Lower Bounds for Approximating the Matching Polytope

Makrand Sinha[*]

Paul G. Allen School of Computer Science & Engineering,
University of Washington, Seattle

makrand@cs.washington.edu

Abstract

We prove that any extended formulation that approximates the matching polytope on n-vertex graphs up to a factor of $(1 + \varepsilon)$ for any $\frac{2}{n} \leq \varepsilon \leq 1$ must have at least $\binom{n}{\alpha/\varepsilon}$ defining inequalities where $0 < \alpha < 1$ is an absolute constant. This is tight as exhibited by the $(1 + \varepsilon)$ approximating linear program obtained by dropping the odd set constraints of size larger than $(1 + \varepsilon)/\varepsilon$ from the description of the matching polytope. Previously, a tight lower bound of $2^{\Omega(n)}$ was only known for $\varepsilon = O\left(\frac{1}{n}\right)$ [Rot14, BP15] whereas for $\frac{2}{n} \leq \varepsilon \leq 1$, the best lower bound was $2^{\Omega(1/\varepsilon)}$ [Rot14]. The key new ingredient in our proof is a close connection to the non-negative rank of a lopsided version of the unique disjointness matrix.

1 Introduction

In recent years, there has been a significant development in understanding whether well-known combinatorial optimization problems or polytopes can be expressed as a linear program with a small number of inequalities. The object of interest in this context has been the notion of *extended formulation* of polytopes. A polytope $Q \subseteq \mathbb{R}^{d'}$ is called an extended formulation of a given polytope $P \subseteq \mathbb{R}^d$, if there is a linear map $\pi : \mathbb{R}^{d'} \to \mathbb{R}^d$ such that $\pi(Q) = P$. The *extension complexity* of P, $\mathsf{xc}(P)$ is defined to be the minimal number of facets in any extended formulation for P. If a polytope P with an exponential number of facets has an extended formulation Q of polynomial size, then one can solve an optimization problem over P by writing a small linear program over Q. There are known examples of polytopes, such as the spanning tree polytope and the permutahedron (see Section 6 in [CCZ10]), which have an exponential number of facets but polynomial size extended formulations.

Yannakakis [Yan91] initiated the study of extended formulations motivated by refuting a purported $\mathsf{P} = \mathsf{NP}$ proof which encoded the Traveling Salesman problem as a polynomial size linear program. He proved that the extension complexity of polytopes is equal to the non-negative rank of the *slack matrix* associated with the polytope. Using this connection he was able to show that any *symmetric* extended formulation (one whose formulation is invariant under permutation of variables) that projects to the Traveling Salesman and matching polytopes in $\mathbb{R}^{\binom{[n]}{2}}$ must have size $2^{\Omega(n)}$.

Here, we will be concerned with the matching polytope. The matching polytope $P_{MAT}(n) \subseteq \mathbb{R}^{\binom{[n]}{2}}$ is defined as the convex hull of all matchings in the complete graph on the vertex set $[n]$ where $[n] := \{1, \ldots, n\}$. We will just write P_{MAT} when n is clear from the context. Denoting the indicator vector for a matching M by $\mathbb{1}_M$, the facets of P_{MAT} are completely described by the following degree and non-negativity constraints, and the *odd set* inequalities, as shown in [Edm65]:

$$P_{MAT}(n) = \text{conv}\left\{ \mathbb{1}_M \in \mathbb{R}^{\binom{[n]}{2}} \;\middle|\; M \subseteq \binom{[n]}{2} \text{ is a matching} \right\}$$

$$= \left\{ x \in \mathbb{R}^{\binom{[n]}{2}} \;\middle|\; \sum_{\substack{e \text{ incident} \\ \text{on } i}} x_e \leq 1 \; \forall i \in [n]; \sum_{\substack{e \text{ inside} \\ U}} x_e \leq \frac{|U| - 1}{2} \; \forall U \subseteq [n], |U| \text{ odd}; x \geq 0 \right\}.$$

Note that the description of the matching polytope has $2^{\Omega(n)}$ facet defining inequalities. However, one can still optimize any linear function in polynomial time over the matching polytope using the algorithm of Edmonds [Edm65]. The question of whether one could optimize over the matching polytope with a small extended formulation remained open until recently, when in a breakthrough result, Rothvoß [Rot14] proved that the extension complexity of P_{MAT} is indeed $2^{\Omega(n)}$.

The central question that we want to answer here concerns whether the matching polytope can be approximated by a small extended formulation. Formally, we say that a polytope P is monotone if for any $x \in P$ and y satisfying $0 \leq y \leq x$, we have that $y \in P$. We call a polytope K to be a $(1 + \varepsilon)$ approximation of a monotone polytope P if $P \subseteq K \subseteq (1 + \varepsilon)P$. Notice that if P is monotone, then the above is equivalent to requiring that for any non-negative vector w,

$$\max\{w^T x \mid x \in P\} \leq \max\{w^T x \mid x \in K\} \leq (1 + \varepsilon)\max\{w^T x \mid x \in P\}.$$

The matching polytope P_{MAT} is indeed monotone and it is well-known[1] that for any $\frac{2}{n} \leq \varepsilon \leq 1$, the following polytope Q_ε with at most $\binom{n}{1+1/\varepsilon} + O(n^2)$ facets approximates the matching polytope up to a factor of $1 + \varepsilon$ (note that for $\varepsilon \leq \frac{1}{n-1}$, Q_ε is the matching polytope itself):

$$Q_\varepsilon(n) = \left\{ x \in \mathbb{R}^{\binom{[n]}{2}} \;\middle|\; \sum_{\substack{e \text{ incident} \\ \text{on } i}} x_e \leq 1 \; \forall i \in [n]; \; x \geq 0; \right.$$

$$\left. \sum_{\substack{e \text{ inside} \\ U}} x_e \leq \frac{|U| - 1}{2} \; \forall U \subseteq [n], |U| \text{ odd}, |U| \leq \frac{1 + \varepsilon}{\varepsilon} \right\}.$$

This gives us a *polynomial type approximation scheme* (PTAS) style extended formulation (size at most n^c for a constant $c = c(\varepsilon)$ depending on ε) that approximates the matching polytope. Building on the ideas of Rothvoß [Rot14], Braun and Pokutta [BP15] showed that any extended formulation that approximates the matching polytope up to a factor of $1 + O\left(\frac{1}{n}\right)$ has size $2^{\Omega(n)}$ ruling out a *fully polynomial type approximation scheme* (FPTAS) type extended formulation (size polynomial in both n and $\frac{1}{\varepsilon}$) for matching. Rothvoß [Rot14] observed that this already implies that for any $\frac{2}{n} \leq \varepsilon \leq 1$, any $(1 + \varepsilon)$ approximating extended formulation must have $2^{\Omega(1/\varepsilon)}$ size.

Theorem 1.1 ([Rot14, BP15]). *For any $\frac{2}{n} \leq \varepsilon \leq 1$ and any polytope $P_{MAT} \subseteq K \subseteq (1 + \varepsilon)P_{MAT}$,*

$$\mathsf{xc}(K) \geq 2^{\Omega(1/\varepsilon)}.$$

---

[1]See Appendix A for a proof.

Note that the above implies a tight lower bound of $2^{\Omega(n)}$ for $\varepsilon \leq \frac{2}{n}$ but leaves a gap between the upper and lower bounds. Moreover, the gap gets larger as ε increases and in particular when $\varepsilon = \Omega(1)$ we do not even have non-trivial lower bounds.

The above theorem is proven using the connection between extension complexity and non-negative rank of slack matrices that was established in the work of Yannakakis [Yan91] and subsequently extended by Braun, Fiorini, Pokutta and Steurer [BFPS15] to handle approximations of polytopes. The lower bound on extension complexity above, then follows from a lower bound on the non-negative rank of the slack matrix associated with the matching polytope.

A matrix that is closely related to the matching slack matrix is the unique disjointness matrix. Let $\mathcal{Y}$ be the collection of all subsets of $[n]$. For a parameter $\rho \in [0,1]$, we say that a non-negative matrix $A \in \mathbb{R}^{\mathcal{Y} \times \mathcal{Y}}$ is a unique disjointness matrix if

$$A_{xy} = \begin{cases} 1, \text{ if } |x \cap y| = 0 \text{ and} \\ \leq 1 - \rho, \text{ if } |x \cap y| = 1. \end{cases} \tag{1.1}$$

Many recent lower bounds in extended formulations (see Section 1.2) follow from a lower bound on the non-negative rank of such matrices. Starting with the breakthrough work of Fiorini, Massar, Pokutta, Tiwary and de Wolf [FMP$^+$15], a sequence of works [BFPS15, BM13, BP16] proved that the non-negative rank of any unique disjointness matrix is $2^{\Omega(\rho n)}$.

For this work, the matrix that will be of relevance is the lopsided version of the unique disjointness matrix where the rows are indexed by k-subsets of $[n]$ where $k \leq \frac{n}{2}$, while the columns are indexed by all subsets of $[n]$. This is useful for us since it turns out that to prove an extension complexity lower bound for any $(1+\varepsilon)$-approximation for the matching polytope it suffices to prove a lower bound on similar lopsided slack matrices. The rows of any such slack matrix (an example is exhibited by the slack matrix for the polytope Q_ε) are indexed by odd sets of size at most $O(\frac{1}{\varepsilon})$ and the columns are indexed by all possible matchings of which there are exponentially many in n.

From the known lower bounds for the unique disjointness matrix, one could infer that the non-negative rank of the lopsided unique disjointness matrix must be $2^{\Omega(\rho k)}$. One could however, hope to improve this bound to $\approx \binom{n}{k}$ (which is much larger than $2^{\Omega(k)}$ when k is small) at least for the case $\rho = 1$ by making use of the lopsided structure. Such lopsided structure has previously been exploited in the setting of communication complexity to prove analogous lower bounds for lopsided versions of disjointness [MNSW98, AIP06, Pat11, NR15].

1.1   Our Results

In this paper, we show that the simple upper bound exhibited by the polytope Q_ε defined above is tight in the sense that any extended formulation of a polytope that $(1 + \varepsilon)$ approximates P_{MAT} must (roughly) have as many facets as Q_ε.

Theorem 1.2. *For any $\frac{2}{n} \leq \varepsilon \leq 1$ and any polytope $P_{MAT} \subseteq K \subseteq (1 + \varepsilon)P_{MAT}$, it holds that*

$$\mathsf{xc}(K) \geq \binom{n}{\alpha/\varepsilon},$$

where $0 < \alpha < 1$ is an absolute constant.

Note that the above theorem also covers the case for which tight bounds were previously known: when $\varepsilon \leq \frac{2}{n}$, we get an asymptotically tight lower bound of $\binom{n}{\alpha n/2} = 2^{\Omega(n)}$.

We also prove a tight lower bound on the non-negative rank of the lopsided unique disjointness matrix. Let $k \leq \frac{n}{2}$ and $\mathcal{X} = \binom{[n]}{k}$ be the collection of k-subsets of $[n]$ and let $\mathcal{Y}$ be the collection of all subsets of $[n]$. For a parameter $\rho \in [0,1]$, let $A \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$ be any non-negative matrix satisfying (1.1).

Then denoting by $\mathsf{rk}_+(A)$ the nonnegative rank of A, we prove that:

Theorem 1.3. *For any* $\frac{3 \cdot 1000^8}{\log(n/k)} \leq \rho^8 \leq 1$, *it holds that* $\mathsf{rk}_+(A) \geq \binom{n}{\alpha \rho^8 k}$ *with* $0 < \alpha < 1$ *an absolute constant.*

From known results (see Section 3), proving a lower bound on the non-negative rank of any lopsided unique disjointness matrix also gives a lower bound on the extension complexity of a lopsided version of the correlation polytope. Formally, we have the following result for the polytope $P_{LCORR} \in \mathbb{R}^{n \times n}$ defined as

$$P_{LCORR} = \mathsf{conv}\left\{ bb^T \mid b \in \{0,1\}^n, \sum_{i=1}^n b_i \leq k \right\}.$$

Corollary 1.4. *Let* $1 \leq \sigma^8 \leq \frac{1}{3 \cdot 1000^8} \log\left(\frac{n}{k}\right)$. *For any polytope* $P_{LCORR}(n) \subseteq K \subseteq \sigma P_{LCORR}(n)$, *it holds that* $\mathsf{xc}(K) \geq \binom{n}{\alpha k/\sigma^8}$ *with* $0 < \alpha < 1$ *an absolute constant.*

1.2 Other Related Work

Here we briefly survey some of the recent related lower bounds on the size of linear programs (LPs) and semidefinite programs (SDPs) for other polytopes and optimization problems. The progress in extension complexity lower bounds was jump-started by the breakthrough work of Fiorini, Massar, Pokutta, Tiwary and de Wolf [FMP+15] who related the extension complexity of the TSP and Correlation polytopes with the non-negative rank of unique disjointness matrix where the entries are 0 if the sets intersect and proved exponential lower bounds for the latter. This implied an exponential lower bound on the extension complexity of the TSP polytope extending the results of Yannakakis [Yan91] which only applied to symmetric extended formulations. Via reductions established in [AT13, PV13], exponential lower bounds for extension complexity also follow for several other NP-hard polytopes such as the 3SAT and Knapsack polytopes.

Braun, Fiorini, Pokutta and Steurer [BFPS15] defined a notion of approximation for combinatorial optimization problems in terms of extended formulations. They proved that even approximating the value of any linear objective function over the Correlation polytope requires extended formulations of exponential size by proving non-negative rank lower bounds for unique disjointness when the intersecting entries are $1 - \rho$. Braverman and Moitra [BM13] (see also [BP16]) used information theoretic techniques to prove a tight lower bound of $2^{\Omega(\rho n)}$ on the non-negative rank of unique disjointness. An implication of this result was that any LP of size $2^{O(n^\varepsilon)}$ for the convex hull of all cliques in all n-node graphs cannot achieve better than an $n^{1-\varepsilon}$ integrality gap. This matches the algorithmic hardness of approximation result of Håstad [Hås96]. Underlying all the works above is a lower bound on the non-negative rank of the unique disjointness matrix. These works, however, did not consider the lopsided version of the unique disjointness matrix.

4

In a parallel line of work, Chan, Lee, Raghavendra and Steurer [CLRS13] proved lower bounds on the size of linear programs for constraint satisfaction problems by relating it to known Sherali-Adams integrality gaps using techniques from Fourier analysis and convex optimization. For instance, they proved that any polynomial sized LP for Max-Cut on n-node graphs cannot beat the trivial approximation factor of $1/2$. Recently, Kothari, Meka and Raghavendra [KMR17] improved their lower bounds to 2^{n^c} for some constant $0 < c < 1$. Lee, Raghavendra and Steurer [LRS15] proved that similar situation arises for SDP relaxations: when it comes to maximum constraint satisfaction problems, SDPs of size polynomial size cannot perform much better than sums of squares relaxations of constant degree. For instance, this implies that for MAX-3-SAT, no SDP of polynomial size can beat the trivial approximation factor of $7/8$.

