February 1, 2017 - June 16, 2017

# Bounding Techniques for Extension Complexity

Johanna Seif

*Supervised by*
Samuel Fiorini

# Contents

# 1 Introduction

This report describes the work done during the internship with Professor Samuel Fiorini from February 1 to June 16, 2017. It took place in the Algebra and Combinatorics research unit of the Université Libre de Bruxelles in Brussels, Belgium.

Many mathematical problems can be expressed as a linear program (LP) . Geometrically, the constraints define a polyhedron in a space of dimension equal to the number of variables. However, computationally solving an LP can take a lot of time if there are many constraints. In addition, there exist some problems that are polytime solvable, can be expressed as an LP, but still this LP have an exponential number of constraints in the corresponding: the minimum spanning tree problem for instance is one of them. So, how can such an easy problem be so geometrically hard to describe?

A way to study the "structural complexity" of a polyhedron (or polytope in the bounded case) is extension complexity. Some problems can be expressed with fewer constraints by extending the number of variables. Since the complexity of an LP is linked to the number of constraints, this leads to faster computations. The extension complexity of a polytope $P$ is the minimal number of facets of a polytope that projects to $P$. The number of facet being linked to the number of inequality, the extension and optimization complexities are linked. Therefore to some extend extension complexity is linked to computation. Figure 1 shows a polytope in 2 dimensions with 8 facets that is the projection of a 6 facet polytope in dimension 3. The case of regular polygon shows that there can be a gap between the number of faces and the extension complexity since the extension complexity of an $n$-gon is $\theta(\log n)$.

So, even though Edmonds' [Edm71] description of the spanning tree polytope has an exponential number of constraints, Martin [Mar91] showed that it can be expressed as the projection of a higher dimensional polytope with $O(n^3)$ inequalities.

It is then natural to wonder if some $NP$-hard problems can have a corresponding polytope with polynomial extension complexity, since this would imply $NP = P/poly$. On the other hand one can wonder which polytopes have high extension complexity. One of the major results is by Fiorini, Massar, Pokutta, Tiwary and de Wolf [FMP+11], where they prove that many polytopes, such as the correlation polytope or the traveling salesman problem polytope have exponential extension complexity. This lower bounding technique also raised another important question: do all polytopes for which the associated optimization problem is in $P$ have low extension complexity? This was solved in 2014 by Rothvoss [Rot14], who showed that the extension complexity of a perfect matching polytope of a complete $n$-node graph is $2^{\Omega(n)}$. But maximum weight matching is known to be polynomial.

All these major results are due to the development of bounding techniques. There are many techniques to bound the extension complexity. Most recent papers directly study a special matrix linked to the polytope, called the slack matrix. This is due to the fact that the extension complexity of a polytope is equal to the non-negative rank of its slack matrix. Many bounds were introduced thanks to this connection. Among them are common information introduced by Braun and Pokutta in [BJLP14] and the hyperplane separation bound that can be found in [Rot14]. Another lower bounding technique is to reduce the study of a polytope to a polytope that is known to have high extension complexity. For instance, if we can project a polytope $P$ to a polytope $Q$ with high extension complexity, then $P$ too has high extension complexity. Upper bounds can be obtained by explicit construction of an extension of the polytope studied. There are other ways to obtain then, but constructive proofs have the advantage to provide a structure on which algorithms can be run.

In this report we consider the study of lower bounds computed on the slack matrix, a reduction proof to lower bound the correlation polytope, and a explicit construction of an extension. We start by defining the notion of extension complexity 2. Then we define to lower bounds 3: the hyperplane separation bound and common information. We show in Section 4 that computing common information is in $NP$ and that computing the hyperplane separation bound is $NP$-complete and even hard to approximate. Then, in Section 4 we study the extension complexity of the correlation polytope of minor closed class of graphs. For these graph we show a tight connection to the treewidth of the graph. For this polytope, the lower bound is obtained by reduction, and the upper bound by an explicit construction of an extension.

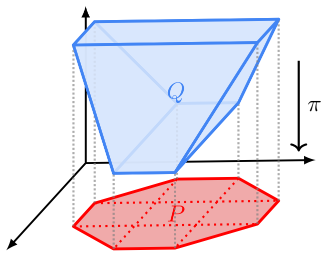



Figure 1: Extension of a polytope with less facets

# 2 Extension complexity and related notions

In this section we formally introduce the notion of extension complexity and some basic notions related to it.

## 2.1 Definitions and notation

We denote by $[n]$ the set $\{1, \ldots, n\}$ where $n \in \mathbb{Z}_{>0}$.

**Definition 1 .** A *polytope* in $\mathbb{R}^d$ is the convex hull of a finite set of points in $\mathbb{R}^d$. The *dimension*, $\dim P$, of a polytope $P$ is the maximal number of affinely independent points in $P$ minus 1. We define a *polyhedron* in $\mathbb{R}^d$ as a subset $P \subseteq \mathbb{R}^d$ defined by $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$, where $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$.

As shown by the Minkowski-Weyl theorem (see [Zie12]) there is an equivalence between being a polytope and being a bounded polyhedron, which means that a polytope can be described by a system of inequalities.

**Definition 2 .** Let $P \subseteq \mathbb{R}^d$ be a polytope. Let $c \in \mathbb{R}^d$, $k \in \mathbb{R}$ and $H = \{x \in \mathbb{R}^d \mid c^\intercal x = k\}$ be a hyperplane. $H$ is a *valid hyperplane* for $P$ if for every $x$ in $P$, $c^\intercal x \leqslant k$. We say $F \subseteq P$ is a *face* of $P$ if $F = P$ or there exists a valid hyperplane $H$ such that $F = P \cap H$. The *dimension of a face* is the dimension of its affine hull. A *facet* of a polytope of dimension $n$ is a $(n-1)$-dimensional face.

**Definition 3 .** The *extension complexity* of a polytope $P$ is the minimal number of facets $f$ for which there exists a polytope $Q$ with at most $f$ facets and an affine map $\pi$ that satisfies $\pi(Q) = P$. The extension complexity of $P$ is denoted $\mathrm{xc}(P)$.

## 2.2 First tools to study extension complexity

We define the notions of slack matrix and non-negative factorization which provide algebraic tools to study the extension complexity of a polytope.

**Definition 4 .** We consider a polytope with two descriptions $P = \mathrm{conv}\{x_1, \ldots, x_n\} = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$ where $A \in \mathbb{R}^{m\times d}$. We define the *slack matrix* $S \in \mathbb{R}^{m\times n}$ by $S_{ij} = b_i - A_i x_j$. The slack matrix is therefore a non-negative matrix.

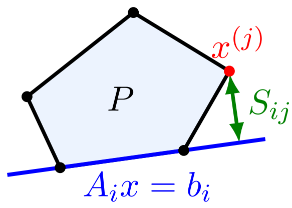
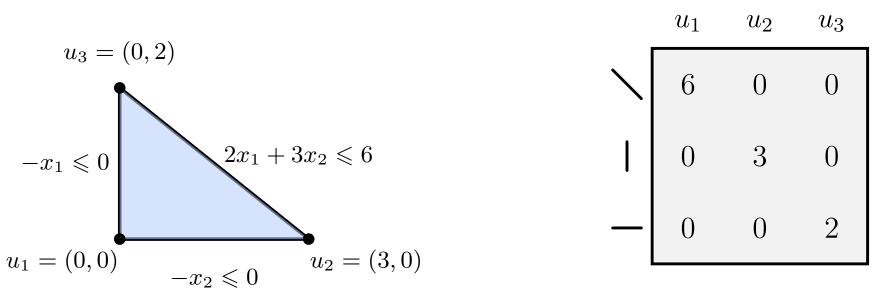



Figure 2: Geometrical interpretation of the slack

The slack matrix of a polytope is not unique since the system of inequalities that describes a polytope is not unique. We call $S_{ij}$ the *slack* of the $j$th vertex $x_j$ with respect to the $i$th facet-defining inequality $A_i x \leqslant b_i$ of $P$. It represents the distance of the vertex $x_j$ from the hyperplane $A_i x = b_i$ when $\|A_i\|_2 = 1$ (see Figure 2). Figure 3 gives an example of slack matrix.




