6.896 Sublinear Time Algorithms	November 4, 2004

Lecture 15

Lecturer: Eli Ben-Sasson — *Scribe: Elena Grigorescu*

1 Lower bounds for property testing

In [1], Ben-Sasson et al. show that random low density parity check (LDPC) codes are hard to test, which implies that there are some very good codes not locally testable. As a corollary, it follows that there are properties defined by 3CNF formulas that require linear query complexity to test. However, in [2], Fischer et al. show that a 2CNF property is testable in $O(\sqrt{n})$.

To obtain the results from [1] we will be using the following techniques:

1. Yao's minimax principle, which provides a general way of proving lower bounds for randomized algorithms.
2. Expansion in random graphs (and application to random LDPC codes).
3. Linear algebra.

1.1 Preliminary definitions

We work over the field of size two, with elements $0, 1$ and addition and multiplication modulo two.
A *linear property* is a linear (vector) space $V \subseteq F$ [1]

Definition 1 *For $x, y \in \{0,1\}^n$ their distance is $\delta(x,y) = \frac{\# x_i \neq \# y_i}{n}$. For a property $P \subseteq \{0,1\}^n$ let $\delta(x,P) = min_{y \in P} \delta(x,y)$. A property P is $(\delta, \epsilon_+, \epsilon_-, q)$-testable if there exists a tester T, which is a randomized algorithm, with query complexity $\leq q$, s.t the following hold:*

- *if $x \in P$ then $Pr[T \text{ accepts } x] \geq 1 - \epsilon_+$. Thus, ϵ_+ is the maximal error on positive inputs.*
- *if $\delta(x,P) \geq \delta$ then $Pr[T \text{ accepts } x] \leq \epsilon_-$, i.e. ϵ_- is the maximal error on negative inputs (that are δ-far from P).*

A test T is *non-adaptive* if the queries depend only on randomness and not on previous queries.

A test T has *perfect completeness* (or *one-sided error*) if $\epsilon_+ = 0$. A *Low Density Parity Check(LDPC)* matrix is a matrix $A_{m \times n}$ s.t $\forall i, |supp(A_i)| \leq c$,[2] where c is a constant. So each row of A contains only a constant number of 1's.

The code defined by A is

$$code(A) = \{x \in F^n | Ax = 0\}.$$

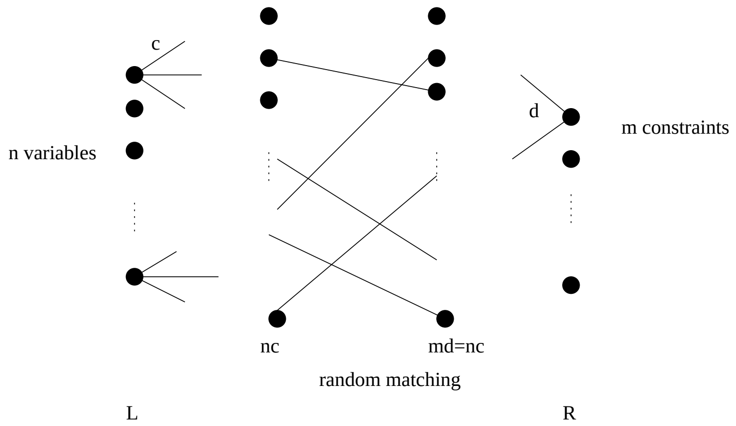
Therefore, $code(A) = kernel(A)$ and note that $code(A)$ is a linear space. The code of an LDPC matrix is called an LDPC code. These codes were introduced by Gallager [3] in the 60's, and are very useful error correcting codes.

A (c, d)- *regular code* is defined by the $m \times n$ adjacency matrix of a bipartite graph with n left vertices, each of degree c, and m right vertices, each of degree d.

A *random* (c, d)- *regular* code is defined by the adjacency matrix $A_{m \times n}$ of the following bipartite graph

[1] V is linear space if $\forall x, y \in V$ then $x + y \in V$.
[2] $supp(A_i)$ is the set of non-zero indices of the ith row of A.

G, as in Figure 1.1. Consider nc left vertices and $md = nc$ right vertices, and a random matching between them. To obtain G, collapse the vertices on the left into n vertices of degree c each, and collapse the right vertices into m vertices of degree d each. This is done by collapsing tuples of c, and d respectively, consecutive vertices into one vertex. Note that multiple edges are allowed.



