Lower bounds for Lovász-Schrijver systems and beyond follow from multiparty communication complexity *

Paul Beame[1], Toniann Pitassi[2], and Nathan Segerlind[1]

[1] Computer Science and Engineering, University of Washington, Seattle, WA 98195-2350
[2] Computer Science Department, University of Toronto, Toronto, ON M5S 1A4

Abstract. We prove that an $\omega(\log^3 n)$ lower bound for the three-party number-on-the-forehead (NOF) communication complexity of the set-disjointness function implies an $n^{\omega(1)}$ size lower bound for tree-like Lovász-Schrijver systems that refute unsatisfiable CNFs. More generally, we prove that an $n^{\Omega(1)}$ lower bound for the $(k+1)$-party NOF communication complexity of set-disjointness implies a $2^{n^{\Omega(1)}}$ size lower bound for all tree-like proof systems whose formulas are degree k polynomial inequalities.

1 Introduction

Linear programming, the problem of optimizing a linear objective function over the points of a given polyhedron, was shown to be polynomial-time solvable over the rationals by Khachian [15]. When integrality constraints are added, however, the resulting integer linear programming problem becomes NP-hard. Many algorithms for such problems attempt to apply efficiencies from rational linear programming to the integral case.

One of the most powerful of such approaches is to begin with the polytope defined by the original linear program without integrality constraints and systematically pare down the polytope by repeatedly refining the linear program with "cutting planes" that remove only nonintegral solutions until we are left with the convex hull of the integral solutions. These are local methods in which the initial polytope Q (expressed by the natural cutting planes constraints) is transformed through a sequence of local operations to smaller and smaller polytopes (each contained in the original one), until the integral hull of Q is reached. (At this point, rational linear programming will find the correct solution.) For decision problems, this sequence terminates with the empty polytope if and only if the initial polytope contains no integral points.

One such method is that of Gomory-Chvátal cuts [6] which derives each new cutting plane as a linear combination and shift of existing facet constraints. There are even more subtle methods available, particularly in the case of 01-programming, which is also NP-complete. In a seminal paper, Lovász and Schrijver [16] introduced a variety of

* Paul Beame's research was supported by NSF grants CCR-0098066 and ITR-0219468. Toniann Pitassi's research was supported by an Ontario Premiere's Research Excellence Award, an NSERC grant, and the Institute for Advanced Study where this research was done. Nathan Segerlind's research was supported by NSF Postdoctoral Fellowship DMS-0303258 and done while at the Institute for Advanced Study.

cutting planes methods that derive new cutting planes by first "lifting" the inequalities to higher degree polynomial inequalities (in particular quadratic inequalities) and then "projecting" them down to linear inequalities using polynomial identities and the fact that $x^2 = x$ for $x \in \{0, 1\}$. These systems are now known as *Lovász-Schrijver systems (LS)*.

It may be too costly to apply these techniques to pare all the way down to the integral hull. However, even applying a smaller number of rounds of the procedure can often lead to a smaller polytope that has good approximability ratio, one for which the best nonintegral solution is not too far away from the best integral solution, so that by rounding we can achieve a good approximation to the optimal value.

Two complexity measures are commonly studied for Lovász-Schrijver and related cutting planes proof systems: *size* and *rank*. Intuitively, rank is the number of intermediate polytopes that must be passed through before arriving at the integral hull. In [16] it was shown that for any (relaxed) polytope P, if the rank of P is d, then the optimization and decision problems for P can be solved exactly deterministically in time $n^{O(d)}$. This very nice algorithmic property of Lovász-Schrijver systems makes them especially appealing for solving or approximating NP-hard optimization problems via linear programming. A variety of rank lower bounds for exact solution are known, even for the case of unsatisfiable systems [4, 8, 11, 7, 12]. Moreover, interesting bounds on the ranks required for good approximations to vertex cover [1] and MaxSAT [5] have been obtained. This, in turn, implies inapproximability results for these problems for *any* polynomial-time algorithm based on rank.

While there is a rich and growing body of results concerning rank, very little is known about the size of LS proofs. Informally, the size of an LS procedure with respect to some polytope P is the smallest number of hyperplanes defining all of the polytopes that we need to pass through before arriving at the integral hull. Clearly size lower bounds imply rank lower bounds, and even tree-size lower bounds imply rank lower bounds, but the converse is not known to be true. The one unconditional (tree-like) size lower bound known for LS [12] is for a family of polytopes for which decision and optimization are trivial and for which the integral hull has a trivial derivation in Chvátal's cutting planes proof system.

Problems in which the facets represent clauses of a CNF formula and a decision algorithm for 01-programming yields a propositional proof system are particularly important to analyze. Proving (tree-like) size lower bounds for such polytopes was given as one of the main open problems in [12]. The only LS size lower bounds known at present for such polytopes formulas are conditional results. First, it is an easy observation that $\mathsf{NP} \neq \mathsf{coNP}$ implies superpolynomial LS size lower bounds for some family of unsatisfiable CNF formulas. It has also been shown by [19, 9, 10] that these lower bounds also hold under other natural complexity assumptions.

In this paper we develop a new method for attacking size lower bounds for LS and for systems that generalize LS. Our main result is a proof that lower bounds on the 3-party communication complexity of set disjointness (in the number-on-forehead model) imply lower bounds on the size of tree-like LS proofs for a particular family of unsatisfiable CNF formulas. We also generalize this result to a much more powerful family of proof systems known as semantic LS^k, where lines are now degree k polynomial

inequalities. All versions of LS are special cases of LS^2, and Chvátal's Cutting Planes proof system is a special case of LS^1.

