The past decade has seen the emergence of datasets of an unprecedented scale, with both large sample sizes and dimensionality. Massive data sets arise in various domains, among them computer vision, natural language processing, computational biology, social networks analysis and recommendation systems, to name a few. In many such problems, the bottleneck is not just the number of data samples, but also the computational resources available to process the data. Thus, a fundamental goal in these problems is to characterize how estimation error behaves as a function of the sample size, number of parameters, and the computational budget available.

In this talk, I present three research threads that provide complementary lines of attack on this broader research agenda: (i) lower bounds for statistical estimation with computational constraints; (ii) interplay between statistical and computational complexities in structured high-dimensional estimation; and (iii) a computationally budgeted framework for model selection. The first characterizes fundamental limits in a uniform sense over all methods, whereas the latter two provide explicit algorithms that exploit the interaction of computational and statistical considerations.

[Joint work with John Duchi, Sahand Negahban, Clement Levrard, Pradeep Ravikumar, Peter Bartlett and Martin Wainwright]