When it comes to approximations, the problems discussed above have exponential size LP lower bounds. The matching problem is very different in the sense that for any $\varepsilon \geq \frac{2}{n}$, it has an LP of size $\approx \binom{n}{1/\varepsilon}$ which achieves a $(1+\varepsilon)$ approximation. We show that for the matching problem this is best possible. A similar situation arises for the Max-Knapsack problem for which Bienstock [Bie08] gave an LP of size $\approx \binom{n}{1/\varepsilon}$ which achieves a $(1-\varepsilon)$ approximation for any $\varepsilon \geq \frac{2}{n}$. An exponential lower bound for exact extended formulations for Max-Knapsack follows by a reduction from the unique disjointness matrix [AT13, PV13] (also see [GJW16] which proves a better lower bound by using a different construction). It is unclear how to extend the aforementioned reduction from unique disjointness to prove a strong lower bound for any $(1-\varepsilon)$ approximation for Max-Knapsack. A lower bound even in the case of $(1-\frac{1}{n})$ approximation remains an interesting open problem.

1.3 Organization

In Section 2 we give the high-level intuition for our results before we delve into technical details. Section 3 discusses the slack matrix for the Matching and Lopsided Correlation polytopes. Section 4 contains the basic notation and preliminaries. In Section 5 we prove lower bounds on the non-negative rank of the lopsided unique disjointness matrix and the matching slack matrix. Section 6 proves a key technical lemma.

2 Overview of Techniques

To give an outline of the proofs, we will assume basic familiarity with the definition of entropy of discrete random variables. Random variables will be denoted by capital letters and the values attained by them will be denoted by smaller letters. We will write $p(x)$ to denote the distribution of X as well as the probability of the event $X = x$ in the probability space p, where the meaning will be clear from context. See section 4 for notational conventions and preliminaries.

Our proof is based on the *common information* approach used by Braun and Pokutta [BP15, BP16]. In this approach, we view a non-negative matrix L indexed by x and y as a probability distribution $p(x, y)$ by normalizing by the total weight of the matrix. Let X and Y be sampled from $p(x, y)$. If $\mathsf{rk}_+(L) = r$, then the corresponding distribution $p(x, y)$ is a convex combination of r product distributions given by the non-negative rank one factors. Viewed this way, a non-negative factorization of L gives us a random variable R with support of size $\log \mathsf{rk}_+(L)$ that then breaks the dependency between X and Y. In other words, X and Y are independent if we know the value of R (see Section 4) or formally, X, R, Y form a Markov chain (denoted by $X - R - Y$).

2.1 Lopsided Unique Disjointness

To give intuition behind the proof of Theorem 1.3, let us take a look at what the corresponding distribution looks like when the matrix we start with is the lopsided unique disjointness matrix A given by (1.1). Let X' and Y' be the random variables sampled from the distribution obtained by normalizing A. Then X' is a random subset of $[n]$ of size at most k and Y' is an arbitrary random subset of $[n]$. It turns out that by using a direct-sum argument (dividing the universe $[n]$ into blocks of size $[\frac{n}{k}]$) and appropriate conditioning we can reduce our question to a problem about random subsets X and Y of the universe $[\frac{n}{k}]$.

In the reduced problem, the set $X \subseteq [\frac{n}{k}]$ is a random subset of size one and $Y \subseteq [\frac{n}{k}]$ is an arbitrary random set while the probability that $X \cap Y = \varnothing$ is at least $\frac{1}{2} + \Omega(\rho)$. We will argue that if the non-negative rank was in fact small, then the probability that $X \cap Y = \varnothing$ can not be much larger than $\frac{1}{2}$ deriving a contradiction.

In the reduced problem, the entropies of X and Y are given by $\mathbf{H}(X|X \cap Y = \varnothing) = \log\left(\frac{n}{k}\right) - O(1)$ and $\mathbf{H}(Y|X \cap Y = \varnothing) = \frac{n}{k} - O(1)$. If $\log \mathsf{rk}_+(A) \ll \gamma^8 k \log\left(\frac{n}{k}\right)$ for a small constant γ, then it turns out that we get a random variable R such that $X - R - Y$ and even after conditioning on R the entropies remain large: $\mathbf{H}(X|R, X \cap Y = \varnothing) \geq \log\left(\frac{n}{k}\right) - \gamma^8 \log\left(\frac{n}{k}\right) - O(1)$ and $\mathbf{H}(Y|R, X \cap Y = \varnothing) \geq \frac{n}{k} - \gamma^8 \log\left(\frac{n}{k}\right) - O(1)$. Note that the conditioning on $X \cap Y = \varnothing$ is needed to carry out the direct-sum argument and is quite essential.

To prove a non-negative rank lower bound for lopsided unique disjointness, we exploit the lopsided structure to prove the following key technical lemma which intuitively says that for most values of R, the probability that the event $X \cap Y = \varnothing$ happens conditioned on R is smaller than $\frac{1}{2} + \gamma$. In the following lemma, we view the set X as an element of $[\frac{n}{k}]$ while we view $Y \in \{0,1\}^{\frac{n}{k}}$ as an indicator vector for a subset of $[\frac{n}{k}]$, so $X \cap Y = \varnothing$ is equivalent to the event that $Y_X = 0$.

Lemma 2.1. *Let m be a large enough integer, $X \in [m], Y \in \{0,1\}^m$ and R be random variables with distribution $p(xyr)$ such that $X - R - Y$. For any γ satisfying $\frac{3}{\log m} \leq \gamma^8 \leq \frac{1}{2^{64}}$ define $\mathcal{B} = \{(x,r)|p(Y_x = 0|r) \geq \frac{1+\gamma}{2}\}$. If*

$$\mathbf{H}(X|R, Y_X = 0) \geq (1 - \gamma^8)\log m - 3 \quad \text{and} \quad \mathbf{H}(Y|R, Y_X = 0) \geq m - \gamma^8 \log m - 3,$$

then, $p((x,r) \in \mathcal{B}) \leq 64\gamma$.

With a little work, one could conclude from the above lemma that if the non-negative rank of A was small, then the probability of the event $X \cap Y = \varnothing$ is at most $\frac{1}{2} + O(\gamma)$. Choosing γ to be sufficiently small, we can ensure that this probability is much smaller than $\frac{1}{2} + \Omega(\rho)$, and so, we derive a contradiction to our initial assumption, that the non-negative rank of the lopsided unique disjointness matrix is small.

To understand Lemma 2.1 in more detail, it is worthwhile to consider some simple examples. If it was the case that $\mathbf{H}(X|R, Y_X = 0) \geq \log m - \alpha$ and $\mathbf{H}(Y|R, Y_X = 0) \geq m - \alpha$ where $\alpha \ll 1$, then using Pinsker's inequality, conditioned on the event $Y_X = 0$ and most values of R, X and Y are close to uniform in statistical distance. Hence, the probability that x, r is such that $p(Y_x = 0|r)$ is significantly larger than $\frac{1}{2}$ is small. Lower bounds on the non-negative rank of the standard unique disjointness matrix (rows and columns indexed by all possible subsets of $[n]$) essentially follow from a variation of this argument since in those cases the entropy loss is small enough so that we can say that the distributions of X and Y (conditioned on R and the event $Y_X = 0$) are close to uniform in statistical distance.

In our case, however, the entropy loss is large enough so that the distributions are quite far from uniform in statistical distance. Let us try to construct an example where the entropy loss is larger. Let R be a random subset of $[m]$ of size $m^{1-\alpha}$ and $T \subseteq R$ be a random subset of size $\alpha \log m$. X will be a uniform index chosen from the subset R while the string Y is chosen uniformly conditioned on the event that the $Y_i = 0$ for every $i \in T$. In this case, R is a random variable that breaks the dependency between X and Y. Also, we have that $\mathbf{H}(X|R, Y_X = 0) = (1-\alpha)\log m$ and $\mathbf{H}(Y|R, Y_X = 0) = m - \alpha \log m$, so the entropy loss is of the same order as in the assumptions of Lemma 2.1. Here, the event that x, r is such that $p(Y_x = 0|r)$ is significantly larger than $\frac{1}{2}$ occurs only when $x \in T$, and hence, the probability of such x, r is at most $|T|/|R| \le \frac{\alpha \log m}{m^{1-\alpha}}$.

Generalizing the intuition gained from the example given above, the proof of Lemma 2.1 proceeds by showing that if the measure of x, r such that $p(Y_x = 0|r) \ge \frac{1+\gamma}{2}$ is large, as well as the entropy $\mathbf{H}(X|R, Y_X = 0)$ is large, then for most values of R, there is a large set of coordinates of Y that must be very biased given R, and hence the entropy $\mathbf{H}(Y|R, Y_X = 0)$ must be small. This intuition is borrowed from the lower bounds on lopsided disjointness in communication complexity [Pat11, NR15], but since the setting of non-negative rank is different, the technical details involved for converting this intuition into proof are more challenging here.

Before moving on, we stress two key points about Lemma 2.1. Firstly, given the assumptions on entropy here, one may hope to say that the probability that x, r is such that $p(Y_x = 0|R = r) \le \frac{1-\gamma}{2}$ must also be small. However, this is not true − the lemma is one-sided and it is fairly easy to construct examples where this is not the case. And secondly, the lopsided structure is crucial in Lemma 2.1. Such a lemma is not true if one considers the case where Y is also a random subset of $[m]$ of size one satisfying $\mathbf{H}(Y|R, X \cap Y = \varnothing) \ge (1 - \gamma^8)\log m - O(1)$ for a constant $\gamma > 0$. So, even though one could hope that the non-negative rank of the small set unique disjointness matrix (where we restrict both rows and columns to be indexed by sets of size at most k) is $\approx \binom{n}{k}$, a common information based approach, as used here, will be unable to prove this.

2.2 Matching Slack Matrix

Recalling the connection between extension complexity and non-negative rank of slack matrices, it turns out that to prove Theorem 1.2 it is sufficient to prove a lower bound on the non-negative rank of an appropriate slack matrix associated with the matching polytope (see Section 3). For the sake of providing intuition, we will work with a slightly simpler slack matrix S in this section. The rows of this slack matrix are indexed by odd cuts (subsets) of $[2n + 6]$ of size at most $1/\varepsilon$ and the columns are indexed by perfect matchings in the complete graph on the vertex set $[2n + 6]$. The entry corresponding to cut u and perfect matching m is $S_{um} = |\delta(u) \cap m| - 1$ where $\delta(u)$ is the set of edges of the complete graph on $[2n + 6]$ crossing u (edges with exactly one end point in u). The true slack matrix whose non-negative rank we need to bound to prove Theorem 1.2 is a noisy version of this slack matrix where a small constant $\beta > 0$ is added to every entry.

Using a direct-sum argument with appropriate conditioning (dividing vertices into $1/\varepsilon$ chunks of size $O(\varepsilon n)$ each), we can reduce our problem to a question about a random odd cut U and a random perfect matching M in a graph with $2t + 6$ vertices where $t := \varepsilon n$. Furthermore, the non-negative rank decomposition gives a random variable R such that $U - R - M$ where the size of support of R is $\log \mathsf{rk}_+(S)$. Denoting by $q(u, m, r)$ the distribution of U, M, R, it turns out that the probability $q(U = u, M = m)$ is proportional to $|\delta(u) \cap m| - 1$. In particular, if only one edge of the matching m crosses the cut u, then it has probability zero under the distribution q.

To describe the high-level idea of the proof we need some notation. We first fix an arbitrary

perfect matching $\mathcal{A}$ in the graph and an arbitrary cut $\mathcal{Z}$ that cuts all edges of $\mathcal{A}$. We pick a uniformly random partition $B = (B_0, B_1, \ldots, B_t)$ of the set $\mathcal{Z}$ such that $|B_0| = 3$ and $|B_i| = 2$ for each $i \in [t]$. We call a cut U *consistent* if $U = B_0 \cup B_j$ for some $j \in [t]$ and note that the size of any such cut is always 5. We call a perfect matching M *consistent* if M always includes the edges of $\mathcal{A}$ touching B_0 and inside the other blocks B_j for $j \in [t]$, either M includes the edges of $\mathcal{A}$ or M matches the block B_j to itself and the neighbors of B_j under $\mathcal{A}$ to itself (see Figure 2.1).
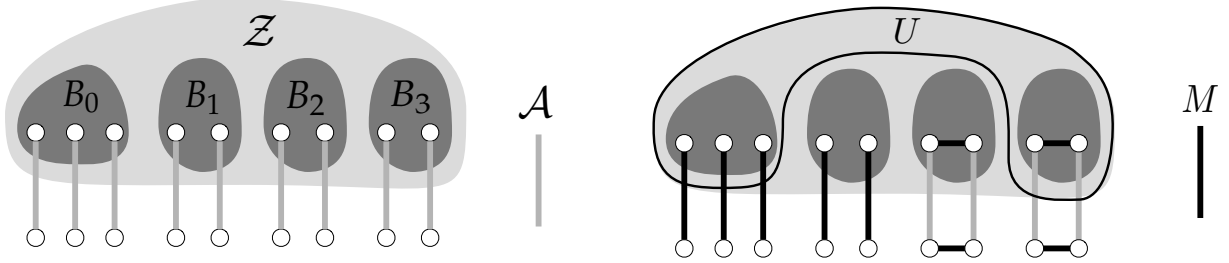

Figure 2.1: An example of $\mathcal{A}, \mathcal{Z}, B$ with $t = 3$ (left) and U, M conditioned on $\mathcal{D}$ (right). Note that U and M are both consistent.

Let $\mathcal{E}$ be the event that U and M are both consistent. Note that given the partition B, one can check whether the cut U is consistent without knowing what the matching M is and vice-versa. Hence, as $U - R - M$, it follows that even conditioned on the event $\mathcal{E}$, U and M are still independent given R and B. Furthermore, when the event $\mathcal{E}$ occurs then either $|\delta(U) \cap M| = 3$ (when U does not cut the edges of M apart from those incident on B_0) or $|\delta(U) \cap M| = 5$ otherwise.

Let $\mathcal{D} \subseteq \mathcal{E}$ denote the event that U and M are consistent and U does not cut the edges of M apart from those incident on B_0. Comparing this setup to the case of lopsided unique disjointness, one can see that apart from choosing B_0, U corresponds to picking a set of size one among the t blocks $(B_1, \ldots, B_t)$, and apart from the edges touching B_0, M corresponds to picking a subset of the same t blocks by considering the elements where M crosses the block B_j to be in the set. The event $\mathcal{D}$ then exactly corresponds to the event that these sets are disjoint. In the "disjoint" case, there are exactly 3 edges of M crossing the cut U where as in the "intersecting" case the number of edges of M crossing U is exactly 5.