More generally, we show that proving lower bounds on the $(k+1)$-party communication complexity of set disjointness implies lower bounds on the size of tree-like semantic LS^k proofs. By a natural extension of the ideas in [2] one can show that the $(k+1)$-party set disjointness problem is "complete" for the $(k+1)$-party communication complexity class $(k+1)$-NP^{cc} and a lower bound showing that it is not in $(k+1)$-RP^{cc} would already given excellent lower bounds for LS^k proofs. Such a result is already known in the case $k=1$ [2] (and was used in [13] to derive tree-like size lower bounds for Chvátal's Cutting Planes system) and set disjointness is one of the most well-studied problems in communication complexity.

Our proof can be seen as a generalization of [13] to arbitrary k but the extension requires a number of new ideas and a substantially more complicated argument that includes a detailed analysis of large sets of vertex-disjoint paths in expander graphs.

2 Definitions

2.1 Multiparty Communication Complexity and Set Disjointness

The *k-party number-on-the-forehead (NOF) model of communication complexity* computes functions (or relations) of input vectors $(x_1, \ldots, x_k) \in X_1 \times \ldots \times X_k$ distributed among k parties, such that party $i \in [k]$ sees all x_j for all $j \in [k]$, $j \neq i$.

The *k-party set disjointness problem* $\textsc{Disj}_{k,n} : (\{0,1\}^m)^k \to \{0,1\}$ is defined by $\textsc{Disj}_{k,n}(\vec{x}) = 1$ iff there is some $j \in [n]$ such that $x_{i,j} = 1$ for all $i \in [k]$. (We follow standard terminology although it might be more appropriate to call this set intersection rather than disjointness.)

A *$(0,\epsilon)$-error k-party NOF communication protocol* for set disjointness is a protocol that for every disjoint input produces output 0 and for intersecting inputs outputs 1 with probability at least $1-\epsilon$.

It is conjectured that for any $k \geq 2$ the k-party set disjointness problem requires nearly linear randomized NOF communication complexity. This conjecture is equivalent showing that nondeterministic k-party communication complexity can be almost optimally separated from randomized k-party communication complexity. The conjecture is proven for $k=2$ [14], but the best known lower bound for $k \geq 3$ is $\Omega(\log n)$ for general models and $\Omega(n^{1/k})$ for more restricted models [3].

2.2 Threshold Logics and the Complexity of a Search Problem

The two most prevalent classes of threshold logics are Gomory-Chvátal cutting planes [6], and matrix cuts, defined by Lovász and Schriver [16]. These proof systems, CP LS, LS_0, and LS_+, are special cases of more general *semantic* threshold logic proof systems.

A *k-threshold formula* over Boolean variables $x_1, \ldots, x_n$ is a formula of the form $\sum_j \gamma_j m_j \geq t$, where γ_j, t are integers, and for all j, m_j is a multilinear monomial of degree at most k. The *size* of a k-threshold formula is the sum of the sizes of γ_j and t, written in binary notation.

Let f_1, f_2, g be k-threshold formulas in the variables $\vec{x}$. We say that *g is semantically entailed* by f_1 and f_2 if for every $0/1$ assignment to $\vec{x}$ that satisfies both f_1 and f_2, g is also satisfied.

Let f be an unsatisfiable CNF formula over $x_1, \ldots, x_n$, and let $t_1, \ldots, t_m$ be the underlying set of clauses of f, written as 1-threshold inequalities. A **Th(k)** *refutation of* f, $\mathcal{P}$, is a sequence of k-threshold formulas, $L_1, \ldots, L_q$, where each L_j is one of the inequalities t_i, $i \in [m]$, or is semantically entailed by two formulas L_i and $L_{i'}$ with $i, i' < j$, and the final formula L_q is $0 \geq 1$. The *size* of $\mathcal{P}$ is the sum of the sizes of all k-threshold formulas occurring in $\mathcal{P}$. The proof is *tree-like* if the underlying directed acyclic graph, representing the implication structure of the proof, is a tree. (That is, every formula in the proof, except for the formulas from f, is used at most once as an antecedent of an implication.)

CP refutations are a special case of **Th(1)** semantic refutations, and thus lower bounds for tree-like **Th(1)** semantic refutations imply similar lower bounds for tree-like CP. (This was already shown in [13].)

As mentioned earlier, since we can assume that any of the Lovász-Schrijver systems can be assumed to have fan-in two, it follows that any of the systems LS_0, LS and LS^+ can easily be converted into **Th(2)** semantic refutations with at most a polynomial increase in size, and if the original proof is tree-like, so is the semantic refutation. Thus, lower bounds for tree-like **Th(2)** semantic refutations imply similar lower bounds for all tree-like Lovász-Schrijver systems.

Let f be an unsatisfiable CNF formula. We will be interested in the following search problem, $Search_f$ associated with f: given a truth assignment α, find a clause from f which is falsified by α. The model for this computation is a decision tree whose nodes evaluate polynomial threshold functions:

A *k-threshold decision tree* is a rooted, directed tree whose vertices are labeled with k-threshold functions and edges are labeled with either 0 or 1. The leaves of the tree are labeled with clauses of f. A k-threshold decision tree solves $Search_f$ in the obvious way: start at the root and evaluate the threshold function; follow the edge that is consistent with the value of the threshold function; continue until the computation reaches a leaf and output the associated clause. The size S of a k-threshold decision tree is the sum of the sizes of all threshold formulas in the tree, where the coefficients are written in binary. The depth of a k-threshold decision tree is the depth of the underlying tree.

