Rate amplification and query-efficient distance amplification for linear LCC and LDC

Gil Cohen, Tal Yankovitz

Tel Aviv University

June 15, 2021

Benny Chor

To the memory of Benny Chor

Overview

Error correcting codes

Definition (Error correcting code)

A set $C \subseteq \Sigma^n$ s.t.

$$\text{dist}(x, y) = |\{i \in [n] : x_i \neq y_i\}| \geq \delta n$$

$\forall x \neq y \in C$ is a code with distance δ.

- We focus on linear codes in which $\Sigma = \mathbb{F}$ is a finite field and C an $\mathbb{F}$-vector space. In such case, the rate of C is given by

$$\rho = \frac{1}{n} \dim C.$$

- A linear code can be defined as a linear map $c : \mathbb{F}^k \to \mathbb{F}^n$ s.t. $C = c(\mathbb{F}^k)$.

Error correcting codes

- A family of codes $\{c : \mathbb{F}^k \to \mathbb{F}^{n(k)}\}_k$ is asymptotically good if both $\delta, \rho = \Omega(1)$.
- Good codes exist.
- A code can be used for communication over imperfect channels; Instead of sending a message $m \in \mathbb{F}^k$ one sends $c(m)$. If the fraction of errors $< \frac{\delta}{2}$, the receiver can infer m.

What if we are only interested in m_i for some $i \in [k]$?

Locally decodable codes

Definition (Locally decodable codes; Katz-Trevisan 2000)

A code $c : \mathbb{F}^k \to \mathbb{F}^n$ is a q-query locally decodable code (LDC) if there is a randomized algorithm

$$D : \mathbb{F}^n \times [k] \to \mathbb{F}$$

(called a decoder) s.t. $\forall i \in [k]$, $m \in \mathbb{F}^k, r \in \mathbb{F}^n$,

$$\mathrm{dist}(r, c(m)) \leq \delta n \quad \Longrightarrow \quad \Pr[D(r, i) = m_i] \geq \frac{2}{3}.$$

Further, the number of queries made by D to r is bounded by q.

Locally correctable codes

Definition (Locally correctable codes)

A code $C \subseteq \mathbb{F}^n$ is a q-query locally correctable code (LCC) if there is a randomized algorithm

$$D : \mathbb{F}^n \times [n] \to \mathbb{F}$$

(called a corrector) s.t. $\forall i \in [n]$, $c \in C$, $r \in \mathbb{F}^n$,

$$\text{dist}(c, r) \leq \delta n \quad \Longrightarrow \quad \Pr[D(r, i) = c_i] \geq \frac{2}{3}.$$

Further, the number of queries made by D to r is at most q.

We will focus on non-adaptive LC in which the choice of which indices the query is done prior to the actual querying.

What is known?

- Much work has been done on the constant query regime.
- More recently, good LC were considered, where we wish to minimize the query complexity.
- Katz-Trevisan proved $q = \Omega(\log n)$ for LDC.
- Linear LCC $\implies$ linear LDC.
- LC need structure - random won't do.

What is the lowest query complexity $q = q(n)$ of good codes?

Explicit constructions

- Reed-Muller codes give $q = n^{\varepsilon}$ for any desired constant ε.
- For a while, no (nontrivial) LC were known with $\rho > \frac{1}{2}$.
- The first construction to break the "rate $\frac{1}{2}$ barrier" is via multiplicity codes (Kopparty-Saraf-Yekhanin 2010).
- Alternative constructions were obtained using expanders (Hemenway-Rafail-Ostrovsky-Wootters 2013) and using lifted RM (Guo-Kopparty-Saraf 2013). All require $q = n^{\varepsilon}$.

Explicit constructions

- The state of the art construction of LC (Kopparty-Meir-Ron-Zewi-Saraf 2016) achieves

$$q = 2^{\widetilde{O}(\sqrt{\log n})} = n^{o(1)}.$$

- Spoiler: In this work we don't improve upon this great result (nor show it is tight).

This work's theme

bad LC $\implies$ good LC.

Overview

Reed-Muller Codes

As an example, we describe an LCC that works assuming no errors, having query complexity $\sqrt{n}$.

