Average case polyhedral complexity of the maximum stable set problem

Gábor Braun[1], Samuel Fiorini[2], and Sebastian Pokutta[3]

[1]ISyE, Georgia Institute of Technology, Atlanta, GA, USA. *Email:* gabor.braun@isye.gatech.edu
[2]Department of Mathematics, Université libre de Bruxelles CP 216, Bd. du Triomphe, 1050 Brussels, Belgium. *Email:* sfiorini@ulb.ac.be
[3]ISyE, Georgia Institute of Technology, Atlanta, GA, USA. *Email:* sebastian.pokutta@isye.gatech.edu

November 19, 2013

Abstract

We study the minimum number of constraints needed to formulate random instances of the maximum stable set problem via LPs (more precisely, linear extended formulations), in two distinct models. In the uniform model, the constraints of the LP are not allowed to depend on the input graph, which should be encoded solely in the objective function. There we prove a $2^{\Omega(n/\log n)}$ lower bound with probability at least $1 - 2^{-2^n}$ for every LP that is exact for a randomly selected set of instances; each graph on at most n vertices being selected independently with probability $p \geqslant 2^{-\binom{n/4}{2}+n}$. In the non-uniform model, the constraints of the LP may depend on the input graph, but we allow weights on the vertices. The input graph is sampled according to the $G(n,p)$ model. There we obtain upper and lower bounds holding with high probability for various ranges of p. The bounds are close as there is only an essentially quadratic gap in the exponent. Finally, we state a conjecture to close the gap.

1 Introduction

In the last three years, extended formulations considerably gained interest in various areas, including discrete mathematics, combinatorial optimization, and theoretical computer science.

The key idea underlying extended formulations is that by choosing the right variables it is possible to efficiently express various combinatorial optimization problems via linear programs (LPs) in higher dimension. There is an ever expanding collection of examples of small size extended formulations. For instance, Williams [2002] has expressed the minimum spanning tree problem on a planar graph with only a linear number of (variables and) constraints, while in the natural edge variables the LP has an exponential number of constraints. There exist numerous other examples, see e.g., the surveys by Conforti et al. [2010] and Kaibel [2011].

On the other hand, extended formulations ask for the intrinsic difficulty of expressing a given combinatorial optimization problem through a single LP, in terms of the minimum number of *constraints* necessary in such an LP. This leads to a complexity measure that we call loosely here 'polyhedral complexity' (precise definitions are given later from Section 2 on).

In fact, the main reason behind the renewed interest for extended formulations is the recent series of breakthroughs in lower bounding techniques [Rothvoß, 2011, Fiorini et al., 2012, Braun et al., 2012, Braverman and Moitra, 2012, Braun and Pokutta, 2013, Chan et al., 2013, Rothvoß, 2013].

These breakthroughs make it now conceivable to quantify the polyhedral complexity of any given combinatorial optimization problem *unconditionally*, that is, independently of conjectures such as P vs. NP, and without extra assumption on the structure of the LP.

Although a polynomial upper bound on the polyhedral complexity yields a polynomial upper bound on the true algorithmic complexity of the problem—provided that the LP can be efficiently constructed and also that the size of the coefficients is kept under control (see Rothvoß [2011] for a discussion of this last issue) e.g., through interior point methods—it is becoming clear that the converse does not hold. Recently, Chan et al. [2013] proved that every LP for MAXCUT with an integrality gap at most $2-\varepsilon$ needs at least $n^{\Omega\left(\frac{\log n}{\log\log n}\right)}$ constraints, while the approximation factor of the celebrated SDP-based polynomial time algorithm of Goemans and Williamson [1995] is close to 1.13. Even more recently, Rothvoß [2013] solved another major open problem in the area by showing a $2^{\Omega(n)}$ lower bound on the size of any LP expressing the perfect matching problem.

In this paper, we consider the problem of determining the *average case* polyhedral complexity of the maximum stable set problem, in two different models: 'uniform' and 'non-uniform', see Section 1.2 below. Roughly, the uniform model asks for a single LP that works for a given set of input graphs. In the non-uniform model the LP can depend on the input graph G but should work for every choice of weights on the vertices of G (in particular, for all induced subgraphs of G).

We show that the polyhedral complexity of the maximum stable set problem remains high in each of these models, when the input graph is sampled according to natural distributions. Therefore, we conclude that the hardness of the problem is not concentrated on a small mass of graphs but it is spread out through all graphs.

1.1 Related work

At core of the study of extended formulations is Yannakakis's famous paper [Yannakakis, 1988, 1991] relating polyhedral complexity and nonnegative rank. In his work, Yannakakis was able to show that the TSP polytope does not admit a symmetric linear programming formulation of polynomial size (incidentally, he did not rely on the connection to nonnegative rank mentioned above but rather on a group-theoretic argument). The symmetry assumption was then removed later in Fiorini et al. [2012] and a formal framework for approximate linear programming formulations was established in Braun et al. [2012].

Our work is most directly related to Fiorini et al. [2012] and Braun et al. [2012], where the foundation for the framework used here are laid out. The first paper identified the *Unique Disjointness* (*UDISJ*) partial matrix as an important source of lower bounds on polyhedral complexity. The second paper laid out a framework for studying the size of approximate linear programming formulations. This framework forms the basis of our uniform model.

Finally, at the core of our lower bound for the uniform model is a theorem from Braun and Pokutta [2013] that provides a strong lower bound on the nonnegative rank of a submatrix of the UDISJ partial matrix obtained by (adversarially) dropping rows and columns.

1.2 Contribution

We present the first strong and unconditional results on the average case size of LP formulations for the maximum stable set problem. In particular, we establish that the maximum stable set problem in two natural average case models and encodings does not admit a polynomial size linear programming formulation, even in the unlikely case that $P = NP$.

Uniform model In the *uniform model* the polytope P containing the feasible solutions to the stable set problem is *independent* of the instances. The instances will be solely encoded into the objective functions. This ensures that no complexity of the problem is leaked into an instance-specific formulation. A good example of a uniform model is the TSP polytope over K_n with which we can test for Hamiltonian cycles in any graph with at most n vertices by choosing an appropriate objective function. In the uniform model, we show that if we sample each graph on at most n vertices with probability $p \geqslant 2^{-\binom{n/4}{2}+n}$ then with probability at least $1 - 2^{-2^n}$ every LP "solving" the resulting set of instances has at least $2^{\Omega(n/\log n)}$ constraints.

Non-uniform model In the *non-uniform model* we consider the stable set polytope for a *specific but random graph*. The polyhedral description may depend heavily on the chosen graph. We sample a graph G in the Erdős–Rényi $G(n, p)$ model. We then analyze the stable set polytope $\mathrm{STAB}(G)$ of G. If p is small enough, so that the obtained graph is sufficiently sparse, it will contain an induced subgraph of sufficient size whose corresponding face of $\mathrm{STAB}(G)$ projects to the correlation polytope. Via this embedding we can then derive strong lower bounds on the size of any LP expressing $\mathrm{STAB}(G)$ that hold with high probability. We establish superpolynomial lower bounds for p ranging between $\omega((\log^{6+\varepsilon} n)/n)$ and $O(log^{-1}n)$ like for $p = n^{-\varepsilon}$ and $\varepsilon < 1/4$ the LP has at least $2^{\Omega(\sqrt{n^{\varepsilon}\log n})}$ constraints, and for $p = c(\log^{6+\varepsilon)} n)n^{-1}$ we get a lower bound of $n^{(\log 3/2)\log^{1+\varepsilon/4} n}$.