Polytope

Slack matrix

Figure 3: A polytope and its slack matrix

**Definition 5 .** Let $M \in \mathbb{R}^{n\times d}_{\geqslant 0}$. The *non-negative rank* of $M$ is defined by:

$$\mathrm{rk}_+(M) = \min\{r \mid \exists U \in \mathbb{R}^{n\times r}_{\geqslant 0}, V \in \mathbb{R}^{r\times d}_{\geqslant 0} : M = UV\}.$$

Such $U$ and $V$ matrices provide a *non-negative factorization* of $M$. Equivalently:

$$\mathrm{rk}_+(M) = \min\left\{ r \;\middle|\; M = \sum_{i=1}^{r} p_i q_i^{\intercal},\ p_1, \ldots, p_r \geqslant 0,\ q_1, \ldots, q_r \geqslant 0 \right\}$$

Any such sum on $r$ elements is a *non-negative decomposition* of $M$.

The following theorem shows how extension complexity can be expressed as non-negative rank.

**Theorem 1** [Yan91]**.** *Let $P$ be a polytope such that* $\dim P \geqslant 1$ *and $S$ any of its associated slack matrices. Then* $\mathrm{xc}(P) = \mathrm{rk}_+(S)$.

However the nonnegative rank is $NP$-hard to compute and even $ETR$-complete as Shitov shows [Shi16]. Therefore, many lower bounding techniques were developed in order to bound the extension complexity.

For all set $A$ we define its the characteristic vector $\chi^A$ as $\chi^A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases}$

**Definition 6 .** The *support* of a matrix $M \in \mathbb{R}^{n\times d}$, denoted by $\mathrm{supp}(M)$, is defined as:

$$\mathrm{supp}(M) = \{(i,j) \in [n] \times [d] \mid M_{ij} \neq 0\}$$

A *rectangle* is a set $R = I \times J \subseteq [n] \times [d]$. Informally $I$ (resp. $J$) is a subset of the rows (resp. columns). We identify the support with the matrix $\chi^{\mathrm{supp}(M)} \in \mathbb{R}^{n\times d}$ and a rectangle $R$ with $\chi^R$. The *rectangle covering number* of a matrix $M$ is the minimal number of rectangles with support contained in $\mathrm{supp}(M)$ that cover $\mathrm{supp}(M)$. We denote it by $\mathrm{rc}(M)$.

**Example 1 .** We consider the following matrix: $M = \begin{pmatrix} 0 & 4 & 8 & 3 & 5 \\ 0 & 4 & 6 & 7 & 2 \\ 0 & 7 & 5 & 0 & 7 \\ 0 & 0 & 4 & 1 & 0 \end{pmatrix}$. Then the support of $M$ is given in Figure 4a. Here the rectangle covering bound is 2 since the two following rectangles cover all the non-zero values. The first one $R_1$ is in red bold in Figure 4b. And the second one $R_2$ is shown in blue bold in Figure 4c.

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

(a) Support $\mathrm{supp}\, M$

$$\begin{pmatrix} 0 & \mathbf{1} & \mathbf{1} & 1 & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{1} & 1 & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{1} & 0 & \mathbf{1} \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

(b) Rectangle $R_1$

$$\begin{pmatrix} 0 & 1 & \mathbf{1} & \mathbf{1} & 1 \\ 0 & 1 & \mathbf{1} & \mathbf{1} & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & \mathbf{1} & \mathbf{1} & 0 \end{pmatrix}$$

(c) Rectangle $R_2$

Figure 4: Biclique cover of the graph associated to $M$

The rectangle covering number provides a lower bound for the extension complexity. Indeed, let $S \in \mathbb{R}^{n\times d}$ be the slack matrix, $r = \mathrm{rk}_+(S)$. Consider a non-negative factorization of $S$, $S = UV$ where $U \in \mathbb{R}^{n\times r}_{\geqslant 0}, V \in \mathbb{R}^{r\times d}_{\geqslant 0}$.

Then the rectangles $R_k = \{i \in [n] \mid U_{ik} > 0\} \times \{j \in [d] \mid V_{kj} > 0\}$ for $k \in [r]$ cover the support of $S$. This gives the following theorem.

**Theorem 2** [Yan91]**.** *Let $P$ be a polytope such that $\dim P \geqslant 1$ and $S$ its associated slack matrix. Then $\mathrm{rc}(S) \leqslant \mathrm{xc}(P)$.*

This bound has been used in [FMP+11] to prove exponential lower bound, for instance for the correlation polytope of a complete graph (see Definition 14). It is the best bound possible when the only given information is the support of a slack matrix. However sometimes to obtain a more precise lower bound, we need to use more than the support of the matrix. The lower bounds introduced in 3 are some of the lower bounds used in this case.

## 2.3 Biclique cover and partition

In this section we link the biclique cover and rectangle cover problems.

**Definition 7 .** A *biclique* is a complete bipartite graph. A *biclique cover* of a graph $G$ is a collection of biclique subgraphs of $G$ that cover all the edges of $G$. The *biclique cover number* $\mathrm{bc}(G)$ is the minimum number of bicliques needed to have a biclique cover. A *biclique partition* of $G$ is a biclique cover in which each edge is covered exactly once. The *biclique partition number* $\mathrm{bp}(G)$ is the minimum number of bicliques needed to have a biclique partition. Similarly, a *clique* is a complete graph, and a *clique partition* of $G$ is a clique cover in which each edge is covered exactly once. The *clique partition number* $\mathrm{cp}(G)$ is the minimum number of cliques needed to have a clique partition.

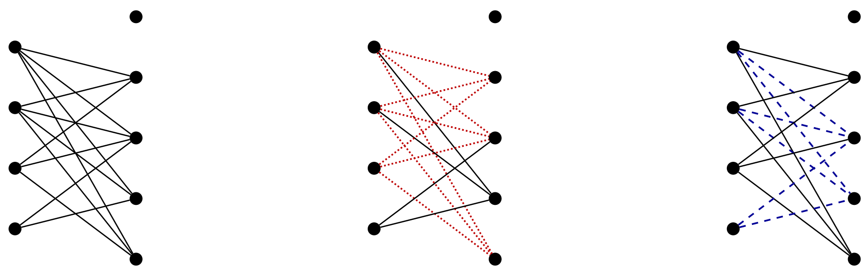
Now consider a non-negative matrix $M$. We construct a bipartite graph $G_M = (R \cup C, E)$ where $R$ is the set of row indices and $C$ is the set of column indices of $M$. We draw an edge $ij$ in $G_M$ for every $ij$ such that $M_{ij} \neq 0$. Then let $R = I \times J$ a rectangle of $M$ included in $\text{supp}(M)$. Then the subgraphs whose edges are $ij$ where $i \in I$ and $j \in J$ corresponds to a biclique in $G_M$, and vice-versa. Therefore a rectangle cover is equivalent to a biclique cover. This proves the following result.

**Theorem 3 .** *For $M$ a non-negative matrix,* $\text{bc}(G_M) = \text{rc}(M)$.

Let $S$ be some slack matrix of a polytope $P$. By Theorems 1 and 2 the biclique cover number of the graph $G_S$ provides a lower bound on the extension complexity of $P$.

**Example 2 .** Consider again the matrix $M$ in Example 1. The associated graph $G_M$ is depicted in Figure 5. We illustrate the biclique cover corresponding to the previous rectangle cover.

The vertices on the left side (resp. right side) are the ones associated to the rows (resp. columns), the upper one being the first row (resp. column). In this example, the red dotted and the blue dashed bicliques are the bicliques obtained by considering the rectangles of the rectangle cover, and provide a biclique cover of the graph.



Graph $G_M$ associated to $M$ — Biclique associated to $R_1$ — Biclique associated to $R_2$

Figure 5: Biclique cover of the graph associated to $M$

# 3 Lower bound definitions

In this section, we define the two lower bounds on which we worked during the internship: the hyperplane separation bound and common information.

## 3.1 Hyperplane separation bound

Let us first introduce the hyperplane separation bound. For two matrices $M, N \in \mathbb{R}^{m \times n}$, we define the Frobenius inner product $\langle M, N \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} M_{ij} \cdot N_{ij}$. For $M$ a matrix, define $\|M\|_\infty = \max_{ij} |A_{ij}|$.

**Definition 8 .** For $W \in \mathbb{R}^{m \times n}$, let $\rho(W) := \max\{\langle W, R \rangle \mid R \in \{0,1\}^{m \times n} \text{ rectangle}\}$. The *hyperplane separation bound* of a matrix $M \in \mathbb{R}_{\geqslant 0}^{m \times n}$ is then defined by:

$$\text{HSB}(M) = \sup_{W \in \mathbb{R}^{n \times d}} \frac{\langle W, M \rangle}{\|M\|_\infty \cdot \rho(W)}$$

In the case where $W \in \mathbb{R}_{\leqslant 0}^{m \times n}$ have that $\rho(W) = 0$, but since numerator is less than zero, we consider that we obtain 0. Similarly, for a zero matrix we consider we have 0. There exists another equivalent definition using the rectangle polytope.

