6.896 Sublinear Time Algorithms November, 16, 2004

Lecture 18

Lecturer: Eli Ben-Sasson *Scribe: John Danaher*

1 The Blum Luby Rubinfeld (BLR) Linearity Test

Throughout our talk today, G, H are groups (not necessarily Abelian), written additively. A function $\varphi : G \to H$ is called a homomorphism if $\forall a, b \in G, \varphi(a + b) = \varphi(a) + \varphi(b)$. In this lecture we will give a test for closeness to a homomorphism. First we define "closeness".

Definition 1 (Distance) *The (relative) Hamming distance between two functions $f, g : G \to H$, is $\delta(f, g) = \Pr_{x\in G}[f(x) \neq g(x)]$. The distance of f from a set of functions F, is $\delta(f, F) = \min_{f'\in F} \delta(f, f')$. Finally, the distance of the set F is $\delta(F) = \min_{f\neq f'\in F} \delta(f, f')$. (If $|F| \leq 1$ we define $\delta(F) = 1$.)*

We now express our testing problem formally:

Definition 2 (Homomorphism Tester) *Given groups G, H, define their set of homomorphisms to be*

$$\mathrm{Hom}(G, H) = \{\varphi : G \to H : \forall a, b \in G, \varphi(a + b) = \varphi(a) + \varphi(b)\}$$

A (q, ϵ_0, c)-tester for $\mathrm{Hom}(G, H)$ is a randomized polynomial time algorithm T^f with oracle access to a function $f : G \to H$, with the following properties.

Operation *T^f tosses some coins, makes q queries to f and outputs* accept/reject.

Completeness *If $f \in \mathrm{Hom}(G, H)$ then $\Pr[T^f$ accepts $] = 1$.*

Soundness $\Pr[T^f$ *rejects* $] > \min(c \cdot \delta(f, \mathrm{Hom}(G, H)), \epsilon_0)$.

We would like a tester with the smallest possible query complexity q, and with the largest possible ϵ_0, c (both parameters are at most 1). Our main Theorem, originally proved by Blum, Luby, Rubinfeld [2], shows the existence of a good tester for any pair of groups G, H. The proof of the Theorem we present today is due to Coppersmith.

Theorem 3 (BLR Linearity Testing) *For every pair of groups G, H, there exists a $(3, 2/9, 1/2)$-tester for $\mathrm{Hom}(G, H)$.*

In the next lecture we will show that for the special case of $G = Z_2^n$ and $H = Z_2$, one achieves a $(3, 1, 1)$-tester, which is pretty much optimal (as it is known any linear property that can be tested with query complexity ≤ 2, actually has a finite number of elements in it [1]). Notice the above Theorem implies that the property of being a homomorphism is locally testable (in sub-linear time, with constant query complexity). Indeed, a function in the property is accepted with probability one, whereas a function that is ϵ-far from $\mathrm{Hom}(G, H)$ is rejected with probability at least $\min(2/9, \epsilon/2)$, and this probability can be boosted to any constant by a constant number of repetitions of the basic (three-query) test.

Before presenting the homomorphism tester and analyzing its performance (and proving Theorem 3), we notice $\mathrm{Hom}(G, H)$ can be used as an error correcting code code with large minimal distance. A function $\varphi : G \to H$ can be naturally interpreted as a string in H^G (i.e. a string of length $|G|$, with entries in H. For $g \in G$, the entry corresponding to g is $\varphi(g)$). Indeed, the Hadamard code is a special case of such a code, corresponding to $\mathrm{Hom}(G, H)$ where $G = Z_2^n$ to $H = Z_2$ (more on this in next lecture). Our next claim shows the code corresponding to $\mathrm{Hom}(G, H)$ has large minimal distance.

Claim 4 *For all groups G, H we have $\delta(\mathrm{Hom}(G, H)) \geq 1/2$.*

Proof $Hom(G, H)$ is closed under addition: if two functions φ, ψ are homomorphisms, then so is their sum $\varphi + \psi$. This implies that the minimum distance of $\mathrm{Hom}(G, H)$ is the minimum distance to zero:

$$\delta(\mathrm{Hom}(G, H)) = \min_{\varphi \in \mathrm{Hom}(G,H), \varphi \not\equiv 0} \Pr_x[\varphi(x) \neq 0]$$

Also, for every homomorphism φ it is well-known that $\ker(\varphi) = \{x \in G : \varphi(x) = 0\}$ is a subgroup of G. This implies $|\ker(\varphi)|$ divides $|G|$. Since $\varphi \not\equiv 0$ we have $|\ker(\varphi)| \leq |G|/2$. This implies that φ is non-zero with probability at least half. This applies to any nontrivial homomorphism φ, completing our proof. ■

We now return to our original problem, of testing whether f is close to $\mathrm{Hom}(G, H)$. We use the following test (denoted T^f), originally suggested and analyzed by Blum, Luby and Rubinfeld [2].

