FPT-Inapproximability of Minimum Codeword Problem over Large Fields

Bingkai Lin
National Institute of Informatics
Parameterized Approximation Algorithms Workshop

This talk is based on Section 6 of the paper Fixed-parameter Approximability of Boolean MinCSPs, Édouard Bonnet, László Egri, Bingkai Lin, Dániel Marx, https://arxiv.org/abs/1601.04935
Minimum Distance Codeword Problem

Notations:
\[ d(v) := \#\{v_i \neq 0 : i \in [m]\} \] is the distance of \( v = (v_1, ..., v_m) \)
\[ \langle w_1, ..., w_r \rangle := \{ \sum_{1 \leq i \leq r} c_i w_i : c_i \in F_q \} \] is the linear space generated by \( w_1, ..., w_r \).

Min Distance Codeword Problem (MDP) over \( F_q \):

**Input:** a set \( W = \{w_1, ..., w_n\} \) of vectors in \( F_q^m \), \( m = n^{O(1)} \) and \( k \)

**Question:** decide if there is a nonzero vector in \( \langle w_1, ..., w_n \rangle \) with distance \( \leq k \).

**k-Even-Set = MDP** over binary field (parameterized by \( k \))
- Does **k-EvenSet** have \( f(k)\text{poly}(n) \) time (FPT) algorithm?
- Long standing open problem! (see http://fptschool.mimuw.edu.pl/opl.pdf)
Outline

- FPT-inapproximability of MDP over large fields
  - From **Min Linear Dependent Set** to MDP
  - From **Biclique** to **Min-Linear-Dependent-Set**
- Reduce Field Size
  - Naïve approach
  - Via Nearest Codeword Problem (**NCP**) 
    Combine with the work of [Arnab et al. ICALP '18]
- Conclusion
FPT-inapproximability of MDP over large fields
Gap-MDP

Gap-MDP\((k,k')\) over \(F_q\) :

**Input:** a set \(W=\{w_1,\ldots,w_n\}\) of vectors in \(F_q^m (m=n^{O(1)})\) and integers \(k'>k\).

**Parameter:** \(k\)

**Question:** distinguish between the following cases:

- (yes) \(<w_1,\ldots,w_n>\) has a nonzero vector with distance at most \(k\).
- (no) Every nonzero vector in \(<w_1,\ldots,w_n>\) has distance at least \(k'\).

The input instance is promised to be in one of these cases.

Polynomial time intractability of **Gap-MDP\((k,ck)\) over \(F_2\) for \(c>1\) (the gap \(c>1\) can be amplified to any constant factor)

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Deterministic?</th>
<th>Elementary?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dumer, Micciancio, and Sudan FOCS’99</td>
<td>RP≠NP</td>
<td>randomized</td>
</tr>
<tr>
<td>Cheng and Wan STOC’09</td>
<td>P≠NP</td>
<td>deterministic</td>
</tr>
<tr>
<td>Austrin and Khot ICALP’13</td>
<td>P≠NP</td>
<td>deterministic</td>
</tr>
</tbody>
</table>
Main Result

Main result: 
Gap-MDP\((k,k\log k)\) over \(F_{\text{poly}}(n)\) is not FPT, assuming FPT\(\neq W[1]\).

Advantage:
• Proof is simple
• Refute super-polynomial time algorithm

Disadvantage:
• Use strong assumption FPT\(\neq W[1]\).
• Large field size.
The Reduction

**Technique**: use One-Side-Gap-Biclique as reduction source

**Two steps:**

- From Gap-Linear-Dependent-Set to Gap-MDP
- From One-Side-Gap-Biclique to Gap-Linear-Dependent-Set
From Gap-Linear-Dependent-Set to Gap-MDP
Gap Linear Dependent Set

Vectors $w_1,...,w_r$ are **linearly dependent** if there exist not all zero $c_1,...,c_r$ s.t.
$$
\sum_{1 \leq i \leq r} c_i w_i = (0,0,...,0)=0
$$

Gap-Linear-Dependent-Set($k,k'$):

**Input:** a set $W=\{w_1,...,w_n\}$ of vectors in $F_q^m$, $m=n^{O(1)}, k'>k$.

**Parameter:** $k$

**Question:** distinguish between the following cases:
- (yes) $W$ contains $k$ linearly dependent vectors.
- (no) any $k'$ vectors in $W$ are not linearly dependent.
From Linear Dependent Set to MDP

Let \( W = \{w_1, ..., w_n\} \) be an instance of **Gap-Linear-Dependent-Set\((k,k')\)**

Choose an integer \( Q > k' \).