Theorem 1. *Suppose that f has a tree-like* **Th(k)***-semantic refutation of size S. Then there exists a $k+1$-party 0-error randomized NOF communication complexity protocol for $Search_f$ (over any partition of the variables into k groups) that communicates $O(\log^3 S)$ bits and produces an answer with probability at least $1 - 1/n$.*

Further, if all k-threshold formulas in the **Th(k)***-semantic refutation have coefficients bounded by a polynomial in n, then the 0-error randomized communication complexity can reduced to $O(\log S(\log\log n)^2)$ or the protocol can be made deterministic using $O(\log S \log n)$ bits.*

Proof (Sketch). First, following ideas similar to the degree 1 case in [13], we recursively search the proof tree using the $\frac{1}{3}$-$\frac{2}{3}$ trick to derive a k-threshold decision tree for $Search_f$ of depth $O(\log S)$ and size $O(S)$. Then, adapting arguments from [18], we show that any relation computed by a shallow k-threshold decision tree can also be efficiently computed by a $k+1$ player communication complexity protocol (number-on-forehead model), over any partition of the variables.

2.3 k-fold Tseitin formulas

Our hard examples are based on the well-known Tseitin graph formulas. Let $G = (V, E)$ be any connected, undirected graph and let $\vec{c} \in \{0,1\}^V$. The *Tseitin formula for G with respect to charge vector $\vec{c}$*, $TS(G, \vec{c})$, has variables $\text{Vars}(G) = \{y_e \mid e \in E\}$. The formula states that for every vertex $v \in V$, the parity of the edges incident with v is equal to the charge, c_v, at node v. It is expressed propositionally as the conjunction of the clauses obtained by expanding $\oplus_{e \ni v} y_e = c_v$ for each $v \in V$. For a graph with maximum degree d, each clause is of width $\leq d$ and the number of clauses is $\leq |V| 2^d$.

$TS(G, \vec{c})$ is satisfiable if and only if $\sum_{v \in V} c_v$ is even. For odd $\vec{c}$, $Search_{TS(G,\vec{c})}$ takes a 0/1 assignment α to $\text{Vars}(G)$ and outputs a clause of $TS(G, \vec{c})$ that is violated. In particular, a solution to $Search_{TS(G,\vec{c})}$ will produce a vertex v such that the parity equation associated with vertex v is violated by α.

To make the search problem hard for k-party NOF communication protocols (and thus, by Theorem 1, hard for $k-1$-threshold decision trees) we modify $TS(G, \vec{c})$ by replacing each variable y_e by the conjunction of k variables, $\bigwedge_{i=1}^{k} y_e^i$, and expanding the result into clauses. We call the resulting *k-fold Tseitin formula*, $TS^k(G, \vec{c})$, and its variable set, $\text{Vars}^k(G) = \{y_e^i \mid e \in E, i \in [k]\}$.

For a fixed graph G and different odd-charge vectors $\vec{c} \in \{0,1\}^{V(G)}$, the various problems $Search_{TS^k(G,\vec{c})}$ are very closely related. Define $\text{ODDCHARGE}^k(G)$ to be the k-party NOF communication search problem which takes as input an odd charge vector $\vec{c} \in \{0,1\}^{V(G)}$, seen by all players, and an assignment α to $\text{Vars}^k(G)$, in which player i sees all values but the assignment α_e^i to y_e^i for $e \in E(G)$, and requires that the players output a vertex v that is a solution to $Search_{TS^k(G,\vec{c})}$.

3 Reduction from Set Disjointness to ODDCHARGE

We give a sequence of reductions to show that for a suitably chosen graph G, an efficient k-party NOF communication complexity protocol for $\text{ODDCHARGE}^k(G)$ will imply an efficient 1-sided error randomized k-party NOF protocol for the set disjointness relation.

We apply the Valiant-Vazirani argument to show that, without loss of generality, it suffices to derive a 1-sided error protocol for a version of set disjointness in which the input has intersection size 0 or size 1, and the job of the players is to distinguish between these two cases. We call this promise problem *zero/one set disjointness*.

Our reduction from zero/one set disjointness to $\text{ODDCHARGE}^k(G)$ goes via an intermediate problem, $\text{EVENCHARGE}^k(G)$, which is the exact analog of $\text{ODDCHARGE}^k(G)$ except that the input charge vector $\vec{c}$ is even rather than odd and the requirement is *either* to find a charge violation or to determine that no charge violation exists.

The reduction from $\text{EVENCHARGE}^k(G)$ to $\text{ODDCHARGE}^k(G)$, which is similar in spirit to a reduction of Raz and Wigderson [20], works by planting a single randomly chosen additional charge violation. This yields a protocol for $\text{EVENCHARGE}^k(G)$ that works well on average for each class of inputs with a given number of charge violations.

The most difficult part of our argument is the reduction from zero/one set disjointness to $\text{EVENCHARGE}^k(G)$ for suitable graphs G. The key idea is that for even $\vec{c}$, charge violations of $TS^k(G, \vec{c})$ come in pairs: Given an instance $\vec{x} \in (\{0,1\}^m)^k$ of

zero/one set disjointness, using the public coins, the players randomly choose an even charge vector $\vec{c}$ and m vertex-disjoint paths in G, $p_1, \ldots, p_m$, for each $j \in [m]$, the players plant the $x_{1,j}, \ldots, x_{k,j}$ as the assignment along each edge of path p_j, in a random solution that otherwise meets the chosen charge constraint. By construction, a charge violation can occur only at the endpoints of a path and only if there is an intersection in the set disjointness problem.

It is tricky to ensure that the resulting problem looks sufficiently like a random instance of EVENCHARGE$^k(G)$ with either 0 or 2 charge violations so that we can apply the average case properties of the protocol for EVENCHARGE$^k(G)$. This places major constraints on the graph G and in particular requires that $m \leq n^{1/3}/\log n$ where $|V(G)| = n$. The bulk of the work is in showing that a small number of specific properties: rapid mixing, modest degree, and high girth – properties all met by a family of expanders constructed in [17] – are sufficient.