1.3 Outline

In Section 2 we recall extended formulations. We introduce the uniform model for the maximum stable set problem in Section 3. We then establish bounds on the average case complexity for the uniform model in Section 4. In Section 5 we consider the non-uniform model and derive lower bounds as well as upper bounds. We conclude with a conjecture in Section 6.

2 Preliminaries

We start by briefly recalling basics of extended formulations, stated in geometric terms. We refer the interested reader to Fiorini et al. [2012] for more details. After that we state the main source of lower bounds in the non-uniform case.

Let $P \subseteq \mathbb{R}^d$ and $L \subseteq \mathbb{R}^e$ be two polyhedra. Then L is called an *extension* (or *lift*) of P if there exists an affine map $\pi\colon \mathbb{R}^d \to \mathbb{R}^e$, so that $\pi(L) = P$. Defining the *size* of polyhedron L as its number of facets, the *extension complexity* of polyhedron P is the minimum size of any of its extensions L, and is denoted by $\mathrm{xc}(P)$. Here we use the notions of extension and extended formulation interchangeably; the latter is simply an equivalent way to describe an extension.

The following monotonicity lemma from Fiorini et al. [2012] provides a reduction mechanism to lower bound the extension complexity.

Lemma 2.1 (Monotonicity of extended formulations)**.** *Let P be a polyhedron. Then the following hold:*

(i) if F is a face of P, then $\mathrm{xc}(F) \leqslant \mathrm{xc}(P)$;

(ii) if L is an extension of P, then $\mathrm{xc}(P) \leqslant \mathrm{xc}(L)$.

As usual, $\mathrm{COR}(n) := \mathrm{conv}\left(\{bb^{\intercal} \in \mathbb{R}^{n\times n} \mid b \in \{0,1\}^n\}\right)$ denotes the *correlation polytope* and $\mathrm{STAB}(G) := \mathrm{conv}\left(\left\{\chi^S \in \mathbb{R}^{V(G)} \,\middle|\, S \text{ stable set of } G\right\}\right)$ is the *stable set polytope* of graph G. (Recall that the characteristic vector χ^S has $\chi^S_v = 1$ if $v \in S$ and $\chi^S_v = 0$ otherwise.) Let log denote the base-2 logarithm.

Theorem 2.2. $\mathrm{xc}(\mathrm{COR}(n)) \geqslant 2^{(\log 3/2)n}$.

The factor $\log 3/2 \approx 0.581$ in the exponent is the current best one due to Kaibel and Weltge [2013]; for various approximate case versions see Braun et al. [2012], Braverman and Moitra [2012], Braun and Pokutta [2013].

The notion of extension directly generalizes to pairs of nested polyhedra. If $P \subseteq Q \subseteq \mathbb{R}^d$ are two polyhedra, an extension of the pair P, Q is a polyhedron $L \subseteq \mathbb{R}^e$ such that $P \subseteq \pi(L) \subseteq Q$ for some affine map $\pi\colon \mathbb{R}^d \to \mathbb{R}^e$. The extension complexity $\operatorname{xc}(P, Q)$ of pair P, Q is the minimum size of an extension of that pair.

3 A uniform model for maximum stable set

Faithful linear encodings were introduced in Braun et al. [2012] to study the polyhedral hardness of approximation of various problems. They are tools leading to a polyhedral pair capturing the complexity of the problem.

Here we recall only the polyhedral pair arising from the standard encoding of the maximum stable set problem. We refer the reader to Appendix A for more details.

We consider a class $\mathcal{G}$ of graphs with vertex set included in $[n] := \{1, \dots, n\}$. For each graph G with $V(G) \subseteq [n]$, we define an objective function $w^G \in \mathbb{R}^{n\times n}$ by letting $w^G_{ij} = 1$ if $i = j$ and $i \in V(G)$, but $w^G_{ij} = w^G_{ji} = -1$ if $ij \in E(G)$, and finally $w^G_{ij} = 0$ otherwise. We define a polyhedron

$$Q(\mathcal{G}) := \left\{ x \in \mathbb{R}^{n\times n}_{+} \;\middle|\; \forall G \in \mathcal{G}\colon \langle w^G, x\rangle \leqslant \alpha(G) \right\},$$

where $\langle w^G, x\rangle = \sum_{i,j} w^G_{i,j} x_{i,j}$ denotes the Frobenius inner product of matrices w^G and x, and $\alpha(G)$ is the stability number of G. We let $\mathrm{STAB}^u(\mathcal{G}, \rho)$ denote the pair of nested polyhedra $(P, (1+\rho)Q(\mathcal{G}))$ with $P = \mathrm{COR}(n)$ and $\rho \geqslant 0$ defining the dilation factor. If $\rho = 0$, we simply denote the pair by $\mathrm{STAB}^u(\mathcal{G})$.

Recall from Braun et al. [2012] that the extension complexity of a polyhedral pair is equal to the nonnegative rank of any of its slack matrices up to a difference of 1, this is called the *factorization theorem* (see Theorem 10 in Appendix A). For the pair $(P, (1+\rho)Q(\mathcal{G}))$ a slack matrix S has rows indexed by all the characteristic vectors $b \in \{0,1\}^n$ of the subsets of $[n]$, corresponding to the vertex $bb^\intercal$ of P, and columns indexed by $\mathcal{G}$. The entries are $S(b, G) = (1+\rho)\alpha(G) - \langle w^G, b\rangle$.

For example, when $\mathcal{G}$ is the set of all cliques, we reindex the cliques by the characteristic vectors $a \in \{0,1\}^n$ of their vertex sets. We obtain a matrix M as a slack matrix with rows and columns indexed by $a, b \in \{0,1\}^n$, and with entries $M(a, b) = (1 - a^\intercal b)^2 + \rho$. Thus, in particular,

$$M(a, b) = \begin{cases} \rho & \text{if } a^\intercal b = 1 \\ 1+\rho & \text{if } a^\intercal b = 0. \end{cases}$$

For $\rho = 0$, this is known as the *unique disjointness (UDISJ)* (partial) matrix. For general $\rho \geqslant 0$, this is called the *ρ-shifted* UDISJ matrix. We shall need the following theorem from Braun and Pokutta [2013] to bound the nonnegative rank of certain submatrices of the (ρ-shifted) UDISJ matrix.

Theorem 3.1. *For the ρ-shifted UDISJ matrix M, let M_k be the submatrix for sets of size k. Let S be any submatrix of M_k obtained by deleting at most an α-fraction of rows and at most a β-fraction of columns for some $0 \leqslant \alpha, \beta < 1$. Then for $0 < \varepsilon < 1$:*

$$\operatorname{rank}_+ S \geqslant 2^{(1/8(\rho+1) - (\alpha+\beta)\mathbb{H}[1/4])n - O(n^{1-\varepsilon})} \quad \textit{for} \quad k = n/4 + O(n^{1-\varepsilon}).$$

We refer the reader to Appendix A.3 for more information on UDISJ.