**Definition 9 .** We define the *rectangle polytope* $\mathrm{RECT}(m,n)$ as the convex hull of all the rectangles of $[m]\times[n]$ matrices.

$$\mathrm{RECT}(m,n) = \mathrm{conv}\{R \mid R_{ij} = x_i y_j, x \in \{0,1\}^m, y \in \{0,1\}^n\}$$

**Theorem 4 .** *Let $M \in \mathbb{R}^{m\times n}$ with $\|M\|_\infty = 1$. Then:*

$$\mathrm{HSB}(M) = \min_{\lambda>0}\{\lambda^{-1} \mid \lambda M \in \mathrm{RECT}(m,n)\}$$

*Proof.*

($\leqslant$) Take $\lambda > 0$ such that $\lambda M \in \mathrm{RECT}(m,n)$. Then there exists non-negative numbers $(\mu_i)_{i\in[k]}$ and rectangles $(R_i)_{i\in[k]}$ such that $\lambda M = \sum_{i=1}^k \mu_i R_i$ and $\sum_{i=1}^k \mu_i = 1$. By definition, for all $W$:

$$\langle W, \lambda M\rangle = \left\langle W, \sum_{i=1}^k \mu_i R_i \right\rangle = \sum_{i=1}^k \mu_i \langle W, R_i\rangle \leqslant \sum_{i=1}^k \mu_i \rho(W) = \rho(W)$$

So $\frac{1}{\lambda} \geqslant \frac{\langle W,M\rangle}{\rho(W)}$.

($\geqslant$) Let $\lambda^*$ be such that $(\lambda^*)^{-1} = \min_{\lambda>0}\{\lambda^{-1} \mid \lambda M \in \mathrm{RECT}(m,n)\}$. By minimality of $\lambda^*)^{-1}$ $\lambda^* M$ is on the boundary of $\mathrm{RECT}(m,n)$ hence in a proper face $F$ of $\mathrm{RECT}(m,n)$. The face $F$ is determined by an inequality $\langle W^*, X\rangle \leqslant b$ where $W^* \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}$. Since $F$ contains $\lambda^* M$, it is nonempty so $b = \rho(W^*)$. Then $\rho(W^*) = \langle W^*, \lambda^* M\rangle$. Therefore $\mathrm{HSB}(M) \geqslant \frac{\langle W^*,M\rangle}{\rho(W^*)} = \frac{1}{\lambda^*}$. $\square$

Figure 6 gives a geometrical illustration of the hyperplane separation bound and the link between the definitions given in Definition 8 and Theorem 4.

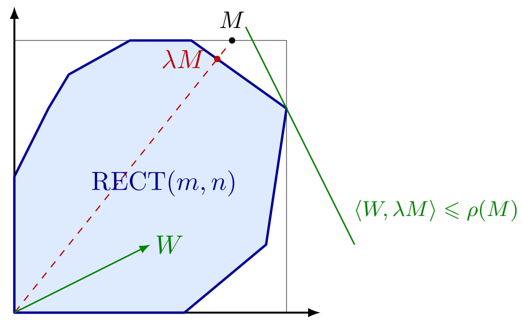



Figure 6: Geometrical interpretation of the hyperplane separation bound

This equivalent definition provides a first upper bound on the hyperplane separation bound .

**Theorem 5 .** *Let $M \in \mathbb{R}^{m\times n}_{\geqslant 0}$. Then* $\mathrm{HSB}(M) \leqslant m$.

**Theorem 6** [Rot14]**.** *Let $S \in \mathbb{R}^{n\times d}_{\geqslant 0}$ be a slack matrix of a polytope $P$. Then:*

$$\mathrm{xc}(P) \geqslant \mathrm{HSB}(S)$$

**Example 3 .** We consider $J_n$ the $n$ by $n$ matrix with ones everywhere, and $I_n$ the identity. Let $M = J_n - I_n$. We define a probability distribution on the rectangles $R = I\times J$ where $I$ and $J$ are disjoint and $I\cup J = \{1,\dots,n\}$ by: for all $i$ in $\{1,\dots,n\}$: $\begin{cases} i\in I & \text{with probability } 1/2 \\ i \in J & \text{with probability } 1/2 \end{cases}$ , independently for all $i$.

We call $\lambda_R$ this distribution. For $i \neq j$, $(i,j) \in R$ if and only if $i \in I, j \notin I$ and the probability of this is then $1/4$. Then:

$$\sum_{R \text{ rectangle}} \lambda_R \chi^R = \frac{1}{4} M$$

Therefore $\mathrm{HSB}(M) \leqslant 4$.

We define the weight matrix:

$$W = \begin{pmatrix} -\infty & & 1 \\ & \ddots & \\ 1 & & -\infty \end{pmatrix}$$

The $-\infty$ means here that we take coefficients really big in absolute value in comparison to the value of the elements of the matrix and of the size. We introduced the disjointness matrix:

$$DISJ_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad DISJ_n = DISJ_1^{\otimes n} \text{ for } n \geqslant 1$$

We have $\langle W, DISJ_n \rangle = n(n-1)$. A rectangle $R = I \times J$ that maximizes the scalar product must avoid each $-\infty$ entry, and so $I \cap J = \emptyset$. Also by maximality, $I \cup J = \{1, \ldots, n\}$ and so $\rho(W) = n^2/4$. Therefore $\mathrm{HSB}(M) \geqslant \frac{n(n-1)}{n^2/4} = 4\frac{n-1}{n}$.

So asymptotically the hyperplane separation bound is 4.

### 3.2 Common information

We introduce common information which provides a new lower bound that is based on an information theory measure. The bound was first introduced by Wyner [Wyn75].

**Definition 10 .** Let $P$ and $Q$ be two probability distributions on a finite set $X$. The *entropy* of $P$ is: $\mathbb{H}[P] = -\sum_x P(x) \log P(x)$. The *conditional entropy* of $P$ condition by $Q$ is: $\mathbb{H}(P|Q) = \sum_y Q(y)\mathbb{H}[P|Q(y)]$. The *mutual information* of $P$ and $Q$ is: $\mathbb{I}(P;Q) = \mathbb{H}[P] - \mathbb{H}[P|Q]$.

**Definition 11 .** For $M \in \mathbb{R}_{\geqslant 0}^{m \times n}$, $M \neq 0$, we consider the row and column joint random variables $(R, C)$. The *induced distribution* is then naturally, for all $i$ in $\{1, \ldots, m\}$ and $j$ in $\{1, \ldots, n\}$:

$$\mathbb{P}[R = i, C = j] = \frac{M_{ij}}{\sum_{k,\ell} M_{k\ell}}$$

A random variable $\Pi$ is a *seed* for $M$ if $R$ and $C$ are independent given $\Pi$.

Now let us define the common information of a matrix.

**Definition 12 .** Let $M \in \mathbb{R}_{\geqslant 0}^{m \times n}$. The *private information* is defined as:

$$\mathbb{W}[M] = \sup_{\Pi \text{ seed for } M} \mathbb{H}[R, C|\Pi] = \sup_{\Pi \text{ seed for } M} \mathbb{H}[R|\Pi] + \mathbb{H}[C|\Pi]$$

The *common information* of $M$ is defined as:

$$\mathbb{C}[M] = \inf_{\Pi \text{ seed for } M} \mathbb{I}[R, C; \Pi] = \mathbb{H}[M] - \mathbb{W}[M]$$

Another way to understand the private information is to see it as the largest entropy among all the rank-1 decompositions of a matrix $M$. Indeed a *rank-1 decomposition* corresponds to a decomposition into rank-1 matrices and therefore matrices for which the rows and the columns are independent as probability distributions. We define a *weighted convex rank-1 decomposition* of a non-negative matrix $M$ as $\sum_{i=1}^{s} \lambda_i p_i q_i^\intercal$ where $\lambda_i \geqslant 0$, $\sum \lambda_i = 1$. Now assume $\|M\|_1 = 1$ ($M$ is a probability distribution matrix). Given a convex rank-1 decomposition: $M = \sum_{i=1}^{s} \lambda_i p_i q_i^\intercal$, we have $\mathbb{H}[M] = \sum_{i=1}^{s} \lambda_i(\mathbb{H}[p_i] + \mathbb{H}[q_i])$.

The following theorem, with Theorem 1, proves that common information can be used to lower bound the extension complexity.

