Approximation Resistance from Pairwise Independent Subgroups

Siu On Chan
UC Berkeley
Max-CSP

**Goal:** Satisfy the maximum fraction of constraints

Examples:

1. **Max-3XOR:**
   \[
   x_1 + x_{10} + x_{27} = 1 \\
   x_4 + x_5 + x_{16} = 0 \\
   \vdots
   \]

2. **Max-3SAT:**
   \[
   x_2 \lor \overline{x_9} \lor x_{31} \\
   x_8 \lor x_{15} \lor \overline{x_{17}} \\
   \vdots
   \]
Max-CSP

**Goal:** Satisfy the maximum fraction of constraints

Examples:

1. **Max-3XOR: \(\left(\frac{1}{2} + \varepsilon\right)\)-hardness** [Håstad 01]
   
   \[x_1 + x_{10} + x_{27} = 1\]
   \[x_4 + x_5 + x_{16} = 0\]
   \[\vdots\]

2. **Max-3SAT: \(\left(\frac{7}{8} + \varepsilon\right)\)-hardness** [Håstad 01]
   
   \[x_2 \lor \bar{x}_9 \lor x_{31}\]
   \[x_8 \lor x_{15} \lor \bar{x}_{17}\]
   \[\vdots\]
Definition (Approximation resistance)

NP-hard to beat a random assignment even when almost satisfiable

That is, NP-hard to decide if an instance of MAX-CSP has value

\[ \geq 1 - \varepsilon \quad \text{or} \quad \leq \text{“random assignment value”} + \varepsilon \]

Examples: MAX-3XOR, MAX-3SAT

Question

Which CSPs are approximation resistant? Why?

Partial answer

If given by a predicate \( C \) that is a “pairwise independent subgroup” [Chan13]
Max-CSP($C$)

Max-CSP($C$) or Max-C:
Each clause

- involves the same number, $k$, of literals
- accepts the same collection $C \subseteq \mathbb{Z}_2^k$ of local assignments

Examples ($k = 3$):

1. $C = \begin{cases} 000 & 001 & 011 & 010 \\ 100 & 101 & 111 & 110 \end{cases}$ \implies \text{Max-C = Max-3XOR}

2. $C = \begin{cases} 000 & 001 & 011 & 010 \\ 100 & 101 & 111 & 110 \end{cases}$ \implies \text{Max-C = Max-3SAT}

Random assignment value $= |C|/2^k$
## Previous work

<table>
<thead>
<tr>
<th>Arity</th>
<th>Approximation resistant $\text{Max-CSP}(C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>none [Goemans–Williamson95, Håstad05]</td>
</tr>
<tr>
<td>3</td>
<td>contains all strings of the same parity [Håstad01, Zwick98]</td>
</tr>
<tr>
<td>4</td>
<td>many examples [Guruswami–Lewin–Sudan–Trevisan98, Hast05]</td>
</tr>
<tr>
<td>$\geq$ 5</td>
<td>scattered results [Håstad01, Samorodnitsky–Trevisan00] [Engebretsen–Holmerin08, Hast05, Håstad11]</td>
</tr>
</tbody>
</table>

\[
\text{Arity} = \#\text{variables per constraint}
\]
Criteria for approximation resistance (red region):
Criteria for approximation resistance (red region):

- [Austrin–Mossel09]: contains pairwise independent subset, assuming Unique-Games Conjecture
  
  - $C$ is pairwise independent if $\forall i \neq j \in [k], \forall a, b \in \mathbb{Z}_2$, 

  $$\Pr_{c \in C}[c_i = a, c_j = b] = 1/|\mathbb{Z}_2|^2$$

  Example: $C = \{k$-bit strings of even parity$\} = kXOR$
Criteria for approximation resistance (red region):