Let $\mathbb{F}_q$ be finite field. Define

$$C = \{f(\mathbb{F}_q^2) \mid f \in \mathbb{F}_q^{<q-1}[x, y]\}.$$

We have $n = q^2$ and

$$k = \frac{q(q-1)}{2} \approx \frac{n}{2}.$$

Hence, $\rho = \frac{1}{2} - o(1)$.

Reed-Muller Codes

The corrector. Given a desired index $p = (a, b) \in \mathbb{F}_q^2$, sample a random line through p,

$$L = \{(At + a, Bt + b) \mid t \in \mathbb{F}_q\},$$

and query on $L \setminus p$.

Analysis. Set $g(t) = f(At + a, Bt + b)$. As $\deg g < q - 1$, the $q - 1$ points we query determine g. Thus, we can compute

$$g(0) = f(a, b) = f(p).$$

Reed-Muller Codes

- As Nati pointed out, if we assume no errors then we have a trivial corrector with $q = 1$ queries.
- The above LCC however has the advantage of being smooth: The marginal distribution of every query is uniform over $\mathbb{F}_q^2 \setminus p$.
- Observe that a smooth LCC has distance $\Omega(\frac{1}{q}) = \Omega(\frac{1}{\sqrt{n}})$.
- In any case, to get a good code, one can restrict to degree αq for a constant $\alpha < 1$.

Multiplicity codes

A first example of multiplicity codes is given by

$$(\mathbb{F}_q^3)^n \supseteq C = \{\left(f(\mathbb{F}_q^2), f_x(\mathbb{F}_q^2), f_y(\mathbb{F}_q^2)\right) \mid f \in \mathbb{F}_q^{<d}[x,y]\}$$

with $d = 2q-2$. The rate is then $\rho \approx \frac{2}{3}$.

Corrector & Analysis.

- Pass three random lines through p. Each gives a linear relation between $f(p), f_x(p), f_y(p)$.
- W.h.p the relations are linearly independent.
- Solve a system of linear equations to get $(f(p), f_x(p), f_y(p))$.

One can consider higher dimension and higher order derivatives.

KMRZS and the AEL distance amplification procedure

The KMRZS construction consists of two steps:

1. Use a multiplicity code with $\delta = \frac{1}{\text{poly}\log n}$ and $q = 2^{\widetilde{O}(\sqrt{\log n})}$.
2. Amplify the distance to a constant.

For (2), the AEL distance amplification procedure is invoked (Alon-Edmonds-Luby 1995) that converts a q-query LC with distance δ to a

$$q_{\text{new}} = q \cdot \text{poly}\frac{1}{\delta}$$

query LC with constant distance.

Overview

Query-efficient distance amplification

$$\text{bad LC} \implies \text{good LC}.$$

Theorem (Query-efficient distance amplification)

One can efficiently transform a q-query LDC with distance δ to a constant distance LDC with query complexity

$$q_{\text{new}} = q \cdot \log(n) \cdot q\left(\frac{1}{\delta}\right),$$

where $q(b)$ is the query complexity of a good LDC on message length b.

Query-efficient distance amplification

Corollary

For any constant $\alpha < 1$ a q-query LDC with distance $\delta = \frac{1}{n^\alpha}$ can be transformed to a constant distance LDC with query complexity

$$q_{\text{new}} = q^{O(\log \log n)}.$$

As another corollary, one can amplify

$$\delta = 2^{-(\log n)^\alpha} \quad \Rightarrow \quad q_{\text{new}} = q^{O(\log \log \log n)}$$

Rate amplification

Our second contribution is a rate amplification procedure.

Theorem (Rate amplification for LCC)

Let $\rho_{\text{desired}} < 1$ constant. Let C be a constant distance q-query LCC with rate ρ. One can transform C to a good LCC with $\rho \geq \rho_{\text{desired}}$ and query complexity

$$q_{\text{new}} = (q \cdot \log n)^{\text{poly}\left(\frac{1}{\rho}\right)}.$$

We will discuss only this result in the remaining of this talk.

