Electronic Colloquium on Computational Complexity, Report No. 158 (2013)

Information-theoretic approximations of the nonnegative rank

Gábor Braun[1], Rahul Jain[2], Troy Lee[3], and Sebastian Pokutta[1]

[1]ISyE, Georgia Institute of Technology, Atlanta, GA 30332, USA. *Email:* gabor.braun@isye.gatech.edu, sebastian.pokutta@isye.gatech.edu
[2]National University of Singapore and Centre for Quantum Technologies *Email:* rahul@comp.nus.edu.sg
[3]Nanyang Technological University and Centre for Quantum Technologies *Email:* troyjlee@gmail.com

November 18, 2013

Abstract

Common information was introduced by Wyner [1975] as a measure of dependence of two random variables. This measure has been recently resurrected as a lower bound on the logarithm of the nonnegative rank of a nonnegative matrix in Jain et al. [2013], Braun and Pokutta [2013]. Lower bounds on nonnegative rank have important applications to several areas such as communication complexity and combinatorial optimization.

We begin a systematic study of common information extending the dual characterization of Witsenhausen [1976]. Our main results are: (i) Common information is additive under tensoring of matrices. (ii) It characterizes the (logarithm of the) amortized nonnegative rank of matrix, i.e., the minimal nonnegative rank under tensoring and small ℓ_1 perturbations. We provide quantitative bounds compared to an analogous asymptotic result by Wyner [1975]. (iii) We deliver explicit witnesses from the dual problem for several matrices leading to explicit lower bounds on common information, which are robust under ℓ_1 perturbations. This includes improved lower bounds for perturbations of the all important unique disjointness partial matrix, as well as new insights into its information-theoretic structure.

ISSN 1433-8092

1 Introduction

Nonnegative matrix factorizations play a crucial role in many disciplines of theoretical computer science and discrete mathematics, including machine learning, communication complexity, and combinatorial optimization. While for machine learning one is often interested in finding a factorization, for communication complexity and combinatorial optimization it often suffices to study the nonnegative rank, i.e., the minimal *size* of a nonnegative factorization. The *nonnegative rank* of a nonnegative matrix M is the smallest r such that M can be written as $M = \sum_{i=1}^{r} u_i v_i^\top$ for $u_i, v_i \geq 0$.

In communication complexity, the logarithm of the nonnegative rank of M provides a lower bound on the deterministic communication complexity of M, which is polynomially tight by Lovász [1990]. In combinatorial optimization, the nonnegative rank of the slack matrix of a polytope P characterizes the linear extension complexity of P, that is the minimum number of facets of a larger dimensional polytope Q that projects linearly to P.

Thus in both fields, it is of great interest to lower bound the nonnegative rank. Unfortunately, lower bounding the nonnegative rank is both conceptually and computationally hard—in fact, computing the nonnegative rank is known to be NP-hard by Vavasis [2009] (see Moitra [2012] for recent positive results on computing the nonnegative rank).

Most existing lower bounds on the nonnegative rank argue only about the support of the matrix, i.e., the zero/nonzero pattern of the matrix (for an interesting exception see the norm based bounds in Fawzi and Parrilo [2012]). Notice that zeros provide a strong constraint on a nonnegative factorization as if $M(x, y) = 0$ then in every factor $uv^\top$ either $u(x) = 0$ or $v(y) = 0$. The most commonly used lower bound on the nonnegative rank is the rectangle covering bound, which also characterizes nondeterministic communication complexity, a bound suggested in the landmark paper of Yannakakis [1991] connecting nonnegative rank and extension complexity. Rectangle covering arguments can show strong lower bounds in interesting cases, for example, for the unique disjointness partial matrix UDISJ with rows and columns labeled by n-bit strings where $\text{UDISJ}(x, y) = 1$ if $x \cap y = \emptyset$ and $\text{UDISJ}(x, y) = 0$ if $|x \cap y| = 1$ and UDISJ is undefined otherwise. Using arguments from the randomized communication complexity lower bound of Razborov [1992], Wolf [2003] showed lower lower bounds exponential in n on the rectangle covering bound of UDISJ. This bound, in turn, played a key role in the exponential lower bounds on the extension complexity of the Traveling Salesman (TSP) polytope in Fiorini et al. [2012].

