Electronic Colloquium on Computational Complexity, Report No. 4 (2010)

Low Rate Is Insufficient for Local Testability

Eli Ben-Sasson* Michael Viderman
Computer Science Department
Technion — Israel Institute of Technology
Haifa, 32000, Israel.
{eli, viderman}@cs.technion.ac.il

January 6, 2010

Abstract

Locally testable codes (LTCs) are error-correcting codes for which membership of a given word in the code can be tested probabilistically by examining it in very few locations. Kaufman and Sudan [2] proved that sparse, low-bias linear codes are locally testable (in particular sparse random codes are locally testable). Kopparty and Saraf [3] conjectured that all linear sparse codes (codes with logarithmic dimension) are locally testable.

In this paper we refute this conjecture by showing that for every $\epsilon > 0$ there exists a code $C_\epsilon \subset \mathbb{F}_2^n$ with relative distance $(1/2 - \epsilon)$ and $\dim(C) = \Theta(\log(n))$ which is not locally testable and not locally decodable. Moreover, our construction can achieve any (non-constant) dimension, e.g. we can construct C s.t. $\dim(C) = \log\log(n)$ and C is not locally testable (decodable). This also shows that the requirement of "low-bias" in the work of Kaufman and Sudan [2] was necessarily.

1 Introduction

Locally testable codes (LTCs) are error correcting codes for which distinguishing, when given oracle access to a purported word w, between the case that w is a codeword and the case that it is very far from all codewords, can be accomplished by a randomized algorithm, called a *tester*, which reads a constant amount of information from w.

On the other hand, *locally decodable codes* (LDCs) allow to recover each message entry with high probability by reading only a few entries of the codeword even if a constant fraction of it is adversely corrupted. The both families of error correcting codes are explicitly studied, for survey see e.g. [5].

Kaufman and Sudan [2] showed that local testability and decodability can be found in random. They showed that all sparse, low-bias linear codes are locally testable. In particular, sparse random linear codes have low-bias and hence are locally testable. This result was later generalized by Kopparty and Saraf [4] to *high error* regime.

Kopparty and Saraf [3] conjectured that all linear sparse codes (codes with logarithmic dimension) are locally testable. We refute this conjecture in Theorem 4.1 by showing a linear sparse code (with linear distance) which are not locally testable (and not locally decodable). This also shows that the "low-bias" requirement in the work of Kaufman and Sudan [2] was necessarily.

*Research of both authors supported by grant number 2006104 by the US-Israel Binational Science Foundation and by grant number 679/06 by the Israeli Science Foundation.

ISSN 1433-8092

2 Preliminary

Let $\mathbb{F}$ be a finite field and $[n]$ be the set $\{1,\ldots,n\}$. In this work, we consider only linear codes. We start with a few definitions.

Let $C \subseteq \mathbb{F}^n$ be a linear code over $\mathbb{F}$. For $w \in \mathbb{F}^n$, let $\mathrm{supp}(w) = \{i \in [n] \mid w_i \neq 0\}$ and $|w| = |\mathrm{supp}(w)|$. We define the relative distance between two words $x, y \in \mathbb{F}^n$ to be $\delta(x,y) = \frac{\Delta(x,y)}{n}$ and let $\Delta(x,y) = \delta(x,y) \cdot n$. The distance of a code is denoted by $\Delta(C)$ and defined to be the minimal value of $\Delta(x,y)$ for two distinct codewords $x, y \in C$. Similarly, the relative distance of the code is denoted $\delta(C) = \frac{\Delta(C)}{n}$. For $x \in \mathbb{F}^n$ and $C \subseteq \mathbb{F}^n$, let $\delta(x,C) = \delta_C(x) = \min_{y \in C}\{\delta(x,y)\}$ denote the relative distance of x from the code C. We note that $\Delta(C) = \min_{c \in C \setminus \{0\}}\{\mathrm{wt}(c)\}$. If $\delta(x,C) \geq \epsilon$, we say that x is ϵ-far from C and otherwise x is ϵ-close to C. Let $\dim(C)$ be the dimension of C. The vector inner product between u_1 and u_2 is denoted by $\langle u_1, u_2 \rangle$. The dual code $C^\perp$ is defined as $C^\perp = \{u \in \mathbb{F}^n \mid \forall c \in C : \langle u, c \rangle = 0\}$. In a similar way we define $C^\perp_{\leq t} = \left\{u \in C^\perp \mid |u| \leq t\right\}$ and $C^\perp_t = \left\{u \in C^\perp \mid |u| = t\right\}$.

For $w \in F^n$ let w^i be the concatenation of w to itself i times.

3 Definition of LTCs

In this section, we formally define Locally Testable Codes (LTCs). We define LTCs following [1], which in particular explains why LTCs can be defined in this way without loss of generality.

Definition 3.1 (LTCs and Testers)**.** Let $C \in \mathbb{F}^n$ be a linear code. Given a distribution $\mathcal{D}$ over set $C^\perp$, we define the support of $\mathcal{D}$ over $C^\perp$ as $\mathcal{D}_S = \left\{u \in C^\perp \mid \mathcal{D}(u) > 0\right\}$. We say that $\mathcal{D}$ is a (q,ϵ,δ)-distribution of the code C, if the following conditions are satisfied:

- $\mathcal{D}_S \subseteq C^\perp_{\leq q}$.
- For all $x \in \mathbb{F}^n$ s.t. $\delta(x,C) \geq \delta$ it holds that $\Pr_{u \sim \mathcal{D}}[\langle u, x \rangle \neq 0] \geq \epsilon$.

