PCP-free Inapproximability of Nearest Codeword and Minimum Distance

Xuandi Ren

UC Berkeley

Based on a joint work with

Vijay Bhattiprolu, *Venkat Guruswami*, and *Euiwoong Lee*

Problem Definitions

- **Nearest Codeword Problem**
  - Input: a linear code $C \subseteq \mathbb{F}_q^n$, and a vector $b \in \mathbb{F}_q^n$.
  - Output: the minimum distance from b to any codeword in C, i.e., $\min_{c \in C} |b - c|_0$.

Problem Definitions

- **Nearest Codeword Problem (An Equivalent View)**
  - Input: an affine subspace $\mathrm{V} \subseteq \mathbb{F}_q^n$.
  - Output: the minimum Hamming weight of a vector $x \in V$.

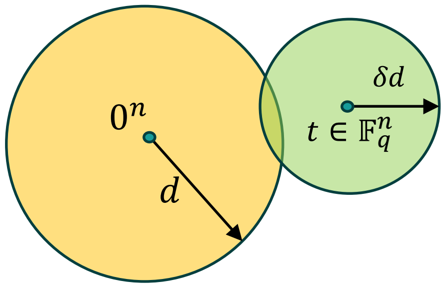
Problem Definitions

- **Nearest Codeword Problem (An Equivalent View)**
  - Input: an affine subspace $\mathrm{V} \subseteq \mathbb{F}_q^n$.
  - Output: the minimum Hamming weight of a vector $x \in V$.

- **Minimum Distance Problem**
  - Input: a linear code (a.k.a. subspace) $C \subseteq \mathbb{F}_q^n$.
  - Output: the minimum Hamming weight of a non-zero codeword in C.

Hardness of NCP and MDP

- **NP-hardness of NCP**:
  - reducing from ExactCover.
- **NP-hardness of MDP**:
  - [Vardy'97], using as a gadget a Reed-Solomon code concatenated with Hadamard code.



- **NP-hardness of approximating NCP**:
  - still from ExactCover, a direct corollary of PCP theorem
- **NP-hardness of approximating MDP:**
  - [Dumer-Micciancio-Sudan'03], a randomized reduction, using *locally dense code* as a gadget
  - derandomized by [Cheng-Wan'12, Austrin-Khot'14, Micciancio'14]

Our Results

- A simple deterministic reduction showing the inapproximability of NCP/MDP within any constant factors assuming NP≠P

- PCP-free
- Deterministic
- Homogenization in a reverse way – MDP -> NCP

Proof Overview

- Starting Point:
  - satisfiability of a system of homogeneous quadratic equations
- Key tools:
  - bound on the 2^{nd} generalized Hamming weight of any code
  - rank-1 testing of a matrix in a tensor code, via Hamming weight
  - an ε-balanced code
- Idea:
  - rewrite QuadEQ as rank-1 testing
  - use ε-balanced code to ensure in the (YES) case we have low weight
  - use the bound on 2^{nd} generalized Hamming weight to argue in the (NO) case, every solution has high weight

2^{nd} Generalized Hamming Weight

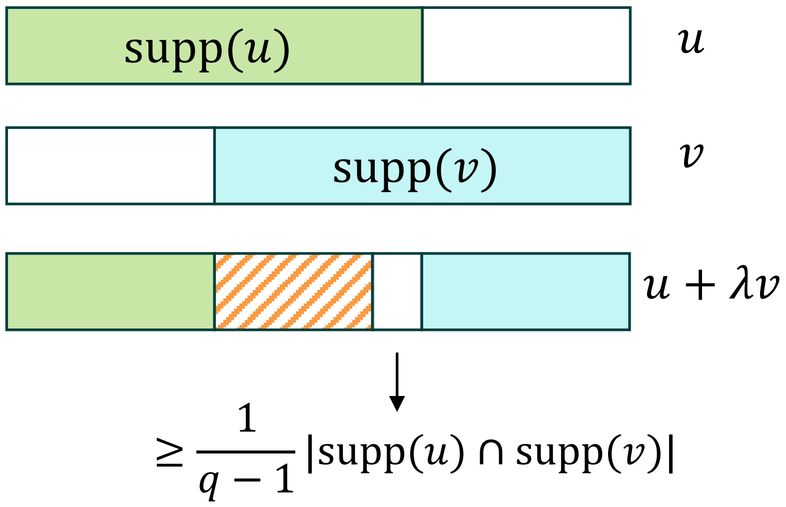
- For any linear code $C \subseteq \mathbb{F}_q^n$ with distance $d(C)$, the ***2^{nd} generalized Hamming weight***, written as $d_2(C)$, is defined as the minimum of
$$|\mathrm{supp}(u) \cup \mathrm{supp}(v)|,$$
for any linearly independent codewords $u, v \in C$.

2^{nd} Generalized Hamming Weight

- For any linear code $C \subseteq \mathbb{F}_q^n$ with distance $d(C)$, the ***2^{nd} generalized Hamming weight***, written as $d_2(C)$, is defined as the minimum of
$$|\mathrm{supp}(u) \cup \mathrm{supp}(v)|,$$
for any linearly independent codewords $u, v \in C$.