Overview

A dual view on Reed-Muller

Claim

If $g \in \mathbb{F}_q[t]$ *is non constant and* $\deg g < q - 1$ *then*

$$\sum_{t \in \mathbb{F}_q} g(t) = 0.$$

Recall that

$$L_{a,b,A,B} = \{(At + a, Bt + b) \mid t \in \mathbb{F}_q\}.$$

If we assume $A \neq 0$ it suffices to consider

$$L_{a,b} = \{(t, at + b) \mid t \in \mathbb{F}_q\}.$$

A dual view on Reed-Muller

Let

$$\mathbf{1}_{a,b}(p) = \begin{cases} 1, & p \in L_{a,b} \\ 0, & \text{otherwise.} \end{cases}$$

We can write

$$\mathbf{1}_{a,b}(x,y) = 1 - (y - (ax+b))^{q-1}.$$

We have that

$$(y - (ax+b))^{q-1} = \sum_{k=0}^{q-1} \sum_{j=0}^{k} (-1)^k \binom{q-1}{k} \binom{k}{j} a^j b^{k-j} x^j y^{q-1-k}.$$

A dual view on Reed-Muller

$$(y-(ax+b))^{q-1} = \sum_{k=0}^{q-1}\sum_{j=0}^{k} c_{k,j}\cdot a^j b^{k-j}\cdot x^j y^{q-1-k}.$$

Observe that having $\{x^j y^i \mid i+j \leq q-1\}$ in $C^\perp$ would suffice for the RM corrector. Now,

$$\dim C^\perp = \frac{q(q+1)}{2}$$

implying

$$\dim C = q^2 - \dim C^\perp = \frac{q(q-1)}{2}.$$

Thence, $C^\perp$ is the dual of RM.

Overview

Structure theorem for LCC

In the dual view on RM, we have

1. The point set $P = \mathbb{F}_q^2$.
2. With every $p \in P$ we associated a distribution D_p over $\mathbb{F}_q^P$ s.t.

$$\Pr_{f \sim D_p}[f(p) = 0] = 0.$$

   Moreover, $|f| = |\{p' \mid f(p') \neq 0\}| \leq q$.
3. $\forall p \neq p' \in P$

$$\Pr_{f \sim D_p}[f(p') \neq 0] \leq \frac{q-1}{q^2-1} = \frac{1}{q+1}.$$

4. $\dim \mathrm{Span} L \leq \frac{q(q+1)}{2}$ where $L = \bigcup_{p \in P} \mathrm{supp}(D_p)$.

Structure theorem for LCC

Definition (Dual LCC)

A (q, τ, ρ)-Dual LCC is a family of distributions $\mathcal{D} = \{D_p\}_{p \in P}$ s.t.

1. $\forall p \in P$
$$\Pr_{f \sim D_p}[f(p) = 0] = 0.$$
2. $\forall f \in L \quad |f| \leq q$, where $L = \bigcup_{p \in P} \mathrm{supp}(D_p)$
3. $\forall p \neq p' \in P$,
$$\Pr_{f \sim D_p}[f(p') \neq 0] \leq \tau.$$
4. $\dim \mathrm{Span} L \leq (1 - \rho)|P|$.

Structure theorem for LCC

Definition (Dual LCC)

A (q, τ, ρ)-Dual LCC is a family of distributions $\mathcal{D} = \{D_p\}_{p \in P}$ s.t.

1. $\Pr_{f \sim D_p}[f(p) = 0] = 0$.
2. $|f| \leq q$.
3. $\Pr_{f \sim D_p}[f(p') \neq 0] \leq \tau$.
4. $\dim \mathrm{Span} L \leq (1 - \rho)|P|$.

We generalize the definition and allow L to contain further elements other than $L_0 = \cup_p \mathrm{supp} D_p$.

Theorem (Informal)

C is a q-query linear LCC with distance d $\iff$ $C^\perp$ is a $(q+1, \tau = \frac{q}{d}, \rho = \frac{1}{n} \dim C)$ dual LCC.