- [Austrin–Mossel09]: contains pairwise independent subset, assuming Unique-Games Conjecture
  - $C$ is pairwise independent if $\forall i \neq j \in [k], \forall a, b \in \mathbb{Z}_2$, 
    \[
    \Pr_{c \in C}[c_i = a, c_j = b] = \frac{1}{|\mathbb{Z}_2|^2}
    \]
  - Example: $C = \{k$-bit strings of even parity$\} = k$XOR

- [Chan13]: contains pairwise independent subgroup
  - Almost all Max-CSP($C$) [Håstad09]
Corollaries

- Optimal $\Theta(k/2^k)$-hardness for Max-$k$CSP, using predicate in [Samorodnitsky–Trevisan09]
- Optimal query-efficient Probabilistically Checkable Proof (PCP) for $\mathsf{NP}$
- Optimal $\Theta(qk/q^k)$-hardness for non-boolean Max-$k$CSP when $k \geq$ domain size $q$, using predicate of [Håstad12]
Corollaries

- Optimal $\Theta(k/2^k)$-hardness for Max-$k$CSP, using predicate in [Samorodnitsky–Trevisan09]
- Optimal query-efficient Probabilistically Checkable Proof (PCP) for $\text{NP}$
- Optimal $\Theta(qk/q^k)$-hardness for non-boolean Max-$k$CSP when $k \geq$ domain size $q$, using predicate of [Håstad12]
- Improved hardness of ALMOST-COLORING, INDEPENDENT-SET on bounded degree graphs, 2PROVER-1ROUND-GAME
  - network connectivity problems [Laekhanukit12]
- Follow-up works: [Khot–Tulsiani–Worah12, Huang13a, Huang13b]

Motivated by integrality gaps in sum-of-square programs (the strongest known semidefinite programs) [Schoenebeck08, Tulsiani09, Chan13]
Proof sketch

**Theorem**

If $C \subseteq \mathbb{Z}_2^k$ is a subgroup that is pairwise independent, then $\text{Max-CSP}(C)$ is approximation resistant

**Definition**

$C$ is pairwise independent if $\forall i \neq j \in [k], \forall a, b \in \mathbb{Z}_2,$

$$\Pr_{c \in C}[c_i = a, c_j = b] = 1/|\mathbb{Z}_2|^2$$
<table>
<thead>
<tr>
<th>LABEL-COVER</th>
<th>composition</th>
<th>MAX-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes:</td>
<td>1</td>
<td>\approx 1</td>
</tr>
<tr>
<td>No:</td>
<td>(o(1))</td>
<td>\approx</td>
</tr>
</tbody>
</table>
Proof overview

<table>
<thead>
<tr>
<th>Label-Cover</th>
<th>Composition</th>
<th>Max-C</th>
<th>XOR</th>
<th>Max-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes:</td>
<td></td>
<td>$\approx 1$</td>
<td></td>
<td>$\approx 1$</td>
</tr>
<tr>
<td>No:</td>
<td>$o(1)$</td>
<td>$\approx</td>
<td>C</td>
<td>/2^k$</td>
</tr>
</tbody>
</table>
$k$ players try to convince a judge that a Max-C instance $M$ is satisfiable

1. Judge picks random clause $(\vec{v}, \vec{b}) = ((v_1, \ldots, v_k), (b_1, \ldots, b_k))$ from Max-C instance $M$ ($\vec{b} \in \mathbb{Z}_2^k$ specifies positive/negative literals)

2. Gets assignments $f_i(v_i) \in \mathbb{Z}_2$ from $k$ players
$k$ players try to convince a judge that a Max-C instance $M$ is satisfiable

1. Judge picks random clause $(\vec{v}, \vec{b}) = (((v_1, \ldots, v_k), (b_1, \ldots, b_k)))$ from Max-C instance $M$ ($\vec{b} \in \mathbb{Z}_2^k$ specifies positive/negative literals)

2. Gets assignments $f_i(v_i) \in \mathbb{Z}_2$ from $k$ players

3. Accepts $\iff \vec{f}(\vec{v}) - \vec{b} \in C$
Two parties try to convince a judge that a CSP instance $L$ is satisfiable

