

Interface of statistics, computing, and data science

XIAOMING HUO

H. Milton Stewart School of Industrial and Systems Engineering, Georgia
Institute of Technology, USA

Email: huo@gatech.edu

Inference (aka predictive modeling) is in the core of many data science problems. Traditional approaches could be either statistically or computationally efficient, however not necessarily both. The existing principles in deriving these models - such as the maximal likelihood estimation principle - may have been developed decades ago, and do not take into account the new aspects of the data, such as their large volume, variety, velocity and veracity. On the other hand, many existing empirical algorithms are doing extremely well in a wide spectrum of applications, such as the deep learning framework; however they do not have the theoretical guarantee like these classical methods. We aim to develop new algorithms that are both computationally efficient and statistically optimal. Such a work is fundamental in nature, however will have significant impacts in all data science problems that one may encounter in the society. Following the aforementioned spirit, I will describe a set of my past and current projects including L1-based relaxation, fast nonlinear correlation, optimality of detectability, and nonconvex regularization. All of them integrates statistical and computational considerations to develop data analysis tools.