As the probability under q is proportional to $|\delta(U) \cap M| - 1$, it is not too hard to see that the probability of the event $q(\mathcal{D}|\mathcal{E}) = \frac{(3-1)}{(3-1)+(5-1)} = \frac{1}{3}$. Furthermore, the entropies $\mathbf{H}(U|B, \mathcal{D}) = \log t - O(1)$ and $\mathbf{H}(M|B, \mathcal{D}) = t - O(1)$. If $\log \mathsf{rk}_+(S) \ll \gamma^8 \cdot \frac{n}{t} \log t$ for a small constant γ, then even after conditioning on R the entropies remain large: $\mathbf{H}(U|RB, \mathcal{D}) \geq \log t - \gamma^8 \log t - O(1)$ and $\mathbf{H}(M|RB, \mathcal{D}) \geq t - \gamma^8 \log t - O(1)$.

We want to proceed similarly to the case of lopsided unique disjointness: we want to conclude that the assumptions on entropy imply that $q(\mathcal{D}|\mathcal{E})$ must be much smaller than $\frac{1}{3}$. In the case of lopsided disjointness, it was enough for us to use Lemma 2.1 and bound the contribution to $p(\mathcal{D}|\mathcal{E}, R = r, B = b)$ by $\frac{1}{2} + \gamma$ for most r, b. But now as we want to prove that the probability is smaller than $\frac{1}{3}$, we need to exploit the combinatorial structure of the matching polytope.

This is where the random partition B, which is the key new idea introduced by Rothvoß [Rot14], is useful. We are going to argue that only certain kinds of partitions can contribute to the probability of the event $\mathcal{D}$ otherwise there is a non-zero probability of sampling a cut U and a matching M that satisfies $|\delta(U) \cap M| = 1$ and any such pair has probability zero in the distribution q, since the corresponding slack matrix entry in S is zero. Then, averaging over all the choices of the random

partition B, we can show the total contribution of these random partitions to $q(\mathcal{D}|\mathcal{E})$ is indeed less than $\frac{1}{3}$.

Let us make some simplifying assumptions first. Define M_j to be the edges of M corresponding to block B_j. Let us assume that for all values r, b the probability that $M_j = \mathcal{A}_j$ (M crosses B_j) is roughly $\frac{1}{2}$ conditioned on $\mathcal{E}, R = r, B = b$ for each $j \in [t]$. Since conditioned on $\mathcal{E}$ either $M_j = \mathcal{A}_j$ or $M_j \neq \mathcal{A}_j$, this implies that

$$q(M_j \neq \mathcal{A}_j|\mathcal{E}, R = r, B = b) \approx q(M_j = \mathcal{A}_j|\mathcal{E}, R = r, B = b) \approx \frac{1}{2}.$$

Also assume that U is almost uniform among the m possible cuts. Then, as U and M are independent conditioned on $\mathcal{E}, R = r, B = b$, we get that

$$q(U = B_0 \cup B_j, M_j \neq \mathcal{A}_j|\mathcal{E}, R = r, B = b) \approx q(U = B_0 \cup B_j, M_j = \mathcal{A}_j|\mathcal{E}, R = r, B = b).$$
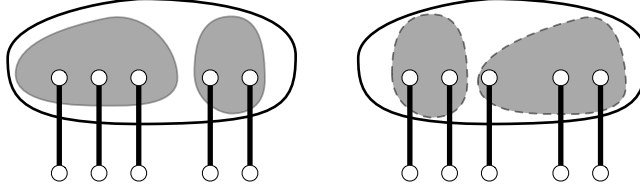

Figure 2.2: An example of a cut and matching corresponding to the event $U = B_0 \cup B_j, M_j = \mathcal{A}_j$ and an example of two different splits of the cut into $B_0 \cup B_j$. By symmetry any such partition is equi-probable.

The event on the right hand side above fixes a cut of size 5 in the graph and a matching that crosses on all the edges. Fix the blocks outside this cut arbitrarily. By symmetry all of the $\binom{5}{3}$ ways of splitting this cut into B_0 and B_j are equally likely (see Figure 2.2). It turns out that the right hand side above can be non-zero only when the vertices chosen in B_0 form a 2-intersecting family (note that this determines B_j as the parts outside are already fixed) and hence averaging over b and r, we get that $q(\mathcal{D}|\mathcal{E}) \approx \frac{\Gamma}{10} q(\overline{\mathcal{D}}|\mathcal{E})$ where Γ is a bound on the size of any such family. It turns out that Γ is small enough so that we can conclude $q(\mathcal{D}|\mathcal{E}) < \frac{1}{3}$ and derive a contradiction.

Why must the probability be zero when there are two partitions b and b' (which are same everywhere outside this cut) such that $|b_0 \cap b_0'| = 1$? This is because we can choose an appropriate cut u that is consistent with b' (See Figure 2.3(a)) and has non-zero probability $q(U = u|\mathcal{E}, R = r, B = b') > 0$ and an appropriate matching m that is consistent with b (see Figure 2.3(a)) and has non-zero probability $q(M = m|\mathcal{E}, R = r, B = b) > 0$. Note that then it must also hold that the cut and the matching has non-zero probability even without conditioning on $\mathcal{E}$: $q(U = u|R = r) > 0$ and $q(M = m|R = r) > 0$. But since U and M are independent given R then such a pair would have non-zero probability even though $|\delta(u) \cap m| = 1$ and any such pair must have zero probability under q.

If it was the case that $\mathbf{H}(U|B, \mathcal{D}) \geq \log t - \alpha$ and $\mathbf{H}(M|B, \mathcal{D}) \geq t - \alpha$ where $\alpha \ll 1$, then one could work with statistical distance as is done in [Rot14, BP15], but since in our case the entropy loss is much larger, we use Lemma 2.1 in conjunction with the random partition idea.

As mentioned before, to prove Theorem 1.2 we have to work with a noisy version of the slack matrix used above, where the probability of a pair U and M satisfying $|\delta(U) \cap M| = 1$ is not zero, but a small constant. This along with the limitations of Lemma 2.1 and the lopsided structure makes translating this intuition into a formal proof considerably more involved.
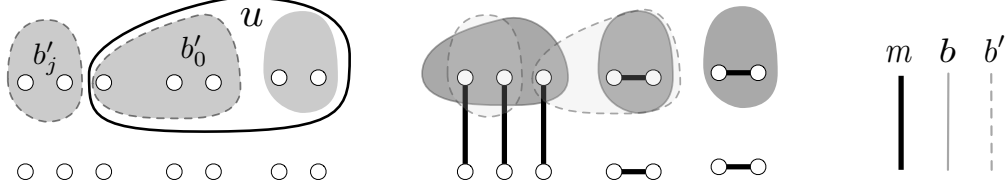
(a) An example of a cut u consistent with b' (left) and a matching m consistent with b (right). Note that b and b' agree on all the blocks except b_0 and b_j.
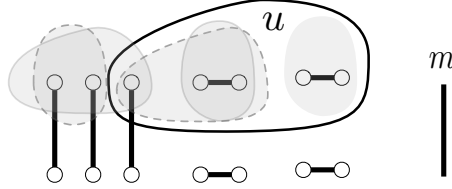
(b) The cut u and matching m satisfy $|\delta(u) \cap m| = 1$.

3   Slack Matrices and Non-Negative Rank

Consider polytopes $P = \mathsf{conv}\{x_1, \ldots, x_s\} \subseteq \mathbb{R}^d$ and $Q = \{x \in \mathbb{R}^n \mid \langle a_i, x \rangle \le b_i, \ \forall i \in [t]\} \subseteq \mathbb{R}^d$ such that $P \subseteq Q$. The *Slack Matrix* $S^{P,Q} \in \mathbb{R}^{s \times t}$ corresponding to polytopes P and Q is a non-negative matrix defined by $S^{P,Q}_{ij} = b_i - \langle a_i, v_j \rangle$.

The *non-negative rank* of a non-negative matrix S is

$$\mathsf{rk}_+(S) = \min\{r | \exists U \in \mathbb{R}^{s \times r}_{\ge 0}, V \in \mathbb{R}^{r \times t}_{\ge 0} : S = UV\}.$$

Braun, Fiorini, Pokutta and Steurer [BFPS15] showed that to lower bound the size of extended formulations for any polytope sandwiched between inner polytope P and outer polytope Q, it suffices to lower bound the non-negative rank of the corresponding slack matrix $S^{P,Q}$.

Theorem 3.1 ([BFPS15]). *Let $P = \mathsf{conv}\{v_1, \ldots, v_s\} \subseteq \mathbb{R}^d$ and $Q = \{x | \langle a_i, x \rangle \le b_i, \forall 1 \le i \le t\} \subseteq \mathbb{R}^d$ be two polytopes. Then, for any polytope $K \subseteq \mathbb{R}^d$ satisfying $P \subseteq K \subseteq Q$, $xc(K) \ge \mathsf{rk}_+(S^{P,Q}) - 1$.*

Slack Matrix for the Matching Polytope

Let $0 < \beta < 1$ be a constant which will be determined by the proof. In proving Theorem 1.2, we may assume without loss of generality that $\frac{c'}{n} \le \varepsilon \le 1$ for a sufficiently large constant c'. We call a subset of vertices a *cut* and for a cut U, we define $\delta(U)$ to be the set of edges with exactly one end point in U and $E(U)$ to be the set of edges inside U.

Consider the following outer polytope $Q'_\varepsilon(n) \subseteq \mathbb{R}^{\binom{[n]}{2}}$ for matching:

$$Q'_\varepsilon(n) = \left\{ x \in \mathbb{R}^{\binom{[n]}{2}} \ \middle| \ \sum_{e \in \delta(\{i\})} x_e \le 1 \,, \forall i \in [n]; \ x \ge 0; \right.$$
$$\left. \sum_{e \in E(U)} x_e \le \frac{|U| + \beta - 1}{2} \,, \forall U \subseteq [n], \ |U| \text{ odd}, \ |U| \le 1 + \frac{\beta}{\varepsilon}. \right\}$$

For any cut U of size at most $1 + \frac{\beta}{\varepsilon}$, $(1 + \varepsilon)(|U| - 1) \leq |U| + \beta - 1$, so we have the following.

Proposition 3.2. $(1 + \varepsilon)P_M(n) \subseteq Q'_\varepsilon(n)$.

Hence, to lower bound the extension complexity of approximating polytopes, it suffices to lower bound the non-negative rank of the slack matrix $S^{P_{MAT}, Q'_\varepsilon}$. Since there are only $O(n)$ degree constraints, let us restrict ourselves to the submatrix S of $S^{P_{MAT}, Q'_\varepsilon}$ given by odd cuts and perfect matchings. Then, the entry corresponding to cut u and perfect matching m is

$$S_{um} = \frac{|u| + \beta - 1}{2} - |E(u) \cap m| = \frac{|\delta(u) \cap m| + \beta - 1}{2}, \tag{3.1}$$

since for any perfect matching m and odd cut u, $|E(u) \cap m| = \frac{|u| - |\delta(u) \cap m|}{2}$. Note that this matrix has $\Theta(1/\varepsilon)$ rows. We prove that

Theorem 3.3.

$$\mathsf{rk}_+(S) \geq \binom{n}{\alpha/\varepsilon}$$

where $\frac{c'}{n} \leq \varepsilon \leq 1$ for a large enough constant $c' = c'(\beta)$ and $0 < \alpha < 1$ is an absolute constant.

Theorems 3.1 and 3.3 then give us Theorem 1.2.

Slack Matrix for the Lopsided Correlation Polytope

Consider the polytope $Q = Q(n) = \{x \in \mathbb{R}^{n \times n} \mid \langle 2\mathsf{diag}(a) - aa^T, x \rangle \leq 1, a \in \{0, 1\}^n\}$ where $\mathsf{diag}(a) \in \mathbb{R}^{n \times n}$ denotes the diagonal matrix which has the vector a on the diagonal and is zero otherwise. It is well-known (see [BP16] for example) that $P_{LCORR} \subseteq Q$ and the slack matrix corresponding to the inner polytope P_{LCORR} and outer polytope σQ is the non-negative matrix σA where A is a lopsided unique disjointness matrix satisfying the conditions in (1.1) with parameter $\rho = \frac{1}{\sigma}$. Corollary 1.4 then directly follows from Theorems 1.3 and 3.1.

4 Notation and Preliminaries

4.1 Probability Spaces and Variables

Unless otherwise stated, logarithms in this text are computed base two. We denote by $[n]$ the set $\{1, 2, \ldots, n\}$ and by $\binom{[n]}{k}$ the collection of subsets of $[n]$ of size k. Random variables are denoted by capital letters (e.g. A) and the values they attain are denoted by lower-case letters (e.g. a). Events in a probability space will be denoted by calligraphic letters (e.g. $\mathcal{E}$). For event $\mathcal{D}$, we use $\mathbb{1}_{\mathcal{D}}$ to denote the corresponding indicator variable, $\overline{\mathcal{D}}$ to denote the complement event and for another event $\mathcal{E}$, we use the shorthand $\mathcal{D}, \mathcal{E}$ to denote the intersection event $\mathcal{D} \cap \mathcal{E}$. Given $a = a_1, a_2, \ldots, a_n$ (resp. $a = a^1, \ldots, a^n$), we write $a_{\leq i}$ (resp. $a^{\leq i}$) to denote $a_1, \ldots, a_i$ (resp. $a^1, \ldots, a^i$). We define $a_{<i}, a_{\geq i}, a_{>i}$ (resp. $a^{<i}, a^{\geq i}, a^{>i}$) similarly.

Given a probability space p and a random variable A in the underlying sample space, we use the notation $p(a)$ to denote both the distribution on the variable a, and the number $\mathbb{P}_p[A = a]$. The meaning will be clear from context. We will often consider multiple probability spaces with the same underlying sample space, so for example $p(a)$ and $q(a)$ will denote the distribution of the random

variable A under the probability spaces p and q respectively with the underlying sample space of p and q being the same. We write $p(a|b)$ to denote either the distribution of A conditioned on the event $B = b$, or the number $\mathbb{P}[A = a|B = b]$. Given a distribution $p(a, b, c, d)$, we write $p(a, b, c)$ to denote the marginal distribution on the variables a, b, c (or the corresponding probability). We often write $p(ab)$ instead of $p(a, b)$ for conciseness of notation. If $\mathcal{E}$ is an event, we write $p(\mathcal{E})$ to denote its probability according to p.

The support of a random variable A is defined to be the set $\mathsf{supp}(A) := \{a \mid p(a) > 0\}$. Given a fixed value c, we denote by $\mathbb{E}_{p(b|c)}[g(a, b, c)] := \sum_b p(b|c) \cdot g(a, b, c)$, the expected value of the function $g(a, b, c)$ under the distribution $p(b|c)$. If the probability space p is clear from the context, then we will just write $\mathbb{E}_{b|c}[g(a, b, c)]$ to denote the expectation. We use $\mathbb{E}_{a \in \mathcal{A}}[g(a)]$ to denote the expected value of $g(a)$ under the uniform distribution over the set $\mathcal{A}$.

We write $A - R - B$ to assert that the random variables A, R, B form a Markov chain, or, in other words, $p(arb) = p(r) \cdot p(a|r) \cdot p(b|r)$. In stating the preliminary lemmas and definitions, p is assumed to be the underlying probability space of the random variables being considered.