Distributions on labeled graphs For the rest of the paper in the Tseitin tautologies we will use a family of graphs H_n that is the union of two edge-disjoint graphs on the same set of n vertices $[n]$, G_n and T_n. G_n will be a Δ-regular expander graph of the form defined by Lubotzky, Phillips, and Sarnak [17] for $\Delta = \Theta(\log n)$. Since $\overline{G_n}$ has degree $> n/2$, there is a spanning tree T_n of maximum degree 2 (a Hamiltonian path) in $\overline{G_n}$. Clearly H_n also has maximum degree $\Theta(\log n)$ and thus $TS^k(H_n, \vec{c})$ has size $n^{O(k)}$.

Let H_n be such a graph and let $\vec{c}$ be an even charge vector. We define $Sol(H_n, \vec{c})$ to be the set of all 0/1 assignments to the edges of H_n so that for each vertex $v \in [n]$, the parity of edges incident with v is equal to c_v. A uniform random distribution over $Sol(H_n, \vec{c})$ can be obtained by first selecting 0/1 values uniformly at random for all edges in G_n and then choosing the unique assignment to the edges of T_n that fulfill the charge constraints given by $\vec{c}$.

Given a bit value b associated with an edge $e \in G_n$, we can define a uniform distribution $\mathcal{L}^k = \mathcal{L}_k(b)$ over the corresponding variables y_e^i, $i \in [k]$. Such an assignment is chosen randomly from $\mathcal{L}_k$ on input b by the following experiment. If $b = 1$ then set all variables associated with edge e, y_e^i, $i \in [k]$ to 1. Otherwise if $b = 0$, set the vector $(\vec{y}_e)_{i \in [k]}$ by choosing uniformly at random from the set of $2^k - 1$ not-all-1 vectors.

Definition 1. *For any $t \geq 0$ let $\mathcal{D}_t$ be a distribution given by the following experiment on input $H_n = G_n \cup T_n$.*

1. *Choose an even charge vector $\vec{c} \in \{0,1\}^n$ uniformly at random.*
2. *Choose some $\beta \in Sol(H_n, \vec{c})$ uniformly at random.*
3. *For each $e \in G_n$, select the values for the vector $(y_e)_{i \in [k]}$ from $\mathcal{L}_k(\beta_e)$ and for each $e \in T_n$, set $y_e^i = \beta_e$ for all $i \in [k]$.*
4. *Select a random subset $U \subseteq [n]$ of $2t$ vertices and produce charge vector $\vec{c}^U$ from $\vec{c}$ by toggling all bits c_v for $v \in U$.*
5. *Return the pair $(\alpha, \vec{c}^U)$ where α is the boolean assignment to the variables y_e^i, $i \in [k]$, $e \in H_n$.*

Reduction from EVENCHARGE **to** ODDCHARGE

Lemma 1. *Let G be any connected graph on n vertices and let $\Delta(G)$ be the maximum degree in G. Suppose that Π_{odd} is a randomized k-party NOF protocol for*

ODDCHARGE$^k(G)$ *that produces an answer with probability at least* $1-\epsilon$*, is correct whenever it produces an answer, and uses at most* s *bits of communication. Then there is a randomized* k*-party NOF protocol* Π_{even} *for* EVENCHARGE$^k(G)$ *that uses* $s + \Delta(G)$ *bits of communication and has the following performance:*

$$\Pr_{(\alpha,\vec{c})\in\mathcal{D}_0}[\Pi_{even}(\alpha,\vec{c}) = \text{true}] = 1$$

$$\Pr_{(\alpha,\vec{c})\in\mathcal{D}_t}[\Pi_{even}(\alpha,\vec{c}) \in \mathit{Err}(\alpha,\vec{c})] \geq 2/3 - \epsilon \text{ for } t \geq 1.$$

Proof. Let Π_{odd} be a protocol for ODDCHARGE$^k(G)$ and assume that $V(G) = [n]$. We give a protocol Π_{even} for EVENCHARGE$^k(G)$. On input $(\alpha, \vec{c})$ and random public string r: Using r, choose a random vertex $v \in [n]$. Check whether the parity equation associated with vertex v is satisfied by α using at most $\Delta(G)$ bits of communication. If it is not, return v. Otherwise, create an odd charge vector, $\vec{c}^{\,\{v\}}$, which is just like $\vec{c}$ except that the value of c_v is toggled. Now run Π_{odd} on input $(\vec{c}^{\,\{v\}}, \alpha)$. If Π_{odd} returns the planted error v or if Π_{odd} does not return a value then return "true"; if Π_{odd} returns $u \neq v$, output u.

Suppose that $(\alpha, \vec{c}) \in \mathcal{D}_0$. Then α satisfies all charges specified by $\vec{c}$, so when Π_{odd} returns a vertex the above protocol must output "true" because Π_{odd} has one-sided error–that is, Π_{odd} will only return a vertex u when there is an error on the parity equation associated with u. Now suppose that $(\alpha, \vec{c}) \in \mathcal{D}_t$ so exactly $2t$ parity equations are violated. If the random vertex v does not satisfy its parity constraints, then the algorithm is correct. The remaining case is when v satisfies the parity equation and in this case we call Π_{odd} on a pair $(\alpha, \vec{c}^{\,\{v\}})$ where exactly $2t+1$ parity equations are violated.

