Seminar on Sublinear Time Algorithms

Lecture 4

April 14, 2010

Lecturer: Robert Krauthgamer | Scribe by: Merav Parter | Updated: May 10, 2010

1 Testing Homomorphism of a Function

Definition 1 *A function* $f : \mathbb{Z}_n \to \mathbb{Z}_n$ *is called homomorphism if*
$\forall\ x, y\ \in \mathbb{Z}_n\ , f(x) + f(y) = f(x + y)$.

Definition 2 *A function* $f : \mathbb{Z}_n \to \mathbb{Z}_n$ *is called* ϵ*-close to homomorphism if it can be changed in at most* ϵ*-fraction of places* $x \in \mathbb{Z}_n$ *to become a homomorphism, otherwise it is* ϵ*-far.*

Task Definition: Given a function f, test whether it is a homomorphism or ϵ*-far* from it.

Theorem 3 *[Ben-Or, Luby, Rubinfeld and Coppersmith]* $\forall\ 0 \le \epsilon \le \frac{1}{3}$ *there is a tester for homomorphism that determines w.h.p if* f *is a homomorphism or* ϵ*-far from it in time* $O(\frac{1}{\epsilon})$.

Key Idea: Relate ϵ to $\delta(f) = P_{\forall x,y \in \mathbb{Z}_n}[f(x) + f(y) \neq f(x + y)]$.

Algorithm Test Homomorphism

1. Repeat $\frac{4}{\epsilon}$ times:
   (a) Choose $x, y \in \mathbb{Z}_n$ at random and check if $f(x) + f(y) = f(x + y)$.
2. Accept if all these hold with equality, otherwise Reject.

Analysis:
Runtime: Obvious.
Correctness: Since the algorithm always accepts a homomorphism function, it is a one-sided error algorithm. For the rest of the discussion, we therefore assume that f is ϵ-far from homomorphism. We wish to show that $P(algorithm\ accepts\ f) \quad \le \frac{1}{3}$.
A simple but important observation in this context is that if f is ϵ-close to homomorphism then there exists a "corrected" function $g : \mathbb{Z}_n \to \mathbb{Z}_n$ which is both homomorphism and ϵ-close to f. This function g is defined as follow: $g(x) = plurality_y f(x + y) - f(y)$. Intuitively, the value $f(x + y) - f(y)$ can be thought of as the "vote" of y on x. If f is ϵ-close to homomorphism, then most of the votes for a given x are the same, resulting in a homomorphism function g. To prove this intuition in a formal manner, we state two auxiliary claims.

Claim 4 *f and g agree on at least $1-2\delta$ values.*

Claim 5 *If $\delta(f) \leq \frac{1}{6}$ then g is homomorphism.*

Assuming these claims to be correct, we turn to prove Theorem 3.

Proof [**of Theorem 3**]
Recall that we consider the case where f is ϵ-far from homomorphism. First, assume that $\delta(f) \leq \frac{\epsilon}{2}$ ($< \frac{1}{6}$). By Claim 4 and Claim 5 we have that f is $2\delta(f) \leq \epsilon$ close to homomorphism, and we end with a contradiction. Next, assume that $\delta(f) > \frac{\epsilon}{2}$. We will see that in this case the algorithm will reject the function w.h.p.:
$P[alg \;\; accepts \;\; f] \leq (1-\delta(f))^{\frac{4}{\epsilon}} \leq (1-\frac{\epsilon}{2})^{\frac{4}{\epsilon}} < e^{-2} < \frac{1}{3}$ as required. ■

It is yet left to prove the supporting claims. We begin with Claim 4.

Proof [**of Claim 4**]
Let $\Delta(f,g)$ denote the fraction of disagreements between f and g.
Let $B = \{x : Pr_y[f(x)+f(y) \neq f(x+y)] \geq \frac{1}{2}\}$. Notice that B contains all the $x's$ where f and g disagree. In addition, $\delta(f) \geq \frac{|B|}{n} \cdot \frac{1}{2}$, where $\frac{|B|}{n}$ is the probability to choose a bad x, and $\frac{1}{2}$ is a lower bound for the probability to chose a bad partner y. Overall we have that $\Delta(f,g) \leq \frac{|B|}{n} \leq 2\delta(f)$ as required. ■

As a step toward proving Claim 5, we state the following claim.

