数学与系统科学研究院

计算数学所学术报告

报告人： **Dr. Yuhua Zhu**

（ *Stanford University* ）

报告题目：

Towards the theoretical understanding of large batch training in stochastic gradient descent

邀请人： 刘歆 副研究员

报告时间：**2019 年 6 月 24 日（周一）**

上午 8:30-9:30

报告地点： 科技综合楼三层

301 报告厅

Abstract：

Stochastic gradient descent (SGD) is almost ubiquitously used for training non-convex optimization tasks. Recently, a hypothesis proposed by Keskar et al. that large batch methods tend to converge to sharp minimizers has received increasing attention by researchers. We theoretically justify this hypothesis by providing new properties of SGD in both finite-time and asymptotic regime. In particular, we give an explicit escaping time of SGD from a local minimum in the finite-time regime and prove that SGD tends to converge to flatter minima in the asymptotic regime (although may take exponential time to converge) regardless of the batch size. We also find that SGD with a larger learning rate to batch size ratio tends to converge to a flat minimum faster, however, its generalization performance could be worse than the SGD with a smaller learning rate to batch size ratio. We include experiments to corroborate these theoretical findings.

欢迎大家参加！