To get familiar with the notation, consider the following example. Let $A \in \{0, 1\}^2$ be a uniformly distributed random variable in a probability space p. Then, $p(a)$ is the uniform distribution on $\{0, 1\}^2$ and if $a = (0, 0)$, $p(a) = 1/4$. Let A_1 and A_2 denote the first and second bits of A, then if $B = A_1 + A_2 \bmod 2$, then when $b = 1$, $p(a|b)$ is the uniform distribution on $\{(0, 1), (1, 0)\}$. If $a = (1, 0)$, and $b = 1$, then $p(a|b) = 1/2$, and $p(a, b) = 1/4$. If $\mathcal{E}$ is the event that $A_1 = B$, then $p(\mathcal{E}) = 1/2$. Let $q(a) = p(a|\mathcal{E})$, then $q(a)$ is the uniform distribution on $\{(0, 0), (1, 0)\}$ and $q(a_2)$ is the distribution over the sample space $\{0, 1\}$ which takes the value 0 with probability 1.

4.2 Entropy and Mutual Information

For a discrete random variable A, the entropy of A is defined as

$$\mathbf{H}(A) = \mathbb{E}_{p(a)}\left[\log \frac{1}{p(a)}\right].$$

For any two random variables A and B, the entropy of A conditioned on B is defined as $\mathbf{H}(A|B) = \mathbb{E}_{p(b)}[\mathbf{H}(A|b)]$. The mutual information between A and B is defined as $\mathbf{I}(A : B) = \mathbf{H}(A) - \mathbf{H}(A|B) = \mathbf{H}(A) - \mathbf{H}(B|A)$. Similarly, the conditional mutual information is defined as $\mathbf{I}(A : B|C) = \mathbf{H}(A|C) - \mathbf{H}(A|BC)$.

We shall often work with multiple probability spaces over the same underlying sample space. To avoid confusion, we shall explicitly write $\mathbf{H}_p(A)$ (and $\mathbf{I}_p(A : B|C)$) to specify the probability space p being used for computing the entropy (and mutual information).

4.3 The Binary Entropy Function

The binary entropy function[2] is defined to be $\mathsf{h}(x) := -x \log x - (1 - x) \log(1 - x)$ for $x \in [0, 1]$. The function h is concave on the interval $[0, 1]$ and is decreasing on the interval $[\frac{1}{2}, 1]$.

The proposition below will be quite useful. A proof is given in Appendix B.

Proposition 4.1. $\mathsf{h}(x) \leq 1 - 2 \log e \left(x - \frac{1}{2}\right)^2$ *for all* $x \in [0, 1]$.

---

[2] We adopt the convention that $x \log x = 0$ at $x = 0$.

4.4 Non-negative Rank and Common Information

Given a non-negative matrix $A \subset \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$, we can view it as a probability distribution $p(xy)$ as follows:

$$p(xy) = \frac{A_{xy}}{\sum_{x',y'} A_{x'y'}}.$$

Let X and Y denote random variables with distribution $p(xy)$. Then, we have the following proposition whose proof can be found in Appendix B.

Proposition 4.2 ([BP16]). *There is a random variable R such that $X - R - Y$ and $|\mathsf{supp}(R)| \leq \mathsf{rk}_+(A)$.*

One can view the above proposition as saying that to prove a lower bound on the non-negative rank, it suffices to lower bound a well-known information theoretic quantity called the *common information* between X and Y. For more details on this interpretation, see [BP16].

4.5 Preliminary Information Theory Lemmas

The proofs of the following basic facts can be found in [CT06]:

Proposition 4.3. *If $A \in \{0,1\}^\ell$, then $\mathbf{I}(A : B) \leq \ell$.*

Proposition 4.4. $\mathbf{H}(A|B) \leq \mathbf{H}(A)$ *where the equality holds if and only if A and B are independent.*

The above implies that if A and B are independent then $\mathbf{I}(A : B) = 0$.

Proposition 4.5 (Chain Rule). *If $A = A_1, \ldots, A_n$, then $\mathbf{H}(A) = \sum_{i=1} \mathbf{H}(A_i|A_{<i})$ and $\mathbf{I}(A_1, \ldots, A_n : B) = \sum_{i=1}^n \mathbf{I}(A_i : B_i|A_{<i})$.*

Proposition 4.6 (Chernoff Bound). *The number of strings in $\{0,1\}^m$ with hamming weight at least $3m/4$ is at most $e^{-m/8} \cdot 2^m$.*

The proof of the following lemma can be found in Appendix B.

Lemma 4.7 ([BR11]). *Let $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$ be random variables such that the n-tuples $(X_1, Y_1), \ldots, (X_n, Y_n)$ are mutually independent. Let R be an arbitrary random variable. Then,*

$$\sum_{i=1}^n \mathbf{I}(X_i : R|X_{<i}Y_{\geq i}) \leq \mathbf{I}(X : R|Y) \ \ and \ \ \sum_{i=1}^n \mathbf{I}(Y_i : R|X_{\leq i}Y_{>i}) \leq \mathbf{I}(Y : R|X).$$

Lemma 4.8. *Let X be a random variable such that $\mathbf{H}(X) \geq \log \ell - a$ where $\ell = |\mathsf{supp}(X)|$ and $a \geq 0$. Define $\mathcal{S} = \{x | p(x) \leq \frac{2^{(a+1)/\gamma}}{\ell}\}$. Then, $p(\mathcal{S}) \geq 1 - \gamma$.*

Proof. Set $b = 2^{(a+1)/\gamma}$. Denoting by $\overline{\mathcal{S}}$ the complement of $\mathcal{S}$, we can write

$$\mathbf{H}(X) = \sum_{x \in \mathcal{S}} p(x) \log \left(\frac{1}{p(x)}\right) + \sum_{x \notin \mathcal{S}} p(x) \log \left(\frac{1}{p(x)}\right)$$

$$\leq \sum_{x \in \mathcal{S}} p(x) \log \left(\frac{1}{p(x)}\right) + p(\overline{\mathcal{S}}) \log \left(\frac{\ell}{b}\right)$$

$$\leq p(\mathcal{S}) \log \left(\sum_{x \in \mathcal{S}} \frac{1}{p(\mathcal{S})}\right) + p(\overline{\mathcal{S}}) \log \left(\frac{\ell}{b}\right),$$

13

where the first inequality follows from the definition of $\mathcal{S}$ and the second from concavity of the log function. We can further upper bound

$$\mathbf{H}(X) \leq p(\mathcal{S}) \log\left(\frac{1}{p(\mathcal{S})}\right) + p(\mathcal{S}) \log|\mathcal{S}| + p(\overline{\mathcal{S}}) \log\left(\frac{\ell}{b}\right)$$

$$\leq 1 + p(\mathcal{S}) \log \ell + p(\overline{\mathcal{S}}) \log\left(\frac{\ell}{b}\right),$$

where we used that $x \log(1/x) \leq 1$ for $0 \leq x \leq 1$. Since $\mathbf{H}(X) \geq \log \ell - a$, we get that

$$\log \ell - a \leq 1 + p(\mathcal{S}) \log \ell + (1 - p(\mathcal{S})) \log\left(\frac{\ell}{b}\right),$$

which gives that $p(\mathcal{S}) \geq 1 - \frac{a+1}{\log b} = 1 - \gamma$. $\qquad\square$

Lemma 4.9. *Let $Y \in \{0,1\}^{\ell}$ and define $\mathsf{bias}_i(Y) := p(Y_i = 0) - p(Y_i = 1)$. If there is a set $\mathcal{S} \subseteq [\ell]$ such that $\mathbb{E}_{i\in\mathcal{S}}[\mathsf{bias}_i(Y)] \geq 2\gamma$ where $\gamma > 0$, then $\mathbf{H}(Y) \leq \ell - \gamma^2|\mathcal{S}|$.*

Proof. We may write

$$\mathbb{E}_{i\in\mathcal{S}}[\mathsf{bias}_i(Y)] = \mathbb{E}_{i\in\mathcal{S}}[p(Y_i = 0) - (1 - p(Y_i = 0))] = 2\mathbb{E}_{i\in\mathcal{S}}[p(Y_i = 0)] - 1,$$

which by the assumption implies that $\mathbb{E}_{i\in\mathcal{S}}[p(Y_i = 0)] \geq \frac{1}{2} + \gamma$.

We can upper bound

$$\mathbb{E}_{i\in S}[\mathbf{H}(Y_i)] = \mathbb{E}_{i\in\mathcal{S}}[\mathsf{h}(p(Yi = 0))] \leq \mathsf{h}(\mathbb{E}_{i\in\mathcal{S}}[p(Y_i = 0)]),$$

where the last inequality follows from the concavity of the binary entropy function h. Since h is a decreasing function on $\left[\frac{1}{2}, 1\right]$ and $\mathsf{h}(\frac{1}{2} + x) \leq 1 - 2\log e \cdot x^2 \leq 1 - x^2$,

$$\mathbb{E}_{i\in\mathcal{S}}[\mathbf{H}(Y_i)] \leq \mathsf{h}\left(\frac{1}{2} + \gamma\right) \leq 1 - \gamma^2.$$

Denoting by $\overline{\mathcal{S}}$ the complement of $\mathcal{S}$ and applying the chain rule we get:

$$\mathbf{H}(Y) \leq \mathbf{H}(Y_{\overline{\mathcal{S}}}) + \mathbf{H}(Y_{\mathcal{S}}) \leq \mathbf{H}(Y_{\overline{\mathcal{S}}}) + \sum_{i\in\mathcal{S}} \mathbf{H}(Y_i)$$

$$\leq (\ell - |\mathcal{S}|) + |\mathcal{S}|(1 - \gamma^2) = \ell - \gamma^2|\mathcal{S}|.$$

$\qquad\square$

Lemma 4.10 (Averaging Lemma)**.** *Let A be a bounded random variable such that $\mathbb{E}[A] \geq \alpha$. For any $\beta < \alpha$, let $\mathcal{S} = \{a \mid A(a) \geq \beta\}$. Then, $p(\mathcal{S}) \geq \frac{\alpha-\beta}{m-\beta}$ where $m = \max_a\{A(a)\}$.*

Proof. We have

$$\mathbb{E}[A] = \sum_{a\in\mathcal{S}} p(a)A(a) + \sum_{a\notin\mathcal{S}} p(a)A(a) \leq p(\mathcal{S})m + (1 - p(\mathcal{S}))\beta.$$

which gives us that $(m - \beta)p(\mathcal{S}) \geq \mathbb{E}[A] - \beta \geq \alpha - \beta$. $\qquad\square$

14

4.6 Intersecting Families

The following lemma will be crucial for the analysis. It is a special case of the Erdős-Ko-Rado Theorem (see [Wil84]) which says that when $n \geq 6$, then the size of any family of $\binom{[n]}{3}$ that intersects in two elements is at most $n - 2$ (Lemma 4.11 follows from the case $n = 6$). We give a self-contained proof in Appendix B.

Lemma 4.11. *Let $\mathfrak{F} \subseteq \binom{[5]}{3}$ be a family of subsets such that any two sets in $\mathfrak{F}$ intersect in two elements. Then, $|\mathfrak{F}| \leq 4$.*

5 Lower Bounds on Non-negative Rank

Let us recall the main technical lemma which we use to derive a lower bound on the non-negative rank of the lopsided unique disjointness matrix as well as the matching slack matrix.

Lemma 2.1. *Let m be a large enough integer, $X \in [m], Y \in \{0,1\}^m$ and R be random variables with distribution $p(xyr)$ such that $X - R - Y$. For any γ satisfying $\frac{3}{\log m} \leq \gamma^8 \leq \frac{1}{2^{64}}$ define $\mathcal{B} = \{(x, r) | p(Y_x = 0 | r) \geq \frac{1+\gamma}{2}\}$. If*

$$\mathbf{H}(X | R, Y_X = 0) \geq (1 - \gamma^8) \log m - 3 \quad and \quad \mathbf{H}(Y | R, Y_X = 0) \geq m - \gamma^8 \log m - 3,$$

then, $p((x, r) \in \mathcal{B}) \leq 64\gamma$.

We will prove the above lemma in Section 6. First we use it to derive non-negative rank lower bounds.

5.1 Non-negative Rank of Lopsided Unique Disjointness

In this section we prove Theorem 1.3.

Theorem 1.3. *For any $\frac{3 \cdot 1000^8}{\log(n/k)} \leq \rho^8 \leq 1$, it holds that $\mathsf{rk}_+(A) \geq \binom{n}{\alpha \rho^8 k}$ with $0 < \alpha < 1$ an absolute constant.*

It will be convenient to assume that n is divisible by k and $\frac{n}{k}$ is a large enough integer. We split the universe $[n]$ into blocks of size $\frac{n}{k}$ where $\{\frac{n}{k}(i-1) + 1, \ldots, \frac{n}{k}i\}$ is the ith block for every $i \in [k]$.

Define a distribution on $X \in \binom{[n]}{k}$ and $Y \in \{0,1\}^n$ given by

$$q(xy) = \frac{A_{xy}}{\sum_{x',y'} A_{x'y'}},$$

where A is the lopsided unique disjointness matrix defined in (1.1) with parameter ρ and we view $y \in \{0,1\}^n$ as the indicator vector for a subset of $[n]$. Let X^i denote the intersection of X with the elements in the ith block, and let Y^i be the projection of Y onto the coordinates in the ith block. For every $x \in [\frac{n}{k}]$, we will use the notation Y_x^i to denote the xth coordinate of Y_i, in other words $Y_{\frac{n}{k}(i-1)+x}$.

Let $\mathcal{D}$ denote the event that x and y are disjoint and x^i has exactly one element for every i.

15

Lemma 5.1. *Let R be any random variable satisfying $X - R - Y$. Then, for every $i \in [k]$ it holds that*

$$\mathbf{I}_q\left(R : X^i | X^{<i} Y^{\geq i} \mathcal{D}\right) + \mathbf{I}_q\left(R : Y^i | X^{\leq i} Y^{>i} \mathcal{D}\right) \geq \left(\frac{\rho}{1000}\right)^8 \log\left(\frac{n}{k}\right).$$

Using Lemma 4.7 and Proposition 4.2 together with the above, we have

$$2 \log \mathsf{rk}_+(M) \geq \sum_{i=1}^{k} \left(\mathbf{I}_q\left(R : X^i | X^{<i} Y^{\geq i} \mathcal{D}\right) + \mathbf{I}_q\left(R : Y^i | X^{\leq i} Y^{>i} \mathcal{D}\right)\right) \geq \left(\frac{\rho}{1000}\right)^8 k \log\left(\frac{n}{k}\right),$$

which proves Theorem 1.3 as $\binom{n}{s} = 2^{\Theta\left(s \log\left(\frac{n}{s}\right)\right)}$.

Proof of Lemma 5.1. Fix i as in the statement of the lemma. Let $\mathcal{E}$ be the event that X^i has exactly one element for every i and $X \cap Y$ is a subset of the ith block. Note that $\mathcal{D} \subset \mathcal{E}$.