4 Average case complexity in the uniform model

We will now establish our main result for the uniform average case complexity model. We obtain that for any random collection of graphs where each graph is picked independently with probability p, the polyhedral complexity of solving the stable set problem over that particular collection of graphs is high, or more precisely, the extension complexity of the corresponding pair is high. This shows in particular that the instances of the stable set problem resulting in high extension complexity are not localized in a set of small density.

Main Theorem 4.1 (Super-polynomial xc of $\mathrm{STAB}^u(\mathcal{G})$ w.h.p.)**.** *Let $n \geqslant 40$ and $p \in [0,1]$ with $p \geqslant 2^{-\binom{n/4}{2}+n}$. Pick a family $\mathcal{G}$ of graphs by selecting each graph G with $V(G) \subseteq [n]$ independently with probability p. Then*

$$\mathbb{P}\left[\mathrm{xc}(\mathrm{STAB}^u(\mathcal{G})) \geqslant 2^{\Omega(n/\log n)}\right] \geqslant 1 - 2^{-2^n}.$$

A crucial point of the proof is a concentration result on $\alpha(G)$. It is well-known that almost all graphs G on n vertices have stability number $\alpha(G) \sim 2\log n$. However, the following rough estimate will be sufficient for our purpose. We include an easy proof for completeness.

Lemma 4.1. *Let $n \geqslant 10$. The probability that a uniformly sampled random graph G with $V(G) = [n]$ has $\alpha(G) \geqslant 3\log n$ is at most n^{-1}.*

Proof. We follow closely [Diestel, 2005, Chapter 11, page 304]. Notice that $G(n, 1/2)$ induces the uniform distribution on all graphs with vertex set $[n]$. So we assume $G = G(n, 1/2)$. Letting $k := 3\log n$, we have $n \geqslant k \geqslant 2$ and

$$\mathbb{P}\left[\alpha(G) \geqslant k\right] \leqslant \binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} \leqslant n^k \left(\frac{1}{2}\right)^{\binom{k}{2}} = n^{\frac{3}{2}(1-\log n)} \leqslant n^{-1}$$

where the first inequality follows from the union bound. □

We are ready to prove the main theorem of this section.

of Main Theorem 1. The main idea of the proof is that, with large enough probability, we have $\max\left\{\langle w^K, x\rangle \;\middle|\; x \in Q(\mathcal{G})\right\} = O(\log n)$ for many cliques K with $V(K) \subseteq [n]$ and $\Theta(n)$ vertices. This implies that some slack matrix of the pair $\mathrm{STAB}^u(\mathcal{G})$ contains the $O(\log n)$-shifted UDISJ as a submatrix obtained by picking a large fraction of the rows (and all columns). We apply Theorem 3.

Consider a clique K with $V(K) \subseteq [n]$, and size $k := \lceil n/4 \rceil$. We say that a graph G is *good* for K if $V(G) = V(K)$ and $\alpha(G) \leqslant 3\log n$. Clique K is said to be *good* if some graph $G \in \mathcal{G}$ is good for K. Otherwise, K is called *bad*.

We claim that, with high probability, the total fraction of bad cliques among all k-cliques K is at most $\alpha := 1/(24\log n)$. By Lemma 4, the total number of graphs G with $V(G) = V(K)$ that are not good for a fixed k-clique K is at most $k^{-1}2^{\binom{k}{2}}$. Thus

$$\begin{aligned}\mathbb{P}\left[K \text{ is bad}\right] &= \mathbb{P}\left[\mathcal{G} \text{ contains no good graph for } K\right] \\ &\leqslant (1-p)^{(1-k^{-1})2^{\binom{k}{2}}} \leqslant \mathrm{e}^{-p(1-k^{-1})2^{\binom{k}{2}}} \leqslant 2^{-\frac{9}{10}2^n \log \mathrm{e}} \leqslant \alpha\, 2^{-2^n}.\end{aligned}$$

where the second inequality follows from $k \geqslant n/4 \geqslant 10$ and $p \geqslant 2^{-\binom{n/4}{2}+n}$. Let X denote the random variable that counts the number of bad k-cliques K. By Markov's inequality,

$$\mathbb{P}\left[X \geqslant \alpha\binom{n}{k}\right] \leqslant 2^{-2^n}.$$

If clique K is good and G is a good graph for K, the inequality $\langle w^G, x\rangle \leqslant 3\log n$ is valid for $Q(\mathcal{G})$. Thus the inequality $\langle w^K, x\rangle \leqslant 3\log n$ is also valid for $Q(\mathcal{G})$, because $x \geqslant 0$ is valid for $Q(\mathcal{G})$, and $w^K \leqslant w^G$.

Suppose that the fraction of cliques K with $V(K) \subseteq [n]$ and size $k = \lceil n/4\rceil$ that are bad is at most α. We have shown that this holds with probability at least $1 - 2^{-2^n}$. By what precedes, we can define a slack matrix for the pair $\mathrm{STAB}^u(\mathcal{G})$ that contains a $(3\log n)$-shift of UDISJ with at most an α-fraction of the rows thrown away. From Theorem 3 and from the factorization theorem, the extension complexity of the pair $\mathrm{STAB}^u(\mathcal{G})$ is at least $2^{(1/8(3\log n+1)-\alpha\mathbb{H}[1/4])\cdot n-O(n^{1-\varepsilon})} = 2^{\Omega(n/\log n)}$. $\square$

A close inspection of the proof of Main Theorem 1 shows that we can immediately apply the framework in Braun et al. [2012] to obtain a lower bound on the average case *approximate* extension complexity. We obtain the following result. The proof is identical, except that we choose $\alpha := 1/24(1+\rho)\log n$, and the inequalities that yield the slack matrix are of the form $\langle w^K, x\rangle \leqslant (1+\rho)3\log n$ for all good cliques K with $k = \lceil n/4\rceil$ vertices, which are all valid for $(1+\rho)Q(\mathcal{G})$.

Corollary 4.2 (Super-polynomial xc of $\mathrm{STAB}^u(\mathcal{G},\rho)$ w.h.p.). *As in Main Theorem 1, let $\mathcal{G}$ be a random set of such that each graph G with $V(G) \subseteq [n]$ is contained in $\mathcal{G}$ with probability $p \geqslant 2^{-\binom{n/4}{2}+n}$ independent of the other graphs. Then the ρ-approximate pair $\mathrm{STAB}^u(\mathcal{G},\rho)$ with $\rho \leqslant \frac{n^{1-\varepsilon}}{\log n}$ for some $0 < \varepsilon < 1/2$ has extension complexity $2^{\Omega(n^{\varepsilon})}$, with probability at least $1 - 2^{-2^n}$.*

Observe that the approximation factor in Corollary 5 can be larger than $3\log n$. The reason why this is possible, contradicting initial intuition, is that the hardness arises from having many different graphs and hence many objective functions to consider simultaneously and the encoding is highly non-monotone. Roughly speaking, graphs with different vertex sets are independent of each other, even if one is an induced subgraph of the other.

5 Average case complexity in the non-uniform model

We now turn our attention to the non-uniform model, where we consider the stable set polytope over a *specific but random* graph G and analyze its extension complexity. Our strategy is to embed certain gadget graphs as induced subgraphs of G, using the probabilistic method. Here we consider the Erdős–Rényi graph model and sample G from $G(n,p)$.

We begin by defining the gadget graphs we use and seeing how an induced gadget forces up the extension complexity of $\mathrm{STAB}(G)$ when the underlying template graph is a complete graph.