For each vector \( w \in W \), construct its corresponding vector \( w' \)

\[
\begin{align*}
    w'_1 &= (w_1, w_1, ..., w_1, 1, 0, ..., 0) \\
    w'_2 &= (w_2, w_2, ..., w_2, 0, 1, ..., 0) \\
    w'_n &= (w_n, w_n, ..., w_n, 0, 0, ..., 1)
\end{align*}
\]

- (yes) If \( W \) has \( k \) linearly dependent vectors, then \( \langle w'_1, ..., w'_n \rangle \) has \( k \)-distance vector.
- (no) If \( \langle w'_1, ..., w'_n \rangle \) contains a \( k' \)-distance vector, then \( W \) has \( k' \) linearly dependent vectors.[**must cancel out the Q copies part**]

**Gap-Linear-Dependent-Set\((k,k')\)** is not FPT => **Gap-MDP\((k,k')\)** is not FPT
From **One-Side-Gap-Biclique** to **Gap-Linear-Dependent-Set**
One-Side-Gap-Biclique

One-Side-Gap-Biclique($k, l, h$)

**Input:** a bipartite graph $G=(L \cup R, E)$ and integers $k, l, h$.

**Parameter:** $k$

**Question:** distinguish between the following cases:

• (yes) there are $k$ vertices in $L$ with at least $h$ common neighbors.

• (no) any $k$ vertices in $L$ have at most $l$ common neighbors.

The input graph $G$ is promised to be in one of these cases.

Theorem [Lin, SODA 2015]

One-Side-Gap-Biclique($k, l, h$) is not FPT for $(\sqrt{k} + 1)! < l < h < n^{1/\sqrt{k}}$, assuming $W[1] \neq \text{FPT}$. 
From Biclique to Linear Dependent Set

Let \( G=(L \cup R, E) \) be an instance of \textbf{One-Side-Gap-Biclique}(\( k, l, h \)).

Choose a prime power \( q>(|L|+|R|) \) s.t. \((L \cup R)\) can be treated as subset of \( F_q-\{0\} \).

Let \( B=h-1>k-1 \). Define a function \( g \) from \( L \cup R \) to \( F_q^B \):

- For every \( v \in R \), \( g(v)=(1, v^1, \ldots, v^{h-2}) \)
- For every \( u \in L \), \( g(u)=(1, u^1, \ldots, u^{k-2}, 0, \ldots, 0) \)

Partition each vector from \( F_q^{qB} \) into \( q \) blocks. Each block has \( B \) elements.

For \( u \in L \) and \( v \in R \), define \( w_{\{u,v\}} \in F_q^{qB} \)

\[
\begin{bmatrix}
0 & 0 & 0 & g(v) & 0 & 0 \\
0 & 0 & 0 & g(u) & 0 & 0 \\
\end{bmatrix}
\]

\( W=\{w_{\{u,v\}}:\{u,v\}\in E\} \) is an instance of \textbf{Gap-Linear-Dependent-Set}(\( kh, \sqrt[k]{\frac{h}{l}} \cdot h \)).
Properties of the function $g$

Let $B=\max\{k,h\}-1$. Define a function $g$ from $L \cup R$ to $F_q^B$.

For every $v$ in $R$, $g(v)=(1, v^1, \ldots, v^{h-2})$
For every $u$ in $L$, $g(u)=(1, u^1, \ldots, u^{k-2},0,\ldots,0)$

(R1) The images of any $h-1$ vertices in $R$ are linearly independent. [Vandermonde matrix]
(R2) The images of any $h$ vertices in $R$ are linearly dependent.

\[
g(v_1)=(1, v_1^1, \ldots, v_1^{h-2})
g(v_2)=(1, v_2^1, \ldots, v_2^{h-2})
\vdots
\]
\[
g(v_n)=(1, v_n^1, \ldots, v_n^{h-2})
\]

(L1) The images of any $k-1$ vertices in $L$ are linearly independent.
(L2) The images of any $k$ vertices in $L$ are linearly dependent.
Yes Case

If \( G = (L \cup R, E) \) is a yes instance of **One-Side-Gap-Biclique\((k,l,h)\).**

There exist \( X \in \binom{L}{k} \) and \( Y \in \binom{R}{h} \) s.t. for all \( u \in X, v \in Y, \{u, v\} \in E \)

**Claim:** \( W' = \{w_{\{u,v\}}: u \in X, v \in Y\} \) is linearly dependent.

Suppose \( X = \{u_1, ..., u_k\}, Y = \{v_1, ..., v_h\} \).

