This paper talks about Variance reduction using a momentum-based method (STORM). Variance reduction technique typically requires - carefully tuned learning rates and use of “mega-batches” to achieve improved results. STORM (The algorithm of this paper) - Does not require any batches and uses adaptive learning rate which enables simpler implementation and lesser hyper parameter tuning. Their technique for removing the batches uses a variant of momentum to achieve variance reduction in non-convex optimization.

SVRG algorithms have improved the convergence rate to critical points of non-convex SGD from to . Despite this improvement, SVRG has not seen as much success in practice in non-convex machine learning problems. Two potential issues that still prevail are the use of non-adaptive learning rates and reliance on giant batch sizes. This is where STORM steps in. STORM that stands for STOchastic Recursive Momentum, achieves variance reduction using a variant of the momentum term. Hence, our algorithm does not require a gigantic batch to compute. Storm achieves the optimal convergence rate of , and it uses an adaptive learning rate schedule that will automatically adjust to the variance values of ∇f(xt, ξt).

The algorithm of storm is as follows:





STORM finds critical points in stochastic, smooth, non-convex problems. It also removes the need for batch-size and incorporates adaptive learning rates. Storm is substantially easier to tune. On CIFAR-10 with a ResNet32 architecture, it seems to be optimizing the objective in fewer iterations than baseline algorithms which has been shown in this paper experimentally.

**References:** *Cutkosky, Ashok, and Francesco Orabona. Momentum-based variance reduction in non-convex sgd. NeurIPS 2020.*