Fix a graph T. This graph serves as a template for defining the *gadget graph* of T, denoted as $T^{\pentagon}$: the graph obtained by replacing each edge ij of T with an *edge gadget* E_{ij}, which is a 5-cycle with possibly additional connecting paths (hairs) as shown in Figure 1. In total, $T^{\pentagon}$ has $v := |V(T)| + (2\ell+3)\,|E(T)|$ vertices and $e := (2\ell+5)\,|E(T)|$ with hairs of length ℓ (i.e., every hair has ℓ many edges). We allow $\ell = 0$, in this case the grey and black vertices of Figure 1 coincide.

The hairs are used to decrease the average degree of induced subgraphs of the gadget graph, as shown in the following lemma.

Lemma 5.1. *For any graph T, the average degree of any induced subgraph of $T^{\pentagon}$ is at most $2 + 4/(2\ell+3)$ for $\ell \geqslant 1$. For $\ell = 0$ the average degree is at most 4.*

Proof. Let G be an induced subgraph of $T^{\pentagon}$. We shall prove the stronger claim that $|E(G)|\,/\,|V(G)\setminus V(T)|$ is upper bounded by 2 if $\ell = 0$ and by $1 + 2/(2\ell+3)$ if $\ell \geqslant 1$ (in other words, we ignore the original vertices of T at the estimation).

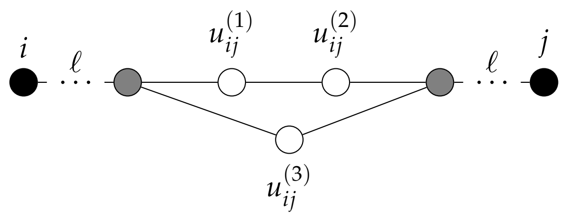

Figure 1: Edge gadget E_{ij} replacing edge ij of T in the gadget graph $T^{\pentagon}$. Black vertices represent vertices of the template graph T, white and grey vertices represent new vertices added to construct $T^{\pentagon}$. There are $\ell \geqslant 0$ many edges between the black and grey vertices.

First we apply some modifications to G which do not decrease the factor $|E(G)| / |V(G) \setminus V(T)|$ if it was already at least 1. We add the original vertices of T to G (together with the edges connecting them to vertices already in G), and then we successively remove degree-1 vertices of G in the edge gadgets. Hence we may assume without loss of generality, that G has no degree-1 vertices of the edge gadgets, and it contains the original vertices of T. So for a fixed E_{ij}, the graph G contains either only the original vertices of T, or both the degree-3 (grey) vertices of the 5-cycle. In the latter case, from every path connecting these and i, j, the graph G contains either the whole path, or only the end points. We claim that if $\ell \geqslant 1$ then adding the missing paths will not decrease the factor $|E(G)| / |V(G) \setminus V(T)|$ below $1 + 2/(2\ell + 3)$ if it was greater than this value. Indeed, for every path, the ratio of added edges and vertices is at least $1 + 2/(2\ell + 3)$, namely, $1 + 1/(\ell - 1)$, 2 or $3/2$. Therefore we may assume that every edge gadget E_{ij} is either completely contained in G or only the two vertices i and j of T are contained in G. Let k denote the number of E_{ij} completely contained in G, then $|E(G)| = k(2\ell + 5)$ and $|V(G) \setminus V(T)| = k(2\ell + 3)$, and their ratio is exactly $1 + 2/(2\ell + 3)$, finishing the proof in case $\ell \geqslant 1$.

If $\ell = 0$ then a similar argument applies, except that adding the shorter path (containing $u_{ij}^{(3)}$) and removing the longer path ($u_{ij}^{(1)}, u_{ij}^{(2)}$) will not decrease the factor $|E(G)| / |V(G) \setminus V(T)|$ below 2 if it were already larger. $\square$

In the next lemma, we denote by $\mathrm{COR}(T)$ the projection of the $|V(T)| \times |V(T)|$ correlation polytope $\mathrm{COR}(|V(T)|)$ on the variables x_{ii} for $i \in V(T)$ and x_{ij} for $ij \in E(T)$. We call this polytope the *correlation polytope of graph T*. In particular, $\mathrm{COR}(K_t) = \mathrm{COR}(t)$.

Lemma 5.2. *If graph G contains $T^{\pentagon}$ (with arbitrary even hair length ℓ) as an induced subgraph, then*

$$\mathrm{xc}(\mathrm{STAB}(G)) \geqslant \mathrm{xc}(\mathrm{COR}(T)).$$

In particular, for $T = K_t$ we get $\mathrm{xc}(\mathrm{STAB}(G)) \geqslant 2^{(\log 3/2)t}$.

Proof. Let F be the face of $\mathrm{STAB}(G)$ whose vertices are the characteristic vectors of stable sets of $T^{\pentagon}$ containing the maximum number vertices in each edge gadget E_{ij}. Thus, F is defined by intersecting $\mathrm{STAB}(G)$ with the (face inducing) hyperplanes $\sum_{v \in V(E_{ij})} x_v = \ell + 2$ for all $ij \in E(T)$. Here x_v is the coordinate for vertex v in $T^{\pentagon}$. For simplicity, we denote by $x_{ij}^{(k)}$ the coordinate for the additional vertex $u_{ij}^{(k)}$ of the 5-cycle in E_{ij}, see Figure 1.

Then it can be easily verified that F is an extension of $\mathrm{COR}(T)$ via the affine map $\pi \colon x \mapsto y = \pi(x)$ where

$$y_{ij} = \begin{cases} x_i & \text{if } i = j, \\ 1 - x_{ij}^{(1)} - x_{ij}^{(2)} & \text{if } i \neq j. \end{cases}$$

In this definition, the y_{ij} are the correlation variables, with $i, j \in V(T)$ and either $i = j$ or $ij \in E(T)$.

Now Lemma 1 gives

$$\mathrm{xc}(\mathrm{STAB}(G)) \geqslant \mathrm{xc}(F) \geqslant \mathrm{xc}(\mathrm{COR}(T)).$$

For $T = K_t$, using Theorem 2, we have

$$\mathrm{xc}(\mathrm{STAB}(G)) \geqslant \mathrm{xc}(\mathrm{COR}(t)) \geqslant 2^{(\log 3/2)t}.$$

□

5.1 Existence of gadgets in random graphs

In this section, we estimate the probability that a random Erdős–Rényi graph $G = G(n, p)$ contains an induced copy of a graph H. Recall that in the $G(n, p)$ model, each of the $\binom{n}{2}$ pairs of vertices is connected by an edge with probability p, independently from the other edges.

Lemma 5.3. *Let H be a graph with v vertices and with all induced subgraphs having average degree at most d. Let $0 < p \leqslant 1/2$ and*

$$g = g(n, p, v) := \frac{v^2 p^{-\frac{d}{2}} (1-p)^{-\frac{v}{2}}}{n - v}.$$

The probability of $G(n, p)$ not containing an induced copy of H satisfies

$$\mathbb{P}\left[H \overset{ind}{\not\subseteq} G(n, p)\right] \leqslant c_0 g^2 \approx 1.23 g^2,$$

where $c_0 := \exp(2W(1/\sqrt{2}))/2$ and W is the Lambert W-function, the inverse of $x \to x \exp x$.

