Factor-11/8 NP-inapproximability for 2-Variable Linear Equations

Sangxia Huang

EPFL
Switzerland

Joint work with Johan Håstad
Rajsekar Manokaran
Ryan O'Donnell
John Wright

2-Variable Linear Equations

Definition ($\text{E2L}\textsc{in}_q$)

System of $\mathbb{F}_q$ linear equations, each containing exactly two variables.

Example of an $\text{E2L}\textsc{in}_2$ instance

$$\begin{aligned} x_1 + x_4 &= 1 \\ x_3 + x_5 &= 0 \\ x_2 + x_5 &= 0 \\ &\dots \end{aligned}$$

2-Variable Linear Equations

Definition (E2LIN$_q$)

System of $\mathbb{F}_q$ linear equations, each containing exactly two variables.

Example of an E2LIN$_2$ instance

$$\begin{aligned} x_1 + x_4 &= 1 \\ x_3 + x_5 &= 0 \\ x_2 + x_5 &= 0 \\ &\dots \end{aligned}$$

Equivalent to Unique-Games

[Khot, Kindler, Mossel, O'Donnell '07], [Mossel, O'Donnell, Oleszkiewicz '10]

Approximation for 2-Variable Linear Equations

Definition ($(\varepsilon, \varepsilon')$-approximation algorithm)

Given an instance in which the best solution falsifies at most an ε-fraction of the equations, the algorithm finds a solution falsifying at most an ε'-fraction of the equations.

Approximation for 2-Variable Linear Equations

Definition ($(\varepsilon, \varepsilon')$-approximation algorithm)

Given an instance in which the best solution falsifies at most an ε-fraction of the equations, the algorithm finds a solution falsifying at most an ε'-fraction of the equations.

- Simple observation: there is a $(0, 0)$-approximation algorithm for E2LIN$_q$.

Approximation for 2-Variable Linear Equations

- $(\varepsilon, \frac{2}{\pi}\sqrt{\varepsilon} + o(\varepsilon))$-approximation algorithm for E2LIN$_2$
  [Goemans, Williamson '94]

Approximation for 2-Variable Linear Equations

- $(\varepsilon, \frac{2}{\pi}\sqrt{\varepsilon} + o(\varepsilon))$-approximation algorithm for $\textsc{E2Lin}_2$
  [Goemans, Williamson '94]
- $(\varepsilon, C_q\sqrt{\varepsilon})$-approximation for $\textsc{E2Lin}_q$, for some $C_q = \Theta(\sqrt{\log q})$
  [Charikar, Makarychev, Makarychev, '06]

Approximation for 2-Variable Linear Equations

- $(\varepsilon, \frac{2}{\pi}\sqrt{\varepsilon} + o(\varepsilon))$-approximation algorithm for $\textsc{E2Lin}_2$
  [Goemans, Williamson '94]
- $(\varepsilon, C_q\sqrt{\varepsilon})$-approximation for $\textsc{E2Lin}_q$, for some $C_q = \Theta(\sqrt{\log q})$
  [Charikar, Makarychev, Makarychev, '06]
- UGC implies that improving on the above algorithms is NP-hard
  [Khot, Kindler, Mossel, O'Donnell '07], [Mossel, O'Donnell, Oleszkiewicz '10]

Approximation for 2-Variable Linear Equations

- $(\varepsilon, \frac{2}{\pi}\sqrt{\varepsilon} + o(\varepsilon))$-approximation algorithm for $\textsc{E2Lin}_2$
  [Goemans, Williamson '94]
- $(\varepsilon, C_q\sqrt{\varepsilon})$-approximation for $\textsc{E2Lin}_q$, for some $C_q = \Theta(\sqrt{\log q})$
  [Charikar, Makarychev, Makarychev, '06]
- UGC implies that improving on the above algorithms is NP-hard
  [Khot, Kindler, Mossel, O'Donnell '07], [Mossel, O'Donnell, Oleszkiewicz '10]
- If there exists $q = q(\varepsilon)$ such that $(\varepsilon, \omega(\sqrt{\varepsilon}))$-approximating $\textsc{E2Lin}_q$ is NP-hard, then UGC holds
  [Rao '11]

NP-hardness for Approximating E2LIN$_2$

Theorem (Håstad '97)

For any $C < \frac{5}{4}$ *and* $0 < \varepsilon < \varepsilon_0 = \frac{1}{4}$, $(\varepsilon, C\varepsilon)$*-approximating* E2LIN$_2$ *is NP-hard.*