1. Judge picks clause $\square$ and variable $\bigcirc$ from $\square$ at random
2. Asks for assignment to $\square$ from one party and assignment to $\bigcirc$ from the other
3. Accepts if the assignments agree at $\bigcirc$

Winning probability 1 or $\approx 0$? NP-hard to tell! (PCP Theorem and Parallel Repetition Theorem)
Label-Cover $\rightarrow$ Max-C (Composition)

$k$ players try to convince a judge that a CSP instance $L$ has a satisfying assignment $A$

1. Judge picks $\overline{\text{\textbullet}}$ and $\text{\textbullet}$ from $L$ as in LABEL-COVER
2. Asks $(\overline{\text{\textbullet}}, z_i)$ or $(\text{\textbullet}, z_i)$ from each player
   - $z_i$: subset of satisfying assignments to clause $\overline{\text{\textbullet}}$ or variable $\text{\textbullet}$
3. Get boolean replies $y_i$ from $k$ players
4. Accept $\iff (y_1 - b_1, \ldots, y_k - b_k) \in C$
Label-Cover $\rightarrow$ Max-C (Composition)

$k$ players try to convince a judge that a CSP instance $L$ has a satisfying assignment $A$

1. Judge picks $\rightarrow$ and $\bullet$ from $L$ as in LABEL-COVER
2. Asks $(\rightarrow, z_i)$ or $(\bullet, z_i)$ from each player $z_i$: subset of satisfying assignments to clause $\rightarrow$ or variable $\bullet$
3. Get boolean replies $y_i$ from $k$ players
4. Accept $\Leftrightarrow (y_1 - b_1, \ldots, y_k - b_k) \in C$

$z_1, \ldots, z_k, b_1, \ldots, b_k$ are correlated, as specified by “dictator test”
Composition barrier

<table>
<thead>
<tr>
<th>LABEL-COVER</th>
<th>composition</th>
<th>MAX-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes:</td>
<td>1</td>
<td>$\approx 1$</td>
</tr>
<tr>
<td>No:</td>
<td>$o(1)$</td>
<td>$\approx</td>
</tr>
</tbody>
</table>

Some players share $\langle \text{ }, z_1 \rangle$, others share $\langle \bullet, z_k \rangle$ replies not random

[Bellare–Goldreich–Sudan98, Sudan–Trevisan98]
XOR

<table>
<thead>
<tr>
<th>LABEL-COVER</th>
<th>composition</th>
<th>MAX-C</th>
<th>XOR</th>
<th>MAX-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>No:</td>
<td>$o(1)$</td>
<td>$\leq 0.9$</td>
<td>$\approx</td>
<td>C</td>
</tr>
</tbody>
</table>

XOR of games:
- Parallel repetition without blowing up alphabet size
- Each player should respond with the XOR of replies to individual games

Game $M \oplus M'$:
1. Judge picks random clauses $(\vec{v}, \vec{b})$ from $M$ and $(\vec{v}', \vec{b}')$ from $M'$
2. Gets boolean assignments $f_i(\vec{v}_i, \vec{v}'_i)$ from $k$ players
3. Accepts $\iff \tilde{f}(\vec{v}, \vec{v}') - \vec{b} - \vec{b}' \in C$

Preserves almost-satisfiability when $C$ is a subgroup
**XOR-lemma?**

*Wishful thinking (XOR-lemma)*

\[ \text{val}(M) \leq 0.9 \implies \text{val}(M \oplus \ldots \oplus M) \rightarrow |C|/2^k \]

*Counterexample: Mermin’s game* [Briët–Buhrman–Lee–Vidick13]
Observation
Correlation can only decrease upon taking XOR

<table>
<thead>
<tr>
<th>LABEL-COVER</th>
<th>composition</th>
<th>MAX-C</th>
<th>XOR</th>
<th>MAX-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>No: o(1)</td>
<td>player j</td>
<td>random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uncorrelated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ M_1 : \]
Observation
Correlation can only decrease upon taking XOR