Writing $W = X^{<i} Y^{>i}$, we will prove that for any fixed value w attained by W, we have that

$$\mathbf{I}_q\left(R : X^i | w Y^i \mathcal{D}\right) + \mathbf{I}_q\left(R : Y^i | w X^i \mathcal{D}\right) \geq \left(\frac{\rho}{1000}\right)^8 \log\left(\frac{n}{k}\right), \tag{5.1}$$

and the proof is completed by averaging over w.

Note that after fixing w and r, X and Y are independent and they can be checked separately to verify that the event $\mathcal{E}$ occurs as for every block $j \neq i$ either X^j or Y^j is fixed given w. It follows that the distribution $p(xyr) := q(xyr | w\mathcal{E})$ satisfies $p(xy|r) = p(x|r)p(y|r)$. Furthermore under the distribution p, $\mathcal{D}$ is equivalent to the event that $Y^i_{X^i} = 0$. We can compute $p(\mathcal{D}) \geq \frac{1}{1 + (1 - \rho)} = \frac{1}{2 - \rho}$ since the matrix entries are given by

$$A_{xy} = \begin{cases} 1 \text{ when } (x, y) \in \mathsf{supp}(p(xy | \mathcal{D})), \\ \leq 1 - \rho \text{ when } (x, y) \in \mathsf{supp}(p(xy | \overline{\mathcal{D}})), \end{cases} \tag{5.2}$$

and the number of entries is the same in both cases.

For the sake of contradiction assume that (5.1) does not hold. Then, we are going to show that $p(\mathcal{D})$ must be significantly smaller than what we computed above. Define $\mathcal{B} = \{(x^i, r) \mid p(Y^i_{x^i} = 0 | r) \geq \frac{1}{2} + \frac{\rho}{2000}\}$. We can upper bound $p(\mathcal{D})$ as follows:

$$p(\mathcal{D}) = \sum_{x^i, r} p(x^i, r) p(Y^i_{x^i} = 0 | r, X^i = x^i) = \sum_{x^i, r} p(x^i, r) p(Y^i_{x^i} = 0 | r)$$

$$\leq \sum_{(x^i, r) \notin \mathcal{B}} p(x^i, r) \left(\frac{1}{2} + \frac{\rho}{2000}\right) + \sum_{(x^i, r) \in \mathcal{B}} p(x^i, r) p(Y^i_{x^i} = 0 | r)$$

$$\leq \left(\frac{1}{2} + \frac{\rho}{2000}\right) + p((x^i, r) \in \mathcal{B}),$$

where the second equality follows since $p(xy | r)$ is product.

We will show that

Claim 5.2. $p((x^i, r) \in \mathcal{B})) \leq \frac{64\rho}{1000}$.

This implies $p(\mathcal{D}) \leq \frac{1}{2} + \frac{\rho}{2000} + \frac{64\rho}{1000} < \frac{1}{2 - \rho}$, which contradicts the fact that $p(\mathcal{D}) \geq \frac{1}{2 - \rho}$. This finishes the proof. Next we prove Claim 5.2. $\qquad \square$

Proof of Claim 5.2. Set $t := \frac{n}{k}$. Conditioned on $x\mathcal{D}$, every coordinate of y^i other than x^i is uniform and hence $\mathbf{H}_p(Y^i|X^i\mathcal{D}) \geq t - 1$. Similarly, conditioned on $y\mathcal{D}$, X^i is uniform on the coordinates of y^i that are zero. We may compute from (5.2) that $p(y^i|\mathcal{D}) = \frac{\Delta(y^i)}{t2^{t-1}}$ where $\Delta(y^i)$ is the number of zeros in y^i. The probability of any string y^i with less than $t/4$ zeros is at most $\frac{1}{2^{t+1}}$ and using Proposition 4.6 their total measure under the distribution $p(y^i|\mathcal{D})$ can be bounded by $\frac{e^{-t/8}}{2}$. Hence, $\mathbf{H}_p(X^i|Y^i\mathcal{D}) \geq (1 - \frac{e^{-t/8}}{2})\log\left(\frac{t}{4}\right) \geq \log t - 3$.

As $p(xyr) = q(xyr|w\mathcal{E})$, if (5.1) is not true, then

$$\mathbf{I}_p\left(R : X^i|Y^i\mathcal{D}\right) = \mathbf{H}_p(X^i|Y^i\mathcal{D}) - \mathbf{H}_p(X^i|RY^i\mathcal{D}) \leq \left(\frac{\rho}{1000}\right)^8 \log t,$$

and a similar statement is obtained by writing $\mathbf{I}_p\left(R : Y^i|X^i\mathcal{D}\right)$ in terms of entropy. It follows that

$$\mathbf{H}_p(X^i|RY^i\mathcal{D}) \geq \left(1 - \left(\frac{\rho}{1000}\right)^8\right)\log t - 3, \text{ and } \mathbf{H}_p(Y^i|RX^i\mathcal{D}) \geq t - \left(\frac{\rho}{1000}\right)^8 \log t - 1.$$

Since entropy can only decrease under conditioning and $\left(\frac{\rho}{1000}\right)^8 \geq \frac{3}{\log t}$, X^i, Y^i and R satisfy the conditions of Lemma 2.1 and the claim follows. $\qquad\square$

5.2   Non-negative Rank of the Matching Slack Matrix

Let us recall the definition of the matching slack matrix S from (3.1). The entry S_{um} corresponding to cut u and perfect matching m is

$$S_{um} = \frac{|\delta(u) \cap m| + \beta - 1}{2},$$

where $0 < \beta < 1$ is a constant that will be determined by the proof and $\delta(u)$ is the set of edges crossing u. In this section we prove Theorem 3.3.

Theorem 3.3.

$$\mathsf{rk}_+(S) \geq \binom{n}{\alpha/\varepsilon}$$

where $\frac{c'}{n} \leq \varepsilon \leq 1$ for a large enough constant $c' = c'(\beta)$ and $0 < \alpha < 1$ is an absolute constant.

For convenience we assume that β/ε is an integer and we work with graphs on $4n + 6$ vertices where n is divisible by β/ε. Set $t = \frac{n}{\beta/\varepsilon}$ and note that $t \geq c$ for a large constant $c = c(\beta)$.

Using the slack matrix S, define a distribution on cuts U and perfect matchings M given by

$$q(um) = \frac{S_{um}}{\sum_{u',m'} S_{u'm'}}.$$

Fix an arbitrary perfect matching $\mathcal{A}$ and an arbitrary cut $\mathcal{Z}$ that cuts all edges of $\mathcal{A}$. Let $B = (B_0, B_1, \ldots, B_n)$ be a uniformly random partition of the set $\mathcal{Z}$ into blocks such that $|B_0| = 3$ and $|B_j| = 2$ for each $j \in [n]$. For $i \in [\frac{n}{t}]$, we call $B^i = \{B_{t(i-1)+1}, \ldots, B_{ti}\}$ the ith *chunk* and for $j \in [t]$ use $B^i_j = B_{t(i-1)+j}$ to denote the jth block of the chunk B^i. Let $\mathcal{A}_j$ denote the edges of $\mathcal{A}$ that touch B_j.
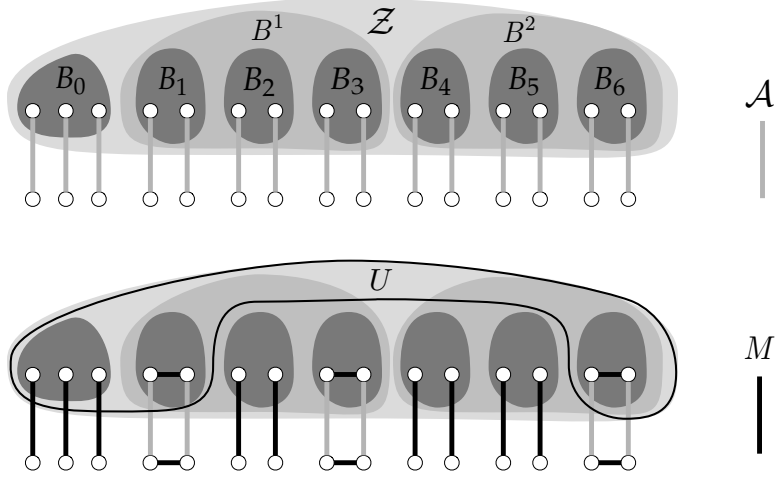
17

Figure 5.1: Top to bottom: An example of $\mathcal{A}, \mathcal{Z}, B$ with $n = 6$ and $t = 3$; U, M conditioned on $\mathcal{D}$.

We say that the cut U is *consistent* with B if $B_0 \subseteq U \subseteq \mathcal{Z}$ and for every $i \in [\frac{n}{t}]$, $U^i := U \cap (B_1^i \cup \cdots \cup B_t^i)$ equals B_j^i for some $j \in [t]$. We say that M is *consistent* with B if $\mathcal{A}_0 \subseteq M$ and for each $j \in [n]$, either $\mathcal{A}_j \subseteq M$ or M matches B_j to itself and matches the neighbors of B_j under $\mathcal{A}_j$ to themselves.

For $i \in [\frac{n}{t}]$, we write M^i to denote the edges of M contained in the vertices of B^i and neighbors of B^i under $\mathcal{A}$. We write M_j^i (and $\mathcal{A}_j^i$) to denote the edges of M^i (and $\mathcal{A}$) corresponding to the vertices of B_j^i and its neighbors under $\mathcal{A}$.

Let U and M be sampled from q and let $\mathcal{D}$ denote the event that U and M are consistent with B and for every $i \in [\frac{n}{t}]$, U^i does not cut the edges of M^i.

Lemma 5.3. *Let R be any random variable satisfying $U - R - M$. Then, for every $i \in [\frac{n}{t}]$ it holds that*

$$\mathbf{I}_q \left(R : U^i | BU^{<i} M^{\geq i} \mathcal{D} \right) + \mathbf{I}_q \left(R : M^i | BU^{\leq i} M^{>i} \mathcal{D} \right) = \Omega(\log t).$$

Using Lemma 4.7 and Proposition 4.2 together with the above, we have

$$2 \log \mathsf{rk}_+(S) \geq \sum_{i=1}^{n/t} \left(\mathbf{I}_q \left(R : U^i | BU^{<i} M^{\geq i} \mathcal{D} \right) + \mathbf{I}_q \left(R : M^i | BU^{\leq i} U^{>i} \mathcal{D} \right) \right) = \Omega \left(\frac{n}{t} \log t \right) = \Omega \left(\frac{1}{\varepsilon} \log(\varepsilon n) \right),$$

which proves Theorem 3.3 as $\binom{n}{s} = 2^{\Theta \left(s \log \left(\frac{n}{s} \right) \right)}$.

Proof of Lemma 5.3. Fix a value of i as in the statement of the lemma. Let $\mathcal{E}$ be the event that U, M are consistent with B and for each $j \neq i$, the edges of M^j are not cut by U^j. Note that $\mathcal{D} \subset \mathcal{E}$, but under $\mathcal{E}$ edges of M^i may be cut by U^i.

For any partition b and for any $(u, m) \in \mathsf{supp}(p(um|b\mathcal{E}))$, the weight of the slack matrix entries are given by

$$S_{um} = \begin{cases} \frac{2+\beta}{2} & \text{when } (u, m) \in \mathcal{D}, \\ \frac{4+\beta}{2} & \text{when } (u, m) \in \overline{\mathcal{D}}, \end{cases} \tag{5.3}$$

and note that the number of entries in both cases above is exactly $t2^{t-1}$.

18

Writing $W = U^{<i}M^{>i}B^{-i}$, we will prove that for any fixed value w attained by W, the following holds

$$\mathbf{I}_q\left(R : U^i | M^i B w \mathcal{D}\right) + \mathbf{I}_q\left(R : M^i | U^i B w \mathcal{D}\right) \geq \beta^8 \log t, \tag{5.4}$$

and the proof is completed by averaging over w. Observe that the partition of the ith chunk B^i and B_0 is still a random variable even after fixing w.

After fixing wrb, U and M are independent and they can be checked separately to verify that the event $\mathcal{E}$ occurs as for every block $j \neq i$ either U^j or M^j is fixed given wb. It follows that the distribution $p(umbr)$ defined as $p(umbr) = q(umbr|w\mathcal{E})$ satisfies $p(um|rb) = p(u|rb)p(m|rb)$. A direct computation using (5.3) then shows that

$$p(\mathcal{D}) = \frac{2+\beta}{(2+\beta)+(4+\beta)} = \frac{2+\beta}{6+2\beta} \geq \frac{1}{3}.$$

For the sake of contradiction, assume that (5.4) does not hold. Then, we are going to argue that $p(\mathcal{D})$ must be significantly smaller than $\frac{1}{3}$ reminiscent to the proof of Lemma 5.1.

Define the random variable $J \in [t]$ to be $J = j$ if $U^i = B^i_j$ and note that the distribution $p(jm|rb)$ is product. Furthermore, note that under the distribution p, the event $\mathcal{D}$ is equivalent to $M^i_J \neq \mathcal{A}^i_J$.

For any $(j, r, b) \in \mathsf{supp}(p(jrb))$, we define the set of blocks correlated with j as

$$\mathcal{J}_{jrb} = \left\{ k \;\middle|\; k \in [t], k \neq j, p(M^i_k \neq \mathcal{A}^i_k | rb, M^i_j \neq \mathcal{A}^i_j) \notin \left(\frac{1}{4}, \frac{3}{4}\right) \right\}.$$

Define sets:

$$\mathcal{S}_1 = \left\{ (j, r, b) \;\middle|\; p(M^i_j \neq \mathcal{A}^i_j | rb) \leq \frac{1}{30} \right\}, \qquad \mathcal{S}_2 = \left\{ (j, r, b) \;\middle|\; p(M^i_j \neq \mathcal{A}^i_j | rb) \geq \frac{1+\beta}{2} \right\},$$

$$\mathcal{S}_3 = \left\{ (j, r, b) \;\middle|\; p(j | rb\mathcal{D}) \geq \frac{2^{1/\beta^2}}{t^{1-\beta^2}} \right\}, \qquad \mathcal{S}_4 = \left\{ (j, r, b) \;\middle|\; |\mathcal{J}_{jrb}| \geq \frac{\beta}{2} \cdot \frac{t^{1-\beta^2}}{2^{1/\beta^2}} - 1 \right\}.$$

For brevity, when $\mathcal{S} \subseteq [t] \times \mathsf{supp}(p(rb))$ we write $p(\mathcal{S})$ to denote the probability of the event $(j, r, b) \in \mathcal{S}$. Using Lemma 2.1 and other entropy related arguments, we will be able to show that the contribution of $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ and $\mathcal{S}_4$ to $p(\mathcal{D})$ is small.

Claim 5.4. *(a)* $p(\mathcal{S}_1, \mathcal{D}) \leq \frac{1}{30}$, *(b)* $p(\mathcal{S}_2, \mathcal{D}) \leq 64\beta$, *(c)* $p(\mathcal{S}_3, \mathcal{D}) \leq 3\beta^2$, *(d)* $p(\mathcal{S}_4, \mathcal{D}) \leq \beta$.