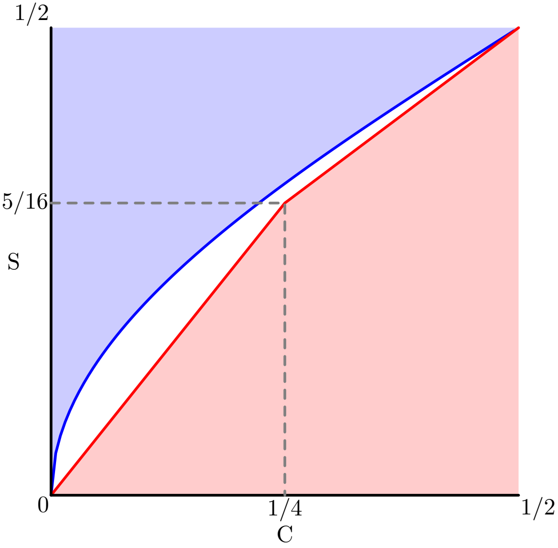
NP-hardness for Approximating E2LIN$_2$

Theorem (Håstad '97)

For any $C < \frac{5}{4}$ and $0 < \varepsilon < \varepsilon_0 = \frac{1}{4}$, $(\varepsilon, C\varepsilon)$-approximating E2LIN$_2$ is NP-hard.

- The same hardness holds for any q, and $\varepsilon_0 \to \frac{1}{2}$ as $q \to \infty$ [O'Donnell, Wright '12]

NP-hardness for Approximating E2LIN$_2$



Our Results

Theorem

For any $C < \frac{11}{8}$, $0 < \varepsilon < \varepsilon_0 = \frac{1}{8}$, *it is NP-hard to* $(\varepsilon, C\varepsilon)$*-approximate* E2LIN$_2$.

Our Results

Theorem

For any $C < \frac{11}{8}$, $0 < \varepsilon < \varepsilon_0 = \frac{1}{8}$, *it is NP-hard to* $(\varepsilon, C\varepsilon)$*-approximate* E2LIN$_2$.

- Proof by gadget reduction from CSP with some predicate ϕ

Our Results

Theorem

For any $C < \frac{11}{8}$, $0 < \varepsilon < \varepsilon_0 = \frac{1}{8}$, *it is NP-hard to* $(\varepsilon, C\varepsilon)$*-approximate* $\text{E2L}\textsc{in}_2$.

- Proof by gadget reduction from CSP with some predicate ϕ
- Our gadget is optimal among all gadget reductions from ϕ

Our Results

Theorem

For any $C < \frac{11}{8}$, $0 < \varepsilon < \varepsilon_0 = \frac{1}{8}$, *it is NP-hard to* $(\varepsilon, C\varepsilon)$*-approximate* E2LIN$_2$.

- Proof by gadget reduction from CSP with some predicate ϕ
- Our gadget is optimal among all gadget reductions from ϕ
- For any predicate ψ whose set of satisfying assignments supports a pairwise independent distribution, no gadget reduction from CSP with predicate ψ can establish NP-hardness factor better than $\frac{1}{1-e^{-1/2}} \approx 2.54$

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2LIN$_2$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2$\text{L}_{\text{IN}_2}$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

Theorem (Håstad '97)

For any k ≥ 3 *and* $\varepsilon > 0$*,* $(\varepsilon, \frac{1}{2} - \varepsilon)$*-approximating* E*k*$\text{L}_{\text{IN}_2}$ *is NP-hard.*

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2LIN$_2$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

Theorem (Håstad '97)

For any $k \geq 3$ and $\varepsilon > 0$, $(\varepsilon, \frac{1}{2} - \varepsilon)$-approximating E$kLIN_2$ *is NP-hard.*

- Take $k = 3$
- Let z be a global auxiliary variable

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2LIN$_2$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

Theorem (Håstad '97)

For any $k \geq 3$ and $\varepsilon > 0$, $(\varepsilon, \frac{1}{2} - \varepsilon)$-approximating E$kLIN_2$ *is NP-hard.*

- Take $k = 3$
- Let z be a global auxiliary variable
- Consider an equation

$$x_1 + x_2 + x_3 = 0.$$

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2Lin$_2$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

Theorem (Håstad '97)

For any $k \geq 3$ and $\varepsilon > 0$, $(\varepsilon, \frac{1}{2} - \varepsilon)$-approximating E$kLin_2$ *is NP-hard.*

- Take $k = 3$
- Let z be a global auxiliary variable
- Consider an equation

$$x_1 + x_2 + x_3 = 0.$$

- Introduce variables y_1, y_2, y_3, y_4, and constraints

$$\begin{aligned}
z + y_1 &= 1, & z + y_2 &= 1, & z + y_3 &= 1, & z + y_4 &= 0,\\
x_1 + y_1 &= 1, & x_1 + y_2 &= 1, & x_1 + y_3 &= 0, & x_1 + y_4 &= 1,\\
x_2 + y_1 &= 1, & x_2 + y_2 &= 0, & x_2 + y_3 &= 1, & x_2 + y_4 &= 1,\\
x_3 + y_1 &= 0, & x_3 + y_2 &= 1, & x_3 + y_3 &= 1, & x_3 + y_4 &= 1.
\end{aligned}$$

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2Lin$_2$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

