Theory Group Seminar Notes

Rishit Dagli

October 2022

Contents

Introduction

These are my notes for the seminars that happen in the Theory Group at The University of Toronto. Many thanks to Professor Allan Borodin for allowing me to attend the Theory Group seminars and helping out.

A PDF of these notes is available at https://rishit-dagli.github.io/cs-theory-notes/main.pdf. An online version of these notes are available at https://rishit-dagli.github.io/cs-theory-notes.

The Theory Group focuses on theory of computation. The group is interested in using mathematical techniques to understand the nature of computation and to design and analyze algorithms for important and fundamental problems.

The members of the theory group are all interested, in one way or another, in the limitations of computation: What problems are not feasible to solve on a computer? How can the infeasibility of a problem be used to rigorously construct secure cryptographic protocols? What problems cannot be solved faster using more machines? What are the limits to how fast a particular problem can be solved or how much space is needed to solve it? How do randomness, parallelism, the operations that are allowed, and the need for fault tolerance or security affect this?

1 Lower Bounds for Locally Decodable Codes from Semirandom CSP Refutation

7th October 2022

The related paper: Combinatorial lower bounds for 3-query LDCs by Alrabiah et al. [1]. Seminar by Peter Manohar. [2]

1.1 Abstract

A code C is a q-locally decodable code (q-LDC) if one can recover any chosen bit b_i of the k-bit message b with good confidence by randomly querying the n-bit encoding x on at most q coordinates. Existing constructions of 2-LDCs achieve blocklength $n = exp(O(k))$, and lower bounds show that this is in fact tight. However, when $q = 3$, far less is known: the best constructions have $n = subexp(k)$, while the best known lower bounds, that have stood for nearly two decades, only show a quadratic lower bound of $n \geq \Omega(k^2)$ on the blocklength.

In this talk, we will survey a new approach to prove lower bounds for LDCs using recent advances in refuting semirandom instances of constraint satisfaction problems. These new tools yield, in the 3-query case, a near-cubic lower bound of $n \geq \tilde{\Omega}(k^3)$, improving on prior work by a polynomial factor in k.

1.2 Locally Decodable Codes

Take codes $b \in 0,1^k \to x \in 0,1^n$
Codes x are read by the decoder, $i \in [k]$, $\hat{b_i} \in 0,1$

Definition 1. *C is a (q, δ, ϵ)-locally decodable if for any x with $\triangle(x, Enc(b)) \leq \delta n$, $Dec^x(i) = b_i$ w.p. $\geq \frac{1}{2} + \epsilon$ for any i.*

Ask the question, what is the best possible rate for a q-LDC given a q?

q	Lower Bound	Upper Bound
2	$2^{\Omega(k)} \leq n$	$n \leq 2^k$
3	$k^2 \leq n$	$n \leq exp(k^{o(1)})$
$O(1)$, even	$k^{\frac{q}{q+1}} \leq n$	$n \leq exp(k^{o(1)})$
$O(1)$, odd	$k^{\frac{q+1}{q-1}} \leq n$	$n \leq exp(k^{o(1)})$

Focus on the case $q = 3$, we have gotten better bounds:

$$k \leq n \leq 2^k \tag{1}$$

$$k^2 \leq n \leq exp(exp(\sqrt{\log k \log\log k}))$$

In [1], they show that a better minimum bound can be found than these existing ones for $q = 3$:

$$k^3 \leq n \tag{2}$$

The main result is that:

Theorem 1. *Let C be a $(3, \delta, \epsilon)$-locally decodable codes. Then $n \geq \tilde{\Omega}_{\delta,\epsilon}(k^3)$.*

Semi-random CSP refutation comes to our aid to prove this! The intuitive way to put this theorem is that q-LDC lower bound is same as refuting "LDC" q-XOR.

1.3 How to prove the Theorem

The idea:

- q-LDC lower bound is same as refuting "LDC" q-XOR
  - CSP Refutation
- Proof of existing q-LDC lower bound for q even
- Proof sketch of k^3 lower bound

1.4 Normally Decodable Codes

We can see that the decoder we have can arbitrary but WLOG we can assume there are q-unif hypergraphs $H_1, H_2, \cdots H_k$ where every H_i is such that:

$$H_i \subseteq \binom{[n]}{q}$$

We can also see that:
Each H_i is a matching such that $|H_i| \geq \delta n$
and, $Dec(i)$ picks $C \leftarrow H_i$ and outputs $\sum_{j \in C} x_j$

One such example is the Hadmard code:

$$b \in 0, 1^k \mapsto f = (\langle b, v \rangle)_{v \in 0,1}^k \tag{3}$$

$$b_i = f(e_i) = f(v) + f(v + e_i)$$

Can think of this as v and $v + e_i$ are connected.

Matching vector codes are $\approx \mathbb{Z}_m^h$

1.5 Proof: Going from LDC to XOR

We suppose that our code is linear and that there exists q-unif hypergraphs $H_1, H_2, \cdots H_k$.

We also know that:
Each H_i is a matching such that $|H_i| \geq \delta n$
and, $Dec(i)$ picks $C \leftarrow H_i$ and outputs $\sum_{j \in C} x_j$

So, we start by considering a q-XOR instance ψ_b:

$$\textit{Vars: } \{x_j\}_{j \in [n]}$$
$$\textit{Over Equations: } \sum_{j \in C} x_j = b_i, \forall i \in [k], C \in H_i$$

We can write down the maximum fraction of satisfiable constraints: $val(\psi_b) = 1$ for any $b \in 0, 1^k$.