Support based bounds have obvious shortcomings: they completely ignore the actual values of the nonzero entries. Thus they are useless for matrices with no zero entries. Exactly this case arises when showing lower bounds on the extension complexity of a polytope that *approximates* a polytope P (see Braun et al. [2012], Braverman and Moitra [2012]). Even for a matrix with zero entries, support based bounds cannot say anything about the nonnegative rank under small perturbations.

It is often the case that optimization problems become easier when a discrete objective function is replaced by a continuous proxy function. This is the approach taken in the information-theoretic framework for nonnegative rank lower bounds initiated by Braverman and Moitra [2012] and further developed by Braun and Pokutta [2013]. We extend this framework to obtain a strong information-theoretic tool to lower bound the nonnegative rank of matrices and partial matrices (such as the UDISJ partial matrix). At the core of our techniques is the information-theoretic notion of *common information*. Common information was introduced in Wyner [1975], and further developed by Witsenhausen [1976] who provided a convex geometry approach to lower bound the common information. Little has been written about the common information outside of these early papers, but it turns out to be the correct notion to capture nonnegative matrix factorization from an information-theoretic point of view. We take it out of the setting of (asymptotic) information theory and turn it into a quantitative tool to lower bound the nonnegative rank of a matrix.

Contribution

Our contribution is threefold: besides extending the dual approach, we apply it to derive not only theoretical properties of common information, but also practical lower bounds with application to concrete matrices.

Common information as amortized log nonnegative rank A relaxed notion, even if not capturing a quantity exactly, can sometimes characterize it in an amortized fashion. Examples are the fractional rectangle covering bound characterizing the amortized rectangle covering bound (see Karchmer et al. [1995]) and information cost characterizing amortized communication (see Braverman and Rao [2011]). We similarly prove that common information is the amortized log nonnegative rank. An asymptotic, qualitative version was already included in Wyner [1975], however, we also establish rate of convergence and provide actual approximations. We give an explicit compression result in Theorem 4.1, stating roughly $\lim_{\ell\to\infty,\delta\to 0,\varepsilon\to 0} \frac{\log \mathrm{rk}_+ M_{\varepsilon,\delta,\ell}}{\ell} = \mathbb{C}\left[M\right]$ where $M_{\varepsilon,\delta,\ell} \approx M^{\otimes \ell}$ and the number of required copies ℓ to obtain an approximation with (total relative) error at most δ and ε deviation from $\mathbb{C}\left[M\right]$ is roughly $\Omega\left(\frac{\log^2(mn/\varepsilon)}{\varepsilon^2 \mathbb{C}[M]^2}\right) \cdot \ln(\delta^{-1})$. From this we also obtain that common information is the limit superior of all measures lower bounding the log nonnegative rank under natural conditions (see Corollary 4.2). Our proof is inspired by a result of Jain et al. [2013] that bounds the nonnegative rank of an approximation of a single matrix (i.e., in the nontensored setting) in terms of common information.

Lower bound of common information via dual programs We extend the framework in Witsenhausen [1976] to obtain strong lower bounds on common information not only of matrices but also of partial matrices using witnesses (i.e., dual certificates). Dual witnesses are central to the behavior of common information under various perturbations of the matrix, e.g., provide an explicit degree of continuity of common information (see Lemma 5.7) for full matrices. We give an example that common information of partial matrices is not continuous in general.