- Take $k = 3$
- Let z be a global auxiliary variable
- Consider an equation
$$x_1 + x_2 + x_3 = 0.$$
- Introduce variables y_1, y_2, y_3, y_4, and constraints
$$\begin{aligned}
z + y_1 = 1, \quad & z + y_2 = 1, \quad & z + y_3 = 1, \quad & z + y_4 = 0, \\
x_1 + y_1 = 1, \quad & x_1 + y_2 = 1, \quad & x_1 + y_3 = 0, \quad & x_1 + y_4 = 1, \\
x_2 + y_1 = 1, \quad & x_2 + y_2 = 0, \quad & x_2 + y_3 = 1, \quad & x_2 + y_4 = 1, \\
x_3 + y_1 = 0, \quad & x_3 + y_2 = 1, \quad & x_3 + y_3 = 1, \quad & x_3 + y_4 = 1.
\end{aligned}$$

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2LIN$_2$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

- Take $k = 3$
- Let z be a global auxiliary variable
- Consider an equation

$$x_1 + x_2 + x_3 = 0.$$

- Introduce variables y_1, y_2, y_3, y_4, and constraints

$$\begin{array}{llll} z + y_1 = 1, & z + y_2 = 1, & z + y_3 = 1, & z + y_4 = 0, \\ x_1 + y_1 = 1, & x_1 + y_2 = 1, & x_1 + y_3 = 0, & x_1 + y_4 = 1, \\ x_2 + y_1 = 1, & x_2 + y_2 = 0, & x_2 + y_3 = 1, & x_2 + y_4 = 1, \\ x_3 + y_1 = 0, & x_3 + y_2 = 1, & x_3 + y_3 = 1, & x_3 + y_4 = 1. \end{array}$$

- Let σ be an assignment to x_1, x_2 and x_3
  - If σ satisfies the equation, then there is an assignment to the auxiliary variables that falsifies 4 equations
  - Otherwise any assignment to the auxiliary variables falsify at least 6 equations

Gadget Reduction for $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness of E2LIN$_2$

[Håstad '97][Trevisan, Sorkin, Sudan, Williamson '00]

- Take $k = 3$
- Let z be a global auxiliary variable
- Consider an equation
$$x_1 + x_2 + x_3 = 0.$$
- Introduce variables y_1, y_2, y_3, y_4, and constraints
$$\begin{array}{llll} z + y_1 = 1, & z + y_2 = 1, & z + y_3 = 1, & z + y_4 = 0, \\ x_1 + y_1 = 1, & x_1 + y_2 = 1, & x_1 + y_3 = 0, & x_1 + y_4 = 1, \\ x_2 + y_1 = 1, & x_2 + y_2 = 0, & x_2 + y_3 = 1, & x_2 + y_4 = 1, \\ x_3 + y_1 = 0, & x_3 + y_2 = 1, & x_3 + y_3 = 1, & x_3 + y_4 = 1. \end{array}$$
- Let σ be an assignment to x_1, x_2 and x_3
  - If σ satisfies the equation, then there is an assignment to the auxiliary variables that falsifies 4 equations
  - Otherwise any assignment to the auxiliary variables falsify at least 6 equations

$\Rightarrow (\frac{4}{16} + \varepsilon, \frac{5}{16} - \varepsilon)$-hardness for E2LIN$_2$

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	...
0	0	0	1	1	1	1	...
0	1	1	1	1	0	0	...
1	0	1	1	0	1	0	...
1	1	0	0	1	1	0	...

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	…
0	0	0	1	1	1	1	…
0	1	1	1	1	0	0	…
1	0	1	1	0	1	0	…
1	1	0	0	1	1	0	…

- For each satisfying assignment to x_1,x_2,x_3, setting the auxiliary variables according to the matrix above achieves *unsat* at most C

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	$\dots$
0	0	0	1	1	1	1	$\dots$
0	1	1	1	1	0	0	$\dots$
1	0	1	1	0	1	0	$\dots$
1	1	0	0	1	1	0	$\dots$

- For each satisfying assignment to x_1,x_2,x_3, setting the auxiliary variables according to the matrix above achieves *unsat* at most C
- For each unsatisfying assignment to x_1,x_2,x_3, any assignment to the auxiliary variable gives *unsat* at least S

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	...
0	0	0	1	1	1	1	...
0	1	1	1	1	0	0	...
1	0	1	1	0	1	0	...
1	1	0	0	1	1	0	...

- For each satisfying assignment to x_1,x_2,x_3, setting the auxiliary variables according to the matrix above achieves *unsat* at most C
- For each unsatisfying assignment to x_1,x_2,x_3, any assignment to the auxiliary variable gives *unsat* at least S
- Key Observation: If two columns y_i and y_j are identical, then merging them gives a new gadget that is as good

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	$\dots$
0	0	0	1	1	1	1	$\dots$
0	1	1	1	1	0	0	$\dots$
1	0	1	1	0	1	0	$\dots$
1	1	0	0	1	1	0	$\dots$

- For each satisfying assignment to x_1,x_2,x_3, setting the auxiliary variables according to the matrix above achieves *unsat* at most C
- For each unsatisfying assignment to x_1,x_2,x_3, any assignment to the auxiliary variable gives *unsat* at least S
- Key Observation: If two columns y_i and y_j are identical, then merging them gives a new gadget that is as good $\Rightarrow$ Need at most $2^4 = 16$ variables in the gadget

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson ’00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	...
0	0	0	1	1	1	1	...
0	1	1	1	1	0	0	...
1	0	1	1	0	1	0	...
1	1	0	0	1	1	0	...

