Abstract

The metric sketching problem is defined as follows. Given a metric on n points, and $\epsilon > 0$, we wish to produce a small size data structure (sketch) that, given any pair of point indices, recovers the distance between the points up to a $1+\epsilon$ distortion. In this paper we consider metrics induced by ℓ_2 and ℓ_1 norms whose spread (the ratio of the diameter to the closest pair distance) is bounded by $\Phi > 0$. A well-known dimensionality reduction theorem due to Johnson and Lindenstrauss yields a sketch of size $O(\epsilon^{-2} \log(\Phi n) n \log n)$, i.e., $O(\epsilon^{-2} \log(\Phi n) \log n)$ bits per point. We show that this bound is not optimal, and can be substantially improved to $O(\epsilon^{-2} \log(1/\epsilon) \cdot \log n + \log \log \Phi)$ bits per point. Furthermore, we show that our bound is tight up to a factor of $\log(1/\epsilon)$. We also consider sketching of general metrics and provide a sketch of size $O(n \log(1/\epsilon) + \log \log \Phi)$ bits per point, which we show is optimal.