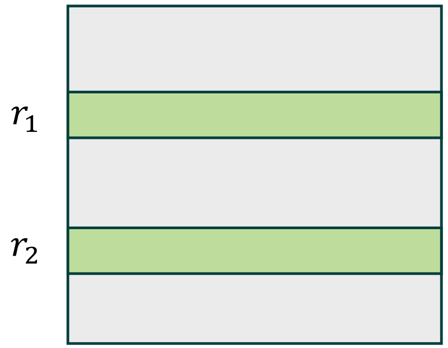
- Fact: $d_2(C) \geq \left(1 + \frac{1}{q}\right) d(C)$.

- Proof:
  - Since C is a linear code, $\forall \lambda \in \mathbb{F}_q$, $u + \lambda v$ also belongs to C
  - The sparsest vector $u + \lambda v$ among all choices of λ has weight $\leq$
  $$|u|_0 + |v|_0 - \left(1 + \frac{1}{q-1}\right) |\mathrm{supp}(u) \cap \mathrm{supp}(v)|$$
  by pigeonhole, but this should be $\geq d(C)$
  - Rearranging gives desired bound



Rank-1 Testing via Hamming Weights

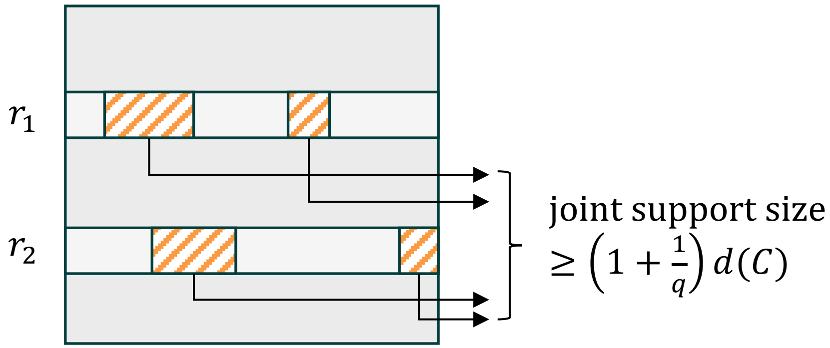
- For any linear code $C \subseteq \mathbb{F}_q^n$ with distance $d(C)$, consider the tensor code $C \otimes C$, we have:
  - (1) Some rank-1 matrix $M \in C \otimes C$ achieves $|M|_0 = d(C)^2$.
  - (2) Every rank-($\geq$2) matrix $M \in C \otimes C$ has $|M|_0 \geq \left(1 + \frac{1}{q}\right) d(C)^2$.



- Proof of (2):
  - Any matrix M of rank ≥ 2 has two linearly independent rows r_1, r_2

Rank-1 Testing via Hamming Weights

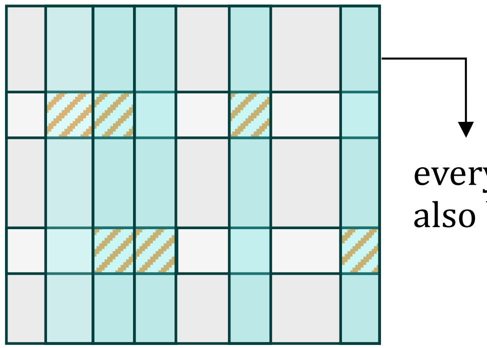
- For any linear code $C \subseteq \mathbb{F}_q^n$ with distance $d(C)$, consider the tensor code $C \otimes C$, we have:
  - (1) Some rank-1 matrix $M \in C \otimes C$ achieves $|M|_0 = d(C)^2$.
  - (2) Every rank-($\geq$2) matrix $M \in C \otimes C$ has $|M|_0 \geq \left(1 + \frac{1}{q}\right) d(C)^2$.



- Proof of (2):
  - Any matrix M of rank ≥ 2 has two linearly independent rows r_1, r_2
  - By the bound on 2^{nd} generalized Hamming weight, r_1, r_2 have joint support size $\geq \left(1 + \frac{1}{q}\right) d(C)$

Rank-1 Testing via Hamming Weights

- For any linear code $C \subseteq \mathbb{F}_q^n$ with distance $d(C)$, consider the tensor code $C \otimes C$, we have:
  - (1) Some rank-1 matrix $M \in C \otimes C$ achieves $|M|_0 = d(C)^2$.
  - (2) Every rank-($\geq$2) matrix $M \in C \otimes C$ has $|M|_0 \geq \left(1 + \frac{1}{q}\right) d(C)^2$.



- Proof of (2):
  - Any matrix M of rank ≥ 2 has two linearly independent rows r_1, r_2
  - By the bound on 2^{nd} generalized Hamming weight, r_1, r_2 have joint support size $\geq \left(1 + \frac{1}{q}\right) d(C)$
  - For each column c in their joint support, c has weight $\geq d(C)$ since $M \in C \otimes C$

ε-Balanced Codes

- A linear code $C \subseteq \mathbb{F}_q^n$ with distance $d(C)$ is said to be ε-balanced if the Hamming weight of any non-zero codeword is in $[d(C), (1+\varepsilon)d(C)]$.