Proof. The proof is via the second-moment method.

Let S be any graph isomorphic to H with $V(S) \subseteq V(G)$. Let X_S be the indicator random variable of S being an induced subgraph of G. Obviously, the total number of induced subgraphs of G isomorphic to H satisfies $X = \sum_S X_S$. We estimate the expectation and variance of X. Let e denote the number of edges of H, and let $\mathrm{Aut}(H)$ denote the automorphism group of H. The expectation is clearly

$$\mathbb{E}[X] = \sum_S \mathbb{E}[X_S] = \binom{n}{v} \frac{v!}{|\mathrm{Aut}(H)|} p^e (1-p)^{\binom{v}{2} - e}.$$

The variance needs more preparations. Let now S and T be two graphs isomorphic to H with $V(S), V(T) \subseteq V(G)$. Using that X_S and X_T are independent and thus $\mathrm{Cov}[X_S, X_T] = 0$ when $|V(S) \cap V(T)| \leqslant 1$ we get

$$\begin{aligned}\mathrm{Var}[X] = \sum_{S,T} \mathrm{Cov}[X_S, X_T] &\leqslant \sum_{|V(S) \cap V(T)| \geqslant 2} \mathbb{E}[X_S X_T] \\ &= \sum_{|V(S) \cap V(T)| \geqslant 2} \mathbb{E}[X_S]\, \mathbb{E}[X_T \mid X_S = 1] = \mathbb{E}[X] \sum_{T\colon |V(S) \cap V(T)| \geqslant 2} \mathbb{E}[X_T \mid X_S = 1].\end{aligned}$$

Note that in the last sum S is fixed, and by symmetry, the sum is independent of the actual value of S. That is why we could factor it out. We obtain via Chebyshev's inequality,

$$\mathbb{P}\left[H \overset{\mathrm{ind}}{\not\subseteq} G(n, p)\right] = \mathbb{P}[X = 0] \leqslant \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} \leqslant \frac{\sum_{T\colon |V(S) \cap V(T)| \geqslant 2} \mathbb{E}[X_T \mid X_S = 1]}{\mathbb{E}[X]}.$$

We shall estimate $\mathbb{E}[X_T \mid X_S = 1]$, which is the probability that H is induced in G provided S is induced in G, as a function of $k := |V(S) \cap V(T)|$. We assume that S and T coincide on $V(S) \cap V(T)$, and therefore have at most $dk/2$ edges in common, as their intersection is isomorphic to an induced subgraph of H, and therefore have average degree at most d by assumption. Hence as $p \leqslant 1/2$

$$\mathbb{E}[X_T \mid X_S = 1] = \mathbb{P}\left[T \overset{\text{ind}}{\subseteq} G \,\middle|\, S \overset{\text{ind}}{\subseteq} G\right] \leqslant p^{e-\frac{d}{2}k}(1-p)^{\binom{v}{2}-e-\binom{k}{2}+\frac{d}{2}k}.$$

This is clearly also true if S and T do not coincide on $V(S) \cap V(T)$, as then the probability is 0. Now we can continue our estimation by summing up for all possible T with $k \geqslant 2$:

$$\begin{aligned}
\frac{\sum_T \mathbb{E}[X_T \mid X_S = 1]}{\mathbb{E}[X]} &\leqslant \frac{\sum_{k=2}^{v} \binom{v}{k}\binom{n-v}{v-k} \frac{v!}{|\operatorname{Aut} H|} p^{e-\frac{d}{2}k}(1-p)^{\binom{v}{2}-e-\binom{k}{2}+\frac{d}{2}k}}{\binom{n}{v} \frac{v!}{|\operatorname{Aut} H|} p^{e}(1-p)^{\binom{v}{2}-e}} \\
&= \sum_{k=2}^{v} \frac{\binom{v}{k}\binom{n-v}{v-k}}{\binom{n}{v}} p^{-\frac{d}{2}k} \Bigg(\underbrace{(1-p)^{\frac{d+1-k}{2}}}_{\leqslant (1-p)^{-\frac{v}{2}}}\Bigg)^{k} \leqslant \sum_{k=2}^{v} \frac{v^k}{2(k-2)!} \left(\frac{v}{n-v}\right)^{k} \left(p^{-\frac{d}{2}}(1-p)^{-\frac{v}{2}}\right)^{k} \\
&= \frac{1}{2} g^2 \sum_{k=2}^{v} \frac{1}{(k-2)!} g^{k-2} \leqslant \frac{1}{2} g^2 \exp(g),
\end{aligned}$$

as

$$\frac{\binom{v}{k}\binom{n-v}{v-k}}{\binom{n}{v}} \leqslant \frac{\binom{v}{k}\frac{(n-v)^{v-k}}{(v-k)!}}{\frac{(n-v)^{v}}{v!}} = \binom{v}{k}^2 \frac{k!}{(n-v)^k} \leqslant \frac{1}{k!}\left(\frac{v}{n-v}\right)^{k}.$$

The lemma follows: the probability of H not being an induced subgraph is at most $e^g g^2/2$. This upper bound is 1 exactly if $g = 2W(1/\sqrt{2})$. For $g \leqslant 2W(1/\sqrt{2})$, we obtain the upper bound in the lemma. For $g \geqslant 2W(1/\sqrt{2})$, the upper bound in the lemma is at least 1, so the statement is obvious. $\square$

5.2 High extension complexity with high probability

In order to obtain lower bounds on the extension complexity of the stable set polytope of $G = G(n,p)$, we use Lemma 8 together with Lemma 7, taking H to be $K_t^{\pentagon}$.

Main Theorem 5.1 (Super-polynomial xc of $\operatorname{STAB}(G(n,p))$ w.h.p.)**.** *For $p \geqslant 1/\sqrt[4]{n}$ and for every constant $c > 0$, we have*

$$\mathbb{P}\left[\operatorname{xc}(\operatorname{STAB}(G(n,p))) \geqslant 2^{c(\log 3/2)\sqrt{\frac{\ln n}{p}}}\right] \geqslant 1 - (c_1 + o(1)) \frac{n^{3c^2/2-2+o(1)} \ln^4 n}{p^8}, \tag{1}$$

where $c_1 := \frac{81 c_0 c^8}{16}$. In particular, for $\varepsilon < 1/4$, taking $c \leqslant \sqrt{2(1-4\varepsilon)/3}$,

$$\mathbb{P}\left[\operatorname{xc}(\operatorname{STAB}(G(n,n^{-\varepsilon})) \geqslant 2^{c(\log 3/2)\sqrt{n^{\varepsilon}\ln n}}\right] \geqslant 1 - (c_1 + o(1)) n^{-1+4\varepsilon+o(1)} \ln^4 n, \tag{2}$$

and (at the other end of the range) for fixed $\delta > 0$, taking $c = 1$,

$$\mathbb{P}\left[\operatorname{xc}(\operatorname{STAB}(G(n,\delta \ln^{-1} n))) \geqslant n^{(\log 3/2)\delta^{-1/2}}\right] \geqslant 1 - (c_1 + o(1)) \frac{n^{-1/2+o(1)} \ln^{12} n}{\delta^8}. \tag{3}$$