Hence, denoting by $\mathcal{G}$ the complement of the event $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4$, we can bound

$$p(\mathcal{D}) \leq p(\mathcal{G}, \mathcal{D}) + \frac{1}{30} + 64\beta + 3\beta^2 + \beta \leq p(\mathcal{G}, \mathcal{D}) + \frac{1}{30} + 2^7\beta. \tag{5.5}$$

To bound the contribution of $\mathcal{G}$, we need to use the combinatorial structure of matchings. For this, we further split $\mathcal{G}$ into two events. We say $(j, r, b) \in \mathcal{G}_2$ if $(j, r, b) \in \mathcal{G}$ and for all partitions b' such that $(j, r, b') \in \mathcal{G}$ and b' agrees with b on all the blocks except b_0 and b^i_j, it holds that $|b_0 \cap b'_0| \geq 2$. We define $\mathcal{G}_1 = \mathcal{G} \setminus \mathcal{G}_2$. Note that by definition if $(j, r, b) \in \mathcal{G}_1$, then there exists another partition b' such that $(j, r, b') \in \mathcal{G}_1$ and b' agrees with b on all the blocks except b_0 and b^i_j.

As discussed in Section 2.2, most of the contribution comes from the event $\mathcal{G}_2$ which we can bound by

Claim 5.5.
$$p(\mathcal{G}_2, \mathcal{D}) \leq \frac{4}{10} p(\overline{\mathcal{D}}) + \beta = \frac{4}{10} \cdot \frac{4+\beta}{6+2\beta} + \beta.$$

Next we want to relate the contribution of $\mathcal{G}_1$ to the probability of the event $|\delta(U) \cap M| = 1$ under the distribution q. Since, this probability is not zero under the slack matrix S, we will need some more notation to bound it more carefully. We call a matching M to be *bad* for B if M includes exactly one edge crossing B_0 which is an edge of $\mathcal{A}_0$ and for other blocks either $\mathcal{A}_j \subseteq M$ or M matches B_j to itself and matches the neighbors of B_j under $\mathcal{A}_j$ to themselves. See Figure 5.2.
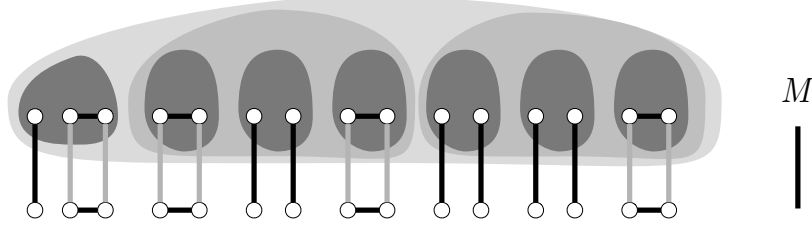

Figure 5.2: A bad matching M

Let $\mathcal{F}$ denote the event that U is consistent with B, M is *bad* for B and for each $j \neq i$, the edges of M^j are not cut by U^j. For any partition b, the entries of the slack matrix when $(u, m) \in \mathsf{supp}(q(um|b\mathcal{F})$ are given by

$$S_{um} = \begin{cases} \frac{2+\beta}{2} & \text{when } |\delta(u) \cap m| = 3 \\ \frac{\beta}{2} & \text{when } |\delta(u) \cap m| = 1. \end{cases} \tag{5.6}$$

Note that the number of entries in the two cases above is exactly $3t2^{t-1}$ since for every fixing of all the edges of M except M_0, there are exactly three matchings that are bad.

With the above, we will be able to show that

Claim 5.6.
$$p(\mathcal{G}_1, \mathcal{D}) \leq 48000 q(|\delta(U) \cap M| = 1 \mid \mathcal{E} \cup \mathcal{F}) + \beta \leq 2^{14}\beta.$$

Plugging the above claims in (5.5) and choosing β to be a sufficiently small constant, we derive that

$$p(\mathcal{D}) \leq \frac{4}{10} \cdot \frac{4+\beta}{6+2\beta} + \frac{1}{30} + 2^7\beta + 2^{14}\beta \leq \frac{9}{30} + 2^{15}\beta,$$

which is a contradiction as $p(\mathcal{D}) = \frac{2+\beta}{6+2\beta} \geq \frac{1}{3}$. This finishes the proof. Next we turn to proving the claims. $\qquad\square$

Proof of Claim 5.5. For any $(j, r, b) \in \mathcal{G}$, we have $p(M_j^i \neq \mathcal{A}_j^i | rb) \leq p(M_j^i = \mathcal{A}_j^i | rb) + \beta$ since

$$p(M_j^i \neq \mathcal{A}_j^i | rb) - p(M_j^i = \mathcal{A}_j^i | rb) = 2p(M_j^i \neq \mathcal{A}_j^i | rb) - 1 \leq 2 \cdot \frac{1+\beta}{2} - 1 = \beta.$$

Recall that $\mathcal{D}$ is equivalent to the event $M_J^i \neq \mathcal{A}_J^i$ under p. As $\mathcal{G}_2 \subset \mathcal{G}$ and $p(jm|rb)$ is product,

$$p(\mathcal{D}, \mathcal{G}_2) = p(M_J^i \neq \mathcal{A}_J^i, \mathcal{G}_2) = \sum_{(j,r,b)\in\mathcal{G}_2} p(jrb)p(M_j^i \neq \mathcal{A}_j^i|rb)$$

$$\leq \sum_{(j,r,b)\in\mathcal{G}_2} p(jrb)p(M_j^i = \mathcal{A}_j^i|rb) + \beta = p(M_J^i = \mathcal{A}_J^i, \mathcal{G}_2) + \beta.$$

Conditioned on the event $J = j, M_j^i = \mathcal{A}_j^i, R = r$, by symmetry the partition B^i is uniform among all partitions which agree on all the blocks except B_0 and B_j^i (recall Figure 2.2). For every fixing of the blocks outside B_0 and B_j^i, there are $\binom{5}{3}$ such partitions out of which at most 4 can be in $\mathcal{G}_2$ by Lemma 4.11. In particular, this means that for any (j, r) we can bound

$$p(\mathcal{G}_2 | M_j^i \neq \mathcal{A}_j^i, J = j, r) = \sum_{b:(j,r,b) \in \mathcal{G}_2} p(b | M_j^i = \mathcal{A}_j^i, J = j, r) \leq \frac{4}{\binom{5}{3}} = \frac{4}{10}$$

Averaging over j and r, we get that $p(\mathcal{G}_2 | M_J^i = \mathcal{A}_J^i) \leq \frac{4}{10}$. Finally,

$$p(\mathcal{G}_2, \mathcal{D}) \leq p(M_J^i = \mathcal{A}_J^i, \mathcal{G}_2) + \beta = p(M_J^i = \mathcal{A}_J^i) p(\mathcal{G}_2 | M_J^i = \mathcal{A}_J^i)$$
$$\leq \frac{4}{10} p(M_J^i = \mathcal{A}_J^i) + \beta = \frac{4}{10} p(\overline{\mathcal{D}}) + \beta.$$

$\square$

Proof of Claim 5.6. For every (r, b) fix an ℓ arbitrarily such that $(\ell, r, b) \in \mathcal{G}_1$ (if such an ℓ does not exist then the contribution of r, b to $p(\mathcal{G}_1, \mathcal{D})$ is zero anyway). Then by definition, there exists a partition b' such that $(\ell, r, b') \in \mathcal{G}_1$, b and b' agree on all the blocks except $b_0 \cup b_\ell^i$ and $|b_0 \cap b_0'| = 1$. Note that b' is determined by r, b. Define $\mathcal{T}_{rb} := \mathcal{J}_{\ell rb} \cup \mathcal{J}_{\ell rb'} \cup \{\ell\}$ and let $\mathcal{T}$ denote the event that $J \in \mathcal{T}_{rb}$. We will show that when $(j, r, b) \in \mathcal{G}_1$ and $j \notin \mathcal{T}_{rb}$:

$$p(J = j, M_j^i \neq \mathcal{A}_j^i, r, b) \leq 48000 q(J = j, M_j^i \neq \mathcal{A}_j^i, M \text{ is bad}, r, b | \mathcal{E} \cup \mathcal{F}). \quad (5.7)$$

The above implies that

$$p(\mathcal{G}_1, \overline{\mathcal{T}}, M_J^i \neq \mathcal{A}_J^i) \leq 48000 q(\mathcal{G}_1, \overline{\mathcal{T}}, M_J^i \neq \mathcal{A}_J^i, M \text{ is bad} | \mathcal{E} \cup \mathcal{F})$$
$$\leq 48000 q(M_J^i \neq \mathcal{A}_J^i, M \text{ is bad} | \mathcal{E} \cup \mathcal{F}). \quad (5.8)$$

Furthermore, since both (ℓ, r, b) and (ℓ, r, b') are in $\mathcal{G}_1 \subseteq \mathcal{G}$, $|\mathcal{T}_{rb}| \leq \beta \cdot \frac{t^{1-\beta^2}}{2^{1/\beta^2}} - 1 \leq \beta \cdot \frac{t^{1-\beta^2}}{2^{1/\beta^2}}$ and also $p(j | rb, \mathcal{D}) \leq \frac{2^{1/\beta^2}}{t^{1-\beta^2}}$ when $(j, r, b) \in \mathcal{G}_1$. So, we have $p(\mathcal{T}, \mathcal{G}_1 | \mathcal{D}) \leq \beta$.

When $M_J^i \neq \mathcal{A}_J^i$ and M is bad, then the cut $U^i = B_0 \cup B_J^i$ satisfies $|\delta(U) \cap M| = 1$ (see Figure 5.3(b)). So, from (5.8) and the above, we get

$$p(\mathcal{G}_1, \mathcal{D}) = p(\mathcal{G}_1, M_J^i \neq \mathcal{A}_J^i) \leq p(\mathcal{G}_1, \overline{\mathcal{T}}, M_J^i \neq \mathcal{A}_J^i) + p(\mathcal{G}_1, \mathcal{T} | \mathcal{D})$$
$$\leq 48000 q(M_J^i \neq \mathcal{A}_J^i, M \text{ is bad} | \mathcal{E} \cup \mathcal{F}) + \beta \leq 48000 q(|\delta(U) \cap M| = 1 \mid \mathcal{E} \cup \mathcal{F}) + \beta.$$

Using (5.3) and (5.6), the first term on the right hand side is $\frac{3\beta}{(2+\beta)+(4+\beta)+3(2+\beta)+3\beta} \leq \frac{\beta}{4}$ and hence $p(\mathcal{G}_1, \mathcal{D}) \leq 2^{14} \beta$.

All that remains is to prove (5.7). Since $(\ell, r, b) \in \mathcal{G}$ and $j \notin \mathcal{T}_{rb}$, $p(M_\ell^i \neq \mathcal{A}_\ell^i | rb) \geq \frac{1}{30}$ and $p(M_j^i \neq \mathcal{A}_j^i | rb, M_\ell^i \neq \mathcal{A}_\ell^i) \geq \frac{1}{4}$, so $p(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i | rb) \geq \frac{1}{120}$. Since probability is always less than one, trivially

$$p(M_j^i \neq \mathcal{A}_j^i | rb) \leq 120 p(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i | rb).$$

Using the above and the fact that $p(jm | rb)$ is product, we get

$$p(J = j, M_j^i \neq \mathcal{A}_j^i, r, b) \leq 120 p(J = j, M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, r, b) \quad (5.9)$$

21

Next we relate the probability on the right hand side above to the probability under q conditioned on the event $\mathcal{E} \cup \mathcal{F}$ as follows:

$$p(J = j, M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, r, b) = \frac{q(J = j, M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E}, r, b \mid \mathcal{E} \cup \mathcal{F})}{q(\mathcal{E} \mid \mathcal{E} \cup \mathcal{F})}$$

$$\leq \frac{20}{6} q(J = j, M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E}, r, b \mid \mathcal{E} \cup \mathcal{F})$$

where the last inequality follows since $q(\mathcal{E} \mid \mathcal{E} \cup \mathcal{F}) = \frac{(4+\beta)+(2+\beta)}{(2+\beta)+(4+\beta)+3(2+\beta)+3\beta} = \frac{6+2\beta}{12+8\beta} \geq \frac{6}{20}$ as $\beta < 1$.

Plugging the above in (5.9),

$$p(J = j, M_j^i \neq \mathcal{A}_j^i, r, b) \leq 400 q(J = j, M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E}, r, b \mid \mathcal{E} \cup \mathcal{F}). \tag{5.10}$$

Note that the distribution $q(jm|rb, \ \mathcal{E} \cup \mathcal{F})$ is still product as J and M are independent given rb and given the partition b one can check if the matching m is bad or consistent without knowing the cut. Next, we relate the probability of the right hand side to the probability under the partition b' where b' is the partition guaranteed from (ℓ, r, b) being in $\mathcal{G}_1$. We will show that

$$q(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E} \mid rb, \ \mathcal{E} \cup \mathcal{F}) \leq 120 q(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E} \mid rb', \mathcal{E} \cup \mathcal{F}). \tag{5.11}$$

Note that the two events above correspond to different matchings since the partitions are different (see Figure 5.3(a)). The probability of any matching that agrees with the event on the right hand side above is zero under $p(m|rb)$, but that is not the case under $q(m|rb, \mathcal{E} \cup \mathcal{F})$. In fact, by symmetry the probability of any such matching is the same under the distribution $q(m|rb, \mathcal{E} \cup \mathcal{F})$ and $q(m|rb', \mathcal{E} \cup \mathcal{F})$. This is because as $q(jm|rb, \mathcal{E} \cup \mathcal{F})$ is product, the probability does not change under conditioning on the event that $J = \ell$ which is the same as conditioning that the cut is $b_0 \cup b_\ell^i$. Then, since the partitions were picked uniformly, both the splits (b_0, b_ℓ^i) and $(b_0', b_\ell'^i)$ are equally likely even when conditioned on $J = \ell, \mathcal{E} \cup \mathcal{F}$ since the matching is either consistent or bad with respect to both of them.

Note that any matching which agrees with the event on the right hand side above is bad for b (see Figure 5.3(b)). From this it follows that

$$q(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E} \mid rb, \ \mathcal{E} \cup \mathcal{F}) \leq 120 q(M_j^i \neq \mathcal{A}_j^i, M \text{ is bad} \mid rb, \mathcal{E} \cup \mathcal{F}).$$

Plugging the above in (5.10) and using that the distribution is product, we get (5.7). To complete the proof, we next show that (5.11) holds.

