Jeremiah Birrell
Texas State University
Reza Ebrahimi
University of South Florida
Rouzbeh Behnia
University of South Florida
Jason Pacheco
University of Arizona

Differentially Private Stochastic Gradient Descent with Fixed-Size Minibatches

DP-SGD with Fixed-size Minibatches: RDP has been a widely-used accountant for DP-SGD with Poisson subsamling. Fixed-size subsampling is preferred due to constant memory usage. Wang et al.[4] provide the best computable bounds in the fixed-size regime for RDP that are practical for application to DP-SGD. We show that there is room for obtaining tighter bounds specific to DP-SGD with Gaussian noise.

RDP with Poisson Subsampling: Given a loss function $\mathcal{L}$, a training dataset D with $|D|$ elements, and a fixed minibatch size, we consider the DP-SGD NN parameter updates with fixed-size minibatches,

$$\Theta_{t+1}^D = \Theta_t^D - \eta_t G_t \,,$$

$$G_t = \frac{1}{|B|}\left(\sum_{i\in B_t^D} \text{Clip}(\nabla_\theta \mathcal{L}(\Theta_t, D_i)) + Z_t\right)$$

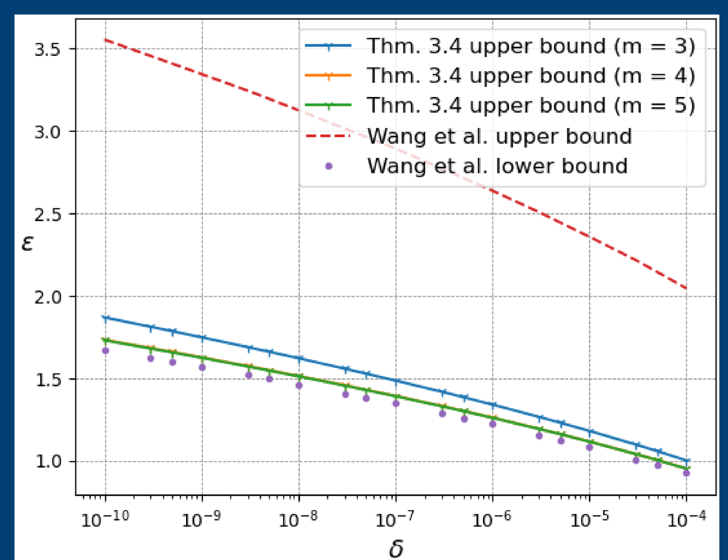
where the noises Z_t are Gaussians with mean 0 and covariance $C^2\sigma_t^2 I$, and the B_t^D are random minibatches with expected size $|B|$.
There are multiple ways to form minibatches, B_t.

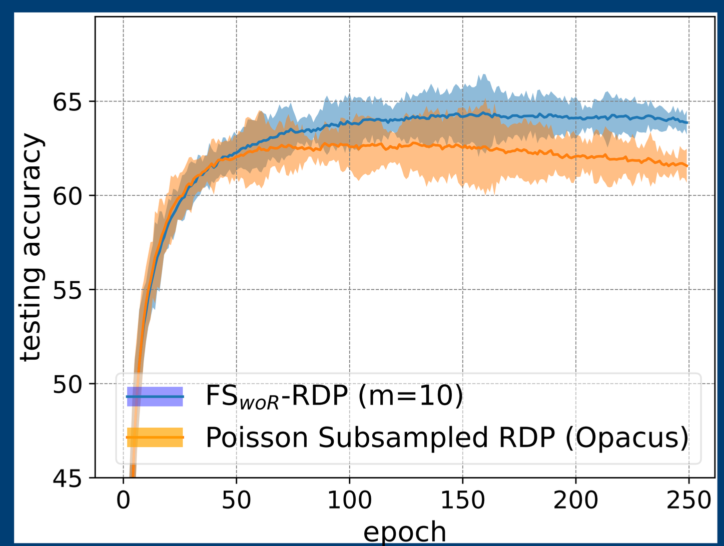
Poisson subsampling: Minibatches are formed by iid Bernoulli random variables (chosen sampling probability q) which decide whether each sample is included in the minibatch or not.

