New Logarithmic Step Size For Stochastic Non-Convex Optimization

Mahsa Soheil SHAMAEE, Sajad Fathi HAFSHEJANI, Zeinab SAEIDIAN

Frontiers of Computer Science, DOI: 10.1007/s11704-023-3245-z

Problems & Ideas

- Cosine step size problems:

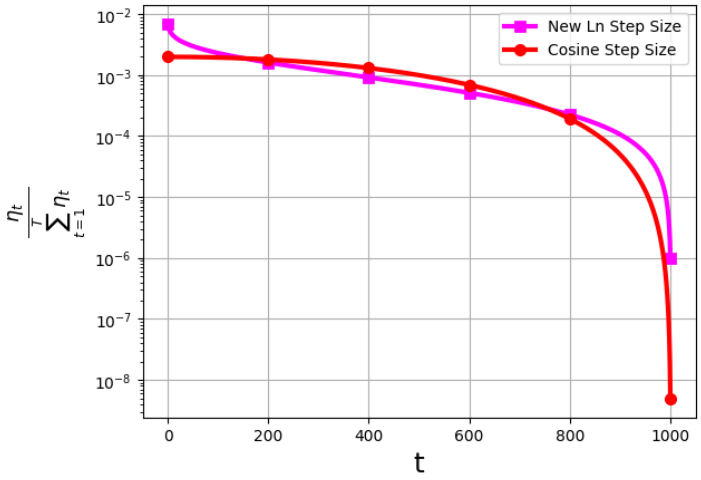
  Assigning excessively low probability distribution values to the final iterations.

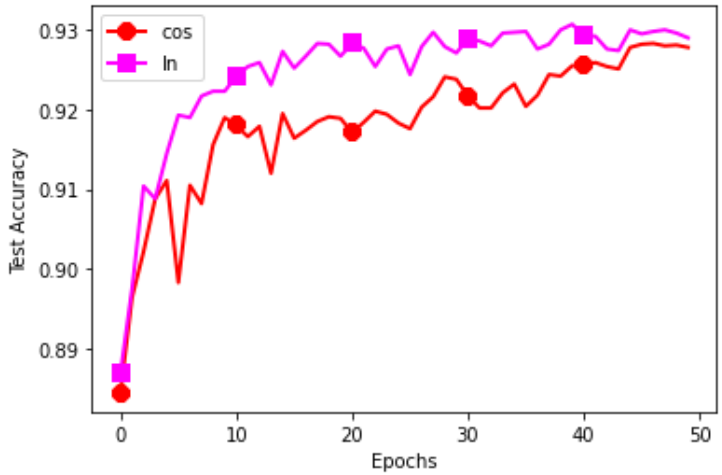
  Subpar performance of this step size during early iterations.

- Ideas:

  Introduction of a new step size designed to perform effectively in both initial and final iterations.

- Introduction of a new step size with improved behavior specifically tailored for the final iterations.

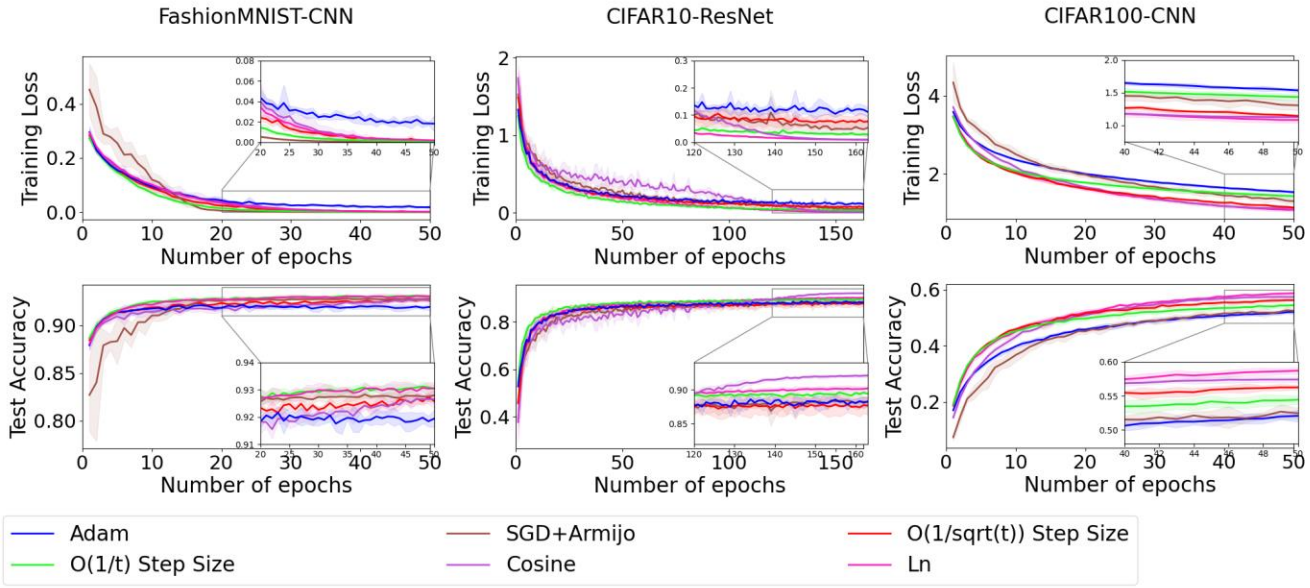




Left: Representation of the value $\frac{\eta_t}{\sum_{t=1}^{T}\eta_t}$ for both the new step size and cosine step size. Right: The comparison of the SGD with the new proposed step size and cosine step size on the FashionMnist dataset.

Main Contributions

- Contributions:
  - Introduction of a superior logarithmic step size for SGD, offering balanced probability distribution, resulting in enhanced performance during final iterations;
  - Establishment of a convergence rate of $O\left(\frac{1}{\sqrt{T}}\right)$ for smooth non-convex functions, aligning with the best-known rates;
  - Comprehensive experiments showing a substantial 0.9% accuracy improvement on the CIFAR100 dataset.



Comparison of the new proposed step size with five other step sizes on the FashionMNIST, CIFAR10, and CIFAR100 datasets.