- A construction of ε-balanced code over $\mathbb{F}_q$:
  - Take Reed-Solomon code with degree εn (it has distance $(1-\varepsilon)n$ and is thus ε-balanced)
  - Concatenate it with Hadamard code over $\mathbb{F}_q$

NP-hardness of QuadEQ

- **(Non-homogeneous) QuadEQ:**
  - Input: a system of m quadratic equations on n variables $\{x_1, \dots, x_n\}$ over $\mathbb{F}_q$:

$$\left\{\sum_{i,j\in[n]} A_{i,j}^{(t)} x_i x_j = b^{(t)}\right\}_{t\in[m]}$$

  - Output: whether there is a solution $\{x_1, \dots, x_n\}$

- NP-hardness:
  - Reduce from Circuit Satisfiability
  - Add an equation $x_i(x_i - 1) = 0$ for each variable to ensure it takes Boolean value
  - Add an equation constraining each gate's computation (e.g. for an AND gate $y_k = y_i \wedge y_j$, add an equation $x_k^2 = x_i x_j$)

NP-hardness of QuadEQ

- **(Homogeneous) QuadEQ:**
  - Input: a system of m quadratic equations on n variables $\{x_1, \dots, x_n\}$ over $\mathbb{F}_q$:

$$\left\{\sum_{i,j\in[n]} A_{i,j}^{(t)} x_i x_j = 0\right\}_{t\in[m]}$$

  - Output: whether there is a *non-zero* solution $\{x_1, \dots, x_n\}$

- NP-hardness:
  - Reduce from non-homogeneous version
  - Add a variable z, replacing the constant 1
  - Add an equation $x_i(x_i - z) = 0$ for each variable, to ensure it takes either 0 or z
  - If z takes 0 in some solution, then every x_i also has to take 0, and this is the all-0 solution
  - Otherwise, $\{x_i z^{-1}\}_{i\in[n]}$ is a solution to the non-homogeneous system

Reducing Homo-QuadEQ to gap MDP

- Take a Homo-QuadEQ instance $\left(n, m, \left\{A_{i,j}^{(t)}\right\}_{i,j\in[n],t\in[m]}\right)$.
- Let $G \in \mathbb{F}_q^{N\times n}$ be the generating matrix of an ε-balanced code C with distance $d(C)$.

- The output MDP instance:
  - the subspace of matrices

$$GXG^T, X \in \mathbb{F}_q^{n\times n},$$

    with constraints
  - $X^T = X$
  - $\forall t \in [m], \sum_{i,j\in[n]} A_{i,j}^{(t)} X_{i,j} = 0$

Reducing Homo-QuadEQ to gap MDP

- $V = \{GXG^T \mid X \in \mathbb{F}_q^{n\times n}, X^T = X, \forall t \in [m], \sum_{i,j\in[n]} A_{i,j}^{(t)} X_{i,j} = 0\}$

- (Completeness)
  - Let $x \in \mathbb{F}_q^n$ be the non-zero solution of Homo-QuadEQ, we take $X = xx^T$.
  - $GXG^T = (Gx)(Gx)^T$, which has weight $(1+\varepsilon)^2 d(C)^2$ by the ε-balanced property of C.

Reducing Homo-QuadEQ to gap MDP

- $V = \{GXG^T \mid X \in \mathbb{F}_q^{n\times n}, X^T = X, \forall t \in [m], \sum_{i,j\in[n]} A_{i,j}^{(t)} X_{i,j} = 0\}$

- (Soundness)
  - Suppose Homo-QuadEQ has no non-zero solution, we argue $d(V) \geq \left(1 + \frac{1}{q}\right) d(C)^2$.
  - If X is rank-1, then $X = xx^T$ for some non-zero solution x of Homo-QuadEQ.
  - If X has rank ≥ 2, then GXG^T is a matrix in $C \otimes C$ with rank ≥ 2. $|GXG^T|_0 \geq \left(1 + \frac{1}{q}\right) d(C)^2$

Corollaries

- Simple tensoring amplifies the inapproximability ratio of MDP to
  - any constant assuming NP≠P;
  - $2^{\log^{1-\varepsilon} n}$ assuming $\mathrm{NP} \not\subseteq \mathrm{DTIME}(2^{\log^{O(1)} n})$
  - $n^{c/\log\log n}$ for some fixed c, assuming $\mathrm{NP} \not\subseteq \cap_{\delta>0} \mathrm{DTIME}(2^{n^{\delta}})$

- Same inapproximability ratios for NCP:
  - the same reduction, but from Non-Homo-QuadEQ, to get a mild constant gap
  - then amplify the gap, which is a bit non-trivial
- An alternative way to get NCP:
  - in our (YES) case of MDP, there is a coordinate which always takes 1
  - don't need to worry about all-0 solutions

Thanks!