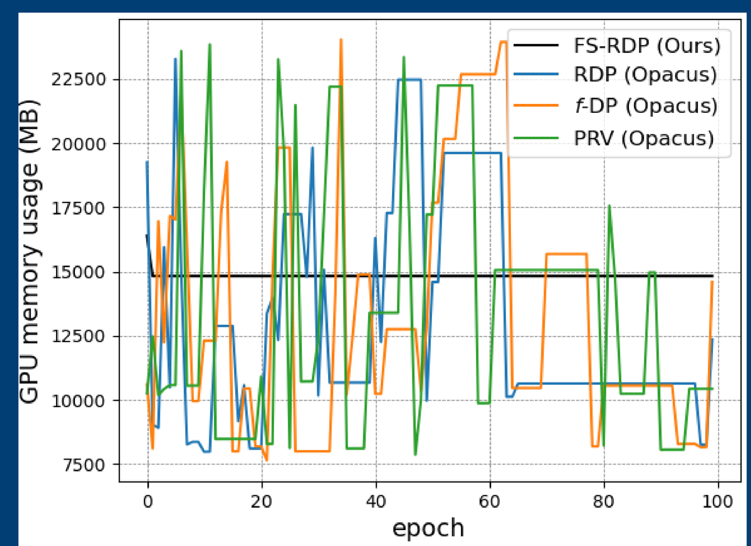
Motivation: DP-SGD with Fixed size subsampling is appealing for its constant memory usage, unlike the variable sized minibatches in Poisson subsampling.

Contribution: We present a new and holistic Rényi differential privacy (RDP) accountant for DP-SGD with fixed-size subsampling without replacement (FSwoR) and with replacement (FSwR).

Results:
(1) FSwoR accounts for both add/remove and replace-one adjacency, and improves on the best current computable bound by a factor of 4.
(2) FSwR includes explicit non-asymptotic upper and lower bounds.
(3) DP-SGD gradients with fixed-size subsampling exhibit lower variance in addition to memory usage benefits.







Disadvantage of Poisson subsampling: Leads to variable sized minibatches and therefore inconsistent memory usage. It also has higher variance.

Fixed-size subsampling: Constant memory usage, but RDP bounds more difficult to obtain. **RDP Bounds for SGD with Poisson Subsampling:** First bounds obtained by Abadi et al.[2] and Mironov et al.[3]

RDP for Fixed-size Subsampling without Replacement: The first general purpose RDP bounds (i.e., for general $\mathcal{M}$) with fixed-size subsampling obtained by Wang et al.[4]

We obtain tighter RDP bounds for fixed-size subsampled DP-SGD using a Taylor expansion method, with precise bounds on the expansion remainder terms[1].

Our RDP SGD under Fixed-size Subsampling: T-step FS$_{\text{woR}}$-RDP Upper Bound under Replace-one Adjacency[1]

$$\epsilon_{[0,T]}(\alpha) \leq$$

$$\sum_{t=0}^{T-1} \frac{1}{\alpha-1} \log\left[1 + q^2\alpha(\alpha-1)\left(e^{4/\sigma_t^2} - e^{2/\sigma_t^2}\right) + O(q^3)\right]$$

- We provide computatable bounds on the $O(q^3)$ term.
- Our result improves on the RDP bound of Wang et al.[2] by approximately a factor of 4 and is close to the theoretical lower bound[2] in practice.

Conclusion: As we showed theoretically and empirically, since FSwoR under replace-one adjacency leads to the same leading-order privacy guarantees as the widely-used Poisson subsampling, we suggest using the former over the latter to benefit from the memory management and reduced variance.

References:
[1] J. Birrell, R. Ebrahimi, arxiv.org/pdf/2408.10456 , 2024
[2] McMahan, H. B., Mironov, I., Talwar, K., Zhang, L., ACM CCS, 2016
[3] Mironov, I., Talwar, K., Zhang, L., arXiv:1908.10530, 2019
[4] Wang, Y.-X., Balle, B., and Kasiviswanathan, S. P., PMLR, 2019