Adaptive stochastic optimization using trajectory cues

Oliver Hinder

Motivation

- Modern machine learning involves minimizing the prediction error of a model using stochastic gradient descent (SGD) and its variants
- SGD is difficult to use because one must select a tuning parameters (the learning rate) for each problem instance

Project Description

- Eliminate the learning rate tuning parameter from SGD
- Eliminate other tuning parameters from more sophisticated variants of SGD

Context

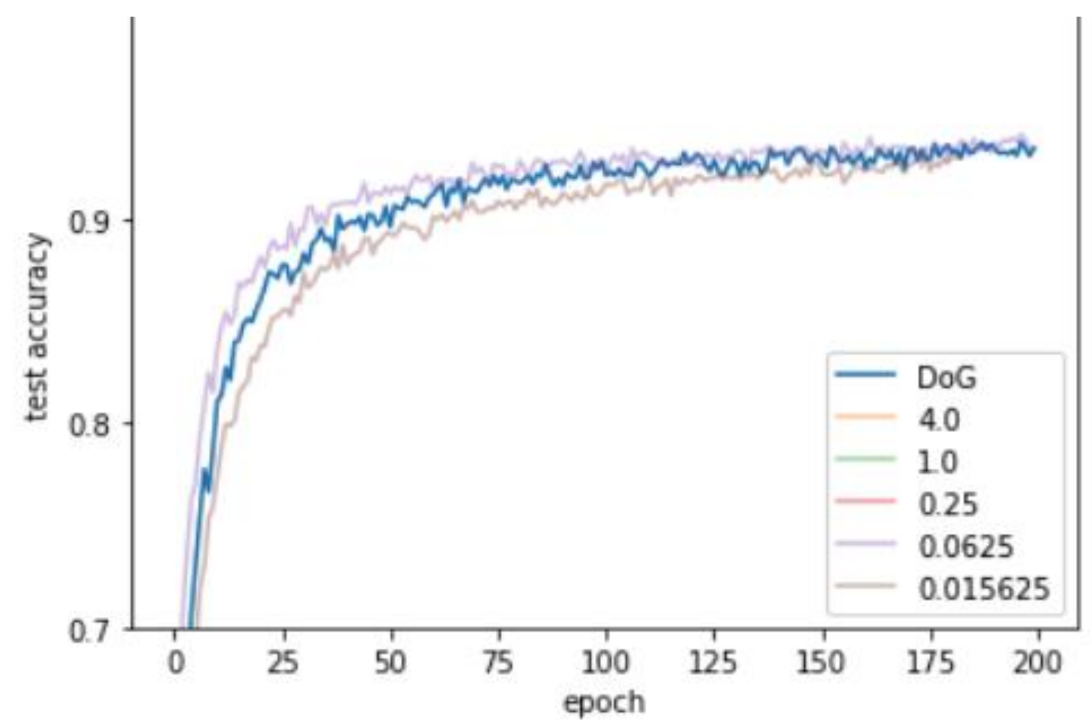
- Current practice for using SGD involves searching over a user specified grid of learning rates
- There is much work exploring the adaptive methods under the assumption that data points are chosen by an adversary instead of stochastically
- By devising an automatic tuning algorithm for the stochastic case we believe we can better results.

What is a trajectory cue?

- The trajectory refers to the path of an algorithm
- Trajectory cues are information that we get from running the algorithm
- An example of a trajectory cue is the distance travelled by the algorithm
- There has been no research explicitly studying how trajectory cues can be used to choose tuning parameters of stochastic optimization algorithms

University of Pittsburgh

- **Training machine learning models requires users to select many tuning parameters.**
- **These parameters are hard to select.**
- **This project aims to eliminate these parameters by using information inferred from training algorithm trajectories.**
- **This will make training machine learning models less time consuming and more user-friendly.**



Preliminary results: training CIFAR-10 with a CNN. Comparison is with a range of different fixed step sizes and with our method called DoG, note step sizes not displayed diverge.

Project Deliverables

- Funding from NSF
- Publication of several papers developing the theory and practice of adaptive stochastic optimization using trajectory cues
- The development of practical software

Potential Impact

- The most efficient way to remove tuning parameters from SGD is a major outstanding problem in optimization theory
- Better understanding of the differences between between adversarial and stochastic optimization
- If we can translate our theory into practice this will make training machine learning models faster and easier