Moreover, for $p = n^{-1+\varepsilon}$ and any fixed $0 < \varepsilon < 8/11$

$$\mathbb{P}\left[\mathrm{xc}(\mathrm{STAB}(G(n, n^{-1+\varepsilon}))) \geqslant 2^{n^{\varepsilon/8}\log 3/2}\right] \geqslant 1 - c_0(1+o(1))\left(\frac{2(2-\varepsilon)}{\varepsilon}\right)^4 n^{-\varepsilon/2}. \tag{4}$$

Finally, for $c, k > 0$ and $p = c(\log^{4k+2} n)/n$

$$\mathbb{P}\left[\mathrm{xc}(\mathrm{STAB}(G(n, c(\log^{4k+2} n)/n))) \geqslant 2^{\log 3/2 \log^k n}\right] \geqslant 1 - c_0(1+o(1))\frac{4}{c^2}. \tag{5}$$

Proof. We apply Lemma 8 to the graph $H := K_t^{\pentagon}$ together with Lemma 7 to obtain:

$$\begin{aligned}
\mathbb{P}\left[\mathrm{xc}(\mathrm{STAB}(G(n,p))) \geqslant 2^{(\log 3/2)t}\right] &\geqslant \mathbb{P}\left[K_t^{\pentagon} \overset{\mathrm{ind}}{\subseteq} G(n,p)\right] \\
&\geqslant 1 - c_0\,\frac{v^4 p^{-d}(1-p)^{-v}}{(n-v)^2} \\
&\geqslant 1 - c_0\,(1+o(1))\frac{v^4 p^{-d} \mathrm{e}^{pv}}{v^2} \qquad \text{if } v = o(n).
\end{aligned}$$

Here v is the number of vertices of H, and every induced subgraph of H should have average degree at most d. We shall use the d provided by Lemma 6.

Now we shall substitute various values for p, t, d, ℓ to obtain the equations of the theorem.

For Equation (1), we choose

$$\ell = 0, \qquad t := \left\lceil c\sqrt{\frac{\ln n}{p}}\right\rceil, \qquad d = 4.$$

Note that $v = t + 3\binom{t}{2} = (1+o(1))\frac{3}{2}t^2 = (1+o(1))\frac{3}{2}c^2\frac{\ln n}{p} = o(n)$ as $p \geqslant 1/\sqrt[4]{n}$, and hence

$$\frac{v^4 p^{-d}\mathrm{e}^{pv}}{n^2} = (1+o(1))\frac{81}{16}c^8 p^{-8} n^{3c^2/2-2+o(1)} \ln^4 n.$$

This finishes the proof of Equation (1). Equations (2) and (3) are special cases of (1).

For Equations (4) and (5) we shall use a positive ℓ and let

$$\gamma := \frac{2\ell+3}{2}$$

to ease computation. Then

$$\begin{aligned}
d &= 2 + \frac{4}{2\ell+3} = 2 + \frac{2}{\gamma}, \\
v &= t + (2\ell+3)\binom{t}{2} = (1+o(1))\gamma t^2.
\end{aligned}$$

Hence

$$\frac{v^4 p^{-d}\mathrm{e}^{pv}}{n^2} = (1+o(1))\gamma^4 \frac{t^8 \mathrm{e}^{p(1+o(1))\gamma t^2}}{p^{2+2/\gamma} n^2}.$$

To prove Equation (4), let $p = n^{-1+\varepsilon}$, $\ell = 2\lceil 2/\varepsilon - 11/4\rceil$ and $t = n^{\varepsilon/8}$, then $\gamma \geqslant 4(1-\varepsilon)/\varepsilon$ and $\gamma < 2(2-\varepsilon)/\varepsilon$ therefore

$$\gamma^4 \frac{t^8 \mathrm{e}^{p(1+o(1))\gamma t^2}}{p^{2+2/\gamma} n^2} = (1+o(1))\left(\frac{2(2-\varepsilon)}{\varepsilon}\right)^4 n^{-\varepsilon/2}$$

proving the claim. Note that $\gamma p t^2 = o(1)$ and $\ell > 0$ as $\varepsilon < 8/11$.

For Equation (5), we choose $t = \lceil \log^k n \rceil$, $p = c(\log^{4k+2} n)/n$ and $\ell = 2\lceil (\log n)/2 - 3/4 \rceil$, then $\gamma \geqslant \log n$ and $\gamma = (1 + o(1)) \log n$ leading to

$$\gamma^4 \frac{t^8 e^{p(1+o(1))\gamma t^2}}{p^{2+2/\gamma} n^2} \leqslant (1 + o(1)) \frac{4}{c^2}$$

proving the claim. □

Main Theorem 2 gives super-polynomial lower bounds all the way from $p = \omega(\frac{\log^{6+\varepsilon} n}{n})$ to $p = O(1/\log n)$. The key for being able to cover the whole regime is to have the gadgets depend on the parameter choice. Notice that for $p < 1/n$ a random graph almost surely will have all its components of size $O(\log n)$, making the stable set problem easy to solve, so that we essentially leave only a small polylog gap.

5.3 Upper bound on extension complexity with high probability

We now complement Main Theorem 2 with an upper bound, which is close to the lower bound: there is only an essentially quadratic gap in the exponent.

Theorem 5.4 (Upper bound on the xc of STAB($G(n, p)$) w.h.p.)**.** *For $0 < p \leqslant 1/2$,*

$$\mathbb{P}\left[\mathrm{xc}(\mathrm{STAB}(G)) \geqslant 2^{\Omega\left(\frac{\ln^2 n}{p}\right)}\right] \leqslant n^{-\Omega\left(\frac{\ln n}{p}\right)}.$$

In particular, for $p = n^{-\varepsilon}$, we obtain $\mathbb{P}\left[\mathrm{xc}(\mathrm{STAB}(G)) \geqslant 2^{\Omega\left(n^{\varepsilon} \ln^2 n\right)}\right] = o(1)$ and similarly for $p = \delta \ln^{-1} n$, we get $\mathbb{P}\left[\mathrm{xc}(\mathrm{STAB}(G)) \geqslant n^{\Omega\left(\frac{\ln^2 n}{\delta}\right)}\right] = o(1)$.

In order to establish the upper bound stated in Theorem 9, we use the following result.

Lemma 5.5. *Every polytope P has an extension complexity at most the number of its vertices.*

Proof. Let V be the set of vertices of P, and let Q be a simplex with $|V|$ vertices. The simplex Q is an extension of P via mapping the vertices of Q one-to-one to V in an arbitrary fashion, and extending to an affine mapping on Q. This extension has size $|V|$. □

We are ready to prove Theorem 9.

Proof of Theorem 9. By standard arguments (see, e.g., [Diestel, 2005, Chapter 11, page 300]), for $G = G(n, p)$ we have

$$\mathbb{P}\left[\alpha(G) \geqslant r\right] \leqslant \left(n\, e^{-p(r-1)/2}\right)^r$$

and thus for $r = 4\frac{\ln n}{p}$ we get

$$\mathbb{P}\left[\alpha(G) \geqslant 4\frac{\ln n}{p}\right] \leqslant \left(\frac{n}{\sqrt{e}}\right)^{-4\frac{\ln n}{p}}.$$

Therefore, with very high probability, we have $\alpha(G) \leqslant 4\frac{\ln n}{p}$.