For this let us recall that since (ℓ, r, b) and (ℓ, r, b') are both in $\mathcal{G}$, when $j \notin \mathcal{T}_{rb}$, we have $p(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i | rb) \geq \frac{1}{120}$ and a similar statement also holds with respect to b'.
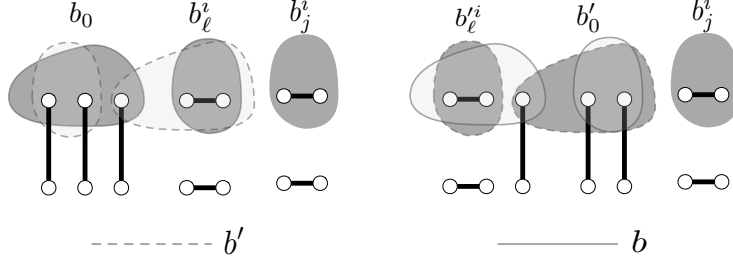
Since $p(m|rb) = q(m|rb, \mathcal{E})$, we can write

$$\frac{p(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i | rb)}{p(M_\ell^i = \mathcal{A}_\ell^i | rb)} = \frac{q(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E} \mid rb, \mathcal{E} \cup \mathcal{F})}{q(M_\ell^i = \mathcal{A}_\ell^i, \mathcal{E} \mid rb', \mathcal{E} \cup \mathcal{F})},$$

and a similar statement holds for b' also.

Similar to the argument before, by symmetry $q(M_\ell^i = \mathcal{A}_\ell^i, \mathcal{E}|rb) = q(M_\ell^i = \mathcal{A}_\ell^i, \mathcal{E}|rb')$ and also $p(M_\ell^i = \mathcal{A}_\ell^i | rb) = p(M_\ell^i = \mathcal{A}_\ell^i | rb')$. It follows that

$$\frac{q(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E} \mid rb, \mathcal{E} \cup \mathcal{F})}{q(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E} \mid rb', \mathcal{E} \cup \mathcal{F})} = \frac{p(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i | rb)}{p(M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i | rb')} \leq 120,$$

(a) Partial Matchings corresponding to the event $M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i$ under partition b(left) and under partition b' (right). Note that b and b' agree in all blocks except b_0 and b_ℓ^i.
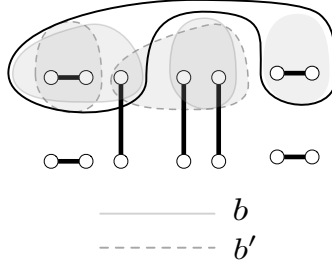
(b) A matching agreeing with the event $M_j^i \neq \mathcal{A}_j^i, M_\ell^i \neq \mathcal{A}_\ell^i, \mathcal{E}$ under partition b' and the cut $U^i = b_0^i \cup b_j^i$. Note that the matching is bad for the partition b.

which proves (5.11). This completes the proof.

$\square$

Proof of Claim 5.4. We have $\mathbf{H}_p(M^i | U^i B \mathcal{D}) \geq t - 1$ since when $U^i = B_j^i$ then M_k^i either equals $\mathcal{A}_k^i$ or is not equal to $\mathcal{A}_k^i$ with probability $\frac{1}{2}$ independently for each $k \neq j$ conditioned on $\mathcal{D}$. Moreover, conditioned on $\mathcal{D}$, U^i is uniform among blocks B_j^i such that $M_j^i \neq \mathcal{A}_j^i$. We may compute from (5.3) that $p(m^i | b\mathcal{D}) = \frac{\Delta(m^i)}{t 2^{t-1}}$ where $\Delta(m^i)$ is the number of blocks k of m^i such that $m_k^i \neq \mathcal{A}_k^i$. It follows that the probability of any m^i with $\Delta(m^i) \leq \frac{t}{4}$ is at most $\frac{1}{2^{t+1}}$ and using Proposition 4.6 (by considering the indicator vector for $M_k^i \neq \mathcal{A}_k^i$ for $k \in [t]$) their total measure under the distribution $p(m^i | b\mathcal{D})$ can be bounded by $\frac{e^{-t/8}}{2}$. Hence, $\mathbf{H}_p(U^i | M^i B \mathcal{D}) \geq (1 - \frac{e^{-t/8}}{2}) \log \left(\frac{t}{4}\right) \geq \log t - 3$.

As $p(umrb) = q(umrb | w\mathcal{E})$, if (5.4) is not true, then

$$\mathbf{I}_p \left(R : U^i | M^i B \mathcal{D} \right) = \mathbf{H}_p(U^i | M^i B \mathcal{D}) - \mathbf{H}_p(U^i | R M^i B \mathcal{D}) \leq \beta^8 \log t,$$

and a similar statement is obtained by writing $\mathbf{I}_p \left(R : M^i | U^i B \mathcal{D} \right)$ in terms of entropy. It follows that

$$\mathbf{H}_p(U^i | R B M^i \mathcal{D}) \geq (1 - \beta^8) \log t - 3, \text{ and } \mathbf{H}_p(M^i | R B U^i \mathcal{D}) \geq t - \beta^8 \log t - 1.$$

In terms of the random variable J (recall that $J = j$ if $U^i = B_j^i$), we get

$$\mathbf{H}_p(J | R B M^i \mathcal{D}) \geq (1 - \beta^8) \log t - 3, \text{ and } \mathbf{H}_p(M^i | J R B \mathcal{D}) \geq t - \beta^8 \log t - 1.$$

For rest of the proof, it will be helpful to keep in mind that under the distribution p, $\mathcal{D}$ is equivalent to the event that $M_J^i \neq \mathcal{A}_J^i$.

(a) Follows from definition of $\mathcal{S}_1$.

(b) In the probability space p, define random variable $Y \in \{0,1\}^t$ as follows: for $k \in [t]$, $Y_k = \mathbb{1}_{M_k^i = \mathcal{A}_k^i}$. Then, it holds that $\mathbf{H}_p(J|RB, Y_J = 0) \geq (1 - \beta^8) \log t - 3$ and $\mathbf{H}_p(Y|RB, Y_J = 0) \geq t - \beta^8 \log t - 3$ and $J - RB - Y$. Applying Lemma 2.1 gives us $p(\mathcal{S}_2, \mathcal{D}) \leq p(\mathcal{S}_2) \leq 64\beta$.

(c) Since removing conditioning only increases entropy, we have that $\mathbf{H}_p(U^i|RB\mathcal{D}) \geq (1 - \beta^8) \log t - 3$. Let $\mathcal{T} = \{(r,b) \mid \mathbf{H}_p(U^i|rb\mathcal{D}) \geq (1 - \beta^4) \log t\}$. As $\mathbf{H}_p(U^i|rb\mathcal{D}) \leq \log t$, Lemma 4.10 then says that $p((r,b) \notin \mathcal{T}|\mathcal{D}) \leq \beta^4 + \frac{3}{\beta^4 \log t} \leq 2\beta^4$ where the last inequality follows as $\beta^8 \log t \geq 3$ when β is a constant and $t \geq c = c(\beta)$ for a sufficiently large constant c.

Lemma 4.8 implies that for any $(r,b) \in \mathcal{T}$, $p(\mathcal{S}_3|rb\mathcal{D}) \leq \beta^2$. By union bound,

$$p(\mathcal{S}_3, \mathcal{D}) \leq p(\mathcal{S}_3|(r,b) \in \mathcal{T}, \mathcal{D}) + p((r,b) \notin \mathcal{T}|\mathcal{D}) \leq \beta^2 + 2\beta^4 \leq 3\beta^2.$$

(d) Set $a = \beta \frac{t^{1-\beta^2}}{2^{1/\beta^2}+1} - 1$. Note that for any $k \in \mathcal{J}_{jrb}$, $\mathbf{H}_p(M_k^i|rb, M_j^i \neq \mathcal{A}_j^i) \leq \mathsf{h}\left(\frac{1}{4}\right) \leq \frac{9}{10}$. When $(j,r,b) \in \mathcal{S}_4$, $|\mathcal{J}_{jrb}| \geq a$ and hence $\mathbf{H}_p(M^i|rb, M_j^i \neq \mathcal{A}_j^i) \leq t - \frac{a}{10}$. Since, $p(jm|rb)$ is product, $\mathbf{H}_p(M^i|rb, J = j, M_j^i \neq \mathcal{A}_j^i) = \mathbf{H}_p(M^i|rb, M_j^i \neq \mathcal{A}_j^i)$ and we can say that

$$\mathbf{H}_p(M^i|JRB\mathcal{D}) = \mathbb{E}_{p(jrb|\mathcal{D})}[\mathbf{H}_p(M^i|rb, M_j^i \neq \mathcal{A}_j^i)] \geq t - \beta^8 \log t - 1.$$

Using Lemma 4.10, we get that $p(\mathcal{S}_4, \mathcal{D}) \leq p(\mathcal{S}_4|\mathcal{D}) \leq \frac{\beta^8 \log t + 1}{a/10} \leq \beta$ as $t \geq c = c(\beta)$ for a large enough c.

$\square$

6 Proof of Main Technical Lemma (Lemma 2.1)

For the sake of contradiction, we assume that $p((x,r) \in \mathcal{B}) \geq 64\gamma$. The proof will proceed by first fixing a value of r, such that the entropies $\mathbf{H}_p(X|r, Y_X = 0)$ and $\mathbf{H}_p(Y|r, Y_X = 0)$ will remain large but the distribution of X and Y conditioned on r will be quite biased (conditioned on r, the probability of the event $Y_X = 0$ will be significantly larger than $\frac{1}{2}$). Then, we will argue that if this was the case, then in fact, the entropy $\mathbf{H}_p(Y|r, Y_X = 0)$ must be much smaller than our assumption.

Let $\mathcal{G}$ denote the set of r such that

(a) $\mathbf{H}_p(X|r, Y_X = 0) \geq (1 - \gamma^4) \log m$ and $\mathbf{H}_p(Y|r, Y_X = 0) \geq m - \gamma^4 \log m$.

(b) $p(x \in \mathcal{S}_r|r) \geq 32\gamma$ where $\mathcal{S}_r = \{x|(x,r) \in \mathcal{B}\}$.

We will be able to argue that

Claim 6.1. $p(r \in \mathcal{G}|Y_X = 0) \geq 256\gamma^2$.

In particular, this means that the set $\mathcal{G}$ is not empty. For the rest of the proof, we will fix some $r \in \mathcal{G}$ and work with the distribution $q(xy) = p(xy|r)$ which is product. Consider the random variable $\mathsf{Bias} = \mathbb{1}_{Y_X=0} - \mathbb{1}_{Y_X=1}$. Note that when $x \in \mathcal{S}_r$,

$$\mathbb{E}_{q(xy|X=x)}[\mathsf{Bias}] = q(Y_x = 0) - q(Y_x = 1) \geq \gamma.$$

We will prove that there exists a *rectangle* with the following properties.

Claim 6.2. *There exists events $\mathcal{S} \subseteq \mathsf{supp}(X)$, $\mathcal{T} \subseteq \mathsf{supp}(Y)$ such that*

(a) $\mathbb{E}_{q(xy|(X,Y)\in\mathcal{S}\times\mathcal{T})}[\mathsf{Bias}] \geq \gamma/2$.

(b) $q(x|X \in \mathcal{S}) \leq \frac{2^{2/\gamma^2}}{m^{1-\gamma^2}}$ *for every* x.

(c) $\mathbf{H}_q(Y|Y \in \mathcal{T}) \geq m - \gamma^2 \log m - \frac{2}{\gamma^2}$.

Lets finish the proof of Lemma 2.1 first. For this, we define

$$\beta_x := q(Y_x = 0|Y \in \mathcal{T}) - q(Y_x = 1|Y \in \mathcal{T}).$$

Since $q(xy|(X,Y) \in \mathcal{S} \times \mathcal{T})$ is a product distribution, Claim 6.2(b) implies that

$$\mathbb{E}_{q(xy|(X,Y)\in\mathcal{S}\times\mathcal{T})}[\mathsf{Bias}] = \sum_x q(x|X \in \mathcal{S})\beta_x \leq \frac{2^{2/\gamma^2}}{m^{1-\gamma^2}} \sum_{x\in\mathcal{S}} \beta_x.$$

So, using Claim 6.2(a), we can say

$$\sum_{x\in\mathcal{S}} \beta_x \geq \frac{\gamma}{2^{2/\gamma^2+1}} m^{1-\gamma^2}, \text{ or equivalently, } \mathbb{E}_{x\in\mathcal{S}}[\beta_x] \geq \frac{\gamma}{2^{2/\gamma^2+1}} \frac{m^{1-\gamma^2}}{|\mathcal{S}|},$$

From Lemma 4.9, it then follows that

$$\mathbf{H}_q(Y|Y \in \mathcal{T}) \leq m - \left(\frac{\gamma}{2^{2/\gamma^2+2}} \frac{m^{1-\gamma^2}}{|\mathcal{S}|}\right)^2 |\mathcal{S}| = m - \frac{\gamma^2}{2^{4/\gamma^2+4}} \frac{m^{2-2\gamma^2}}{|\mathcal{S}|}.$$

As Claim 6.2(b) also implies that $|\mathcal{S}| \geq \frac{1}{2^{2/\gamma^2}} m^{1-\gamma^2}$, we have

$$\mathbf{H}_q(Y|Y \in \mathcal{T}) \leq m - \frac{\gamma^2}{2^{2/\gamma^2+4}} m^{1-\gamma^2},$$

which contradicts Claim 6.2(c) if

$$\frac{\gamma^2}{2^{2/\gamma^2+4}} m^{1-\gamma^2} > \gamma^2 \log m + \frac{2}{\gamma^2}.$$

One can check that if $\frac{3}{\log m} \leq \gamma^8 \leq \frac{1}{2^{64}}$, then the left hand side above is always at least $\Omega(\sqrt{m})$ while the right hand side is $O(\log m)$. This proves Lemma 2.1.

Next we turn to proving Claims 6.1 and 6.2. We will need the following proposition.

Proposition 6.3. *For any events $\mathcal{A}$ and $\mathcal{B}$, $p(\mathcal{A}|\mathcal{B}, Y_X = 0) \geq p(\mathcal{A}|\mathcal{B})p(Y_X = 0|\mathcal{A}\mathcal{B})$.*

Proof of Claim 6.1. Let

$\mathcal{G}_1 = \{r \mid \mathbf{H}_p(X|r, Y_X = 0) \geq (1 - \gamma^4)\log m \text{ and } \mathbf{H}_p(Y|r, Y_X = 0) \geq m - \gamma^4 \log m\}$, and
$\mathcal{G}_2 = \{r \mid p(x \in \mathcal{S}_r|r) \geq 32\gamma\}$.

As $\mathbf{H}_p(X|r, Y_X = 0) \leq \log m$ and $\mathbf{H}_p(Y|r, Y_X = 0) \leq m$ for every r, Lemma 4.10 and union bound imply that $p(r \notin \mathcal{G}_1|Y_X = 0) \leq 2\left(\gamma^4 + \frac{3}{\gamma^4 \log m}\right) \leq 4\gamma^4$ as $\gamma^8 \geq \frac{3}{\log m}$.