We show the probability bound separately for each $T \in [n]^{(2t+1)}$. Because the events $\text{Err}(\alpha, \vec{c'}) = T$ partition the probability space, this proves the claim. By symmetry, for $T \in [n]^{(2t+1)}$ and *any* function g with codomain T, we have that $\Pr_{\alpha,\vec{c},v}[g(\alpha, \vec{c}^{\,\{v\}}) = v \mid \text{Err}(\alpha, \vec{c}^{\,\{v\}}) = T] = 1/(2t+1)$ since it is equally likely for $\vec{c'} = \vec{c}^{\,\{v\}}$ to be generated as $\vec{c}^{\,\{u\}}$ for any $u \in T$. Thus we obtain:

$$\begin{aligned}
&\Pr_{\alpha,\vec{c},v}[\Pi_{even}(\alpha, \vec{c}^{\,\{v\}}) \text{ errs} \mid \text{Err}(\alpha, \vec{c}^{\,\{v\}}) = T] \\
&= \Pr_{\alpha,\vec{c},v}[\Pi_{odd}(\alpha, \vec{c}^{\,\{v\}}) = v \text{ or } \Pi_{odd}(\alpha, \vec{c}^{\,\{v\}}) \text{ is not defined } \mid \text{Err}(\alpha, \vec{c}^{\,\{v\}}) = T] \\
&\leq 1/(2t+1) + \epsilon \;\leq\; 1/3 + \epsilon \qquad \text{for } t \geq 1.
\end{aligned}$$

Reduction from Zero/One Set Disjointness to EVENCHARGE**:** We now show how to use a k-party NOF communication complexity protocol Π_{even} for EVENCHARGE$^k(H_n)$ as guaranteed by Lemma 1 to produce a k-party NOF protocol for the zero/one set disjointness problem which uses the following definition.

Definition 2. *Let* $P_l^{(m)}$ *be the set of all sequences of* m *vertex-disjoint length* l *paths in* G_n*.*

Lemma 2. *Let* $m = n^{1/3}/\log n$*. For sufficiently large* n *and for any even charge vector* $\vec{c}$*, if there is a probabilistic* k*-party NOF communication complexity protocol,* Π_{even} *for* EVENCHARGE$^k(H_n)$ *using* s *bits, satisfying the conditions in Lemma 1 for* $\mathcal{D}_0$ *and*

$\mathcal{D}_1$, then there is a randomized $(0, 1/3 + \epsilon + o(1))$ error k-party NOF communication complexity protocol Π_{01disj} for zero/one set disjointness on input $\vec{x} \in (\{0,1\}^m)^k$ that uses s bits of communication.

Proof. Let $\vec{x}$ be an instance of zero/one set disjointness. Protocol Π_{01disj} will call Π_{even} on the graph H_n, on a pair $(\alpha, \vec{c})$ chosen according to the following distribution/experiment:

1. On input $\vec{x}$ with public coins r:
   (a) Using public coins r, choose a random even charge vector $\vec{c} \in \{0,1\}^n$.
   (b) Using public coins r, choose a sequence of m vertex-disjoint length l paths, $p_1, \ldots p_m$ uniformly at random from $P_l^{(m)}$.
   (c) Using the public coins r, choose $\beta \in Sol(H_n - \bigcup_{j=1}^m p_j, \vec{c})$
2. For all edges $e \in H_n$, all players other than player i compute α_e^i as follows:
   (a) If $e \in p_j$ for $j \in [m]$, set $\alpha_e^i = x_{i,j}$
   (b) If $e \in G_n$ and $e \notin \bigcup_{j=1}^m p_j$, choose the vector $\alpha_e^1 \ldots \alpha_e^k$ according to the distribution $\mathcal{L}_k(\beta_e)$.
   (c) For the remaining edges $e \in T_n$, set all variables α_e^i for $i \in [k]$ equal to β_e.
3. Return $(\alpha, \vec{c})$

We write $\mathcal{R}(\vec{x})$ to denote the distribution on assignment/charge pairs produced by reduction Π_{01disj} when given an input $\vec{x}$. The following lemma, proven in section 4, has the main technical argument and shows that for $t = |\cap \vec{x}| \in \{0,1\}$, although $\mathcal{R}(\vec{x})$ is not the same as $\mathcal{D}_t$, $\mathcal{R}(\vec{x})$ is close to the distribution $\mathcal{D}_t$ in the ℓ_1 norm.

Lemma 3. *Let $\vec{x} \in (\{0,1\}^m)^k$ and $|\cap \vec{x}| = 1$. Then $||\mathcal{R}(\vec{x}) - \mathcal{D}_1||_1$ is $o(1)$.*

Protocol Π_{01disj} will output 0 if Π_{even} returns "true" and 1 otherwise. If $\cap\vec{x} = \emptyset$, by the above construction, the support of $\mathcal{R}(\vec{x})$ is contained in that of $\mathcal{D}_0$ and thus on $\mathcal{R}(\vec{x})$, Π_{even} must answer "true" and the vector $\vec{x}$ is correctly identified as being disjoint. In the case that $\cap\vec{x}$ contains exactly one element, $\Pr[\Pi_{01disj}(\vec{x})) = 0] \geq 2/3 - \epsilon - o(1)$. This completes the proof of the Lemma 2.

Reduction from Set disjointness to Zero/One Set disjointness

Lemma 4. *If there is an $(0, \epsilon)$ randomized NOF protocol for the k-party zero-one-promise set-disjointness problem that uses s bits of communication where ϵ is a constant < 1, then there is a $(0, \frac{1}{3})$ randomized NOF protocol for the k-party set-disjointness problem that uses $O(s \log n)$ bits of communication.*

Naturally, our starting point is the well-known result of Valiant and Vazirani [21].

Lemma 5 (Valiant-Vazirani). *Let a be a positive integer. Fix a nonempty $S \subseteq \{0,1\}^a$, and choose $w_1, \ldots w_a \in \{0,1\}^a$ independently and uniformly. With probability at least $1/4$, there exists $j \in \{0, \ldots, a\}$ so that $|\{x \in S \mid \forall i \leq j,\ x \cdot w_i = 0\}| = 1$.*

Proof (of Lemma 4). Let Π be the protocol for the promise problem. Set $a = \lceil \log n \rceil$. Using public coins, independently and uniformly choose $w_1, \ldots w_l \in \{0,1\}^a$. For $j \in \{0, \ldots a\}$, the players run the protocol Π, using the following rule for evaluating the

input $x_{i,r}$ for $i \in [k], r \in [m]$: interpret r as a vector in $\{0,1\}^a$, and replace the value of $x_{i,r}$ by zero if for some $j' \leq j$, $w_{j'} \cdot r \neq 0$, and use the value $x_{i,r}$ if for all $j' \leq j$, $w_{j'} \cdot r = 0$. If the protocol Π returns 1, the players halt and output 1, otherwise, the players proceed to round $j+1$. If no intersection is found after all $a+1$ rounds, the players announce that the inputs are disjoint.