Claim 6 *The function g is a strong plurality (if δ is small) in the following sense,*
$\forall\, x, P_y[g(x) = f(x+y) - f(y)] \geq 1 - \delta(f)$.

Proof [**of Claim 6**]
We first analyze for an arbitrary $x \in \mathbb{Z}_n$ the "collision probability" of two votes and then relate it to the "plurality probability" as required by the claim.
Fix x and choose y_1, y_2 at random and independently. Then we have that
$P_{y_1,y_2}[f(x+y_1) - f(y_1) = f(x+y_2) - f(y_2)] =$

$P_{y_1,y_2}[f(x+y_1) + f(y_2) = f(x+y_2) + f(y_1)] \geq$

$P_{y_1,y_2}[f(x+y_1) + f(y_2) = f(x+y_2) + f(y_1) = f(x+y_1+f(y_2)] \geq 1 - 2\delta(f)$

Where the last inequality follows by the union-bound. To show that the "collision probability" is at most the "plurality probability", we consider an experiment A with n possible outcomes, where $o(A) \in [1,n]$ denotes the outcome of A. Let $p_i = P[o(A) = i]$, i.e., the probability where that the experiment ended with outcome i., where $p_i > 0$ for every $1 \leq i \leq n$, and $\sum_{i=1}^{n} p_i = 1$. The probability that two independent experiments A, B ended with the same outcome (i.e., collision occurred) is given by

$$P[o(A) = o(B)] = \sum_{i=1}^{n} p_i^2 \leq \max_j(p_j) \cdot \sum_{i=1}^{n} p_i = \max_j(p_j) \cdot 1 = \max_j(p_j).$$

Since $\max_j(p_j)$ is the "plurality probability" the claim is established. ■

We are now ready to complete the proof for Claim 5.

Proof [**of Claim 5**]
Fix x, z. By applying Claim 6 three times, first for x, then for z and finally for $x + z$, we get

1) $P_y[g(x) \neq f(x + y - x) - f(y - x)] \leq 2\delta(f) < \frac{1}{3}$

2) $P_y[g(z) \neq f(z + y) - f(y)] \leq 2\delta(f) < \frac{1}{3}$

3) $P_y[g(x + z) \neq f(x + z + y - x) - f(y - x)] \leq 2\delta(f) < \frac{1}{3}$

With positive probability none of these events happen, implying that
$\exists y$ such that $g(x) + g(z) = [f(y) - f(y - x)] + [f(z + y) - f(y)] = g(x + z)$
where the first equality is followed by (1,2) and the second equality is followed by (3). The Claim follows. ■

2 Testing a Dense Graph for Bipartiteness

Definition 7 *Graph $G = (V, E)$ is ϵ-far from bipartite if it is necessary to remove more than $\epsilon|V|^2$ edges so that it becomes bipartite.*

<u>Task definition</u>: Given a dense graph $G = (V, E)$, determine w.h.p if it is bipartite or ϵ-far from it.

Theorem 8 *(Goldreich-Goldwasser-Ron) There is a tester for bipartiteness that determines whether G is bipartite or ϵ-far from it in time $(\frac{1}{\epsilon})^{O(1)}$.*

In particular, the tester we present always accepts bipartite graphs and rejects ϵ-far instances with probability at least $\frac{2}{3}$.

<u>Key Idea</u>: Sampling small number of vertices is in fact representative.

Algorithm Test-Bipartite

1. Uniformly and independently select $m = \Theta(\frac{\log(\frac{1}{\epsilon})}{\epsilon^2})$ vertices.
2. Accept iff the subgraph induced on them is bipartite (by BFS)

Analysis:
Runtime: $\Theta(\frac{\log^2(\frac{1}{\epsilon})}{\epsilon^4})$. Quadratic in the size of the sample due to construction of the induced subgraph.
Correctness: If G is bipartite then clearly so is every subgraph of it. Hence, this is a one-sided error tester. We next assume that G is ϵ-far from bipartite and wish to show it is rejected by the algorithm with probability greater than $\frac{2}{3}$.
Let R denote the set of sampled vertices. It is convenient to view R as composed of two parts that are sampled one after the other, namely U and S respectively. Let $|U| = O(\frac{1}{\epsilon}\log\frac{1}{\epsilon})$ and $|S| = O(\frac{|U|}{\epsilon})$. One can check that indeed $m = |U| + |S|$. Note, that since the vertices are selected independently, repetitions may occur (e.g., U and S may overlap). We first provide some definitions.

