privacy". In: *SIGSAC (CCS)*. 2016.

[2] Yuxin Bu et al. *SPARTA: Fine-Tuning Large Language Models with Differential Privacy*. arXiv:2503.12822. 2025.

[3] Zhiqi Bu et al. *Differentially Private Bias-Term Fine-tuning of Foundation Models*. arXiv:2210.00036. 2024.

[4] Soham De et al. *Unlocking High-Accuracy Differentially Private Image Classification through Scale*. arXiv:2204.13650. 2022.

[5] A. Ganesh, B. McMahan, and A. Thakurta. *On Design Principles for Private Adaptive Optimizers*. arXiv:2507.01129. 2025.

[6] Ionut-Vlad Modoranu et al. *MicroAdam: Accurate Adaptive Optimization with Low Space Overhead and Provable Convergence*. NeurIPS. 2024.