Clearly, this protocol uses $O(s \log n)$ bits of communication, and by the 0-error property of Π on disjoint inputs, it never outputs 1 when the inputs are disjoint. When the inputs are non-disjoint, the Valiant-Vazirani construction ensures that with probability at least $1/4$, at some round j the protocol Π is used on an input with a unique intersection, and therefore, conditioned on this event, the correct answer is returned with probability at least $1-\epsilon$. Therefore, the correct answer is returned with probability at least $\frac{1}{4} - \frac{\epsilon}{4}$. Because ϵ is bounded away from 1 and the error is one-sided, a constant number of repetitions decreases the probability of error to $1/3$.

Combining the reductions

Theorem 2. *Let $k \geq 2$ and $m = n^{1/3}/\log n$. For each n there is an odd charge vector $\vec{c} \in \{0,1\}^n$ such that for any $\epsilon < 1/2$ the size of any tree-like* **Th(k-1)** *refutation of $TS^k(H_n, \vec{c})$ is at least $2^{\Omega((R^k_\epsilon(\mathrm{DISJ}_{k,m})/\log n)^{1/3})}$. Further if the coefficients in the* **Th(k-1)** *refutations are bounded by a polynomial in n then the refutation size must be at least $2^{\Omega(R^k_\epsilon(\mathrm{DISJ}_{k,m})/(\log n(\log\log n)^2))}$ or at least $2^{\Omega(D^k_\epsilon(\mathrm{DISJ}_{k,m})/\log^2 n)}$.*

Proof (Sketch). By Theorem 1 and the definition of $\mathrm{ODDCHARGE}^k(H_n)$, if for every $\vec{c} \in \{0,1\}^n$ there is tree-like **Th(k-1)** refutation of $TS^k(H_n, \vec{c})$ of size at most S, then there is a $1/n$-error randomized k-party NOF communication complexity protocol for $\mathrm{ODDCHARGE}^k(H_n)$ in which at most $O(\log^3 S)$ bits are communicated. By sending one more bit the players can check that the answer is correct and only output it in this case. Then applying Lemmas 1, 2, and 4 in turn yields an error $1/3$ randomized k-party NOF protocol for $\mathrm{DISJ}_{k,m}$ of complexity $O(\log^3 S \log n + \log^2 n)$ bits in total. Applying a similar reduction using the other parts of Theorem 1 yields the claimed result.

In the full paper we prove that the same lower bounds as Theorem 2 hold for *every* odd charge vector $\vec{c} \in \{0,1\}^n$.

4 Proximity of distributions $\mathcal{D}_1$ and $\mathcal{R}(\vec{x})$ when $|\cap \vec{x}| = 1$

In this section we prove Lemma 3 that for $|\cap \vec{x}| = 1$ the distributions $\mathcal{R}(\vec{x})$ and $\mathcal{D}_1$ are close in the ℓ_1 norm. Let $\mu_{\mathcal{D}_1}$ and $\mu_{\mathcal{R}(\vec{x})}$ be their associated probability measures. We will show that for all but a set of $(\alpha, \vec{c})$ with $\mu_{\mathcal{D}_1}$ measure $o(1)$, $\mu_{\mathcal{D}_1}(\alpha, \vec{c}) = (1 \pm o(1))\mu_{\mathcal{R}(\vec{x})}(\alpha, \vec{c})$.

Given an instance of the set disjointness variables, $\vec{x} = (\{0,1\}^m)^k$, for $j \in [m]$ we say that the *color* of j is the tuple $(x_{1,j}, \ldots, x_{k,j}) \in \{0,1\}^k$. By construction, the assignment $\mathcal{R}(\vec{x})$ produced by R on this instance has color $(x_{1,j}, \ldots, x_{k,j})$ on each edge of the path p_j.

Definition 3. *Given an ordered sequence of paths $\vec{p} \in P_l^{(m)}$, an $\vec{x} \in (\{0,1\}^m)^k$, and an assignment α, write $\chi(\alpha_{\vec{p}}) = \vec{x}$ if and only if every edge on path p_j has color $(x_{1,j}, \ldots, x_{k,j})$ for every $j \in [m]$.*

We first observe that for any $(\alpha, \vec{c})$ with $|\text{Err}(\alpha, \vec{c})| = 2$ the probability $\mu_{\mathcal{D}_t}(\alpha, \vec{c})$ depends only on the number of edges $e \in G_n$ having color 1^k in α.

Definition 4. *Let* $\phi(a, b) = 2^{-a}(2^k - 1)^{-(a-b)}$.

Lemma 6. *For any* $(\alpha, \vec{c})$ *with* $|Err(\alpha, \vec{c})| = 2t$ *and* $m_1 = |\{e \in E(G_n) \mid \alpha_e = 1^k\}|$, $\mu_{\mathcal{D}_1}(\alpha, \vec{c}) = \phi(|E(G_n)|, m_1)/(2^{n-1}\binom{n}{2})$.

Proof. Let $U = \text{Err}(\alpha, \vec{c})$. The probability under $\mathcal{D}_1$ that U is chosen to be flipped is $1/\binom{n}{2t}$ and, given U, all of the 2^{n-1} even charge vectors $\vec{c}^U$ are equally likely. Conditioned on these events, the chance that α labels the edges for the randomly selected element of $Sol(H_n, \vec{c})$ is $2^{-|E(G_n)|}(2^k - 1)^{-(|E(G_n)|-m_1)}$.