Definition 9 *A vertex v is high-degree if its degree is greater than $\frac{\epsilon n}{3}$.*

Definition 10 *Set $U \in V$ is good if all but at most $\frac{\epsilon n}{3}$ of the high-degree vertices of V are adjacent to U.*

Let $\Gamma(w) = \{v : (w, v) \in E\}$ (the neighbors of w).
Let $\Gamma(W) = \cup_{w \in W}\Gamma(w)$.

Claim 11 *With probability of at least $\frac{5}{6}$ over the choice of U, the set U is good.*

Proof
Let $v \in V$ be a high-degree vertex. The probability that U contains none of $v's$ neighbors is at most

$$(1 - \frac{\epsilon}{3})^{|U|} < e^{\frac{\epsilon}{3} \cdot |U|} \tag{1}$$

If we sample $|U| = \frac{3}{\epsilon} \cdot \ln(\frac{18}{\epsilon})$, we get that this probability is at most $(\frac{\epsilon}{18})$. By linearity of expectation, the expected number of such $v's$ is $\leq \frac{\epsilon n}{18}$. Finally by Markov's inequality, the probability that there are more than $\frac{\epsilon n}{3}$ such $v's$ (high-degree vertices with no neighbor in U) is at most $\frac{1}{6}$ as required. ∎

Definition 12 *An edge is said to disturb a partition $U = U_1 \bigcup U_2$ if its endpoints are in the same $\Gamma(U_i)$ for $i \in [1, 2]$.*

Claim 13 *If G is ϵ-far from bipartite then for every good U and for every partition of $U = U_1 \bigcup U_2$ there are at least $\frac{\epsilon n^2}{3}$ disturbing edges.*

Proof
Assume U is indeed good and consider a fixed partition $U = U_1 \bigcup U_2$. Let $N = \Gamma(U)$ and $C = V \backslash N$. Since U is good, we have that C contains at most $\frac{\epsilon n}{3}$ high-degree vertices. We next use the partition of U to induce a partition of N and eventually on V in the following manner:
$N_1 = \Gamma(U_1)$ and $N_2 = \Gamma(U) \backslash N_1$. Let C_1, C_2 be any partition of C such that $(C \bigcap U_1) \subseteq C_1$

and $(C \bigcap U_2) \subseteq C_2$. The final partition of $V = V_1 \bigcup V_2$ is $V_1 = N_1 \bigcup C_2$ and $V_2 = N_2 \bigcup C_1$. Observe that since G is ϵ-far from bipartite, *every* partition of V has more than ϵn^2 "disturbing" edges. In particular this is correct for the partition (V_1, V_2). We next show that many of these "disturbing" edges are incident to vertices in U.
Q: How many disturbing edges at-most can be incident to C (i.e., not incident to N)?
Ans.: C contains at-most n edges from each of at-most $\frac{\epsilon n}{3}$ high-degree vertices. In addition, it contains at-most $\frac{\epsilon n}{3}$ edges from each of at-most n non-high-degree vertices.
Overall, we get that the are at least $\frac{\epsilon n^2}{3}$ disturbing edges that are incident to N. ■

We are now ready to complete the proof of Theorem 8.

Proof [**of Theorem 8**]
Let $G[R]$ be the graph induced by the the selected set R. For $G[R]$ to be bipartite we must have either:

1) U is not good (w.p $\leq \frac{1}{6}$)

2) U is good and $\exists$ a partition $U = U_1 \bigcup U_2$ such that none of its disturbing edges occur in $G[R]$. Applying the union-bound over the possible $2^{|U|}$ partitions of $|U|$ and combining Claim 13, we get that the probability for such an event is at most
$2^{|U|}(1 - \frac{\epsilon}{3})^{\frac{|S|}{2}} = 2^{|U|} \cdot e^{\frac{-\epsilon|S|}{6}} < \frac{1}{6}$.
Overall, the probability to accept an ϵ-far graph G is at-most $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ as required. ■