Seminar on Algorithms and Geometry – Handout 7

Robert Krauthgamer

June 11, 2009

Today's topics

We will investigate dimension reduction in ℓ_1 from two perspectives. The first is a communication complexity lower bound (mentioned last week, but we did not get to it).

Theorem 1 (Woodruff, 2004)**.** *For every $\varepsilon > 0$, one-way distance estimation within factor $1+\varepsilon$ (in the decision version) requires communication $\Omega(1/\varepsilon^2)$.*

The second perspective is a negative answer to the question whether there is an ℓ_1-analogue of the Johnson-Lindenstrauss lemma.

Theorem 2 (Brinkman and Charikar, 2003)**.** *For every $D \geq 1$ and integer $n \geq 1$, there is an n-point subset $X \subset \ell_1$ such that embedding X with distortion D into ℓ_1^k requires dimension $k \geq n^{\Omega(1/D^2)}$.*

Along the way, we will get another embedding lower bound, which matches an embedding of planar metrics into ℓ_2 by [Rao, 1999] up to constant factors.

Theorem 3 (Newman and Rabinovich, 2002)**.** *For every integer $n \geq 1$, there is an n-point planar metric embedding of which into ℓ_2 requires distortion $\Omega(\sqrt{\log n})$.*

We will see in class a more elementary proof of Theorem 1 due to Jayram, Kumar, and Sivakumar [Theory of Computing, 2008], and a short proof of Theorem 2 due to Lee and Naor [GAFA, 2004].

Research Directions

Here are some open problems related to today's class.

Q1. What is the least distortion required to embed planar metrics into ℓ_1? It is conjectured to be $O(1)$. The best bound currently known is $O(\sqrt{\log n})$, which follows from [Rao, 1999] mentioned above.

Q2. What is the communication complexity of distance estimation in ℓ_1 within approximation $1+\varepsilon$ (for randomized protocols without restricting the number of rounds)? The aforementioned lower bound (for one-way protocols) was recently extended to a small number of rounds [Brody and Chakrabarti, 2009].

Q3. Let $e_1, \ldots, e_n \in \ell_1^n$ be the standard unit basis vectors, and let $\varepsilon > 0$. What is the least dimenion $k = k(n, \varepsilon)$ such that they embed with distortion $1 + \varepsilon$ into ℓ_1^k? The current upper bound is $k \le O(\varepsilon^{-2} \log n)$, e.g. by choosing random vectors in the hypercube. Essentially no lower bound on the dependence on ε is known (there is a near-tight lower bound by [Alon, 2003] for the "special case" of ℓ_2).

Homework

1. Consider the distance estimation problem for the path metric (i.e. the set $\mathbb{Z}$ with distance function $d(i, j) = \|i - j\|$). Design for this problem a randomized protocol achieving approximation $1 + \varepsilon$ with communication better than $O(1/\varepsilon^2)$ (i.e. better than what follows immediately for the more general case of ℓ_1).

   Recall this is a decision problem, i.e. to distinguish between the cases $d(x, y) \le R$ and $d(x, y) > (1 + \varepsilon)R$ for a parameter $R > 0$.

   Mention explicitly how many rounds your protocol needs.

2. Let T be a complete binary tree with n leaves ($2n - 1$ nodes), and let $\varepsilon > 0$. Show that the (shortest-path metric) of T embeds with distortion $1 + \varepsilon$ into ℓ_1^k for $k \le (\frac{\log n}{\varepsilon})^{O(1)}$.