Definition 5. *For* $U \subset V$ *with* $|U| = 2$ *let* $P_l^{(m)}(U)$ *be the set of all elements of* $P_l^{(m)}$ *that have a path whose endpoints are* U.

Now consider the measure $\mu_{\mathcal{R}(\vec{x})}(\alpha, \vec{c})$. Let $\{i\} = \cap \vec{x} \subseteq [n]$, $U = \text{Err}(\alpha, \vec{c})$ with $|U| = 2$, and $m_1 = |\{e \in E(G_n) \mid \alpha_e = 1^k\}|$. By the definition of R,

$$\begin{aligned}\mu_{\mathcal{R}(\vec{x})}(\alpha, \vec{c}) &= \Pr_{\vec{p} \in P_l^{(m)}}[\text{Ends}(p_i) = \text{Err}(\alpha, \vec{c}) \wedge \chi(\alpha_{\vec{p}}) = \vec{x}] \\ &\quad \times \Pr_{\vec{c'} \in \{0,1\}^n,\ \alpha' \in \mathcal{L}_k(Sol(H_n - \vec{p}, \vec{c'}))}[\alpha' = \alpha_{G_n - \vec{p}} \text{ and } \vec{c'} = \vec{c}] \\ &= \Pr_{\vec{p} \in P_l^{(m)}}[\text{Ends}(p_i) = \text{Err}(U)] \times \Pr_{\vec{p} \in P_l^{(m)}(U)}[\chi(\alpha_{\vec{p}}) = \vec{x}] \\ &\quad \times\ \phi(|E(G_n)| - ml, m_1 - l)/2^{n-1}.\end{aligned}$$

Observe that p_i is a uniformly chosen element of P_l and we can analyze the first term using the following property of random paths on LPS expanders proven in the full paper.

Lemma 7. *For* $u \neq v \in V(G_n)$ *and* $l \geq c_1 \log n / \log\log n$,
$$\Pr_{p \in P_l}[\textit{Ends}(p) = \{u, v\}] = (1 \pm o(1))/\tbinom{n}{2}.$$

$$\begin{aligned}\text{Thus } \mu_{\mathcal{R}(\vec{x})}(\alpha, \vec{c}) &= (1 \pm o(1))\frac{\phi(|E(G_n)| - ml, m_1 - l)}{\binom{n}{2}2^{n-1}} \cdot \Pr_{\vec{p} \in P_l^{(m)}(U)}[\chi(\alpha_{\vec{p}}) = \vec{x}] \\ &= (1 \pm o(1))\frac{\mu_{\mathcal{D}_1}(\alpha, \vec{c})}{\phi(ml, l)} \cdot \Pr_{\vec{p} \in P_l^{(m)}(U)}[\chi(\alpha_{\vec{p}}) = \vec{x}].\end{aligned}$$

It follows that we will obtain the desired result if we can show that for all but a $o(1)$ measure of $(\alpha, \vec{c})$ under $\mu_{\mathcal{D}_1}$,

$$\Pr_{\vec{p} \in P_l^{(m)}(U)}[\chi(\alpha_{\vec{p}}) = \vec{x}] = (1 \pm o(1))\phi(ml, l) = (1 \pm o(1))2^{-ml}(2^k - 1)^{-(m-1)l}$$

where $U = \text{Err}(\alpha, \vec{c})$. In the case that this happens, we say that $(\alpha, \vec{c})$ is *well-distributed for* $\vec{x}$.

Using the second moment method we prove the following lemma which shows that for all but a $o(1)$ measure of $(\alpha, \vec{c})$ under $\mu_{\mathcal{D}_1}$, $(\alpha, \vec{c})$ is indeed well-distributed for $\vec{x}$. The detailed proof is given in the full paper; the proof uses the fact that $\Theta(\log n)$-degree LPS expanders have $O(\log n / \log\log n)$ mixing time and $\Omega(\log n / \log\log n)$ girth.

Lemma 8. *Let $m \leq n^{1/3}/\log n$ and $l = 2\lceil c_1 \log n/\log\log n\rceil$ and $\vec{x} \in (\{0,1\}^m)^k$ with $|\cap\vec{x}| = 1$. For almost all $U \subset [n]$ with $|U| = 2$,*

$$\Pr_{(\alpha,\vec{c})\in\mathcal{D}_1}[(\alpha,\vec{c}) \text{ is well-distributed for } \vec{x} \mid \mathit{Err}(\alpha,\vec{c}) = U] = 1 - o(1).$$

Lemma 3 follows from this almost immediately.

Proof (of Lemma 3). Let $\vec{x} \in (\{0,1\}^m)^k$ and $|\cap\vec{x}| = 1$. By Lemma 8 and the preceding argument, for all but a set B of U that forms $o(1)$ fraction of all subsets $[n]$ of size 2, $\Pr_{(a,\vec{c})\in\mathcal{D}_1}[\mu_{\mathcal{R}(\vec{x})}(\alpha,\vec{c}) = (1 \pm o(1))\mu_{\mathcal{D}_1}(\alpha,\vec{c}) \mid \mathrm{Err}(\alpha,\vec{c}) = U] = 1 - o(1)$. By Lemma 7, $\Pr_{(\alpha,\vec{c})\in\mathcal{D}_1}[\mathrm{Err}(\alpha,\vec{c}) \in B] = o(1)$. Therefore by summing over distinct choices of U, we obtain that with probability $1 - o(1)$ over $(\alpha,\vec{c}) \in \mathcal{D}_1$, $\mu_{\mathcal{R}(\vec{x})}(\alpha,\vec{c}) = (1 \pm o(1))\mu_{\mathcal{D}_1}(\alpha,\vec{c})$. This is equivalent to the desired conclusion that $||\mathcal{D}_1 - \mathcal{R}(\vec{x})||_1$ is $o(1)$.