Using the inequality $\sum_{i=0}^{k} \binom{n}{i} \leqslant (n+1)^k$, we get

$$\#(\text{stable sets in } G) \leqslant (n+1)^{\alpha(G)} = 2^{\log(n+1)\alpha(G)} = 2^{\left(\frac{1}{\ln 2} + o(1)\right) \ln(n)\alpha(G)}.$$

The result then follows directly from Lemma 11. □

6 Concluding remarks

We conclude with the following conjecture that would further strengthen the results above as well as establishing truly exponential lower bounds on the extension complexity of further combinatorial problems.

Conjecture 6.1 (Sparse Graph Conjecture)**.** *There exists an infinite family $(T_k)_{k\in\mathbb{N}}$ of template graphs such that, denoting by t_k the number of vertices of T_k: (i) $t_k = 2^{O(k)}$; (ii) T_k has bounded average degree; (iii) $\mathrm{xc}(\mathrm{COR}(T_k)) = 2^{\Omega(t_k)}$.*

The existence of such a family would have various consequences.

Exact case. Assuming the Sparse Graph Conjecture we would obtain that the extension complexity of polytopes for important combinatorial problems considered in Fiorini et al. [2012], Avis and Tiwary [2013], Pokutta and Van Vyve [2013] including (among others) the stable set polytope, knapsack polytope, and the 3SAT polytope would have truly exponential extension complexity, that is $2^{\Omega(n)}$ extension complexity, where n is the *dimension* of the polytope.

In particular we would obtain a polyhedral variant of the exponential time hypothesis (ETH), i.e., in the worst case any LP expressing the 3SAT polytope would require an exponential number of inequalities *independently of P vs. NP*. Currently, for all those problems the best lower bound is $2^{\Omega(\sqrt{n})}$ due to the fact that the template graph used is the complete graph, which has linear average degree. Moreover, truly exponential lower bounds would imply that for these problems the extended formulation arising from the convex hull of the vertices is essentially optimal, that is, up to a constant factor in the exponent.

The recent groundbreaking result of Rothvoß [2013] gives $2^{\Omega(n)}$ bounds for the extension complexity of the matching polytope and TSP polytope. These bounds are also tight up to constants, but this time the upper bound does not come from the number of vertices but rather from the number of facets and dynamic programming algorithms, respectively. Notice that the dimension of both polytopes is $d = \Theta(n^2)$, thus the bounds are in fact $2^{\Omega(\sqrt{d})}$.

Average case. As observed above, there is a quadratic gap in the best current lower and upper bounds on the worst-case extension complexity of the stable set polytope: $2^{\Omega(\sqrt{n})}$ versus 2^n respectively. This is reflected in the results we obtain here. Assuming the Sparse Graph Conjecture we could reduce the gap between upper and lower bounds to a logarithmic factor. Moreover, our results could be strengthened to establish super-polynomial lower bounds on the average-case extension complexity up to constant probability p.

References

D. Avis and H. R. Tiwary. On the extension complexity of combinatorial polytopes. *ArXiv e-prints*, Feb. 2013.

G. Braun and S. Pokutta. Common information and unique disjointness. *accepted for FOCS / available at* `http://eccc.hpi-web.de/report/2013/056/`, 2013.

G. Braun, S. Fiorini, S. Pokutta, and D. Steurer. Approximation Limits of Linear Programs (Beyond Hierarchies). In *53rd IEEE Symp. on Foundations of Computer Science (FOCS 2012)*, pages 480–489, 2012. ISBN 978-1-4673-4383-1. doi: 10.1109/FOCS.2012.10.

M. Braverman and A. Moitra. An information complexity approach to extended formulations. *Electronic Colloquium on Computational Complexity (ECCC)*, 19(131), 2012.

S. Chan, J. Lee, P. Raghavendra, and D. Steurer. Approximate constraint satisfaction requires large LP relaxations. To appear in FOCS'13, 2013.

M. Conforti, G. Cornuéjols, and G. Zambelli. Extended formulations in combinatorial optimization. *4OR*, 8:1–48, 2010. doi: 10.1007/s10288-010-0122-z.

R. Diestel. *Graph Theory*. Springer-Verlag Heidelberg, New York, electronic edition edition, 2005.

S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf. Linear vs. Semidefinite Extended Formulations: Exponential Separation and Strong Lower Bounds. *Proceedings of STOC 2012*, 2012.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42:1115–1145, 1995. doi: 10.1145/227683.227684.

V. Kaibel. Extended formulations in combinatorial optimization. *Optima*, 85:2–7, 2011.

V. Kaibel and S. Weltge. A Short Proof that the Extension Complexity of the Correlation Polytope Grows Exponentially. *ArXiv e-prints*, July 2013.

K. Pashkovich. *Extended Formulations for Combinatorial Polytopes*. PhD thesis, Magdeburg Universität, 2012.

S. Pokutta and M. Van Vyve. A note on the extension complexity of the knapsack polytope. *Operations Research Letters*, 41:347–350, 2013.

A. A. Razborov. On the distributional complexity of disjointness. *Theoret. Comput. Sci.*, 106(2): 385–390, 1992.

T. Rothvoß. Some 0/1 polytopes need exponential size extended formulations, 2011. arXiv:1105.0036.

T. Rothvoß. The matching polytope has exponential extension complexity. *ArXiv e-prints*, 2013.

J. Williams. A linear-size zero-one programming model for the minimum spanning tree problem in planar graphs. *Networks*, 39:53–60, 2002.

R. d. Wolf. Nondeterministic quantum query and communication complexities. *SIAM Journal on Computing*, 32(3):681–699, 2003.

M. Yannakakis. Expressing combinatorial optimization problems by linear programs (extended abstract). In *Proc. STOC 1988*, pages 223–228, 1988.

M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. System Sci.*, 43(3):441–466, 1991. doi: 10.1016/0022-0000(91)90024-Y.

A Faithful linear encodings and the uniform model

As mentioned before, faithful linear encodings were introduced in Braun et al. [2012] to study the polyhedral hardness of approximation of various problems. The purpose of such encodings is to recast instances as linear programs over a polytope. The crucial feature of faithful linear encodings is that this polytope is not allowed to depend on the input, which is solely encoded in the objective function. The same polytope is used for all inputs of the same size.

In general, a *(faithful) linear encoding* of a combinatorial problem consists of a set $\mathcal{F} \subseteq \{0,1\}^*$ of *feasible solutions* and a set $\mathcal{O} \subseteq \mathbb{R}^*$ of *admissible objective functions*. An *instance* of the linear encoding specifies a dimension d and an admissible objective function w of dimension d. Solving such an instance means finding $x \in \mathcal{F} \cap \{0,1\}^d$ such that $w^\intercal x$ is maximum.

For every fixed dimension d, a linear encoding defined by $\mathcal{F} \subseteq \{0,1\}^*$ and $\mathcal{O} \subseteq \mathbb{R}^*$ yields an inner 0/1-polytope

$$P := \operatorname{conv}\left(\{x \in \{0,1\}^d \mid x \in \mathcal{F}\}\right)$$

and an outer convex set

$$Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathcal{O} \cap \mathbb{R}^d : w^\intercal x \leqslant \max\{w^\intercal y \mid y \in P\}\}.$$

such that $P \subseteq Q$. Roughly speaking, the inner 0/1-polytope P encodes the feasible solutions to the problem and the outer polyhedron Q encodes the admissible objective functions.

