CS 2429 - Foundations of Communication Complexity

Lecturer: Mika Göös

1 The Clique vs. Independent Set Problem

The *Clique vs. Independent Set* (CIS) is a problem in two-player communication complexity. The CIS game is played on an undirected n vertex graph $G = (V, E)$ that is known to both the players Alice and Bob. As input, Alice gets a set $C \subseteq V$ that spans a clique in G while Bob holds an independent set $I \subseteq V$. The objective is to decide whether $C \cap I = \varnothing$, i.e., compute the value

$$\mathrm{CIS}_G(C, I) := |C \cap I|.$$

Thus, the CIS problem is a highly structured special case of the disjointness problem; for example, we always have that $\mathrm{CIS}_G(C, I) \in \{0, 1\}$.

Interestingly, for many graphs G the deterministic communication complexity $D(\mathrm{CIS}_G)$ is not well-understood; in this note we explain why the CIS problem is important and survey some of the known results.

2 Original Motivation

The CIS problem was first defined in a seminal work of Yannakakis [11] that studies the minimum number of linear inequalities that are needed to *express* the feasible polytope $F \subseteq \mathbb{R}^n$ of a combinatorial optimization problem.

Example Denote by $x_S \in \{0, 1\}^V$ the characteristic vector of a subset $S \subseteq V$. The *vertex packing polytope* $F \subseteq \mathbb{R}^n$ of G is the convex hull of the points x_I where $I \subseteq V$ is an independent set.

Here, a linear program on variables $x_1, \ldots, x_n, y_1, \ldots, y_m$ *expresses* F if the feasible set of the program projects onto F as we ignore the variables $y_1, \ldots, y_m$. Whenever one can express the polytope F as a polynomial sized LP (sometimes called an *extended formulation* for F), one can solve linear optimization problems on F in polynomial time using any of the polynomial-time LP solvers. However, in [11] it was shown that many NP-hard optimisation problems do not admit such polynomial-size extended formulations.

The connection to communication complexity is the following. First, define the *non-negative rank*, $\mathrm{rk}^+(A)$, of a non-negative real $n \times n$ matrix A as the minimum r such that A can be written as PQ, where P and Q are non-negative and of dimensions $n \times r$ and $r \times n$, respectively. (Non-negative rank is a slippery concept: it is not obvious that $\mathrm{rk}^+(A)$ is even computable [2], and determining it exactly is NP-hard [10].)

If A is a boolean matrix, we can bound $\mathrm{rk}^+(A)$ from below in terms of the non-deterministic communication complexity of A:

Proposition 1 *$N^1(A) \le \log \mathrm{rk}^+(A)$.*

This follows because every non-negative factorization $A = PQ$ gives us a non-deterministic cover of the 1-entries of A: the i-th rectangle in the cover is defined by the non-zero entries in the i-th column of P and the i-th row of Q.

Yannakakis [11, Thm 3] characterised the minimum size of an LP expressing a polytope F as the non-negative rank of a certain *slack matrix* related to F. In case of the vertex packing polytope above, the communication matrix of CIS_G appears in connection to this slack matrix. In particular, superlogarithmic lower bounds on the communication complexity of CIS_G imply superpolynomial lower bounds on the size of extended formulations for the vertex packing polytope.

3 Unambiguous Covers

If $C \cap I \neq \varnothing$, it is easy to prove this fact by non-deterministically guessing the unique vertex in the intersection; this yields a non-deterministic protocol having complexity

$$N^1(\mathrm{CIS}_G) = O(\log n).$$

What is interesting about this protocol is that the *yes*-certificate is *unambiguous* in that there is at most one *yes*-certificate for any input pair (C, I). In terms of the communication matrix of CIS_G this protocol corresponds to a *non-overlapping* cover of the 1-entries of CIS_G.

By analogy, in classical complexity theory the class UP (unambiguous polynomial-time) consists of problems that can be solved by an NP machine where every *yes*-instance has exactly one accepting computation path [1].

In the following, we will argue that, in a certain sense,

The CIS *problem is complete for* UP.

Suppose A is any boolean communication matrix that admits a non-overlapping cover of its 1-entries by rectangles $R_1, \ldots, R_n$. We show how this can be converted into a CIS instance, which we call CIS_A: The vertices of the underlying graph correspond to the rectangles $R_1, \ldots, R_n$, and we let R_i and R_j be connected by an edge iff R_i and R_j intersect in rows. Now, the input for the communication problem defined by A is given by a pair (a, b) where a indexes a row and b a column of A. In case of the CIS_A instance this corresponds to Alice being given the clique $C_a = \{R_i : R_i \text{ intersects row } a\}$ and Bob being given the independent set $I_b = \{R_i : R_i \text{ intersects column } b\}$. By inspecting the construction we have that $\mathrm{CIS}_A(C_a, I_b) = A_{ab}$. This proves the following theorem.

Theorem 2 *$D(A) \le D(\mathrm{CIS}_A)$.*

4 Basic Bounds

Yannakakis gave the following upper bound for any G, which is still the best one known.

Theorem 3 ([11]) *$D(\mathrm{CIS}_G) = O(\log^2 n)$.*

The proof of Theorem 3 is very similar to the "$\mathsf{P} = \mathsf{NP} \cap \mathsf{coNP}$" type result that we saw in lecture 2 which states that $D(f) \le O(N^0(f) N^1(f))$; note how this upper bound is small only if *both* the 0-entries and 1-entries have small covers. By contrast, Theorems 2 and 3 above capture a "$\mathsf{P} = \mathsf{UP}$" type result where only the 1-entries are required to have small non-overlapping covers.