New lower bounds for (U)DISJ As an example of the dual approach to partial matrices, we improve lower bounds from Braun and Pokutta [2013] on the conditional common information of the UDISJ partial matrix under perturbations (see Corollary 6.6), closing the gap between the exact and the approximate case. Moreover, we obtain bounds under arbitrary perturbations as long as the total variation is not too large. Finally, following Kaibel and Weltge [2013], we provide a new lower bound on the conditional common information of UDISJ of $n \log 3/2$ (see Theorem 6.2) under a non-direct sum disjointness conditional indicating that breaking direct sums is necessary for obtaining the optimal estimation.

2 Preliminaries

We introduce the notation and review the information-theoretic background that will be used in the sequel; see [Cover and Thomas, 2006, §2] for an in-depth treatment.

We use $\log x$ for the base 2 logarithm and $\ln x$ for the natural logarithm. We use the shorthand $[n] := \{1, \dots, n\}$.

The entropy of a discrete probability distribution P is roughly the expected number of bits needed to encode P.

Definition 2.1 (Entropy)**.** Let P be a discrete probability distribution. The *entropy* of P is

$$\mathbb{H}\left[P\right] := \sum_x P(x) \log \frac{1}{P(x)}.$$

If P is a Bernoulli probability distribution over $\{0, 1\}$ where $p = P(x = 1)$ and $1 - p = P(x = 0)$ we use the notational shorthand $\widetilde{\mathbb{H}}[p] := -p \log p - (1-p)\log(1-p)$ for the *binary entropy* of p. Further, we will use $\widetilde{\mathbb{H}}[X = 0 \mid \ldots]$ to denote $\widetilde{\mathbb{H}}[\mathbb{P}[X = 0 \mid \ldots]]$. For estimating the entropy, the following alternative forms of the well-known inequality $\ln x \leq x - 1$ will be useful:

$$\log ex \leq x \log e, \qquad\qquad \widetilde{\mathbb{H}}[p] \leq p \log(e/p).$$

The second one follows by substituting $x = 1/(1-p)$.

Definition 2.2 (Conditional Entropy)**.** The *conditional entropy* of P conditioned on Q is

$$\mathbb{H}[P \mid Q] = \mathbb{E}_{x \sim Q}\left[\mathbb{H}[P \mid Q = x]\right].$$

We are ready to define mutual information, the key quantity behind common information.

Definition 2.3 (Conditional Mutual Information)**.** The *conditional mutual information* between P and Q given R is $\mathbb{I}[P; Q \mid R] = \mathbb{H}[P \mid R] - \mathbb{H}[P \mid Q, R]$.

Note that mutual information is symmetric: $\mathbb{I}[P; Q \mid R] = \mathbb{I}[Q; P \mid R]$.

2.1 Common information

In this section we will recall the basic properties of common information with a view towards nonnegative factorizations. For $M \in \mathbb{R}^{m \times n}$ a nonnegative matrix, its *induced distribution* on the row/column joint random variable (A, B) is defined by $\mathbb{P}[A = a, B = b] = \frac{M(a,b)}{\sum_{x \in [m], y \in [n]} M(x,y)}$ for all $a \in [m], b \in [n]$. We call a discrete random variable Π a *seed* for A, B (or M) if A, B are independent given Π. For Π coming from a factorization, that is $M = \sum_{\pi \in \Pi} M_\pi$, this is the case if and only if all the factors M_π have rank at most 1. Conversely, every seed with finite range comes from a factorization.

Every nonnegative factorization $M = \sum_{\pi \in \Pi} M_\pi$ refines the distribution (A, B) as

$$q_M(a, b, \pi) = \mathbb{P}[A = a, B = b, \Pi = \pi] := \frac{M_\pi(a,b)}{\sum_{x,y} M(x,y)}.$$

We shall use the shorthand q_M for this distribution.

Definition 2.4 (Common information)**.** Let M be a nonnegative matrix and let A, B be the row and column variable in the induced distribution. Then the *common information of A, B (or M)* is defined as

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

where $\mathbb{W}[M] = \mathbb{W}[A; B] := \sup_{\Pi \text{ seed for } A, B} \mathbb{H}[A, B \mid \Pi] = \sup_{\Pi \text{ seed for } A, B} \mathbb{H}[A \mid \Pi] + \mathbb{H}[B \mid \Pi]$ is the *private information of A, B (or M)*.