By \( L2 \) and \( R2 \), there exist \( a_i, b_j \in F_q \) \((i \in [k], j \in [h])\) s.t.

\[
\sum_{i \in [k]} a_i g(u_i) = 0 \quad \sum_{j \in [h]} b_j g(v_j) = 0
\]

It is easy to check

\[
\sum_{i \in [k], j \in [h]} a_i b_j w_{\{u_i,v_j\}} = 0
\]

\( a_i b_j \neq 0 \) [by \( L1 \) and \( R1 \)]
No Case

If $G=(L \cup R, E)$ is a no instance of One-Side-Gap-Biclique($k, l, h$).

**Claim:** any linearly dependent set $W' \subseteq W$ has $|W'| \geq k \frac{h}{l} \cdot h$

Let
• $X=\{u \in L : \text{there exists } v \in R \text{ s.t. } w_{\{u,v\}} \in W'\}$.
• $Y=\{v \in R : \text{there exists } u \in L \text{ s.t. } w_{\{u,v\}} \in W'\}$.
• $E'=\{\{u,v\} : w_{\{u,v\}} \in W'\}$.

Consider the graph $G'=(X \cup Y, E')$. It satisfies the following conditions:
1. Every vertex in $X$ has at least $h$ neighbors [By R1]
2. Every vertex in $Y$ has at least $k$ neighbors [By L1]
3. Every $k$-vertex set of $X$ has at most $l$ common neighbors [G is a no-instance]

From 1,2,3, we can deduce that $|X| \geq k \frac{h}{l}$

$|W'|=|E'| \geq h|X| \geq h^k \frac{h}{l}$
Gap Linear Dependent Set is Hard

On input an $n$-vertex instance $G$ of **One-Side-Gap-Biclique**($k,l,h$), $(\sqrt{k} + 1)! < l < h < n^{1/\sqrt{k}}$, one can construct a set $W$ of $O(n^2)$ vectors in $F_{\Theta(n)}^{poly}$ in $poly(n)$ time such that

- **(yes)** if $G$ is yes instance, then $W$ has exactly $kh$ linearly dependent vectors;
- **(no)** if $G$ is no instance, then any $\sqrt{k\frac{h}{l}} \cdot h-1$ vectors are linearly independent.

**Small gap, FPT lower bound:**

Set $l = (\sqrt{2k}+1)!$, $h = k^k$ and $k' = kh$ so that $k'\log k'< \sqrt\frac{h}{l} h$

**One-Side-Gap-Biclique**($k,l,h$) has no $f(k)n^{O(1)}$-time algorithm
$\Rightarrow$ **Gap-Linear-Dependent-Set**($k',k'\log k'$) has no $f(k')n^{O(1)}$-time algorithm

**Large gap, polynomial lower bound:**

Set $l = (\sqrt{2k}+1)!$, $h = n^{1/\sqrt{2k}}$

**One-Side-Gap-Biclique**($k,l,h$) has no $f(k)n^{O(1)}$-time algorithm $\Rightarrow$

**Gap-Linear-Dependent-Set**($n^{O(\frac{1}{\sqrt{2k}})}$, $n^{O(\frac{1}{\sqrt{2k}} + \frac{1}{k\sqrt{2k}})}$) has no $n^{O(1)}$-time algorithms.
Reduce Field Size
From $F_{2d}$ to $F_2^d$

**Observation:** elements in $F_{2d}$ can be written as vectors in $F_2^d$.

There exist $e_1, e_2, ..., e_d$ in $F_{2d}$, s.t. every element of $F_{2d}$ can be expressed as a unique linear combination of $e_1, e_2, ..., e_d$.