- For each satisfying assignment to x_1,x_2,x_3, setting the auxiliary variables according to the matrix above achieves *unsat* at most C
- For each unsatisfying assignment to x_1,x_2,x_3, any assignment to the auxiliary variable gives *unsat* at least S
- Key Observation: If two columns y_i and y_j are identical, then merging them gives a new gadget that is as good $\Rightarrow$ Need at most $2^4 = 16$ variables in the gadget
- Search of optimal gadget as LP

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	...
0	0	0	1	1	1	1	...
0	1	1	1	1	0	0	...
1	0	1	1	0	1	0	...
1	1	0	0	1	1	0	...

- For each satisfying assignment to x_1,x_2,x_3, setting the auxiliary variables according to the matrix above achieves *unsat* at most C
- For each unsatisfying assignment to x_1,x_2,x_3, any assignment to the auxiliary variable gives *unsat* at least S
- Key Observation: If two columns y_i and y_j are identical, then merging them gives a new gadget that is as good $\Rightarrow$ Need at most $2^4 = 16$ variables in the gadget
- Search of optimal gadget as LP
  - $\sim \binom{16}{2}$ variables and $\sim 2^{16}$ constraints

Constructing Gadgets

[Trevisan, Sorkin, Sudan, Williamson '00]

x_1	x_2	x_3	y_1	y_2	y_3	y_4	…
0	0	0	1	1	1	1	…
0	1	1	1	1	0	0	…
1	0	1	1	0	1	0	…
1	1	0	0	1	1	0	…

- For each satisfying assignment to x_1, x_2, x_3, setting the auxiliary variables according to the matrix above achieves *unsat* at most C
- For each unsatisfying assignment to x_1, x_2, x_3, any assignment to the auxiliary variable gives *unsat* at least S
- Key Observation: If two columns y_i and y_j are identical, then merging them gives a new gadget that is as good $\Rightarrow$ Need at most $2^4 = 16$ variables in the gadget
- Search of optimal gadget as LP
  - $\sim \binom{16}{2}$ variables and $\sim 2^{16}$ constraints
  - Certificate of optimality via dual LP

Stronger Hardness Guarantees

Definition (Had_r)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_\emptyset + x_S = \sum_{i \in S}(x_\emptyset + x_{\{i\}})$.

Stronger Hardness Guarantees

Definition (Had_r)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_\emptyset + x_S = \sum_{i \in S}(x_\emptyset + x_{\{i\}})$.

- The Had_r predicate has 2^{r+1} satisfying assignments.
- The Had_2 predicate is exactly E4LIN$_2$.

Stronger Hardness Guarantees

Definition (*Had_r*)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_\emptyset + x_S = \sum_{i \in S}(x_\emptyset + x_{\{i\}})$.

- The Had_r predicate has 2^{r+1} satisfying assignments.
- The Had_2 predicate is exactly E4LIN$_2$.

Theorem (Chan '13)

For every $r \geq 2$ and $\varepsilon > 0$, $(\varepsilon, 1 - 2^{r+1}/2^{2^r} - \varepsilon)$-approximating Had_r-CSP is NP-hard.

Stronger Hardness Guarantees

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Predicate $\phi : \{0,1\}^k \to \{0,1\}$

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Predicate $\phi : \{0, 1\}^k \to \{0, 1\}$
- Variable set V

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Predicate $\phi : \{0, 1\}^k \to \{0, 1\}$
- Variable set V
- Distribution of ϕ-constraints $\mathcal{C} \sim \mathcal{I}$, where

$$\mathcal{C} = ((x_1, b_1), \ldots, (x_k, b_k)), \quad x_i \in V, b_i \in \{0, 1\}$$

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Predicate $\phi : \{0,1\}^k \to \{0,1\}$
- Variable set V
- Distribution of ϕ-constraints $\mathcal{C} \sim \mathcal{I}$, where

$$\mathcal{C} = ((x_1, b_1), \ldots, (x_k, b_k)), \quad x_i \in V, b_i \in \{0,1\}$$

- Assignment $A : V \to \{0,1\}$ satisfies constraint $\mathcal{C}$ if

$$\phi(b_1 + A(x_1), \ldots, b_k + A(x_k)) = 1.$$

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Predicate $\phi : \{0,1\}^k \to \{0,1\}$
- Variable set V
- Distribution of ϕ-constraints $\mathcal{C} \sim \mathcal{I}$, where

$$\mathcal{C} = ((x_1, b_1), \ldots, (x_k, b_k)), \quad x_i \in V, b_i \in \{0,1\}$$

- Assignment $A : V \to \{0,1\}$ satisfies constraint $\mathcal{C}$ if

$$\phi(b_1 + A(x_1), \ldots, b_k + A(x_k)) = 1.$$

- The value of A on $\mathcal{I}$: $unsat(A; \mathcal{I}) = \Pr_{\mathcal{C} \sim \mathcal{I}}[A \text{ falsifies } \mathcal{C}]$.