Similarly to $\mathbb{C}[M]$, in the following we will also use the shorthand $\mathbb{I}[M; \Pi]$ for $\mathbb{I}[A, B; \Pi]$, and $\mathbb{H}[M]$ for $\mathbb{H}[A, B]$. We recall the following easy facts about common information (see e.g., Wyner [1975], Witsenhausen [1976], Jain et al. [2013], Braun and Pokutta [2013]).

Fact 2.5. Let M be a nonnegative matrix and let A, B be the row and column variable in the induced distribution. Then

1. *General bounds:* $\mathbb{I}[A; B] \leq \mathbb{C}[A, B] \leq \min\{\mathbb{H}[A], \mathbb{H}[B]\}$
2. *Infimum achieved and Π has small domain:* The infimum in the definition of common information is achieved by a Π with $|\Pi| \leq mn$.
3. *Bounds nonnegative rank:* $\mathbb{C}[M] \leq \mathbb{H}[\Pi] \leq \log \mathrm{rk}_+ M$, where Π is realizer of the infimum.

3 Comparison of common information with other bounds

In this section, we compare common information with the rectangle covering bound and also with information cost, a similar quantity in communication complexity.

For a matrix M, let $\mathrm{supp}(M)$ be the boolean matrix which is zero wherever M is zero and one wherever M is nonzero. Yannakakis [1991] observed that the rectangle covering bound of the support of a matrix M is a lower bound on the nonnegative rank of M, and this technique has been the source of many nonnegative rank lower bounds. We now see that common information is incomparable with the logarithm of the rectangle covering bound, even for a boolean matrix, as the following examples show. In fact, they show that common information is also incomparable with the logarithm of the fractional rectangle bound, defined below.

For a matrix $M \in \{0,1\}^{m\times n}$ its *rectangle covering bound* is, by definition, the minimum number of 1-monochromatic combinatorial rectangles (i.e., submatrices with all entries being 1) needed to cover the 1 entries of M. Let A be a matrix with rows indexed by $(i,j) \in [m]\times[n]$ and columns indexed by 1-monochromatic rectangles of M and let c denote the number of columns of M. Then the rectangle bound is the optimal value of the following integer optimization problem.

$$\begin{aligned}\mathrm{rc}\,(M) = \min\ & \mathbb{1}_{1,c}\, x\\ & Ax \geq \mathbb{1}_{mn,1}\\ & x \in \{0,1\}^c\end{aligned}$$

The fractional rectangle covering bound is obtained by relaxing this integer program to a linear program.

$$\begin{aligned}\mathrm{frc}\,(M) = \min\ & \mathbb{1}_{1,c}\, x\\ & Ax \geq \mathbb{1}_{mn,1}\\ & x \geq 0\end{aligned}$$

Clearly $\mathrm{frc}\,(M) \leq \mathrm{rc}\,(M)$ and by Lovász [1975] it follows that $\mathrm{rc}\,(M) = O(\mathrm{frc}\,(M)\log(mn))$

Lemma 3.1 ($\log \mathrm{rc}\,(\cdot) \not\geq \mathbb{C}\,[\cdot]$, $\mathbb{C}\,[\cdot] \not\geq \log \mathrm{frc}\,(\cdot)$ and hence $\mathbb{C}\,[\cdot] \not\geq \log \mathrm{rc}\,(\cdot)$)**.** *Let*

$$M := \begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}, \qquad N := \begin{pmatrix}1 & 1 & 0\\ 1 & 1 & 1\\ 0 & 1 & 1\end{pmatrix}.$$