<table>
<thead>
<tr>
<th>LABEL-COVER</th>
<th>composition</th>
<th>MAX-C</th>
<th>XOR</th>
<th>MAX-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>No:</td>
<td>$o(1)$</td>
<td>player $j$</td>
<td>random</td>
<td></td>
</tr>
<tr>
<td>uncorrelated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$M_1$: 🙄 😊 😊 😊 😊

$M_2$: 😊 🙄 😊 😊 😊

$M_3$: 😊 😊 🙄 😊 😊

$M_4$: 😊 😊 😊 🙄 😊

⊕ 🙄 🙄 🙄 🙄 🙄
\[ \vec{f}(\vec{v}) - \vec{b} \triangleq (f_1(\vec{v}_1) - b_1, \ldots, f_k(\vec{v}_k) - b_k) \in \mathbb{Z}_2^k \]

\[ \|M\|_\chi \triangleq \max_{\vec{f}:\vec{v} \rightarrow \mathbb{Z}_2^k} \left| \mathbb{E}_{(\vec{v},\vec{b})} \chi(\vec{f}(\vec{v}) - \vec{b}) \right| , \quad \chi \in \mathbb{Z}_2^k \]

**Lemma**

\[ \|M \oplus M'\|_\chi \leq \min \{ \|M\|_\chi, \|M'\|_\chi \} \]
\[ f(\vec{v}) - \vec{b} \triangleq (f_1(\vec{v}_1) - \vec{b}_1, \ldots, f_k(\vec{v}_k) - \vec{b}_k) \in \mathbb{Z}_2^k \]

\[ \|M\|_\chi \triangleq \max_{\tilde{f}:\vec{v} \to \mathbb{Z}_2^k} \left| \mathbb{E}_{(\vec{v},\vec{b})} \chi(\tilde{f}(\vec{v}) - \vec{b}) \right|, \quad \chi \in \mathbb{Z}_2^k \]

**Lemma**

\[ \|M \oplus M'\|_\chi \leq \min\{\|M\|_\chi, \|M'\|_\chi\} \]

\[
\left| \mathbb{E}_{(\vec{v},\vec{b})} \mathbb{E}_{(\vec{v}',\vec{b}')} \chi(\tilde{f}(\vec{v},\vec{v}') - \vec{b} - \vec{b}') \right| \\
\leq \mathbb{E}_{(\vec{v},\vec{b})} \mathbb{E}_{(\vec{v}',\vec{b}')} \chi(\tilde{f}(\vec{v},\vec{v}') - \vec{b} - \vec{b}') \\
\square
g(\vec{v}')
| Label-Cover | \( o(1) \) | \( \| \cdot \|_\chi = o(1) \) | \( \forall \chi : \chi_j \neq 1 \) | Max-C | \( |C|/2^k + o(1) \) |

Uses pairwise independence and invariance principle

Conclusion

- New gap-amplification technique: XOR/direct sum
- Optimal hardness of Max-$k$CSP and optimal query-efficient PCP
- General criteria for approximation resistance

Open problems

1. Optimal hardness of satisfiable Max-$k$CSP?
   - Progress by [Huang13] in the next talk

2. Derandomizing XOR/direct sum

Thank you
Conclusion

- New gap-amplification technique: XOR/direct sum
- Optimal hardness of $\text{Max-}k\text{-CSP}$ and optimal query-efficient PCP
- General criteria for approximation resistance

Open problems

1. Optimal hardness of satisfiable $\text{Max-}k\text{-CSP}$?
   - Progress by [Huang13] in the next talk

2. Derandomizing XOR/direct sum

Thank you 😊

Emoticons modified from
http://www.texample.net/tikz/examples/emoticons/
Gavel from
http://openclipart.org/detail/69745/judge-hammer-by-bocian