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Assignment $A : V \to \{0, 1\}$

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Assignment $A : V \to \{0, 1\}$
- Distribution $\mathcal{D}(A, \mathcal{I})$
  1. Sample $((x_1, b_1), \ldots, (x_k, b_k)) \sim \mathcal{I}$
  2. Output $(b_1 + A(x_1), \ldots, b_k + A(x_k))$

Stronger Hardness Guarantees

A ϕ-CSP instance $\mathcal{I}$

- Assignment $A : V \to \{0, 1\}$
- Distribution $\mathcal{D}(A, \mathcal{I})$
  1. Sample $((x_1, b_1), \ldots, (x_k, b_k)) \sim \mathcal{I}$
  2. Output $(b_1 + A(x_1), \ldots, b_k + A(x_k))$

Example:

$$\phi(x_1, x_2, \overline{x_3})$$
$$\phi(x_3, x_1, x_4)$$
$$\phi(\overline{x_2}, x_3, x_4)$$

Assignment $x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 1$.

$$\Pr_{(b_1,b_2,b_3)\sim\mathcal{D}}[(b_1, b_2, b_3) = (0, 1, 1)] = 1/3$$
$$\Pr_{(b_1,b_2,b_3)\sim\mathcal{D}}[(b_1, b_2, b_3) = (0, 0, 1)] = 2/3.$$

Stronger Hardness Guarantees

Theorem (Håstad '97, restated)

For any $\varepsilon > 0$*, given a* $\text{E}k\text{L}\textsc{in}_2$ *instance, it is NP-hard to distinguish between the following two cases*

Completeness *unsat*$(\mathcal{I}) \leq \varepsilon$

Soundness *For every assignment A,* $d_{TV}(\mathcal{D}(A,\mathcal{I}),\mathcal{U}_k) \leq \varepsilon$

Stronger Hardness Guarantees

Theorem (Håstad '97, restated)

For any $\varepsilon > 0$, given a $\textsc{E}k\textsc{Lin}_2$ instance, it is NP-hard to distinguish between the following two cases

Completeness *unsat*$(\mathcal{I}) \leq \varepsilon$

Soundness *For every assignment A,* $d_{TV}(\mathcal{D}(A, \mathcal{I}), \mathcal{U}_k) \leq \varepsilon$

Theorem (Chan '13, restated)

For every $r \geq 2$ and $\varepsilon > 0$, given an Had_r-CSP instance $\mathcal{I}$, it is NP-hard to distinguish between the following two cases:

Completeness *unsat*$(\mathcal{I}) \leq \varepsilon$

Soundness *For every assignment A,* $d_{TV}(\mathcal{D}(A, \mathcal{I}), \mathcal{U}_{2^r}) \leq \varepsilon$.

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

$$x_1 + x_2 + x_3 + x_4 = 0$$

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

$$x_1 + x_2 + x_3 + x_4 = 0$$

$$\Rightarrow$$

$$x_1 + y = 0, x_2 + y = 0, x_3 + y = 0, x_4 + y = 1$$

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

$$x_1 + x_2 + x_3 + x_4 = 0$$
$$\Rightarrow$$
$$x_1 + y = 0, x_2 + y = 0, x_3 + y = 0, x_4 + y = 1$$

- If $x_1 + x_2 + x_3 + x_4 = 0$, we can set y such that only 1 of the equations are violated

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

$$x_1 + x_2 + x_3 + x_4 = 0$$

$$\Rightarrow$$

$$x_1 + y = 0, x_2 + y = 0, x_3 + y = 0, x_4 + y = 1$$

- If $x_1 + x_2 + x_3 + x_4 = 0$, we can set y such that only 1 of the equations are violated
  $\Rightarrow$ If $unsat(\mathcal{I}) \leq \varepsilon$, then $unsat(\mathcal{J}) \leq \frac{1}{4} + \frac{3}{4}\varepsilon$

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

$$x_1 + x_2 + x_3 + x_4 = 0$$

$$\Rightarrow$$

$$x_1 + y = 0, x_2 + y = 0, x_3 + y = 0, x_4 + y = 1$$

- If $x_1 + x_2 + x_3 + x_4 = 0$, we can set y such that only 1 of the equations are violated
  $\Rightarrow$ If $\mathit{unsat}(\mathcal{I}) \leq \varepsilon$, then $\mathit{unsat}(\mathcal{J}) \leq \frac{1}{4} + \frac{3}{4}\varepsilon$
- If $x_1 = x_2 = x_3 = 1 + x_4$ then $y = x_1$ satisfies all equations