Proof [of Theorem 3] We describe an $O(\log^2 n)$ protocol for CIS_G. The protocol performs a binary search for the intersection. It proceeds in $O(\log n)$ rounds where each round involves $O(\log n)$ bits of communication.

At the start of a round

Alice checks whether her clique C contains a node v with $\deg(v) < n/2$.
Bob checks whether his independent set I contains a node u with $\deg(u) \geq n/2$.

Alice and Bob communicate their findings by exchanging the names ($\log n$ bits) of the nodes v and u if such were found.

If both of the players fail at finding a node satisfying the above, this proves that $C \cap I = \varnothing$ and the protocol concludes (with output 0). Otherwise, suppose Alice finds a node v with $\deg(v) < n/2$ (the case of Bob finding an u is analogous). In this case we must have that $C \subseteq \Gamma(v)$, where $\Gamma(v)$ is the (inclusive) neighbourhood of v. Thus, both Alice and Bob can reduce the size of the problem from n to $n/2$ by executing the next round recursively on the subgraph induced on the vertices $\Gamma(v)$.

The bound $D(f) \leq O(N^0(f)N^1(f))$ is known to be tight as is demonstrated by the problem $\mathrm{DISJ}_{\log n}$, i.e., the disjointness problem where the inputs are $\log n$ sized subsets of $\{1, \ldots, n\}$ [7, Example 2.12]. By contrast, Theorem 3—the best existing upper bound—is not known to be tight!

Indeed, the trivial lower bound $D(\mathrm{CIS}_G) \geq \log n$ follows from considering the case $|C| = |I| = 1$, i.e., the *equality* function. Currently, the only non-trivial lower bound for a specific G is

$$D(\mathrm{CIS}_G) \geq (2 - o(1)) \log n$$

given by Kushilevitz et al. [6]. Interestingly, this lower bound necessarily requires methods stronger than the common *partition bound*: the authors show that for their choice of G, CIS_G has a monochromatic partition of size $C^D(\mathrm{CIS}_G) = O(n)$ which yields only $D(\mathrm{CIS}_G) \geq \log C^D(\mathrm{CIS}_G) = \log n + O(1)$.

5 Recent Progress

Kushilevitz and Weinreb [8, 9] have recently studied special cases and relaxations of the CIS problem. In [9] they consider the deterministic communication complexity of the *promise* version of $\mathrm{DISJ}_{\log n}$ where the input sets have at most one element in common. This setting models the CIS_G problem on *random* graphs G since random graphs have maximum cliques and maximum independent sets of size $\Theta(\log n)$ with high probability. Here, they obtain a near-tight lower bound of

$$D(\text{Promise-}\mathrm{DISJ}_{\log n}) = \Omega(\log^2 n/\mathsf{polylog}(\log n)).$$

In [8] they study the CIS problem in case $|C| = |I| = 2$, i.e., Alice gets an edge as input and Bob gets a non-edge. For these restricted problems one always has the trivial bounds $\log n \leq D(\mathrm{CIS}_G) \leq 2 \log n$. They use the Karchmer–Wigderson [5] approach to circuit depth to show that proving lower bounds larger than $\log n$ for explicit graphs G requires proving lower bounds on "graph complexity" (see, e.g., Jukna [4, §1.7]). They argue that this might explain part of the difficulty of obtaining non-trivial lower bounds for the unrestricted CIS problem.

The CIS problem is also related to the so-called Alon–Saks–Seymour conjecture in graph theory. Recently, this conjecture was disproved [3] implying a lower bound for the *non-deterministic* complexity $N^0(\mathrm{CIS}_G) = N^1(\overline{\mathrm{CIS}}_G) \geq \frac{6}{5}\log n$, where $\overline{\mathrm{CIS}}_G$ is the function $1 - \mathrm{CIS}_G$.

References

[1] S. Aaronson, G. Kuperberg, and C. Granade, *Complexity zoo*. `http://qwiki.stanford.edu/index.php/Complexity_Zoo`, Oct. 2012.

[2] J. E. Cohen and U. G. Rothblum, *Nonnegative ranks, decompositions, and factorizations of nonnegative matrices*, Linear Algebra and its Applications, 190 (1993), pp. 149 – 168.

[3] H. Huang and B. Sudakov, *A counterexample to the Alon–Saks–Seymour conjecture and related problems*, Combinatorica, 32 (2012), pp. 205–219.

[4] S. Jukna, *Boolean Function Complexity: Advances and Frontiers*, vol. 27 of Algorithms and Combinatorics, Springer, 2012.

[5] M. Karchmer and A. Wigderson, *Monotone circuits for connectivity require super-logarithmic depth*, in STOC, 1988, pp. 539–550.

[6] E. Kushilevitz, N. Linial, and R. Ostrovsky, *The linear-array conjecture in communication complexity is false*, Combinatorica, 19 (1999), pp. 241–254.

[7] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, 1997.

[8] E. Kushilevitz and E. Weinreb, *The communication complexity of set-disjointness with small sets and 0-1 intersection*, in FOCS, 2009, pp. 63–72.

[9] ——, *On the complexity of communication complexity*, in STOC, 2009, pp. 465–474.

[10] S. A. Vavasis, *On the complexity of nonnegative matrix factorization*, SIAM J. on Optimization, 20 (2009), pp. 1364–1377.

[11] M. Yannakakis, *Expressing combinatorial optimization problems by linear programs*, J. Comput. Syst. Sci., 43 (1991), pp. 441–466.