Define $f : F_{2d} \to F_2^d$:

For all $c_1 e_1 + c_2 e_2 +, ..., + c_d e_d \in F_{2d}$

$$f(c_1 e_1 + c_2 e_2 +, ..., + c_d e_d) = (c_1, c_2, ..., c_d)$$

Extend $f$ to $F_{2d}^m \to F_2^{dm}$

$$f(v_1, v_2, ..., v_m) = f(v_1) ++ f(v_2) ++ ... ++ f(v_m)$$

$++$: concatenation
Reduce Field Size: Naïve Approach

Let $W$ be an instance of $\textbf{Gap-Linear-Dependent-Set}(k, k')$ over field $F_n$

Choose an integer $Q > k'$. For each $w_i \in W$, introduce $d = \log n$ vectors

$Q$ copies

$w_{i1} = (f(e_1 w_i), f(e_1 w_i), ..., f(e_1 w_i), 0...0,1,0...,0,...,0)$

$w_{i2} = (f(e_2 w_i), f(e_2 w_i), ..., f(e_2 w_i), 0...0,0,1...,0,...,0)$

$w_{id} = (f(e_d w_i), f(e_d w_i), ..., f(e_d w_i), 0...0,0,0...,1,...,0)$

If $W$ has $k$ linear dependent vectors, then $<w_{11}, ..., w_{nd}>$ has a vector with distance at most $k \log n$.

If $W$ is a no-instance, any vector in $<w_{11}, ..., w_{nd}>$ has distance at least $k'$.

$\textbf{Gap-Linear-Dependent-Set} \left( n^{O\left(\frac{1}{\sqrt{2k}}\right)}, n^{O\left(\frac{1}{\sqrt{2k} + \frac{1}{k\sqrt{2k}}}\right)} \right)$ over $F_n$ has no $n^{O(1)}$-time algorithm

$\Rightarrow \textbf{Gap-MDP} \left( n^{O\left(\frac{1}{\sqrt{2k}}\right)} \log n, n^{O\left(\frac{1}{\sqrt{2k} + \frac{1}{k\sqrt{2k}}}\right)} \right)$ over binary field has no $n^{O(1)}$-time algorithm
Reduce Field Size: Going Through NCP

Nearest Codeword Problem (NCP):

**Input:** a set $W=\{w_1,\ldots,w_n\}$ of vectors in $F_2^m$ and $t$ in $F_2^m$, $m=n^{O(1)}$ and $k$.

**Parameter:** $k$

**Question:** decide if there is a vector $x$ in $<w_1,\ldots,w_n>$ with $d(x-t)<=k$.

FPT-inapproximability of $k$-NCP=>W[1]-hardness of $k$-EvenSet (under randomized reduction)

Parameterized Intractability of Even Set and Shortest Vector Problem from Gap-ETH [Arnab et al. ICALP 2018]

To prove FPT-inapproximability of $k$-NCP, it suffices to refute FPT-algorithm for Gap-OddSet.
Gap-Odd-Set Problem

Gap-OddSet(k,k'):

**Input:** a set $W=\{w_1,\ldots,w_n\}$ of vectors in $F_2^m$ ($m=n^{O(1)}$) and integers $k'>k$.

**Parameter:** $k$

**Question:** distinguish between the following cases:

- (yes) $W$ has $k$ vectors whose sum is $(1,1,\ldots,1)$.
- (no) For all subset $X$ of $W$ with $|X|<k'$, $\sum_{x \in X} x \neq (1,1,\ldots,1)$.

Gap-Odd(k+1,3k+1) has no FPT-algorithm

**Proof:**

- Reduction from Gap-Linear-Dependent-Set-Col(k,klogk)
Color version of Gap-Linear-Dependent-Set

Gap-Linear-Dependent-Set-C\textnormal{ol}(k,k'):

\textbf{Input}: a set $W=\{w_1,\ldots,w_n\}$ of vectors in $F_q^m$, $m=n^{O(1)}, k'>k$, a coloring $c:W\rightarrow[k]$

\textbf{Parameter}: $k$

\textbf{Question}: distinguish between the following cases:
• (yes) $W$ contains \textbf{exactly} $k$ linearly dependent vectors with distinct colors under $c$
• (no) any $k'$ vectors in $W$ are not linearly dependent.