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

$$x_1 + x_2 + x_3 + x_4 = 0$$
$$\Rightarrow$$
$$x_1 + y = 0, x_2 + y = 0, x_3 + y = 0, x_4 + y = 1$$

- If $x_1 + x_2 + x_3 + x_4 = 0$, we can set y such that only 1 of the equations are violated
  $\Rightarrow$ If $unsat(\mathcal{I}) \leq \varepsilon$, then $unsat(\mathcal{J}) \leq \frac{1}{4} + \frac{3}{4}\varepsilon$
- If $x_1 = x_2 = x_3 = 1 + x_4$ then $y = x_1$ satisfies all equations
- In all other cases, the best assignment to y violates 2 equations

Stronger Hardness Guarantees

A simple $(\varepsilon, \frac{5}{4}\varepsilon)$-hardness gadget reduction from a E4LIN$_2$ instance $\mathcal{I}$ to a E2LIN$_2$ instance $\mathcal{J}$:

$$x_1 + x_2 + x_3 + x_4 = 0$$

$$\Rightarrow$$

$$x_1 + y = 0, x_2 + y = 0, x_3 + y = 0, x_4 + y = 1$$

- If $x_1 + x_2 + x_3 + x_4 = 0$, we can set y such that only 1 of the equations are violated
  $\Rightarrow$ If $unsat(\mathcal{I}) \leq \varepsilon$, then $unsat(\mathcal{J}) \leq \frac{1}{4} + \frac{3}{4}\varepsilon$
- If $x_1 = x_2 = x_3 = 1 + x_4$ then $y = x_1$ satisfies all equations
- In all other cases, the best assignment to y violates 2 equations
- If for every assignment A, $d_{TV}(\mathcal{D}(A, \mathcal{I}), \mathcal{U}_k) \leq \varepsilon$, then $unsat(\mathcal{J}) \geq \frac{2}{16} \cdot \frac{0}{4} + \frac{8}{16} \cdot \frac{1}{4} + \frac{6}{16} \cdot \frac{2}{4} - \varepsilon = \frac{5}{16} - \varepsilon$

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

Definition (Had_r)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_S + x_\emptyset = \sum_{i \in S}(x_i + x_\emptyset)$.

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

Definition (Had_r)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_S + x_\emptyset = \sum_{i \in S}(x_i + x_\emptyset)$.

- The gadget needs at most 2^{2^r} variables (primary and auxiliary), $V := \{0,1\}^{2^r}$
  - V as the set of Boolean functions on r bits $b_1, \ldots, b_r$

Constructing Gadgets for $\mathit{Had}_r \to$ E2LIN$_2$

Definition (Had_r)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_S + x_\emptyset = \sum_{i \in S}(x_i + x_\emptyset)$.

- The gadget needs at most 2^{2^r} variables (primary and auxiliary), $V := \{0,1\}^{2^r}$
  - V as the set of Boolean functions on r bits $b_1, \ldots, b_r$
- Identify the primary variables with linear functions on r bits

$$P := \left\{\sum_{i \in S} b_i \mid S \subseteq [r]\right\}.$$

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

Definition (Had_r)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_S + x_\emptyset = \sum_{i \in S}(x_i + x_\emptyset)$.

- The gadget needs at most 2^{2^r} variables (primary and auxiliary), $V := \{0,1\}^{2^r}$
  - V as the set of Boolean functions on r bits $b_1, \ldots, b_r$
- Identify the primary variables with linear functions on r bits

$$P := \left\{ \sum_{i \in S} b_i \mid S \subseteq [r] \right\}.$$

- For $i \in \{0,1\}^r$, define the *i-th dictator assignment* $D_i(f) := f(i)$

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

Definition (Had_r)

The Had_r predicate has $k = 2^r$ input variables, one for each subset $S \subseteq [r]$. The input $\{x_S\}_{S \subseteq [r]}$ satisfies Had_r if for each S, $|S| \geq 2$, $x_S + x_\emptyset = \sum_{i \in S}(x_i + x_\emptyset)$.

- The gadget needs at most 2^{2^r} variables (primary and auxiliary), $V := \{0,1\}^{2^r}$
  - V as the set of Boolean functions on r bits $b_1, \ldots, b_r$
- Identify the primary variables with linear functions on r bits

$$P := \left\{ \sum_{i \in S} b_i \mid S \subseteq [r] \right\}.$$

- For $i \in \{0,1\}^r$, define the *i*-th dictator assignment $D_i(f) := f(i)$
  - The value of the primary variables under D_i and $1 + D_i$ satisfy Had_r

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

Definition

A (c, s)-gadget reducing Had_r to E2LIN$_2$ is a gadget $\mathcal{G}$ satisfying the following properties:

Completeness For every dictator assignment D_i, $unsat(D_i, \mathcal{G}) \leq c$

Soundness The expected $unsat(\cdot; \mathcal{G})$ over uniformly random assignments to the primary variables is at least s.