**Theorem 7** [BP16]**.** *For any non-negative matrix $M$, we have:* $\mathbb{C}[M] \leqslant \log(\mathrm{rk}_+(M))$.

This article also provides another theorem that is useful for the computation of the common information of tensor matrices.

**Theorem 8 .** *Let $M$ and $N$ be two non-negative matrices. Then* $\mathbb{C}[M \otimes N] = \mathbb{C}[M] + \mathbb{C}[N]$.

**Example 4 .** We consider the matrix $M$ defined in 3. We will upper bound the common information of this matrix with techniques similar to the ones used in 3. We consider $M' = M/\|M\|_1$. By using the same rectangle distribution, with $|I| = |J|$, we obtain that $\mathbb{W}[M'] \geqslant \sum_I \mathbb{P}(I) \log(n^2/4) = \log(n^2/4)$. Therefore:

$$\begin{aligned}
\mathbb{C}[M] &\leqslant \mathbb{H}[M] - \mathbb{W}[M] \\
&\leqslant \log(n(n-1)) - \log\left(\frac{n^2}{4}\right) \\
&= \log\left(4\frac{n-1}{n}\right)
\end{aligned}$$

So asymptotically the bound is at most 4 in log scale.

### 3.3 Comparison of the bounds

These two bounds have a lot of similarities. In [BJLP14], the main open question is whether the logarithm of the hyperplane separation bound and common information are polynomially related. We attempted to link the bounds by expressing rectangles as products of distributions, since a rectangle matrix is a rank-1 matrix. In many simple examples the two bounds seem to behave in the same way.

In Examples 3 and 4, we compute the two bounds in a similar manner. The rank of $M$ in these examples is $n$, so $\mathrm{rk}_+(M) \geqslant n$. Therefore we see that both lower bounds give results that are close to each other, but asymptotically far from the non-negative rank, and that are therefore poor lower bounds in these particular case.

We now give a family of matrices for which the two bounds are asymptotically different. However, even in this example the bounds are still polynomially related. We introduced the disjointness matrix in Example 3. Using the fact the common information tensors (see Theorem 8, [BJLP14] shows that $\mathbb{C}[DISJ_n] = 2n/3$.

Moreover all the diagonal elements of $DISJ_n$ are ones and it is upper triangular. So consider the following weight matrix in $\mathbb{R}^{2^n \times 2^n}$:

$$W = \begin{pmatrix} & \text{\Large 0} & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & \text{\Large -}\infty & \\ 1 & & & & \end{pmatrix}$$

Here $\langle W, DISJ_n\rangle = 2^n$, $\|DISJ_n\|_\infty = 1$. Every rectangle with more than 1 non-zero element will meet at least one of the $-\infty$ entries, which will give a small $\langle W, R\rangle$. Therefore $\rho(W) = 1$. This and Theorem 5 give $\mathrm{HSB}(DISJ_n) = 2^n$: in this case $\log(\mathrm{HSB}(DISJ_n)) = n > 2n/3 = \mathbb{C}[DISJ_n]$.

This gives an example where we have a gap between the two values, but they are still within a constant factor of each other.

# 4 Complexity of the bounds

One interesting question that can give a better understanding of the bounds is to know what is the complexity of computing them. That is the subject of this section. We prove that the hyperplane separation bound is $NP$-hard to approximate and also that the common information is in $NP$.

## 4.1 The hyperplane separation bound is NP-complete

To show the $NP$-completeness of the hyperplane separation bound we first formally define the decision problem.

**Definition 13 .** The hyperplane separation bound problem is the following decision problem:
**Input:** A matrix $M$ in $\mathbb{Q}_+^{m\times n}$ and $k \in \mathbb{Q}_{>0}$.
**Question:** Do we have that $\mathrm{HSB}(M) \leqslant k$?

Before discussing further the hyperplane separation bound we first need to define the correlation polytope of a graph.

**Definition 14 .** Let $G = (V, E)$ be a graph where $V = [n]$ and $|E| = m$. For a vector $\varepsilon \in \{0,1\}^n$, let $u^\varepsilon$ be the vector of $\mathbb{R}^{n+m}$ such that: $u_i^\varepsilon = \varepsilon_i$ for all $i \in [n]$ and $u_{ij}^\varepsilon = \varepsilon_i\varepsilon_j$ for all $ij \in E$. We define the *correlation polytope* as:

$$\mathrm{COR}(G) = \mathrm{conv}\{u^\varepsilon \mid \varepsilon \in \{0,1\}^n\}$$

We denote $\mathrm{COR}(n) = \mathrm{COR}(K_n)$ which is more simply the convex hull of the Boolean matrices $M = xx^\intercal$ where $x \in \{0,1\}^n$. The *correlation polytope membership problem* is the following decision problem:
**Input:** A matrix $M$ in $\mathbb{Q}_+^{m\times n}$ and $k \in \mathbb{N}$.
**Question:** Is $M$ in $\mathrm{COR}(G)$?

From this definition, we can see that the correlation polytope of a graph $G$ is the convex hull of the characteristic vectors of induced subgraphs of $G$.

**Theorem 9** [Pit91]**.** *Deciding if a matrix is in* $\mathrm{COR}(n)$ *is NP-complete.*

The NP-completeness of this membership problem is what gives us the following theorem.

**Theorem 10 .** *The hyperplane separation bound problem is NP-complete.*

*Proof.* First we show that the hyperplane separation bound is in NP.

Let $\frac{1}{\lambda} = \mathrm{HSB}(M)$. By Carathéodory's theorem, $\lambda M$ can be expressed as a convex combination of at most $mn + 1$ rectangles. In other words, there exist $\ell \leqslant mn + 1$, $\lambda_1, \ldots, \lambda_\ell$ such that $\sum_{i=1}^{\ell} \lambda_i = 1$, and there exist rectangles $R_1, \ldots, R_\ell$ such that $\lambda M = \sum_{i=1}^{\ell} \lambda_i R_i$. Moreover because the coefficients of $M$ are rationals, the $\lambda_i$'s are too. Now let us the $\lambda_i$'s and the $R_i$'s can be provided as a certificate i.e. that they can be provided with a polynomial number of bits. These $\lambda_i$'s are solutions of the system with unique solution:

$$\begin{cases} \lambda M - \sum_{i=1}^{\ell} \lambda_i R_i = 0 \\ \sum_{i=1}^{\ell} \lambda_i = 1 \end{cases} \quad \text{which is equivalent to} \begin{pmatrix} | & | & & | \\ v_M & v_{R_1} & \dots & v_{R_\ell} \\ | & | & & | \\ 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \lambda \\ \lambda_1 \\ \vdots \\ \lambda_\ell \end{pmatrix} = 0,$$

where for a matrix $N$, $v_N$ is an associated vector obtained by just sticking all the lines. Moreover Hadamard proved that a $n \times n$ $\{0,1\}$-matrix has a determinant that is in $O(n^n)$ which can be encoded with $O(n \log n)$ bits. By Cramer's rule of solving the system, and because the determinants can always be obtained by expanding on the $v_m$'s column, we obtain the wanted result.

Then to verify if the hyperplane separation bound is less than $k$, compute the $\sum_{i=0}^{\ell} \lambda_i R_i$ and verify that there exists a $\lambda$ such that this sum is equal to $\lambda M$. Then compare $\lambda$ to $k$. This is a polytime verification so the hyperplane separation bound problem is in NP.

Now we show that the problem is NP-hard. To do so we will show that the rectangle polytope membership problem is NP-hard. Indeed for $M$ in $\mathbb{Q}^{m \times n}$, $\mathrm{HSB}(M) \leqslant k$ if and only if $\frac{M}{k} \in \mathrm{RECT}(m, n)$. We reduce the hyperplane separation bound problem from the correlation polytope membership problem.

Let $M$ be a matrix, instance of the correlation polytope membership problem. We define a new matrix $N$ for the rectangle polytope membership problem as follow:

$$M = \begin{pmatrix} 1 & M_{11} & \cdots & M_{nn} \\ M_{11} & & & \\ \vdots & & M & \\ M_{nn} & & & \end{pmatrix}$$

For the sake of simplicity, we index the coefficient of $N$ from 0 to $n$.

Let's show that: $M \in \mathrm{COR}(n) \Leftrightarrow N \in \mathrm{RECT}(n+1, n+1)$.