We say that $C \subseteq \mathbb{F}^n$ is a (q,ϵ,δ)-LTC if it has a (q,ϵ,δ)-distribution $\mathcal{D}$.

We say that a code C is locally testable when C is a (q,ϵ,δ)-LTC, where $q, \epsilon, \delta > 0$ are constants.

Remark 3.2. Usually we assume that $\delta \leq 1/3$ (see discussion in [1]). Hence, we will say that C is **not** locally testable if there exists a word w s.t. $\delta(w,C) \geq 1/3$ and w is accepted with probability 1 by every tester (distribution) for C (from definition 3.1).

4 Main Result

In the next theorem we refutes the conjecture of [3]. We show a linear code with logarithmical dimension and linear distance which is not locally testable. It can be readily verified that this construction works for any dimension, e.g. in the same way we can construct a code $C \subset F_2^n$ s.t. $\dim(C) = \log\log(n)$ and $\delta(C) \geq 0.49$ which will not be locally testable.

Theorem 4.1 (Bad Rate does not imply LTC)**.** *For every $\epsilon > 0$ there exists $C_\epsilon \subset \mathbb{F}_2^n$ s.t. $\delta(C) \geq \frac{1}{2} - \epsilon$, $\dim(C) = \log(n)$ and C is not locally testable with constant query complexity.*

Proof. Let $\epsilon > 0$ be a constant and $m = \log(n)$. Let $R_\epsilon \subset F_2^m$ be a linear code s.t. $\delta(R_\epsilon) \geq \frac{1}{2} - \epsilon$, $\delta(R_\epsilon{}^\perp) \geq \Omega(1)$ and $\dim(R_\epsilon) = \Theta(m)$ (even random code will have such properties). Let $w \in F_2^m$ be a word s.t. $\delta(w, R_\epsilon) \geq 1/3$ (note that we could use any other threshold less than 1/2, not only 1/3). Let $C_\epsilon \subset \mathbb{F}^n$ be a linear code, s.t. $c \in C_\epsilon$ if and only if $c = r^{(n/m)}$, where $r \in R_\epsilon$, i.e., every codeword of C_ϵ is a codeword of R_ϵ concatenated to itself n/m times.

We have $\dim(C_\epsilon) = \Theta(m) = \Theta(\log(n))$ and $\delta(C_\epsilon) \geq 1/2 - \epsilon$. Moreover, $\delta(w^{(n/m)}, C_\epsilon) = \delta(w, R_\epsilon) \geq 1/3$. Assume by a way of contradiction that C_ϵ is locally testable and $\mathcal{D}$ is a tester(distribution over constant weight dual words) for C_ϵ.

We argue that $w^{(n/m)}$ will be accepted by $\mathcal{D}$ with probability 1.

First, note that for all $u \in C_\epsilon{}^\perp$ if $|u| = 2$ then $\mathrm{supp}(u) = \{m \cdot i, m \cdot j\}$ for some integers $i \neq j$ and thus $\langle u, w^{(n/m)} \rangle = 0$.

Second, note that for all $u \in C_\epsilon{}^\perp$ if $|u| \neq 2$ then $|u| \geq \Omega(m) = \Omega(\log(n))$ by construction of C_ϵ. It follows that for every constant weight dual word $u \in C_\epsilon{}^\perp$ we have $\langle u, w^{(n/m)} \rangle = 0$, i.e., w is accepted with probability 1. □

Remark 4.2. It can be readily verified that the same construction demonstrates a code which is not locally decodable (and in particular not self-correctable) by a constant number of queries. This holds because local correction (decoding) can be done only by using a constant weight dual words. So, if there exists a non-codeword w, which is close to the code, but satisfies all constant weight constraints (dual words) the local decoding (correction) is impossible. To obtain such a word pick any codeword $r \in R_\epsilon$, and let $r_{\text{offset}} = 0^{(m-1)}1$ and $w = (r + r_{\text{offset}})^{(n/m)}$. It holds that $\delta(w, C_\epsilon) = 1/m$, but w satisfies all constant weight constraints.

5 Acknowledgement

We thank Tali Kaufman and Swastik Kopparty for helpful discussions.

References

[1] E.Ben-Sasson, V. Guruswami, T.Kaufman, M.Sudan and M.Viderman. Locally testable codes require redundant testers. In Proceedings of CCC 2009.

[2] T.Kaufman and M.Sudan. Random Sparse linear codes are locally testable and decodable. FOCS 2007.

[3] S.Kopparty and S.Saraf. Tolerant linearity testing and local testability. RANDOM 2009.

[4] S.Kopparty and S.Saraf. Local List-Decoding and Testing of Sparse Random Linear Codes from High-Error. Note in ECCC - TR09-115.

[5] Luca Trevisan. Some Applications of Coding Theory in Computational Complexity. Note in ECCC - TR04-043.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de