Some statistical results on structured array factorization problems

Emile Richard (Stanford)

Abstract:
We consider a class of array factorization problems under specific structural assumptions on the factors. A first example of such problems is for matrices where estimating a vector from noisy quadratic observations is the goal. This arises naturally in many contexts, from dimensionality reduction, to synchronization and phase retrieval problems.
In case of quadratic measurements, the unknown vector is constrained to belong to a cone.
While optimal estimation appears to be intractable for the general problems in this class, we provide evidence that it is tractable when the cone is convex with efficient projection (e.g. the positive orthant). This is surprising, since the corresponding optimization problem is non-convex and --from a worst case perspective-- often NP hard. We characterize the resulting minimax
risk in terms of the statistical dimension of the cone. This quantity is already known to control the risk of estimation from linear observations, but its relevance to the case treated here was far from obvious.
We will consider two extensions of the last problem. First we discuss the case where the rank is larger than 1 and explore the construction of convex criteria for encouraging matrices with parsimonious representations using structured factors. We will discuss statistical properties of the norm encouraging sparse factors and propose an optimization algorithm.
Finally we extend our rank-1 analysis to the Principal Component Analysis problem for large tensors of arbitrary order k under a single-spike (or rank-one plus noise) model.
We establish conditions under which the principal component can be estimated using unbounded computational resources. We also analyse several polynomial-time estimation algorithms. We show that, unless the signal-to-noise ratio diverges in the system dimensions, none of these approaches succeeds. This is possibly related to a fundamental limitation of computationally tractable estimators for this problem.