($\Rightarrow$) We assume $M \in \mathrm{COR}(n)$. Then there exists $(\mu_i)_{i=1}^{l}$ that sum to one, and $x_i \in \{0,1\}^n$ such that $M = \sum_{i=1}^{l} \mu_i x_i x_i^t$. We define one rectangle for each $x_i$ as follow: let $I$ be the index set of the non-zero elements of $x_i$, then $R_i = (\{0\} \cup I) \times (\{0\} \cup I)$. For $k, l \in \{1, \dots, n\}^2$, $\sum_{i=1}^{l} \mu_i (R_i)_{kl} = \sum_{i=1}^{l} \mu_i (x_i x_i^t)_{kl} = M_{kl}$. Since $(0,0)$ is in all the rectangle, $\sum_{i=0}^{l} \mu_i (R_i)_{00} = 1$. For $j \in \{1, \dots, n\}$, $N_{0j}$ (resp. $N_{j0}$) appears in all the rectangle in which $N_{jj} = M_{jj}$ appears. Therefore $\sum_{i=1}^{l} \mu_i (R_i)_{0j} = \sum_{i=1}^{l} \mu_i (R_i)_{jj} = M_{jj}$.
This gives: $N = \sum_{i=0}^{l} \mu_i R_i$.

($\Leftarrow$) We assume $N \in \mathrm{RECT}(n+1, n+1)$. There exist $\lambda_1, \dots, \lambda_l$ such that $\sum_{i=1}^{l} \lambda_i = 1$, and there exists rectangles $R_1, \dots, R_l$ such that $N = \sum_{i=0}^{l} \lambda_i R_i$. Since $\sum_{i=1}^{l} \lambda_i = 1$ and $N_{00} = 1$ we have: $\forall i \in \{1, \dots, l\}, (0,0) \in R_i$. This implies that all the rectangles that contain $N_{ii}$ for $1 \leqslant i \leqslant n$ contain $N_{0i}$ and $N_{i0}$. And since $N_{ii} = N_{0i} = N_{i0}$, these three coefficients are exactly in the same rectangles. Therefore, if $R_i = I_i \times J_i$, we have $I_i = J_i$, which means that there exists $x_i$ in $\{0,1\}^{n+1}$ such that $R_i = x_i x_i^t$.
Therefore $M \in \mathrm{COR}(n)$. □

## 4.2 Non-approximability of the hyperplane separation bound

In this section we show that we cannot achieve a good approximation of the hyperplane separation bound problem. The idea is to use the non-approximability of the biclique cover and biclique partition problems from [CHHK14]. Recall that these problems are defined in Section 2.3.

To be forthright with the reader, note that some results stated in [CHHK14] are not proved. They should appear in the final version, which is not available yet. We managed to reconstruct some of the proofs. We also sent an email to the authors to inquire about the journal version. The answer was that some details were not written yet, but they will send them to us as soon as they are ready.

**Theorem 11** [CHHK14]**.** *Let $G$ be a graph and $k$ an integer. There exist an algorithm that runs in time $|V(G)|^{O(k)}$ and constructs a bipartite graph $H$ such that $|V(H)| = \Theta(|V(G)|^k)$ and:*

$$\left(\frac{\mathrm{cp}(G)}{\log|V(G)|}\right)^k \leqslant \mathrm{bc}(H) \leqslant \mathrm{bp}(H) \leqslant \mathrm{cp}(G)^k |V(G)|^3$$

Feige and Kilian proved that it is NP-hard to approximate the chromatic number of a graph which implies the following theorem.

**Theorem 12** [FK96]. *Let $\varepsilon > 0$ and $G$ be a graph. Unless P=NP, it is NP-hard to approximate the clique partition problems within a factor of $|V(G)|^{1-\varepsilon}$.*

We define the *fractional biclique cover* $\mathrm{bc}^*(G)$ of a graph $G$ as the solution of the following LP:

$$\min_w \sum_{(A,B)\text{ biclique}} x_{(A,B)}$$

$$\text{subject to} \quad \sum_{\substack{e \ni (A,B) \\ (A,B)\text{ biclique}}} x_{(A,B)} \leqslant 1, \forall e \in E(G) \text{ and } x_{(A,B)} \geqslant 0$$

**Lemma 13 .** *Let $H$ be a bipartite graph and the associated matrix $M$ constructed as in 2.3. Then* $\mathrm{HSB}(M) \leqslant \mathrm{bp}(H)$ *and* $\mathrm{bc}(H) \leqslant (1 + \log|E(H)|)\,\mathrm{HSB}(M)$.

*Proof.* Let $k = \mathrm{bp}(H)$ and $(A_1, B_1), \ldots, (A_k, B_k)$ be a biclique partition. We identify the bicliques and the associated rectangle in the matrix. Then since we have a partition we have:

$$\frac{M}{k} = \sum_{i=1}^{k} \frac{1}{k} \chi^{A_i \times B_i}$$

So $\mathrm{HSB}(M) \leqslant \mathrm{bp}(H)$.

Moreover there exists rectangles $R_1, \ldots, R_l$ and $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$, where $\sum_{i=1}^{l} \lambda_i = 1$, such that $M = \mathrm{HSB}(M) \sum_{i=1}^{l} \lambda_i R_i$. We write $R_i = A_i' \times B_i'$. Then $A_i' \times B_i'$ is a biclique since we have a rectangle. Now let $(A, B)$ be a biclique. If there is an $i$ such that $(A, B) = (A_i', B_i')$ then we define $x_{(A,B)} = \mathrm{HSB}(M)\lambda_i$, otherwise $x_{(A,B)} = 0$. For $e \in E(H)$ we have, since $M$ is a $(0, 1)$-matrix:

$$\sum_{e \ni (A,B)} x_{(A,B)} = \sum_{e \ni (A_i', B_i')} \mathrm{HSB}(M)\lambda_i = \mathrm{HSB}(M) \sum_{e \ni (A_i', B_i')} \lambda_i = 1$$

And $\sum_{(A,B)\text{ biclique}} x_{(A,B)} = \mathrm{HSB}(M)$ so $\mathrm{HSB}(M) \geqslant \mathrm{bc}^*(H)$. Moreover $\mathrm{bc}(H) \leqslant (1 + \log|E(H)|)\,\mathrm{bc}^*(H)$ (see [Lov75]) so $\mathrm{bc}(H) \leqslant (1 + \log|E(H)|)\,\mathrm{HSB}(M)$. □

Because $|V(H)| = \Theta(|V(G)|^k)$ we obtain the following theorem.

**Theorem 14 .** *Let $G$ be a graph and $k$ an integer. There exist a constant $a$ and an algorithm that runs in time $|V(G)|^{O(k)}$ and construct a $(0, 1)$-matrix $M$ of size $\Theta(|V(G)|^k)$.*

$$\frac{1}{ak\log|E(G)|}\left(\frac{\mathrm{cp}(G)}{\log|V(G)|}\right)^k \leqslant \mathrm{HSB}(M) \leqslant \mathrm{cp}(G)^k |V(G)|^3$$

The following theorem and its proof are inspired by [CHHK14].

**Theorem 15 .** *Let $\varepsilon > 0$. It is NP-hard to approximate the hyperplane separation bound problem for a matrix $M$ within factor $(mn)^{1-\varepsilon}$ where $M$ is an $m \times n$ matrix.*

*Proof.* We use Theorem 12 to deduce the hardness of approximation of the hyperplane separation bound : this Theorem implies the existence of an algorithm $\mathcal{A}$ that takes a SAT instance $\varphi$ as an input and produce a graph such that:

- If $\varphi$ is satisfiable, then $\mathrm{cp}(G) \leqslant c = |V(G)|^{\varepsilon}$
- If $\varphi$ is not satisfiable, then $\mathrm{cp}(G) \geqslant s = |V(G)|^{1-\varepsilon}$

And the gap is $g = s/c = |V(G)|^{1-2\varepsilon}$.

