Abstract

A conditional sampling oracle for a probability distribution D returns samples from the conditional distribution of D restricted to a specified subset of the domain. A recent line of work [CFGM13, CRS14] has shown that having access to such a conditional sampling oracle requires only polylogarithmic or even constant number of samples to solve distribution testing problems like identity and uniformity. This significantly improves over the standard sampling model where polynomially many samples are necessary. Inspired by these results, we introduce a computational model based on conditional sampling to develop sublinear algorithms with exponentially faster runtimes compared to standard sublinear algorithms. We focus on geometric optimization problems over points in high dimensional Euclidean space. Access to these points is provided via a conditional sampling oracle that takes as input a succinct representation of a subset of the domain and outputs a uniformly random point in that subset. We study two well studied problems: k-means clustering and estimating the weight of the minimum spanning tree. In contrast to prior algorithms for the classic model, our algorithms have time, space and sample complexity that is polynomial in the dimension and polylogarithmic in the number of points. Finally, we comment on the applicability of the model and compare with existing ones like streaming, parallel and distributed computational models.