Overview

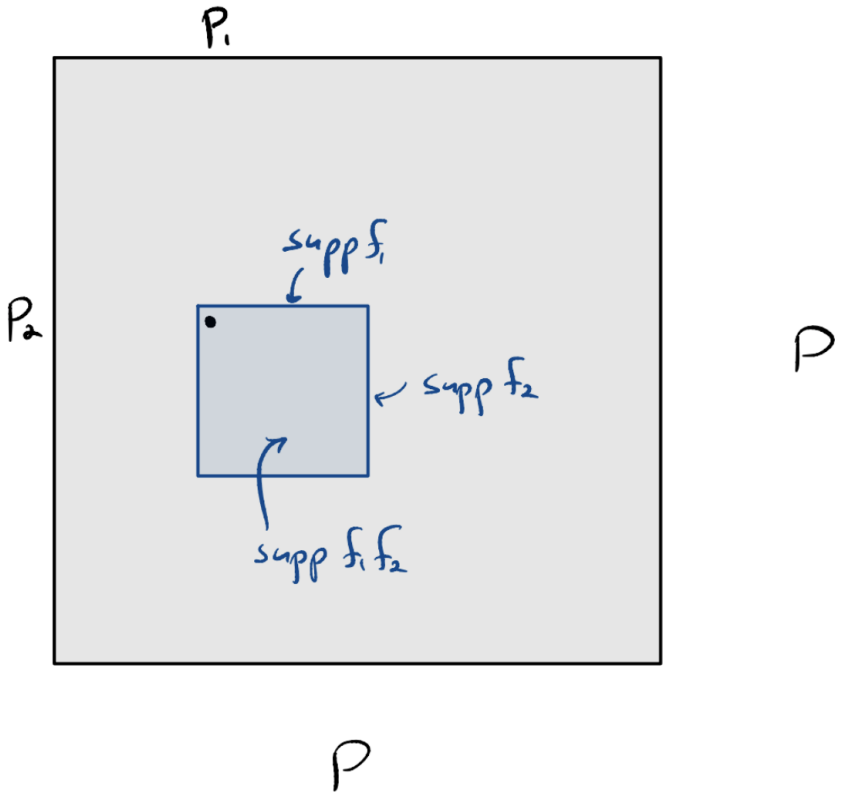
Rate amplification for LCC

The natural way of amplifying the rate of an LCC is to tensor the corresponding dual LCC with itself.

Given a (q, τ, ρ)-Dual LCC $(L, \mathcal{D})$ with $L \subseteq \mathbb{F}^P$ we define $(L', \mathcal{D}')$ as follows:

- The point set is P^2.
- $\mathcal{D}' = \{D'_{(p_1,p_2)} \mid (p_1, p_2) \in P^2\}$ where to sample $f \sim D'_{p_1,p_2}$ we do the following:
  1. Sample $f_1 \sim D_{p_1}$, $f_2 \sim D_{p_2}$ independently.
  2. Output $f_1 f_2$ that is defined by $f_1 f_2(p'_1, p'_2) = f_1(p'_1) f_2(p'_2)$.
- $L' \subseteq \mathbb{F}^{P^2}$ is the smallest subspace containing all functions supported by $\mathcal{D}'$.