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

Definition

A (c, s)-gadget reducing Had_r to E2LIN$_2$ is a gadget $\mathcal{G}$ satisfying the following properties:

Completeness For every dictator assignment D_i, $unsat(D_i, \mathcal{G}) \leq c$

Soundness The expected $unsat(\cdot; \mathcal{G})$ over uniformly random assignments to the primary variables is at least s.

Proposition

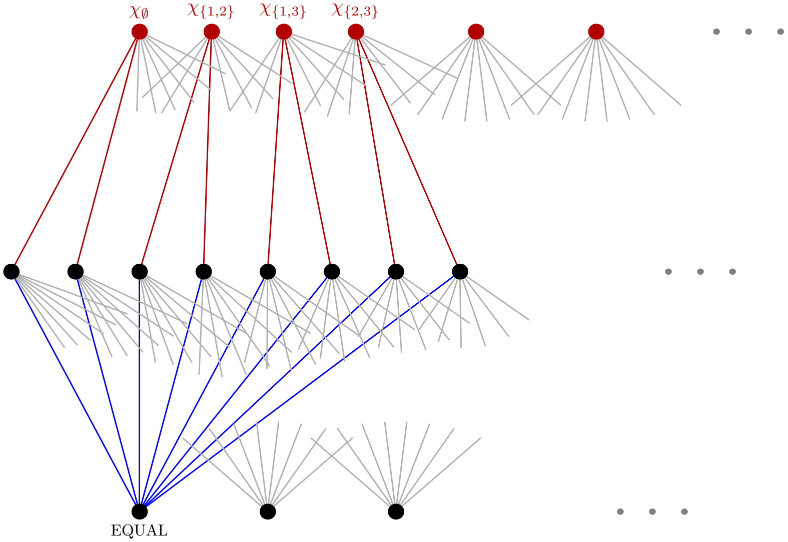
Suppose there exists a (c, s)-gadget reducing Had_r to E2LIN$_2$*, then given an* E2LIN$_2$ *instance $\mathcal{I}$, it is NP-hard to distinguish*

Completeness $unsat(\mathcal{I}) \leq c + \varepsilon$

Soundness $unsat(\mathcal{I}) \geq s - \varepsilon$

Constructing Gadgets for $Had_r \rightarrow$ E2LIN$_2$

Our $\left(\frac{1}{8}, \frac{11}{64}\right)$-Gadget



Constructing Gadgets for $Had_r \rightarrow$ E2LIN$_2$

- Had_2: $s/c = 5/4$ $(= 1.25)$

Constructing Gadgets for $Had_r \rightarrow$ E2LIN$_2$

- Had_2: $s/c = 5/4$ $(= 1.25)$
- Had_3: $s/c = 11/8$ $(= 1.375)$

Constructing Gadgets for $\mathit{Had}_r \to \textsc{E2Lin}_2$

- Had_2: $s/c = 5/4$ $(= 1.25)$
- Had_3: $s/c = 11/8$ $(= 1.375)$

  $\dots$

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

- Had_2: $s/c = 5/4$ $(= 1.25)$
- Had_3: $s/c = 11/8$ $(= 1.375)$

  $\ldots$
- Had_∞: ?

Constructing Gadgets for $Had_r \to$ E2LIN$_2$

- Had_2: $s/c = 5/4$ $(= 1.25)$
- Had_3: $s/c = 11/8$ $(= 1.375)$

  ...

- Had_∞: ?
  $s/c \geq 3/2$ if a certain Game Show Conjecture is true

Limitation of Gadget Construction

Theorem

Let $\mathcal{G}$ be a (c, s)-gadget reducing Had_r to $\text{E2L}\textsc{in}_2$. *Then*

$$\frac{s}{c} \leq \frac{1}{1 - e^{-1/2}} \approx 2.54$$

Limitation of Gadget Construction

Theorem

Let $\mathcal{G}$ be a (c, s)-gadget reducing Had_r to $\text{E2L}\textsc{in}_2$. Then

$$\frac{s}{c} \leq \frac{1}{1 - e^{-1/2}} \approx 2.54$$

Remark

We can assume that all equations in the gadget has the form $x + y = 0$, where x and y are at Hamming distance 1, as long as we only consider assignments A such that $A(1 + x) = 1 + A(x)$.

Limitation of Gadget Construction

Limitation of Gadget Construction

Let $K := 2^{r+1}$ be the number of satisfying assignments of Had_r.

Limitation of Gadget Construction

Let $K := 2^{r+1}$ be the number of satisfying assignments of Had_r.
Let $P^{\pm}$ be the set of primary variables and their negations.