\textbf{Gap-Linear-Dependent-Set-Col}(k',k'\log k') has no FPT-algorithm

\textbf{Proof}: 
• \textbf{color-coding} + reduction from \textbf{One-Side-Gap-Biclique}
From Color Linear Dependent Set to Odd Set

Let $W$ be an instance of $\text{Gap-Linear-Dependent-Set-Col}(k,k\log k)$ over $F_{2^d}$, $d=O(\log n)$.

For each $w \in W$ and $a \in F_{2^d} - \{0\}$, define

$$c(w_i)$$

$$F(a,w) = (0, f(aw), 0, 0, \ldots, 1, \ldots, 0)$$

The target Odd Set instance $W'$ contains all $F(a,w)$ and $w'=(1,1,\ldots,1,0,0,\ldots,0)$

- If $W$ is a yes-instance, then $W'$ has $k+1$ vectors whose sum is $(1,1,\ldots,1)$
- If $W$ is a no-instance, then any subset $X$ of $W'$ whose sum is $(1,1,\ldots,1)$ must have $|X| > 3k$. 
Yes Case

If \( W \) is a yes instance of \textbf{Gap-Linear-Dependent-Set-Col}(k,k\log k). There exist \( w_1, w_2, \ldots, w_k \) in \( W \) and \( a_1, a_2, \ldots, a_k \) in \( F_{2^d} - \{0\} \), so that

\[
\sum_{i \in [k]} a_i w_i = (0, 0, \ldots, 0)
\]

And \( c(w_i) = i \) for all \( i \) in \([k]\).

It is easy to check

\[
w' + \sum_{i \in [k]} F(a_i, w_i) = (1, 1, \ldots, 1)
\]

\[
F(a_1, w_1) = (0, f(a_1 w_1), 1, 0, \ldots, 0)
\]

\[
F(a_2, w_2) = (0, f(a_2 w_2), 0, 1, \ldots, 0)
\]

\[
F(a_k, w_k) = (0, f(a_k w_k), 0, 0, \ldots, 1)
\]

\[
w' = (1, 1, 1, \ldots, 1, 0, 0, \ldots, 0)
\]
No Case

If $W$ is a no instance of $\textbf{Gap-Linear-Dependent-Set-Col}(k, k\log k)$.

Let $X$ be a subset of $W'$ with

$$\sum_{x \in X} x = (1, 1, \ldots, 1)$$

Need to show $|X| > 3k + 1$

1. $X$ must contain $w'$

2. $\sum_{j \in [k']} a_j x_j = (0, 0, \ldots, 0)$, does not imply \{\(x_1, \ldots, x_{k'}\)\} is linearly dependent

**Caution:** \([x_1, \ldots, x_{k'}]\) may contain duplicated elements!
No Case

If \( W \) is a no instance of \textbf{Gap-Linear-Dependent-Set-Col}(k,k\log k).

Let \( X \) be a subset of \( W' \) with
\[
\sum_{x \in X} x = (1, 1, \ldots, 1)
\]

Need to show \(|X| > 3k + 1\)

1. \( X \) must contain \( w' \)
2. Let \( F(a,w) \) and \( w \) have the same color. For every \( i \) in \([k]\), \( X \) must contain odd number of vectors with color \( i \).
3. If for all \( i \) in \([k]\), \( X \) has 3 vectors with color \( i \), then \(|X| \geq 3k + 1\), Done.
4. If for some \( i \) in \([k]\), \( X \) contains 1 vector with color \( i \), then
   \[
a_i x_i + \sum_{j \in [k']-\{i\}} a_j x_j = (0, 0, \ldots, 0)
   \]
   \( \{x_1, \ldots, x_{k'}\} \) is linearly dependent

\( k' > k \log k \)
Conclusions

For large finite field $F_{\text{poly}}(n)$:
Assuming $\text{FPT} \neq \text{W}[1]$, there is no $f(k)n^{O(1)}$-time algorithm for
• Gap-Linear-Dependent-Set
• Gap-MDP

For binary field:
Assuming $\text{FPT} \neq \text{W}[1]$, Gap-MDP has no polynomial-time algorithms
Assuming $\text{FPT} \neq \text{W}[1]$, Gap-OddSet has no FPT-algorithms
Combine the result of Arnab et al. [ICALP '18], we prove Even-Set is $\text{W}[1]$-hard
(under randomized reduction)

Open questions:
1. Find a deterministic reduction for $\text{W}[1]$-hardness of Even Set.
2. Reduce field size without going through Nearest Codeword.
Thank You!