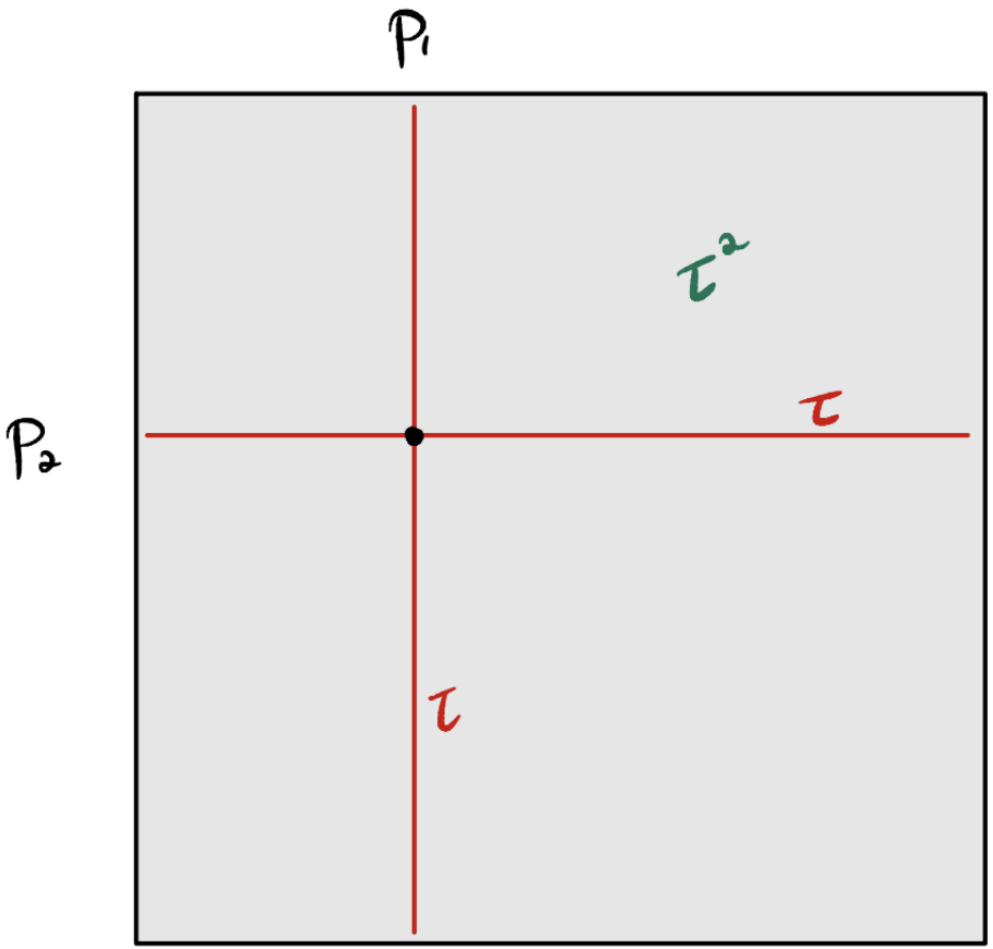
Rate amplification for LCC



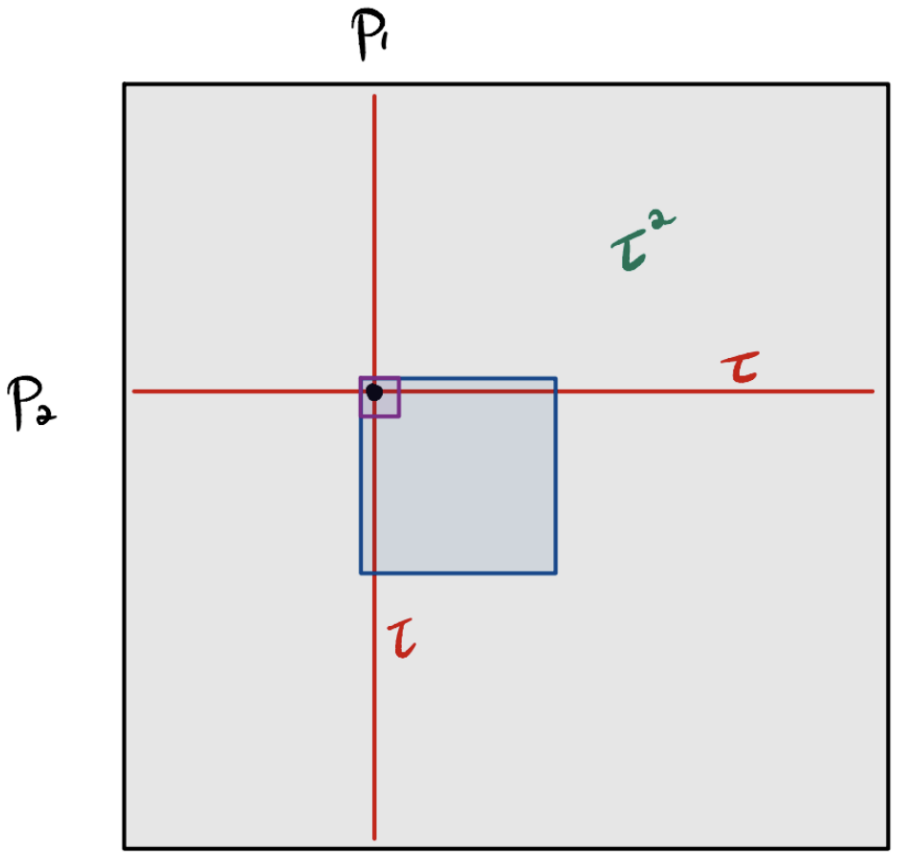
Claim

$(L', \mathcal{D}')$ *is a* (q^2, τ, ρ')*-Dual LCC with* $\rho' = 2\rho - \rho^2$.