For $k$ an integer, we build a new algorithm $\mathcal{A}_k$ that take a SAT instance $\varphi$ as input, runs $\mathcal{A}$ on it to obtain a graph $G$, and produce a matrix $M$ with the algorithm of 14. We compute the gap $g_k$ of $\mathcal{A}_k$. If $\varphi$ is satisfiable then $\mathrm{HSB}(M) \leqslant \mathrm{cp}(G)^k |V(G)|^3 \leqslant c^k |V(G)|^3$. If $\varphi$ is not satisfiable then $\mathrm{HSB}(M) \geqslant \frac{1}{ak \log |E(G)|} \left(\frac{\mathrm{cp}(G)}{\log |V(G)|}\right)^k \geqslant \frac{1}{2ak \log |V(G)|} \left(\frac{s}{\log |V(G)|}\right)^k$. Moreover $|M| = \alpha |V(G)|^k$ where $\alpha$ is a constant and $M$ is the number of entries of the matrix. So we have:

$$g_k \geqslant \frac{s^k}{2akc^k |V(G)|^3 \log^k |V(G)|} = \frac{|V(G)|^{k-2k\varepsilon}}{2ak|V(G)|^3 \log^k |V(G)|} \geqslant \frac{\alpha' |M|^{1-2\varepsilon}}{|M|^{3/k} \log^k |V(G)|}$$

where $\alpha'$ is a constant. And for $k = \lceil 1/\varepsilon \rceil$, $g_k \geqslant \frac{\alpha' |M|^{1-5\varepsilon}}{\log^k |V(G)|} \geqslant \alpha'' |M|^{1-6\varepsilon}$ . Moreover the Feige-Kilian algorithm runs in polynomial time and $|V(G)| = \varphi^{O(1)}$. The algorithm from Theorem 14 runs in $|V(G)|^{O(1/\varepsilon)} = |\varphi|^{O(1)}$. So this reduction show that the hardness result of approximating SAT holds for the hyperplane separation bound problem. $\square$

### 4.3 The common information bound is in NP

We first formally define the decision problem associated to the common information bound.

**Definition 15 .** The common information bound problem is the following decision problem:

**Input:** A matrix $M$ in $\mathbb{Q}_+^{m\times n}$ in and $k > 0$.

**Question:** Do we have that $\mathbb{C}[M] < k$?

**Theorem 16 .** *The common information problem is in NP.*

*Proof.* Let $M \in \mathbb{Q}^{m\times n}$. First we show by contradiction that there exist a decomposition that minimize the entropy with less that $mn + 1$ terms in the sum.

Let $M = \sum_{i=1}^{l} \mu_i p_i q_i^t$ be a decomposition of the matrix $M$ into matrices of rank 1 that minimize $\mathbb{H}[M|\Pi]$ where $\Pi$ is the distribution of the $(\mu_i)$. Among all such decomposition we choose the one with the smallest $l$ possible. Let's show that $l \leqslant mn + 1$. If not then the $p_i q_i^t$ are dependent. So there exists $(\alpha_i)_{i=1}^{l}$ a family of non-all zero elements such that $\sum_{i=1}^{l} \alpha_i p_i q_i^t = 0$ and $\sum_{i=1}^{l} \alpha_i = 0$. Without loss of generality we assume $\alpha_l \neq 0$.

The first case is $\sum_{i=1}^{l} \alpha_i \mathbb{H}(p_i q_i^t) \leqslant 0$. Let $\varepsilon = -\frac{\mu_l}{\alpha_l}$. We still have:

$$M = \sum_{i=1}^{l} \mu_i p_i q_i^t + \varepsilon \sum_{i=1}^{l} \alpha_i p_i q_i^t = \sum_{i=1}^{l-1} (\mu_i + \varepsilon\alpha_i) p_i q_i^t$$

Because $\sum_{i=1}^{l} \alpha_i = 0$ we have a new decomposition of $M$ with rank one matrices and another distribution $\Pi'$. Then:

$$\mathbb{H}[M|\Pi'] = \sum_{i=1}^{l} (\mu_i + \varepsilon\alpha_i)(\mathbb{H}[p_i] + \mathbb{H}[q_i]) = \mathbb{H}[M|\Pi] + \varepsilon \sum_{i=1}^{l} \alpha_i \mathbb{H}(p_i q_i^t) \leqslant \mathbb{H}[M|\Pi]$$

But this decomposition has one less element which contradicts the minimality of $l$: contradiction.

If $\sum_{i=1}^{l} \alpha_i \mathbb{H}(p_i q_i^t) \leqslant 0$ we take $\varepsilon = \frac{\mu_l}{\alpha_l}$ and that leads to the same contradiction. So we have a decomposition with $mn+1$ that minimize the entropy.

Assume $\mathbb{C}[M] < k$. Then there exist a distribution $\Pi$ associated to the decomposition $M = \sum_{i=1}^{mn+1} \mu_i p_i q_i^t$ (where $\mu_i \in \mathbb{Q}$, $p_i \in \mathbb{Q}^m$ and $q_i \in \mathbb{Q}^n$) and such that $\mathbb{H}[M] - \mathbb{H}[M|\Pi] < k$. Providing the $\mu_i, p_i, q_i$ gives a certificate that can be verified in polynomial time. Indeed computing $\mathbb{H}[M]$ is polytime, and same for $\mathbb{H}[M|\Pi]$ since the sum contains at most $mn+1$ elements. $\square$

# 5 An example of bounding via reduction

The previous bounding techniques involve a direct study of the slack matrix of the polytope. However other techniques can be used to study the extension complexity. That is what we did to bound the extension complexity of the correlation polytope of some class of graphs.

## 5.1 Background and definitions

Let $G = (V, E)$ be a graph. We defined the correlation polytope (see 14).

**Definition 16 .** A graph $H$ is a *minor* of a graph $G$ if $H$ can be obtained from a subgraph of $G$ by contracting edges. A class $\mathcal{C}$ of graphs is *minor-closed* if $G \in \mathcal{C}$ and $H$ a minor of $G$ implies that $H \in \mathcal{C}$.

We determine the extension complexity of the correlation polytopes of proper minor-closed classes almost exactly. In order to state our main result, we also need the notion of *tree-width*, which we now define.

**Definition 17 .** A *tree-decomposition* of a graph $G$ is a pair $(T, \mathcal{B})$ where $T$ is a tree and $\mathcal{B} := \{B_t \mid t \in V(T)\}$ is a collection of subsets of vertices of $G$ satisfying:

- $V(G) = \bigcup_{t \in V(T)} B_t$,
- for each $uv \in E(G)$, there exists $t \in V(T)$ such that $u, v \in B_t$, and
- for each $v \in V(G)$, the set of all $w \in V(T)$ such that $v \in B_w$ induces a connected subtree of $T$.

We call each member of $\mathcal{B}$ a *bag*. The *width* of $(T, \mathcal{B})$ is $\max\{|B_t| - 1 \mid t \in V(T)\}$. The *tree-width* of $G$, denoted $\mathsf{tw}(G)$, is the minimum width taken over all tree-decompositions of $G$.

We can now state our main theorems.

**Theorem 17 .** *For every $n$-vertex graph $G$, the extension complexity of the correlation polytope of $G$ is $2^{O(\mathsf{tw}(G)+\log n)}$.*

For proper minor-closed classes, we prove that this bound is tight.

**Theorem 18 .** *For every proper minor-closed class $\mathcal{C}$, there exist a constant $c$ such that for every $n$-vertex graph $G \in \mathcal{C}$,*

$$2^{c(\mathsf{tw}(G)+\log n)} \leqslant \operatorname{xc}(\operatorname{COR}(G)).$$

## 5.2 Upper bound

Wainright and Jordan [WJ04] proved that COR($G$) has an extended formulation of size $n^{O(\mathsf{tw}(G))}$ using hierarchy techniques. In this section, we prove a better upper bound.

**Definition 18 .** A tree-decomposition of width $k$ is *smooth* if all bags of the decomposition are of size $k+1$ and two adjacent bags share exactly $k$ vertices.

**Definition 19 .** An *induced subgraph* of a graph $G = (V, E)$ is a graph formed by a subset of $V$ as vertices, and all the edges between these vertices in that are in $E$.

**Definition 20 .** The *stable set polytope* of a graph $G$, denoted $\mathsf{STAB}(G)$, is the convex hull of the incidence vectors of its stable sets.

It is well-known (see [Bod96]) that every graph $G$ has a smooth tree-decomposition of width $\mathsf{tw}(G)$. We also require the following [Bod96, Lemma 2.5].

**Theorem 19 .** *If $(T, \mathcal{B})$ is a width $k$ smooth tree-decomposition of a graph $G$, then $|\mathcal{B}| = |V(G)| - k$.*

**Theorem 20 .** *For all graphs $G$ with $n$ vertices,*

$$\mathrm{xc}(\mathrm{COR}(G)) = 2^{O(\mathsf{tw}(G)+\log n)}.$$

*Proof.* We start from a smooth, minimal width tree-decomposition $(T, \mathcal{B})$ for $G$, where $\mathcal{B} = (B_v)_{v\in V(T)}$ denotes the collection of bags of the tree-decomposition.