It is sufficient now if we can argue that ψ_b is unsat with high probability for some random b when $n \ll k^{\frac{q}{q-2}}$.

Now we need to refute XOR, there are many ways to argue unsatisfiability of an XOR instance. One reason why we can not use probablistic approaches here is that ψ_b only has k bits of randomness.

One way we can have some success here is to use a refutation algorithm

$$\psi \to A \to algval(\psi)$$

With this the guarantee then would be $val(\psi) \leq algval(\psi)$ which is similar to saying that if $algval(\psi) < 1$ then A refutes ψ. The ideal goal would be to refute random ψ with m constraints with high probability

However, we take a look at semi-random XOR. Our refutation algorithm and the guarantee will still be the same:

$$\psi \to A \to algval(\psi)$$

with the guarantee that $val(\psi) \leq algval(\psi)$.

So, now we generate semi-random $\psi w/m$ constraints:

- The worst case would be random q-unif hypergraph
- Random RHS b_c for each $C \in H$

The equation we have is:

$$\sum_{j \in C} x_j = b_c \tag{4}$$

And we also already know that

$$\psi_b \ is \sum_{j \in C}$$

And, $xj = b_i, i \in [k], C \in H_i$.
ψ_b is almost semi-random.

Thus, we have shown 1.3 Part 1 of Proof.

1.6 Proof: Existing q-LDC lower bound for q even

q-LDC XOR instance ψ_b is encoded by:

- q-uniform hypergraph matchings $\{H_1 \cdots H_k\}$
- right-hand sides are random $b_i \in \{\pm 1\}$
- We have constraints $\prod_{j \in C} x_j = b_i$ for all i and $C \in H_i$

We now have a goal to argue that ψ_b unsat with high probability for random when b when $n \ll k^{q/(q-2)}$

frac. constraints satisfied by $x \in \{\pm 1\}^n$ is $\frac{1}{2} + \frac{f(x)}{2}$.

Here $f(x)$ is:

$$f(x) = \frac{1}{m} \sum_i b_i \sum_{C \in H_i} \prod_{j \in C} x_j \tag{5}$$

$$m = k \cdot \delta n$$

This makes our goal to be to certify with high probability that:

$$\max_{x \in \{\pm 1\}^n} f(x) < 1 \ when \ n \ll k^{\frac{q}{q-2}} \tag{6}$$

We will now try to refute ψ_b. With Equation 5 and Equation 6 to refute ψ_b is like showing:

$$w.h.p. \quad \max_{x \in \{\pm 1\}^n} f(x) < 1 \ where \ f(x) = \frac{1}{m} \sum_i b_i \sum_{C \in H_i} \prod_{j \in C} x_j \tag{7}$$

$when \ n \ll k^{\frac{q}{q-2}}$.

The idea is to design a matrix $A \in \mathbb{R}^{N \times N}$ so that:

$$f(x) \leq ||A||_{\infty \to 1} = \max_{z,w \in \{\pm 1\}^N} z^T A w$$

As shown by Wein et al. [3] the matrix A can be indexed by

$$S \in \binom{[n]}{l}$$

Assign $x \mapsto y$ such that $y^T A y \propto f(x)$
and $y_s := \prod_{j \in S} x_j$ which is simply the tensor product.

We need to now be able to answer how to set $A(S,T)$

$$y^T A y = \sum_{S,T} y_S y_T A(S,T) = \sum_{S,T} A(S,T) \prod_{j \in S \oplus T} x_j \tag{8}$$

Which shows that we are actually using symmetric difference here.

We say that if $S \oplus T = C \in h_i$ then $\prod_{j \in S \oplus T} x_j = b_i$
$\implies A(S,T) = b_i$ if $S \oplus T = C \in H_i$

$$y^T A y = \sum_{i=1}^{k} b_i \sum_{C \in h_i} \sum_{S \oplus t = C} \prod_{j \in C} x_j = Dmf(x) \tag{9}$$

Here $D =$ number of S, T where $S \oplus T = C$.

Simplifying an earlier statement we can also say from here that: $A_C(S,T) = 1$ if $S \oplus T = C$.

For which $A_i = \sum_{C \in h_i} A_C$ and $A = \sum_{i=1}^{k} b_i A_i$

Set $y_S := \prod_{j \in S} x_j$

$y^T A y = Dmf(x) \implies Dmf(x) \leq ||A||_{\infty \to 1}$

Note that the way we defined D here it only depends on $|C| = q$, we can say:

$$D = \binom{q}{\frac{q}{2}} \binom{n-q}{l-\frac{q}{2}}$$

Also we know $A_c \in \mathbb{R}^{N \times N}$ and $N = \binom{n}{l}$.

We have already proven that $||A||_{\infty \to 1} \geq Dm \max_x f(x) \geq Dm \geq D\delta nk$

It is also interesting to note that $||A||_{\infty\to 1} \leq N||A||_2$ and we still need to be able to show that with high probability that $||A||_{\infty\to 1}$ is not too large.

References

[1] Omar Alrabiah, Venkatesan Guruswami, Pravesh Kothari, and Peter Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom CSP refutation. Technical Report TR22-101, Electronic Colloquium on Computational Complexity (ECCC), July 2022.

[2] Arnab Bhattacharyya, L. Sunil Chandran, and Suprovat Ghoshal. Combinatorial lower bounds for 3-query ldcs, 2019. URL https://arxiv.org/abs/1911.10698.

[3] Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. The kikuchi hierarchy and tensor pca, 2019. URL https://arxiv.org/abs/1904.03858.