The *dual space* $V^\perp$ of a linear space $V \subseteq F_2^n$ is defined as $V^\perp = \{x \in F^n | x \perp V\}$.[3]

A test for a linear property V is a *linear test* if it selects a set of indices $I \subseteq [n]$, (according to some distribution) and accepts $w \in F^n$ iff $w \perp V_I^\perp$, where $V_I^\perp = \{u \in V^\perp | supp(u) \subseteq I\}$.

We are now ready to state the main theorem.

1.2 The Main Theorem

Informally, the following Theorem states that there are linear properties, such as random LDPC codes, that cannot be tested with a small number of queries, using linear tests.

Theorem 2 *There exist integers (c, d) and constants $\delta > 0, \epsilon < \frac{1}{2}, \gamma > 0$ s.t. for all sufficiently large n, a random (c, d)-regular LDPC code is w.h.p. s.t. every $(\delta, 0, \epsilon, q)$-linear test has query complexity $q \geq \gamma n$.*

The theorem extends to any test (not just linear ones), since every test for a linear property can be reduced to a 1-sided error linear test, with parameters similar to the initial ones. We now state this result more formaly.

Theorem 3 *If $V \subseteq F^n$ is $(\delta, \epsilon_+, \epsilon_-, q)$- testable, then V is $(\delta, 0, \epsilon_+ + \epsilon_-, q)$-testable by a linear test.*

Note that the linear test has the same amount of error, only the error is shifted from the positive to the negative answers.

Before attempting to prove Theorem 2, we need to introduce a technical tool, Yao's minimax principle.

Definition 4 (Yao's minimax principle) *Let*
$\mathcal{I} =$ *finite set of inputs,*
$\mathcal{A} =$ *finite set of deterministic algorithms,*
$C : \mathcal{I} \times \mathcal{A} \to R^+ =$ *function measuring the cost of an algorithm on an instance.*
Then, $\forall$ distributions p over $\mathcal{I}$ and q over $\mathcal{A}$, we have

$$min_{A \in \mathcal{A}} E_p[C(I_p, A)] \leq Max_{I \in \mathcal{I}} E[C(I, A_q)].$$

[3] $\alpha \perp \beta$ iff $< \alpha, \beta >= 0$, and $\alpha \perp U$ iff $\alpha \perp u, \forall u \in U$.

Therefore, Yao's minimax principle implies that in order to show that for all randomized algorithms there is an input on which they fail, it is enough to consider deterministic tests and produce a distribution on the inputs on which each such test fails.

We start the proof of Theorem 1 by making the following observations. We need to show that for each randomized linear tests T, $\exists w$ that is δ-far from our property V s.t. $Pr[T \text{ accepts } w] \geq 1-\epsilon$. From the minimax principle it follows that, in order to show a lower bound on the query complexity of a randomized linear test, it is enough to find one bad distribution over a subset $B \subset \{0,1\}^n$, such that each $w \in B$ is δ-far from V, and every fixed deterministic q-query linear test will accept inputs from B with high probability, or equivalently, s.t. $Pr_{w \in B}[w \perp V_I^{\perp}] < \epsilon$.

Formally, let
$\mathcal{I} \stackrel{def}{=} \{w \mid \delta(w,V) > \delta\}$,
$\mathcal{A} = \{$ deterministic q query linear tests,$\}$
$C(w,T) = 1$ if T accepts, and 0 otherwise.
The minimax principle translates into

$$\forall p,q \ \mathrm{Max}_{\{w,\ \delta\ far\ from\ V\}} E_q[Pr[T_q \text{ accepts w}]] \geq \min_{\{det\ lin\ q-query\ T\}} E_p[Pr[T \text{ accepts } w_p]].$$

The proof will continue in the next lecture.

References

[1] E. Ben-Sasson, P. Harsha, S. Raskhodnikova *Some 3CNF Properties are Hard to Test*

[2] E. Fischer, E. Lehman, I. Newman, S. Raskhodnikova, R. Rubinfeld, A. Samorodnitsky *Monotonicity testing over general poset domains*

[3] R. G. Gallager, *Low Density Parity-Check Codes.* MIT Press, Cambridge, MA, 1963.