Lemma 4.10 implies that $p(r \in \mathcal{G}_2) \geq 32\gamma$. Furthermore, for any $r \in \mathcal{G}_2$, we have

$$p(Y_X = 0|r) \geq p(x \in \mathcal{S}_r|r)p(Y_X = 0|r, X \in \mathcal{S}_r) \geq 16\gamma.$$

Since $\gamma \leq \frac{1}{2^8}$, using Proposition 6.3,

$$p(r \in \mathcal{G}|Y_X = 0) \geq p(r \in \mathcal{G}_2|Y_X = 0) - p(r \notin \mathcal{G}_1|Y_X = 0)$$
$$\geq p(r \in \mathcal{G}_2) \cdot 16\gamma - p(r \notin \mathcal{G}_1|Y_X = 0) \geq 512\gamma^2 - 4\gamma^4 \geq 256\gamma^2.$$

$\square$

Proof of Claim 6.2. By our choice of $r \in \mathcal{G}$, we have $q(x \in \mathcal{S}_r) \geq 32\gamma$ where $\mathcal{S}_r = \{x|q(Y_x = 0) \geq \frac{1+\gamma}{2}\}$. This also implies:

$$q(Y_X = 0) \geq q(X \in \mathcal{S}_r)q(Y_X = 0|X \in \mathcal{S}_r) \geq 16\gamma. \tag{6.1}$$

We define

$$\mathcal{S}_u = \left\{x \mid q(x|Y_X = 0) \leq \frac{m^{\gamma^2}2^{1/\gamma^2}}{m}\right\} \text{ and } \mathcal{S} = \mathcal{S}_r \cap \mathcal{S}_u,$$

$$\mathcal{T}_b = \left\{y \mid \mathbb{E}_{q(xy|Y=y,X\in\mathcal{S})}[\mathsf{Bias}] \geq \frac{\gamma}{2}\right\}, \mathcal{T}_u = \left\{y \mid q(y|Y_X = 0) \leq \frac{m^{\gamma^2}2^{1/\gamma^2}}{2m}\right\} \text{ and } \mathcal{T} = \mathcal{T}_b \cap \mathcal{T}_u.$$

By definition, we have $\mathbb{E}_{q(xy|(X,Y)\in\mathcal{S}\times\mathcal{T})}[\mathsf{Bias}] \geq \frac{\gamma}{2}$ which establishes (a).

We shall prove the following propositions.

Proposition 6.4. *For $x \in \mathcal{S}_r$, we have $q(x) \leq 2q(x|Y_X = 0)$. Also, $q(x \in \mathcal{S}|Y_X = 0) \geq 8\gamma$.*

Proposition 6.5. *For $y \in \mathcal{T}_b$, we have $q(y) \leq \frac{1}{4\gamma}q(y|Y_X = 0)$. Also, $q(y \in \mathcal{T}|Y_X = 0) \geq \gamma^2$.*

Proposition 6.4 and (6.1) then implies, $q(x \in \mathcal{S}) \geq q(Y_X = 0)q(x \in \mathcal{S}|Y_X = 0) \geq 128\gamma^2$, and since $\gamma \leq \frac{1}{2^8}$, we get that

$$q(x|X \in \mathcal{S}) = \frac{q(x)}{q(x \in \mathcal{S})} \leq \frac{2q(x|Y_X = 0)}{q(x \in \mathcal{S})} \leq \frac{m^{\gamma^2}2^{1/\gamma^2}}{64\gamma^2 m} \leq \frac{2^{2/\gamma^2}}{m^{1-\gamma^2}}.$$

holds for every $x \in \mathcal{S}$ which proves (b).

With a similar argument, using Proposition 6.5 we get that for $y \in \mathcal{T}$, the following holds

$$q(y|Y \in \mathcal{T}) \leq \frac{m^{\gamma^2}2^{1/\gamma^2}}{64\gamma^4 2m} \leq \frac{m^{\gamma^2}2^{2/\gamma^2}}{2m}.$$

which implies that $\mathbf{H}_q(Y|Y \in \mathcal{T}) \geq m - \gamma^2\log m - \frac{2}{\gamma^2}$ which gives (c). $\square$

Now we turn to proving Propositions 6.4 and 6.5.

Proof of Proposition 6.4. Recall that $q(Y_x = 0) \geq \frac{1+\gamma}{2}$ for $x \in \mathcal{S}_r$ and $q(x \in \mathcal{S}_r) \geq 32\gamma$. Since $q(xy) = p(xy|r)$, using Proposition 6.3, it follows that $q(x) \leq 2q(x|Y_X = 0)$ for $x \in \mathcal{S}_r$ and hence $q(x \in \mathcal{S}_r|Y_X = 0) \geq 16\gamma$.

The entropy bound $\mathbf{H}_q(X|Y_X = 0) \geq (1 - \gamma^4) \log m$ implies that $q(x \in \mathcal{S}_u|Y_X = 0) \geq 1 - \gamma^2$ using Lemma 4.8.

Since $\gamma \leq \frac{1}{2^8}$, we get

$$q(x \in \mathcal{S}|Y_X = 0) \geq q(x \in \mathcal{S}_r|Y_X = 0) - q(x \in \overline{\mathcal{S}}_u|Y_X = 0) \geq 16\gamma - \gamma^2 \geq 8\gamma.$$

$\square$

Proof of Proposition 6.5. Note that $\mathbb{E}_{q(xy|Y=y,X\in\mathcal{S})}[\mathsf{Bias}] \geq \gamma/2$ is equivalent to saying that $q(Y_X = 0|y, X \in \mathcal{S}) \geq \frac{1}{2} + \frac{\gamma}{4}$. As $q(xy) = p(xy|r)$ using Proposition 6.3, it follows that $q(y|X \in \mathcal{S}) \leq 2q(y|X \in \mathcal{S}, Y_X = 0)$ for $y \in \mathcal{T}_b$. Since $q(xy)$ is product, Bayes rule and Proposition 6.4 imply that for $y \in \mathcal{T}_b$, the following holds

$$q(y) = q(y|X \in \mathcal{S}) \leq 2q(y|X \in \mathcal{S}, Y_X = 0)$$
$$= \frac{2q(y|Y_X = 0)q(x \in \mathcal{S}|y, Y_X = 0)}{q(x \in \mathcal{S}|Y_X = 0)} \leq \frac{1}{4\gamma}q(y|Y_X = 0).$$

By definition of $\mathcal{S}$, $p(Y_X = 0|X \in \mathcal{S}) \geq \frac{1}{2} + \frac{\gamma}{2}$, or equivalently, $\mathbb{E}_{q(xy|X\in\mathcal{S})}[\mathsf{Bias}] \geq \gamma$. Lemma 4.10 then implies that $q(y \in \mathcal{T}_b) \geq \gamma/2$ and hence $q(y \in \mathcal{T}_b|Y_X = 0) \geq 4\gamma q(y \in \mathcal{T}_b) \geq 2\gamma^2$.

Also, since $\mathbf{H}_q(Y|Y_X = 0) \geq m - \gamma^4 \log m$, using Lemma 4.8 we have $q(y \in \mathcal{T}_u|Y_X = 0) \geq 1 - \gamma^2$. Hence, we can conclude

$$q(y \in \mathcal{T}|Y_X = 0) \geq q(y \in \mathcal{T}_b|Y_X = 0) - q(x \in \overline{\mathcal{T}}_u|Y_X = 0) \geq \gamma^2.$$

$\square$

7   Acknowledgments

References

[AIP06]   Alexandr Andoni, Piotr Indyk, and Mihai Patrascu. On the optimality of the dimensionality reduction method. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '06, pages 449–458, Washington, DC, USA, 2006. IEEE Computer Society.

[AT13]   David Avis and Hans Raj Tiwary. On the Extension Complexity of Combinatorial Polytopes. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part I*, pages 57–68, 2013.

[BFPS15]   Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer. Approximation Limits of Linear Programs (Beyond Hierarchies). *Math. Oper. Res.*, 40(3):756–772, 2015.

[Bie08]   Daniel Bienstock. Approximate formulations for 0-1 knapsack sets. *Oper. Res. Lett.*, 36(3):317–320, 2008.

[BM13]   Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 161–170, 2013.

[BP15]   Gábor Braun and Sebastian Pokutta. The Matching Problem Has No Fully Polynomial Size Linear Programming Relaxation Schemes. *IEEE Trans. Information Theory*, 61(10):5754–5764, 2015.

[BP16]   Gábor Braun and Sebastian Pokutta. Common Information and Unique Disjointness. *Algorithmica*, 76(3):597–629, 2016.

[BR11]   Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011.

[CCZ10]   Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. Extended formulations in combinatorial optimization. *4OR*, 8(1):1–48, Mar 2010.

[CLRS13]   Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate Constraint Satisfaction Requires Large LP Relaxations. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 350–359, 2013.

[CT06]   Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[Edm65]   Jack Edmonds. Maximum Matching and a Polyhedron with $0, 1$ Vertices. *J. of Res. the Nat. Bureau of Standards*, 69 B:125–130, 1965.

[FMP+15]   Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential Lower bounds for Polytopes in Combinatorial Optimization. *J. ACM*, 62(2):17, 2015.

[GJW16]   Mika Göös, Rahul Jain, and Thomas Watson. Extension Complexity of Independent Set Polytopes. *CoRR*, abs/1604.07062, 2016.

[Hås96]   Johan Håstad. Clique is Hard to Approximate Within $n^{1-\varepsilon}$. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 627–636, 1996.

[KMR17]   Pravesh K. Kothari, Raghu Meka, and Prasad Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of csps. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 590–603, 2017.

[LRS15]   James R. Lee, Prasad Raghavendra, and David Steurer. Lower Bounds on the Size of Semidefinite Programming Relaxations. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 567–576, 2015.

[MNSW98] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57(1):37 – 49, 1998.

[NR15] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. How to Compress Asymmetric Communication. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 102–123, 2015.

[Pat11] Mihai Patrascu. Unifying the landscape of cell-probe lower bounds. *SIAM J. Comput.*, 40(3):827–847, 2011.

[PV13] Sebastian Pokutta and Mathieu Van Vyve. A note on the extension complexity of the knapsack polytope. *Oper. Res. Lett.*, 41(4):347–350, 2013.

[Rot14] Thomas Rothvoß. The matching polytope has exponential extension complexity. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 263–272, 2014.

[Wil84] Richard M. Wilson. The exact bound in the Erdős-Ko-Rado theorem. *Combinatorica*, 4(2):247–257, 1984.

[Yan91] Mihalis Yannakakis. Expressing Combinatorial Optimization Problems by Linear Programs. *J. Comput. Syst. Sci.*, 43(3):441–466, 1991.

A   Approximating the Matching Polytope

We show that the polytope $Q_\varepsilon(n)$ is a $(1 + \varepsilon)$ approximation for the matching polytope.

Claim A.1. $P_M(n) \subseteq Q_\varepsilon(n) \subseteq (1 + \varepsilon)P_M(n)$.

Proof. First inclusion is trivial since any vector $x \in P_M(n)$ is also in $Q_\varepsilon(n)$. To see the second inclusion, let $x \in Q_\varepsilon(n)$ be an arbitrary vector. We only need to show that for any odd set $U \subseteq [n]$, $|U| > \frac{1+\varepsilon}{\varepsilon}$, $x(E(U)) \leq (1 + \varepsilon)\frac{|U|-1}{2}$ since the other inequalities are already satisfied. Note that $x(E(U)) \leq \frac{1}{2}\sum_{v \in U} x(\delta(v)) \leq |U|/2 \leq (1 + \varepsilon)\frac{|U|-1}{2}$ where the last inequality holds when $|U| \leq \frac{1+\varepsilon}{\varepsilon}$. $\qquad\square$

B   Proofs of Preliminary Propositions and Lemmas

Proof of Proposition 4.1. Define the function $\mathsf{g}(x) = 1 - 2\log e\left(x - \frac{1}{2}\right)^2 - \mathsf{h}(x)$ for $x \in [0, 1]$. Then, the first and second derivatives of $\mathsf{g}(x)$ for $x \in [0, 1]$ are

$$\mathsf{g}'(x) = -4\log e\left(x - \frac{1}{2}\right) - \log(1 - x) - \log x \quad \text{and} \quad \mathsf{g}''(x) = -4\log e + \frac{\log e}{x(1 - x)}.$$

As $x(1 - x) \leq \frac{1}{4}$ for $x \in [0, 1]$, $\mathsf{g}''(x) \geq 0$ for $x \in [0, 1]$. It follows that the function g is convex on the interval $[0, 1]$ and hence has a unique minima. Furthermore, the derivative $\mathsf{g}'(\frac{1}{2}) = 0$, so the minimum value of $\mathsf{g}(x)$ is attained at $x = \frac{1}{2}$. Hence, $\mathsf{g}(x) \geq \mathsf{g}(\frac{1}{2}) = 0$. $\qquad\square$

Proof of Proposition 4.2. If $\mathsf{rk}_+(A) = r$, then $A = \sum_{i=1}^{r} u_i^T v_i$ where $u_i \in \mathbb{R}_{\geq 0}^{\mathcal{X}}$ and $v_i \in \mathbb{R}_{\geq 0}^{\mathcal{Y}}$ are non-negative (column) vectors. Then, $p(xy)$ can be expressed as a convex combination of r product distributions by setting

$$p(xy|r) = \frac{u_r(x)v_r(y)}{\sum_{x',y'} u_r(x')v_r(y')} \text{ and } p(r) = \frac{\sum_{x',y'} u_r(x')v_r(y')}{\sum_{x',y'} M_{x'y'}}.$$

$\square$

Proof of Lemma 4.7. Using the chain rule,

$$\sum_{i=1}^{n} \mathbf{I}\left(X_i : R|X_{<i}Y_{\geq i}\right) \leq \sum_{i=1}^{n} \mathbf{I}\left(X_i Y_{<i} : R|X_{<i}Y_{\geq i}\right) = \sum_{i=1}^{n} \mathbf{I}\left(X_i : Y_{<i}|X_{<i}Y_{\geq i}\right) + \sum_{i=1}^{n} \mathbf{I}\left(X_i : R|X_{<i}Y\right)$$

$$= \sum_{i=1}^{n} \mathbf{I}\left(X_i : R|X_{<i}Y\right) = \mathbf{I}\left(X : R|Y\right).$$

where the second last equality follows since $\mathbf{I}\left(X_i : Y_{<i}|X_{<i}Y_{\geq i}\right) = 0$. The second bound follows similarly. $\square$

Proof of Lemma 4.11. In case $|\cup_{\mathcal{F} \in \mathfrak{F}} \mathcal{F}| \leq 4$, the size of the family $\mathfrak{F}$ is bounded by $\binom{4}{3} = 4$. We will prove that if $|\cup_{\mathcal{F} \in \mathfrak{F}} \mathcal{F}| = 5$, then size of $\mathfrak{F}$ can at most be 3. Take sets $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ such that $|\mathcal{F}_1 \cup \mathcal{F}_2| = 4$ while $|\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3| = 5$. Then, $\mathcal{F}_1 \cap \mathcal{F}_2 \subseteq \mathcal{F}_3$. But now there can be no more sets in the family because such a set must intersect $\mathcal{F}_1 \cap \mathcal{F}_2$ in just one element but in this case it will intersect with one of $\mathcal{F}_1, \mathcal{F}_2$ or $\mathcal{F}_3$ in exactly one element. $\square$