

NUMERICAL ANALYSIS AND SCIENTIFIC COMPUTING
SEMINAR

*ADAHESIAN: An Adaptive Second Order Optimizer for
Machine Learning*

Michael W. Mahoney
ICSI and Department of Statistics, UC Berkeley

Abstract: Second order optimization algorithms have a long history in scientific computing, but they tend not to be used much in machine learning. This is in spite of the fact that they gracefully handle step size issues, poor conditioning problems, communication-computation tradeoffs, etc., all problems which are increasingly important in large-scale and high performance machine learning. A large part of the reason for this is that their implementation requires some care, e.g., a good implementation isn't possible in a few lines of python after taking a data science boot camp, and that a naive implementation typically performs worse than heavily parameterized/hyperparameterized stochastic first order methods. We describe ADAHESIAN, a second order stochastic optimization algorithm which dynamically incorporates the curvature of the loss function via ADaptive estimates of the Hessian. ADAHESIAN includes several novel performance-improving features, including: (i) a fast Hutchinson based method to approximate the curvature matrix with low computational overhead; (ii) a spatial averaging to reduce the variance of the second derivative; and (iii) a root-mean-square exponential moving average to smooth out variations of the second-derivative across different iterations. Extensive tests on natural language processing, computer vision, and recommendation system tasks demonstrate that ADAHESIAN achieves state-of-the-art results. The cost per iteration of ADAHESIAN is comparable to first-order methods, and ADAHESIAN exhibits improved robustness towards variations in hyperparameter values.

Friday, March 5, 2021, 1:30 pm
<https://emory.zoom.us/j/95900585494>

MATHEMATICS
EMORY UNIVERSITY