5 Discussion

There are a couple of interesting open problems related to our work beyond the natural problem of the communication complexity of $\textsc{Disj}_k$. First, does semantic LS^k have a separation oracle, as LS does? This is closely related to whether or not LS^k is automatizable and we conjecture that the answer to both questions is negative. Secondly, is it possible to extend our lower bounds to other tautologies that would imply inapproximability results for polynomial-time LS^k-based algorithms? (For example, if we could prove superpolynomial lower bounds for tree-like LS^k proofs of random 3CNF formulas, this would imply inapproximability results for LS^k-based linear programming algorithms for MaxSAT [5].)

Finally we would like to point out a connection between our main result and the complexity of disjoint NP pairs. An open question in complexity theory is whether or not all pairs of disjoint NP sets can be separated by a set in P. This is known to be false under the assumption $\mathsf{P} \neq \mathsf{UP}$ and also by the assumption $\mathsf{P} \neq \mathsf{NP} \cap \mathsf{coNP}$. It is an open question whether or not it is implied by $\mathsf{P} \neq \mathsf{NP}$. Let us consider the same question with respect to communication complexity rather than polynomial time: can every pair of relations with small nondeterministic k-party communication complexity be separated by a small probabilistic/deterministic protocol? In [20] the answer is shown to be unconditionally false for $k = 2$. In particular, they give a pair of disjoint properties on $3m$-vertex graphs G, a matching on $2m$ vertices of G and an independent set of $2m + 1$ vertices of G, and show that this pair cannot be separated by any small probabilistic/deterministic protocol. In this paper, we have shown that for any k, the question is still false, under k-$\mathsf{RP}^{cc} \neq k$-NP^{cc}.

Acknowledgements

We are indebted to Avi Wigderson for helpful discussions and insights.

References

1. S. Arora, B. Bollobás, and L. Lovász. Proving integrality gaps without knowing the linear program. In *Proceedings 43nd Annual Symposium on Foundations of Computer Science*, pages 313–322, Vancouver, BC, November 2002. IEEE.

2. L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, Toronto, Ontario, October 1986. IEEE.
3. P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A direct sum theorem for corruption and the multiparty NOF communication complexity of set disjointness. In *Proceedings Twentieth Annual IEEE Conference on Computational Complexity*, San Jose, CA, June 2005.
4. A. Bockmayr, F. Eisenbrand, M.E. Hartmann, and A.S. Schulz. On the Chvatal rank of polytopes in the 0/1 cube. *Discrete Applied Mathematics*, 98(1-2):21–27, 1999.
5. J. Buresh-Oppenheim, N. Galesi, S. Hoory, A. Magen, and T. Pitassi. Rank bounds and integrality gaps for cutting planes procedures. In *Proceedings 44th Annual Symposium on Foundations of Computer Science*, pages 318–327, Boston, MA, October 2003. IEEE.
6. V. Chvátal. Edmonds polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics*, 4:305–337, 1973.
7. V. Chvátal, W. Cook, and M. Hartmann. On cutting-plane proofs in combinatorial optimization. *Linear Algebra and its Applications*, 114/115:455–499, 1989.
8. W. Cook, C. R. Coullard, and G. Turan. On the complexity of cutting plane proofs. *Discrete Applied Mathematics*, 18:25–38, 1987.
9. S. Dash. *On the matrix cuts of Lovász and Schrijver and their use in Integer Programming*. PhD thesis, Department of Computer Science, Rice University, March 2001.
10. S. Dash. An exponential lower bound on the length of some classes of branch-and-cut proofs. In W. Cook and A. S. Schulz, editors, *IPCO*, volume 2337 of *Lecture Notes in Computer Science*, pages 145–160. Springer-Verlag, 2002.
11. F. Eisenbrand and A. S. Schulz. Bounds on the Chvatal rank of polytopes in the $0/1$-cube. *Combinatorica*, 23(2):245–261, 2003.
12. D. Grigoriev, E. A. Hirsch, and D. V. Pasechnik. Complexity of semi-algebraic proofs. In *(STACS) 2002: 19th Annual Symposium on Theoretical Aspects of Computer Science*, volume 2285 of *Lecture Notes in Computer Science*, pages 419–430, Antibes, France, February 2002. Springer-Verlag.
13. R. Impagliazzo, T. Pitassi, and A. Urquhart. Upper and lower bounds on tree-like cutting planes proofs. In *9th Annual IEEE Symposium on Logic in Computer Science*, pages 220–228, Paris, France, 1994.
14. B. Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. In *Proceedings, Structure in Complexity Theory, Second Annual Conference*, pages 41–49, Cornell University, Ithaca, NY, June 1987. IEEE.
15. L. G. Khachian. A polynomial time algorithm for linear programming. *Doklady Akademii Nauk SSSR, n.s.*, 244(5):1093–1096, 1979. English translation in *Soviet Math. Dokl. 20*, 191–194.
16. L. Lovasz and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optimization*, 1(2):166–190, 1991.
17. A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
18. N. Nisan. The communication complexity of threshold gates. In V.S.D. Mikl'os and T. Szonyi, editors, *Combinatorics: Paul Erdös is Eighty, Volume I*, pages 301–315. Bolyai Society, 1993.
19. P. Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *Journal of Symbolic Logic*, 62(3):981–998, September 1997.
20. R. Raz and A. Wigderson. Monotone circuits for matching require linear depth. *Journal of the ACM*, 39(3):736–744, July 1992.
21. L. Valiant and V. Vazirani. NP is as easy as detecting unique solutions. *Theoretical Computer Science*, pages 85–93, 1986.