Algorithm 1 (BLR Test) *Pick $x, y \in G$ uniformly at random. ACCEPT iff $f(x + y) = f(x) + f(y)$.*

Proof **[of Theorem 3]:** We prove only for G, H Abelian, but the proof can be generalized to the non Abelian case. By definition of the BLR test, the query complexity is three, as claimed. Additionally, f is a homomorphism iff $\Pr[T^f \text{accepts}] = 1$. This proves the Completeness condition. Finally, we need to analyze the soundness, which amounts to proving the following claim.

$$\text{If} \Pr[T^f \text{ rejects}] = \epsilon < 2/9, \text{ then } \delta(f, \mathrm{Hom}(G, H)) \leq 2 \cdot \epsilon. \tag{1}$$

We define $\phi(x) = \mathrm{maj}_{y \in G}(f(x + y) - f(y))$. We will shortly prove the following pair of claims.

Claim 5 *$\delta(\phi, f) \leq 2\epsilon$.*

Claim 6 *If $\epsilon < 2/9$, then $\phi \in \mathrm{Hom}(G, H)$.*

Our theorem is implied immediately by the pair of claims. Indeed, by Claim 6 ϕ is a homomorphism, and by Claim 5 f is 2ϵ-close to ϕ. This completes our proof (modulo the proofs of the claims, which we attend to next). ■

Proof **[of Claim 5]** Let $Bad = \{x \in G : \Pr_y[f(x) \neq f(x + y) - f(y)] > 1/2]\}$. If $x \notin Bad$, then the majority of y's "vote" for assigning $\phi(x)$ the value $f(x)$ and in this case $\phi(x) = f(x)$. Thus, it suffices to bound $\frac{|Bad|}{|G|}$. For this we have $\epsilon = \Pr[T^f fails] \geq \frac{|Bad|}{|G|} \cdot \frac{1}{2}$ because T^f fails with probability over $\frac{1}{2}$ if it picks $x \in Bad$. Thus, $\frac{|Bad|}{|G|} \leq 2\epsilon$ and our proof is complete. ■

Proof **[of Claim 6]** We need to show that for all $x, z \in G$, $\phi(x) + \phi(z) = \phi(x + z)$. To do this, we fix x and show that all pairs y_1, y_2 agree to vote for the same value of $\phi(x)$ with very high probability. Pick random $y_1, y_2 \in G$, and we have:

$$\Pr_{y_1,y_2}[f(x + y_1) - f(y_1) = f(x + y_2) - f(y_2)] = \Pr_{y_1,y_2}[f(x + y_1) + f(y_2) = f(x + y_2) + f(y_1)].$$

Recall $x + y_1$ is a random element of G, and so is y_2. Thus, by assumption we have $\Pr_{y_1,y_2}[f(x + y_1) + f(y_2) \neq f(x+y_1+y_2)] = \epsilon < 2/9$. Similarly we have $\Pr_{y_1,y_2}[f(x+y_2)+f(y_1) \neq f(x+y_1+y_2)] = \epsilon < 2/9$. Taking a union bound we get

$$\Pr_{y_1,y_2}[f(x + y_1) - f(y_1) = f(x + y_2) - f(y_2)] > 1 - 2\epsilon \geq 5/9.$$

Since we have two identical random trials that are equal with probability $> 5/9$, this implies that the most common value of a single trial occurs with probability greater than $2/3$ (since to agree they must both choose either the most common value or some other value). So,

$$\forall x \in G, \quad \Pr_y[\phi(x) \neq f(x + y) - f(y)] < 1/3. \tag{2}$$

Recall we have to show for all $x, z \in G$ that $\phi(x) + \phi(z) = \phi(x+z)$. Fix x and z and apply Equation (2) to x, z and $x+z$ to get

$$\begin{aligned}
\Pr_y[\phi(x) \neq f(y) - f(y-x)] &< 1/3 &(3)\\
\Pr_y[\phi(z) \neq f(z+y) - f(y)] &< 1/3 &(4)\\
\Pr_y[\phi(x+z) \neq f(z+y) - f(y-x)] &< 1/3 &(5)
\end{aligned}$$

We want to find some y for which none of the above inequalities are true (i.e. all of the complimentary equalities are true). The probabilities don't add up to 1, so as we throw away the bad values of y for each inequality, there must be some y left over which is what we want. For this fixed y we have

$$\phi(x) + \phi(z) = (f(y) - f(y-x)) + (f(z+y) - f(y)) = f(z+y) - f(y-x) = \phi(x+z).$$

The above equality holds for every $x, z \in G$, showing $\phi \in \mathrm{Hom}(G, H)$ and completing the proof of Claim 6. ■

Looking at this proof by treating $\mathrm{Hom}(G, H)$ as a code, we depend on the fact that if a codeword has only a small amount of noise, then we can locally correct any (possibly corrupted) entry of the codeword, using the rest of the code. The function ϕ is precisely the "local correction" of the noisy f. In the other direction, if f is very noisy, then our local correction process must fail with high probability, thus allowing us to test f with a very small number of queries.

Next week we will show a tighter (Fourier based) analysis for the special Boolean case of $G = Z_2^n$ and $H = Z_2$, corresponding to the Hadamard code.

References

[1] E. Ben-Sasson, O. Goldreich, M. Sudan. Bounds on 2-Query Codeword Testing. *RANDOM-APPROX 2003*: 216-227

[2] Blum, M., Luby, M., and Rubinfeld, R. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences 47*, 3 (Dec. 1993), 549–595. (Preliminary Version in *22nd STOC*, 1990).