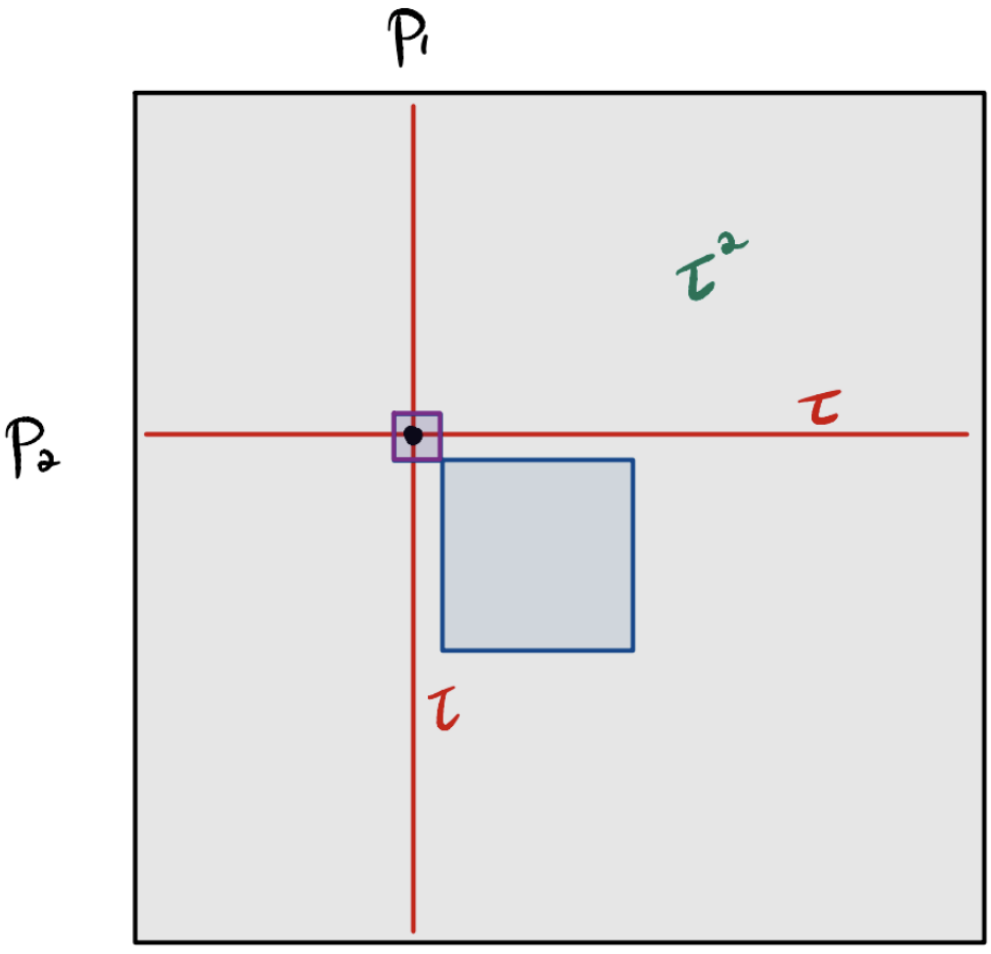
Rate amplification for LCC



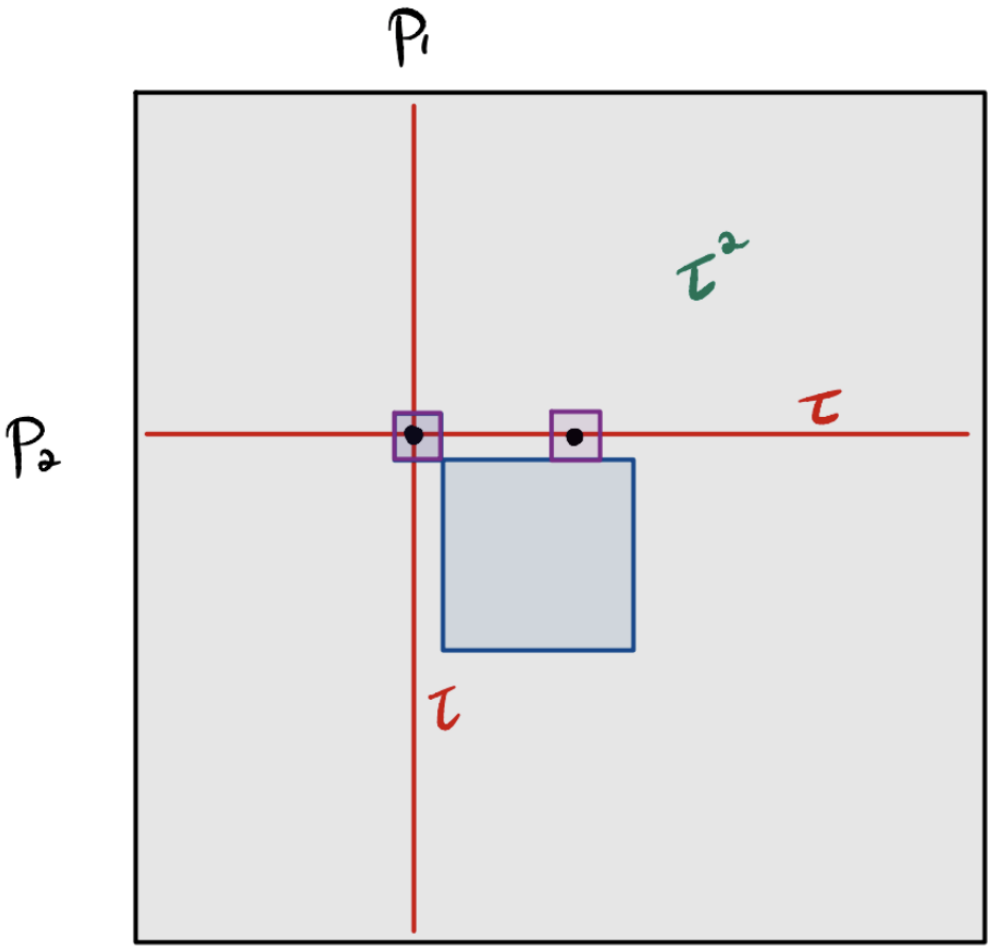
Rate amplification for LCC



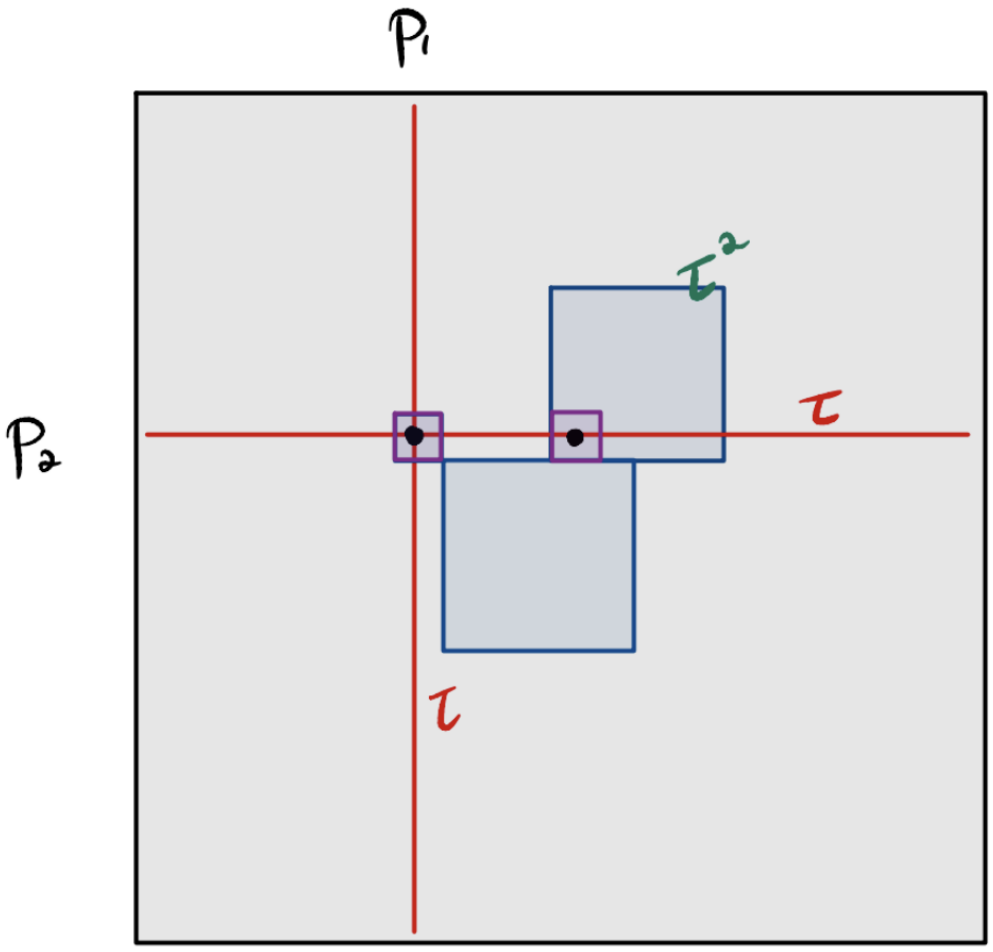
Rate amplification for LCC



Rate amplification for LCC

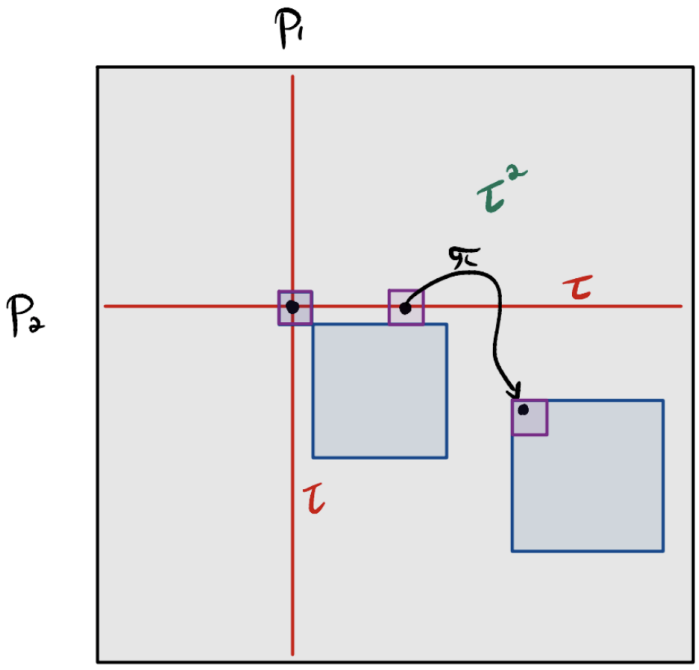


Rate amplification for LCC



Rate amplification for LCC

Let $\pi : P^2 \to P^2$ be a matching. Adjoin $\{e_p + e_{\pi(p)}\}_{p \in P^2}$ to L.



π should be "axis evasive": If $\mathcal{X}_p$ is the set of points of distance 1 from p then

$$\forall p \neq p' \in P^2 \quad |\pi(\mathcal{X}_p) \cap \mathcal{X}_{p'}| \leq s.$$

Rate amplification for LCC

More generally, one can consider a partition π of P^2 s.t.

1. Each part has size c.