A.1 Faithful encoding for the maximum stable set problem

In the case of the maximum stable set problem, *the feasible solutions* are all subsets of vertices, as potential stable sets, encoded as $bb^\intercal$ where $b \in \{0,1\}^n$ is a characteristic vector of a set. That is, we let

$$\mathcal{F} := \{bb^\intercal \mid b \in \{0,1\}^n,\ n = 0,1,\dots\}.$$

Recall that for each graph G with $V(G) \subseteq [n]$, the objective function $w^G \in \mathbb{R}^{n\times n}$ is defined by letting $w^G_{ij} = 1$ if $i = j$ and $i \in V(G)$, $w^G_{ij} = w^G_{ji} = -1$ if $ij \in E(G)$ and $w^G_{ij} = 0$ otherwise. Notice that $\max\{\langle w^G, y\rangle \mid y \in \mathcal{F} \cap \{0,1\}^{n^2}\} = \alpha(G)$, the stability number of G. We collect all vectors $w^G \in \mathbb{R}^{n^2}$ to form the set of admissible objective functions. For technical reasons, we add the vectors $-e_{ij}$ for $i,j \in [n]$ to the set of admissible objective functions. In other words, we let

$$\mathcal{O} = \{w^G \mid G \text{ graph }\} \cup \{-e_{ij} \mid i,j \in [n]\}.$$

The sets $\mathcal{F}$ and $\mathcal{O}$ define a faithful linear encoding of the maximum stable set problem. The corresponding polyhedral pair (P,Q) consists of $P = \mathrm{COR}(n)$ for the inner 0/1-polytope and $Q = \{x \in \mathbb{R}^{n\times n}_+ \mid \forall \text{ graphs } G \text{ such that } V(G) \subseteq [n] : \langle w^G, x\rangle \leqslant \alpha(G)\}$ for the outer polyhedron.

A.2 Extension complexity of a pair

Let $P \subseteq Q \subseteq \mathbb{R}^d$ denote any nested pair of polyhedra. Possibly, this pair of polyhedra corresponds to the d-dimensional vectors of a linear encoding $\mathcal{F}, \mathcal{O}$. Recall that the extension complexity $\mathrm{xc}(P,Q)$ is defined as the minimum number of facets of a polyhedron affinely projecting between P and Q.

In order to analyze the extension complexity of the pair (P,Q), we consider an inner description of P and an outer description of Q

$$P := \operatorname{conv}(\{v_1,\dots,v_n\}) + \operatorname{cone}\{r_1,\dots,r_k\} \qquad Q := \{x \in \mathbb{R}^d \mid Ax \leqslant b\},$$

where the system $Ax \leqslant b$ consists of m inequalities $A_i x \leqslant b_i$ with $i \in [m]$. The *slack matrix of the pair* (P, Q) (w.r.t. these inner and outer descriptions) is the $m \times (n+k)$ matrix $S^{P,Q} = \left[S^{P,Q}_{\text{vertex}}, S^{P,Q}_{\text{ray}}\right]$ given by block decomposition into a vertex and ray part:

$$\begin{aligned} S^{P,Q}_{\text{vertex}}(i,j) &:= b_i - A_i v_j, && i \in [m],\ j \in [n], \\ S^{P,Q}_{\text{ray}}(i,j) &:= -A_i r_j, && i \in [k],\ j \in [n]. \end{aligned}$$

Braun et al. [2012] prove the following characterization of $\operatorname{xc}(P,Q)$ in terms of the nonnegative rank $\operatorname{rank}_+(S^{P,Q})$ of the slack matrix of the pair (P,Q). It generalizes Yannakakis seminal factorization theorem (see Yannakakis [1991]); see also Pashkovich [2012].

We recall the definition of the nonnegative rank of a matrix $M \in \mathbb{R}^{m \times n}$. A *rank-$r$ nonnegative factorization of M* is a factorization of $M = TU$ where $T \in \mathbb{R}^{m \times r}_+$ and $U \in \mathbb{R}^{r \times n}_+$. This is equivalent to $M = \sum_{i \in [r]} u_i v_i^\intercal$ for some (column vectors) $u_i \in \mathbb{R}^m_+, v_i \in \mathbb{R}^n_+$ with $i \in [r]$. The *nonnegative rank of M*, denoted by $\operatorname{rank}_+ M$, is the minimum r such that there exists a rank-r nonnegative factorization of M.

Theorem A.1 (Factorization theorem). *For every slack matrix $S^{P,Q}$ of the pair (P,Q), we have* $\operatorname{rank}_+(S^{P,Q}) - 1 \leqslant \operatorname{xc}(P,Q) \leqslant \operatorname{rank}_+(S^{P,Q})$. *If the affine hull of P is not contained in Q and the recession cone of Q is not full-dimensional, we have* $\operatorname{xc}(P,Q) = \operatorname{rank}_+(S^{P,Q})$. *In particular, this holds when P and Q are polytopes of dimension at least* 1.

A.3 Unique Disjointness

As we have seen, the hardness of the maximum stable set problem arises from having the unique disjointness (partial) matrix UDISJ (or its shifted variant) as a submatrix. Recall that the UDISJ matrix M has 2^n rows and 2^n columns indexed by $0/1$-vectors a and b with entries:

$$M(a,b) = \begin{cases} 0 & \text{if } a^\intercal b = 1 \\ 1 & \text{if } a^\intercal b = 0. \end{cases} \tag{6}$$

Formally, M is only a partial matrix as not all of its entries are defined; we will refer to it as matrix from here on. The fact that it is only partial does not matter for our purpose, as we only care for whether this (partial) matrix occurs as an induced submatrix. UDISJ has been studied in many disciplines, arguably the most notable being communication complexity.

The seminal work of Razborov [1992] together with an observation in Wolf [2003] is at the core of the first results establishing high extension complexity for the correlation polytope, cut polytope, stable set polytope, and the TSP polytope in Fiorini et al. [2012].

Braun et al. [2012] prove that *any* $2^n \times 2^n$ matrix M with rows and columns indexed by vectors in $\{0,1\}^n$ satisfying (6) has superpolynomial nonnegative rank, and that this remains true even if we shift the entries of the matrix M by some number $\rho = O(n^{1/2-\varepsilon})$. This results was then strengthened to $\rho = O(n^{1-\varepsilon})$ in Braverman and Moitra [2012] which then immediately leads to a polyhedral inapproximability of CLIQUE (for a specific linear encoding!) of $O(n^{1-\varepsilon})$, matching the Håstad's hardness result for approximating CLIQUE.

The *ρ-shifted unique disjointness (UDISJ) matrix* is any $2^n \times 2^n$ matrix indexed by pairs (a,b) where $a, b \in \{0,1\}^n$ such that

$$M_{ab} = \begin{cases} \rho & \text{if } a^\intercal b = 1 \\ 1 + \rho & \text{if } a^\intercal b = 0. \end{cases}$$

In Braun and Pokutta [2013] a new information-theoretic approach for studying the nonnegative rank (and hence the extension complexity) has been developed. This approach allows to lower bound the nonnegative rank of various 'deformations' of the UDISJ matrix. Theorem 3 above is one of the quantitative results from Braun and Pokutta [2013], which informally speaking shows that the UDISJ matrix has high nonnegative matrix almost everywhere.