We now define a new graph $H$ in which the induced subgraphs of $G$ are represented by certain stable sets of $H$. For each $v \in V(T)$, we create $2^{|B_v|}$ vertices that represent all the ways to select a subset of $B_v$. We name these vertices $(v, X)$, where $X \subseteq B_v$. Then, we create an edge between $(v, X)$ and $(v', X')$ whenever $vv' \in E(T)$ and the sets $X$, $X'$ are incompatible in the sense that $B_v \cap B_{v'} \cap X \neq B_v \cap B_{v'} \cap X'$. Moreover, for every fixed $v \in V(T)$ the vertices of the form $(v, X)$ for $X \subseteq B_v$ form a clique in $H$, that we denote $K_v$.

In the graph $H$, we consider the stable sets that contain exactly one vertex of each $K_v$. They form the vertices of a face of $\mathsf{STAB}(H)$ which we denote by $F$. Thus, $F = \{y \in \mathsf{STAB}(H) \mid \forall v \in V(T) : \sum_{X\subseteq B_v} y_{(v,X)} = 1\}$. For $vw \in E(T)$ we let $H^{vw}$ be the induced subgraph of $H$ induced by vertices of the formto vertices of the form $(v, X)$ or $(w, Y X)$, where $X$ and $Y$ are arbitrary. We let $y^{vw}$ be the projection of $y \in F$ to the vertices of $H^{vw}$.

Notice that $K_v$ is a cutset, for each non-leaf vertex $v \in V(T)$. Moreover Chvátal's clique cutset lemma [Chv75, Theorem4.1] states that if a graph $G$ has a clique $K$ such that $G - V[K]$ has two components $G_1$ and $G_2$, then we can reconstruct the stable sets of $G$ by taking stable sets of $G_1$ and $G_2$. So we can write $F = \bigcap_{vw\in E(T)} F_{vw}$, where

$$F_{vw} := \left\{ y \in \mathbb{R}^{V(H)}_{\geqslant 0} \,\middle|\, y^{vw} \in \mathsf{STAB}(H^{vw}) \right\}.$$

Now by the smoothness of the tree-decomposition, we get $\mathrm{xc}(F_{vw}) \leqslant |\mathsf{vert}(F_{vw})| \leqslant 2^{\mathsf{tw}(G)+2}$, where $\mathsf{vert}(F_{vw})$ is the set of vertices of $F_{vw}$. Therefore,

$$\mathrm{xc}(F) \leqslant |E(T)| \cdot 2^{\mathsf{tw}(G)+2} = (n - \mathsf{tw}(G) - 1)2^{\mathsf{tw}(G)+2} \leqslant n2^{\mathsf{tw}(G)+2}.$$

To finish the proof it suffices to show that $F$ are equivalent COR($G$) up to an affine map. To see this, let $((x_i)_{i\in V(G)}, (x_{ij})_{ij\in E(G)}) \in \mathrm{COR}(G)$. For each $i \in V(G)$ and $ij \in E(G)$, let $v(i)$ and $v(ij)$ be vertices of $T$ such that $i \in B_{v(i)}$ and $\{i, j\} \subseteq B_{v(ij)}$. Then note that

$$x_i = \sum_{i\in S\subseteq B_{v(i)}} y_{(v(i),S)} \qquad x_{ij} = \sum_{\{i,j\}\subseteq S\subseteq B_{v(ij)}} y_{(v(ij),S)}.$$

This defines a map from $F$ to $\mathrm{COR}(G)$.

Now we show that all vertices of $F$ are mapped on vertices of $\mathrm{COR}(G)$. Let $y$ be a vertex of $F$. For all $v$ in $V(T)$ there exists a set $X_v$ in $B_v$ such that $y_{(v,X_v)}$. Moreover since we have a stable set, if $vv'$ is an edge of $T$, $B_v \cap B_{v'} \cap X_v = B_v \cap B_{v'} \cap X_{v'}$: we will say that $X_v$ and $X_{v'}$ are compatible. Then $x_i = 1$ if and only if then there exist a $B_{v(i)}$ and an $X_{v(i)} \in B_{v(i)}$ such that $y_{(v_i,X_{v_i})} = 1$. Assume that $x_i = 1 = x_j$ for $ij \in E(T)$. Then there exists a bag $B_v$ that contains $ij$ and an $X_v \in B$ such that $y_{(v,X_v)} = 1$. But $i$ is in all the bags in the path from $v(i)$ to $v$. This bags contain a set on which $y$ values 1 and by compatibility $i$ is in this sets. Therefore $i \in X_v$. By the same reasoning $j \in X_v$. So $x_{ij} = 1$. Now if $x_i = 0$ but for an edge $ij$, $x_{ij} = 1$, then by compatibility all bags containing $i$ contain $i$ in the set that is valued to one: contradiction since this would mean that $x_i = 1$. Therefore all vertices of $F$ map on vertices of $\mathrm{COR}(G)$.

Now let $V' \in V(G)$ and $G' = G[V]$ the induced subgraph. For each bag $B_v$ of $T$ we define $X_v = B_v \cap V'$. Let $y_{(v,X)} = 1$ if and only if $X = X_v$. Then if $vv' \in E(T)$, $B_v \cap B_{v'} \cap X_v = B_v \cap B_{v'} \cap X_{v'}$. We constructed a vertex of $F$ that maps to $G'$.

Therefore $F$ and $\mathrm{COR}(G)$ are the same polytope up to an affine map. This provides the upper bound. □

### 5.3 Lower bound

In this section we prove the lower bound. Since the extension complexity of a polytope is at least its dimension, we have the following easy observation.

**Theorem 21 .** *For all graphs $G$ with $n$ vertices,*

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

**Theorem 22 .** *Let $G$ and $H$ be graphs such that $H$ is a minor of $G$. Then* $\mathrm{xc}(\mathrm{COR}(H)) \leqslant \mathrm{xc}(\mathrm{COR}(G))$.

*Proof.* If $uv \in E(G)$, then $\mathrm{COR}(G \setminus uv)$ can be obtained from $\mathrm{COR}(G)$ by projecting out $x_{uv}$ and $\mathrm{COR}(G/uv)$ is obtained from $\mathrm{COR}(G)$ by setting $x_u = x_v$. If $w$ is an isolated vertex of $G$, then $\mathrm{COR}(G - w)$ is obtained from $\mathrm{COR}(G)$ by setting $x_w = 0$. □

**Theorem 23 .** *Let $\mathcal{C}$ be a proper minor-closed class of graphs. For every $n$-vertex graph $G \in \mathcal{C}$,*

$$\mathrm{xc}(\mathrm{COR}(G)) = 2^{\Omega(\mathsf{tw}(G))}.$$

*Proof.* We first show that it is sufficient to prove the theorem for a grid with gadgets $G'$ of size $\Omega(\mathsf{tw}(G)) \times \Omega(\mathsf{tw}(G))$ (see Figure 7). In this grid all the interior vertices are replaced by a constant size planar gadget, and the left vertex of the gadget is linked to the bottom vertex. The gadget of the grid is constant size and planar so it can be embedded in a grid of constant size. Therefore this grid with gadgets can be embedded in a grid of size $\Omega(\mathsf{tw}(G)) \times \Omega(\mathsf{tw}(G))$ by just extending the embedding of the gadget. As shown by Demaine and Hajiaghayi [DH08] (see also Kawarabayashi and Kobayashi [KK12]), since our initial graph $G$ belongs to a proper minor-closed class, it has a grid minor $G_{\ell,\ell}$ where $\ell = \Omega(\mathsf{tw}(G))$. If the grid gadget is chosen carefully we therefore have: $G' \leqslant_m G_{\ell,\ell} \leqslant_m G$. Because the extension complexity of the correlation polytope is minor monotone, this implies that $\mathrm{xc}(\mathrm{COR}(G')) \leqslant \mathrm{xc}(\mathrm{COR}(G))$. Therefore it is enough to show the theorem for the grid with gadgets.