2. $\forall p \neq p' \in P^2 \quad |\pi(\mathcal{X}_p) \cap \mathcal{X}_{p'}| \leq s$.

Using such a partition

$$\rho_{\text{new}} = 2\rho - \rho^2 - \frac{1}{c}$$
$$q_{\text{new}} = O(cq^3)$$
$$\tau_{\text{new}} = O(csq \cdot \tau^2).$$

Taking $c = o(\frac{1}{\rho})$ and applying this for $\approx \log \frac{1}{\rho}$ times yields a constant rate.

Axis-evasive partitions

- A probabilistic argument only gives axis-evasive partitions for $c = \Omega(\log |P|)$, with $s = c$.
- We give an explicit construction for certain values of $c, |P|$ with $s = O(c^2)$.
- More precisely, we require $|P| = q$ to be a prime power and $c + 1 \mid q^2 - 1$.

Axis-evasive partitions

- Identify $\mathbb{F}_{q^2}$ with $\mathbb{F}_q^2$.
- Fix $\alpha \in \mathbb{F}_{q^2} \setminus \mathbb{F}_q$.
- Let $\beta \in \mathbb{F}_{q^2}$ be an element of order $c+1$ in $\mathbb{F}_{q^2}^\times$.
- The partition is given by the quotient group $\mathbb{F}_{q^2}^\times / \langle \beta \rangle$.
- There are some technical conditions on α, β that can always be met.

Summary

This work was about

$$\text{Bad LC} \implies \text{good LC}$$

Future research

1. Construct a bad (but not so bad) LC.
2. Can it be that the current $2^{\widetilde{O}(\sqrt{\log n})}$ bound is tight?
3. Devise an improved distance amplification procedure for LCC.
4. Improve our rate amplification procedure.
5. Amplify rate under more general conditions.

Thank you!