Limitation of Gadget Construction

Let $K := 2^{r+1}$ be the number of satisfying assignments of Had_r.
Let $P^{\pm}$ be the set of primary variables and their negations.
For $x, y \in \{0, 1\}^K$, define their similarity

$$sim(x, y) := Pr_{i \sim [K]}[x_i = y_i],$$

and

$$sim(x, P^{\pm}) = \max_{y \in P^{\pm}} sim(x, y).$$

Limitation of Gadget Construction

Let $K := 2^{r+1}$ be the number of satisfying assignments of Had_r.
Let $P^{\pm}$ be the set of primary variables and their negations.
For $x, y \in \{0, 1\}^K$, define their similarity

$$sim(x, y) := Pr_{i \sim [K]}[x_i = y_i],$$

and

$$sim(x, P^{\pm}) = \max_{y \in P^{\pm}} sim(x, y).$$

Observation

For any $x \in V$, if there exists $y \in P^{\pm}$, such that $sim(x, y) > \frac{3}{4}$, then for all other $y' \in P^{\pm}$, $sim(x, y') < \frac{3}{4}$.

Limitation of Gadget Construction

Let $\mathcal{D}$ be a distribution over $\left[\frac{3}{4}, 1\right]$ with probability density function $\mathcal{D}(t) = C \cdot e^{2t}$, for $t \in \left[\frac{3}{4}, 1\right]$.

Limitation of Gadget Construction

Let $\mathcal{D}$ be a distribution over $\left[\frac{3}{4}, 1\right]$ with probability density function $\mathcal{D}(t) = C \cdot e^{2t}$, for $t \in \left[\frac{3}{4}, 1\right]$.

1. Pick a number $t \sim \mathcal{D}$

Limitation of Gadget Construction

Let $\mathcal{D}$ be a distribution over $\left[\frac{3}{4}, 1\right]$ with probability density function $\mathcal{D}(t) = C \cdot e^{2t}$, for $t \in \left[\frac{3}{4}, 1\right]$.

1. Pick a number $t \sim \mathcal{D}$
2. Pick a uniformly random index $i \in [K]$, and let D_i be the corresponding dictator

Limitation of Gadget Construction

Let $\mathcal{D}$ be a distribution over $\left[\frac{3}{4}, 1\right]$ with probability density function $\mathcal{D}(t) = C \cdot e^{2t}$, for $t \in \left[\frac{3}{4}, 1\right]$.

1. Pick a number $t \sim \mathcal{D}$
2. Pick a uniformly random index $i \in [K]$, and let D_i be the corresponding dictator
3. For variable $x \in V$, if $sim(x, P^{\pm}) > t$, assign x the value of its closest primary variable, otherwise assign x according to D_i.

Limitation of Gadget Construction

Consider an equation $x + y = 0$.

Limitation of Gadget Construction

Consider an equation $x + y = 0$.

$sim(x, P^{\pm}) \leq 3/4$, $sim(y, P^{\pm}) \leq 3/4$

Expected *unsat* is $1/K$.

Limitation of Gadget Construction

Consider an equation $x + y = 0$.

$sim(x, P^{\pm}) \leq 3/4$, $sim(y, P^{\pm}) \leq 3/4$

Expected *unsat* is $1/K$.

$sim(x, P^{\pm}) > sim(y, P^{\pm}) =: s$, $sim(x, P^{\pm}) > 3/4$

We have $sim(x, P^{\pm}) = s + 1/K$, and $s \geq 3/4$.

Limitation of Gadget Construction

Consider an equation $x + y = 0$.

$sim(x, P^{\pm}) \leq 3/4$, $sim(y, P^{\pm}) \leq 3/4$

Expected *unsat* is $1/K$.

$sim(x, P^{\pm}) > sim(y, P^{\pm}) =: s$, $sim(x, P^{\pm}) > 3/4$

We have $sim(x, P^{\pm}) = s + 1/K$, and $s \geq 3/4$.

If $t \in [3/4, s)$, then $x + y = 0$ always satisfied.

Limitation of Gadget Construction

Consider an equation $x + y = 0$.

$sim(x, P^{\pm}) \leq 3/4$, $sim(y, P^{\pm}) \leq 3/4$

Expected *unsat* is $1/K$.

$sim(x, P^{\pm}) > sim(y, P^{\pm}) =: s$, $sim(x, P^{\pm}) > 3/4$

We have $sim(x, P^{\pm}) = s + 1/K$, and $s \geq 3/4$.
If $t \in [3/4, s)$, then $x + y = 0$ always satisfied.
If $t \in [s, s + 1/K)$, expected *unsat* is $1/2$.

Limitation of Gadget Construction

Consider an equation $x + y = 0$.

$sim(x, P^{\pm}) \leq 3/4$, $sim(y, P^{\pm}) \leq 3/4$

Expected *unsat* is $1/K$.

$sim(x, P^{\pm}) > sim(y, P^{\pm}) =: s$, $sim(x, P^{\pm}) > 3/4$

We have $sim(x, P^{\pm}) = s + 1/K$, and $s \geq 3/4$.

If $t \in [3/4, s)$, then $x + y = 0$ always satisfied.

If $t \in [s, s + 1/K)$, expected *unsat* is $1/2$.

If $t \in [s + 1/K, 1]$, expected *unsat* is $1/K$.

Summary

- Optimal gadget reduction for E2LIN$_2$ from stronger hardness guarantees for Had_3-CSP
- Limitations of gadget construction for showing hardness of E2LIN$_2$

Thank you!