The gadget used in the proof is inspired by the crossover gadget of the reduction from 3-SAT to Planar 3-SAT in [Lic82]. Each diamond (see Figure 8a) in the grid with gadgets will be replaced by the planar graph in Figure 8b. In this gadget the square vertices represent the clauses of a SAT formula and the round vertices represent the variables. When a round

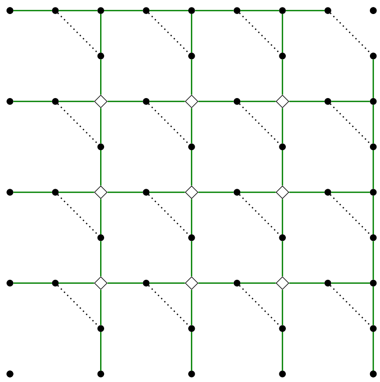
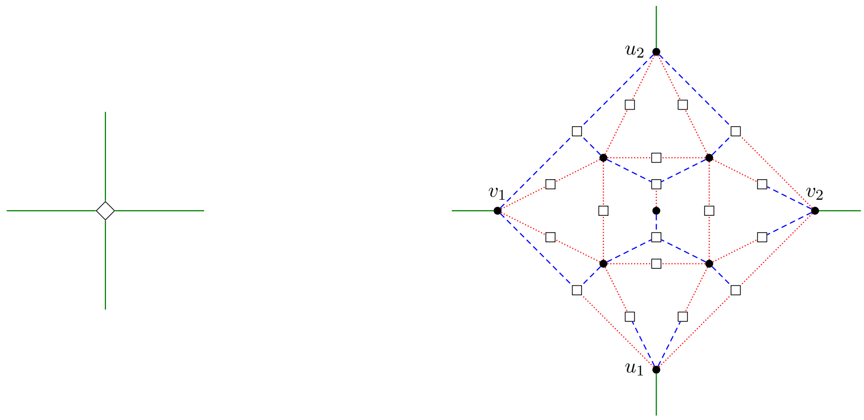


Figure 7: Grid with gadgets

vertex is adjacent to a square via a blue-dashed (resp. red-dotted) edge, this means that the corresponding variable (resp. negation of the variable) appears in the corresponding clause.




(a) (b)

Figure 8: Gadget in the grid

To obtain $G'$, we first replace the degree-2 square vertices as in Figure 9b. Then we replace the red-dotted edges as in Figure 10b. Then we replace the degree-3 square vertices as in Figure 11b.


(a) (b)

Figure 9: Simulation of $a \vee b$ with a stable set of a graph


(a) (b)

Figure 10: Transformation of the NOT edge

The idea is to reduce the study of the correlation polytope of $G'$ to the study of the correlation polytope of a complete bipartite graph by projecting on some face of $\mathrm{COR}(G')$. This projection should transmit the value of the variable that is on the left of the gadget to


Figure 11: Simulation of $i \vee j \vee k$ with a stable set of a graph

the right and the one that is at the bottom should be transmitted up. This then reduces the problem to the complete bipartite graph where the stable sets are the bottom and left vertices, and the edges are given by the dotted edges of Figure 8.

Now let us define the faces on which we project. For each vertex $i$ (resp. edge $ij$) we define the associated non-negative variable $x_i$ (resp. $x_{ij}$).

If $ij$ is a green edge of $G'$ for which both vertices are not in a gadget, then we project on $x_i = x_i = x_{ij}$ (this is valid since we have an induced subgraph).

In Figure 9b, the vertices $i$ and $j$ correspond to the negation of $a$ and $b$. Since $a \vee b = \neg(\neg a \wedge \neg b)$ we set $x_{ij} = 0$.

In Figure 10b, we either take $i$ or $i^{'}$. The projection is then $x_i + x_{i'} - 2x_{ii'} = 1$.

In Figure 11b we want to simulate an OR, which means that one of the variables $x_i$, $x_j$ or $x_k$ should be set to 1. This is described by the following equations: $x_i + x_{i'} - 2x_{ii'} = 1$, $x_j + x_{j'} - 2x_{jj'} = 1$, $x_k + x_{k'} - 2x_{kk'} = 1$, $x_{ii'} = 0$, $x_{ii'} = 0$, $x_{ii'} = 0$ and $x_{i''} + x_{i''} + x_{i''} - 2x_{i''j''} - 2x_{j''k''} - 2x_{k''i''} = 1$.

The projections are defined so that all the clauses are satisfied. Because of the properties of the gadget (see [Lic82]), we have in Figure 8, $x_{u_1} = x_{u_2}$ and $x_{v_1} = x_{v_2}$.

Therefore, these projections reduce the study of the correlation polytope of $G'$ to the correlation polytope of $K_{\ell',\ell'}$ where $\ell'$ is $\Omega(\mathsf{tw}(G))$. Moreover $K_{\ell'}$ is a minor of $K_{\ell',\ell'}$ so $\mathrm{xc}\, K_{\ell'} \leqslant \mathrm{xc}\, K_{\ell',\ell'}$. In [FMP+11], Fiorini, Massar, Pokutta, Tiwary and de Wolf show that $\mathrm{xc}\, K_{\ell'} = 2^{\Omega(\ell')}$: this gives the lower bound we wanted. $\square$

## 6 Conclusion and open problems

In this report we study some bounding technique for extension complexity.

We formulated and studied the decision problems of some bound that are used to study directly the slack matrix of the polytopes. This provided some complexity results on the hyperplane separation bound and the common information bound. We provided the hardness of approximation for the hyperplane separation bound .

Some interesting question we are still working on remain: is common information $NP$-complete? If yes is it hard to approximate? And for both bounds can we design efficient algorithm if we restrict to some families of slack matrices? Can we solve the hyperplane separation bound in polytime if we know the maximal biclique of the graph associated to the support of the slack matrix? Is there parameters that would provide a good FPT for the computation of these bounds?

This study started because of the interest in comparing the two bounds. So an interesting extension to this work could be to the question raised in [BP16]: are the two bound polynomially related?

As for the correlation polytope problem, we have linked tightly the extension complexity with the treewidth for minor closed families of graphs. Now we are working on extending the result for a certain family of graphs, the expander graphs.

**Acknowledgements**

## References

[BJLP14] Gábor Braun, Rahul Jain, Troy Lee, and Sebastian Pokutta. Information-theoretic approximations of the nonnegative rank. *computational complexity*, pages 1–51, 2014.

[Bod96] Hans L Bodlaender. A linear-time algorithm for finding tree-decompositions of small treewidth. *SIAM Journal on computing*, 25(6):1305–1317, 1996.

[BP16] Gábor Braun and Sebastian Pokutta. Common information and unique disjointness. *Algorithmica*, 76(3):597–629, 2016.

[CHHK14] Parinya Chalermsook, Sandy Heydrich, Eugenia Holm, and Andreas Karrenbauer. Nearly tight approximability results for minimum biclique cover and partition. In *European Symposium on Algorithms*, pages 235–246. Springer, 2014.

[Chv75] Václav Chvátal. On certain polytopes associated with graphs. *Journal of Combinatorial Theory, Series B*, 18(2):138–154, 1975.

[DH08] Erik D Demaine and Mohammadtaghi Hajiaghayi. Linearity of grid minors in treewidth with applications through bidimensionality. *Combinatorica*, 28(1):19–36, 2008.

[Edm71] Jack Edmonds. Matroids and the greedy algorithm. *Math. Programming*, 1:127–136, 1971.

[FK96] Uriel Feige and Joe Kilian. Zero knowledge and the chromatic number. In *Computational Complexity, 1996. Proceedings., Eleventh Annual IEEE Conference on*, pages 278–287. IEEE, 1996.

[FMP$^+$11] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *arXiv preprint arXiv:1111.0837*, 2011.

[KK12] Ken-ichi Kawarabayashi and Yusuke Kobayashi. Linear min-max relation between the treewidth of h-minor-free graphs and its largest grid. In *STACS'12 (29th Symposium on Theoretical Aspects of Computer Science)*, volume 14, pages 278–289. LIPIcs, 2012.

[Lic82] David Lichtenstein. Planar formulae and their uses. *SIAM journal on computing*, 11(2):329–343, 1982.

[Lov75] László Lovász. On the ratio of optimal integral and fractional covers. *Discrete mathematics*, 13(4):383–390, 1975.

[Mar91] R Kipp Martin. Using separation algorithms to generate mixed integer model reformulations. *Operations Research Letters*, 10(3):119–128, 1991.

[Pit91] Itamar Pitowsky. Correlation polytopes: their geometry and complexity. *Mathematical Programming*, 50(1):395–414, 1991.

[Rot14] Thomas Rothvoß. The matching polytope has exponential extension complexity. In *Proceedings of the 46th annual ACM symposium on theory of computing*, pages 263–272. ACM, 2014.

[Shi16] Yaroslav Shitov. The complexity of positive semidefinite matrix factorization. *arXiv preprint arXiv:1606.09065*, 2016.

[WJ04] Martin J Wainwright and Michael I Jordan. Treewidth-based conditions for exactness of the Sherali-Adams and Lasserre relaxations. Technical report, Technical Report 671, University of California, Berkeley, 2004.

[Wyn75] Aaron Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975.

[Yan91] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991.

[Zie12] G.M. Ziegler. *Lectures on Polytopes*. Graduate Texts in Mathematics. Springer New York, 2012.