Then for all $n \geq 1$ *we have* $\mathbb{C}\,[M^{\otimes n}] = 2/3 \cdot n < \log \mathrm{frc}\,(M^{\otimes n}) = n$ *and*

$$\mathbb{C}\,[N^{\otimes n}] \geq (\log 7 - 1.79115)n \approx 1.01621 \cdot n > n \geq \log \mathrm{rc}\,(N^{\otimes n}) \geq \log \mathrm{frc}\,(N^{\otimes n})\,.$$

Proof of Lemma 3.1. The matrix M has both rectangle covering bound and fractional rectangle covering bound 2 as $(1,2),(2,1)$ is a fooling set. Thus $\log \mathrm{frc}\,(M) = 1$. Moreover, as shown in Karchmer et al. [1995], we have that $\mathrm{frc}\,(\cdot)$ tensors, so we get $n \geq \log \mathrm{rc}\,(M^{\otimes n}) \geq \log \mathrm{frc}\,(M^{\otimes n}) = n$. On the other hand, $\mathbb{C}\,[M] = 2/3$ by [Witsenhausen, 1976, Theorem 7], hence $\mathbb{C}\,[M^{\otimes n}] = 2/3 \cdot n$ by Lemma 5.5.

It remains to show the statement for N. Clearly $\mathrm{rc}\,(N) = 2$, hence $\log \mathrm{rc}\,(N^{\otimes n}) \leq n$. As $\mathbb{C}\,[N^{\otimes n}] = n\mathbb{C}\,[N]$ by Lemma 5.5, it is enough to prove the lower bound on $\mathbb{C}\,[N^{\otimes n}]$ for $n = 1$.

We establish a lower bound on the common information of N by means of (4). We consider

$$\sup_{\substack{p,q\geq 0\\ \|p\|_1=\|q\|_1=1}} \mathbb{H}\,[p] + \mathbb{H}\,[q] - q^T\Lambda p + \mathrm{Tr}[\Lambda N], \tag{1}$$

which is an upper bound on the private information for any Λ. We will use a Λ determined by numerical optimization:

$$\Lambda := \begin{pmatrix} 1/2 & 0 & \infty \\ 0 & 2.7245 & 0 \\ \infty & 0 & 1/2 \end{pmatrix}.$$

To be precise, we put large negative values instead of ∞, and we consider the limit of (1), when these values tend to ∞. Note that as p, q are chosen from a compact set, the maximizers will have an accumulation point $\tilde{p}, \tilde{q}$ with 0 entries in $\tilde{p}\tilde{q}^T$ at the ∞ entries of Λ. Thus we obtain a lower bound

$$\sup_{\substack{p,q\geq 0 \\ \|p\|_1=\|q\|_1=1 \, p_j q_i = 0 \text{ if } \Lambda_{i,j}=\infty}} \mathbb{H}[p] + \mathbb{H}[q] - q^T\Lambda p + \mathrm{Tr}[\Lambda N].$$

The matrix Λ has been chosen ensuring that

$$f(p,q) = \mathbb{H}[p] + \mathbb{H}[q] - q^T\Lambda p + \mathrm{Tr}[\Lambda N]$$

is piece-wise concave in p, q. The maximal value for (1) is 1.79115 realized by the rank-1 matrix

$$pq^T := \begin{pmatrix} 0.622036 \\ 0.377964 \\ 0 \end{pmatrix} \begin{pmatrix} 0.622036 & 0.377964 & 0 \end{pmatrix}.$$

(Note that the above are *not* simply numerical approximations of $2/3$ and $1/3$ as they lead to a value for (1) of 1.77229 whereas the factor above leads to 1.79115).

We analyze the concavity of the core function f of (1). For simplicity, we use parameters for the entries of Λ:

$$\Lambda = \begin{pmatrix} -a & 0 & \infty \\ 0 & -c & 0 \\ \infty & 0 & -a \end{pmatrix},$$

i.e., $a = -1/2, c = -2.7245$. Let $p = [p_1, p_2, p_3]^T$ and $q = [q_1, q_2, q_3]^T$, i.e., the p_i, q_i be the entries of p and q. The restriction that pq^T is 0 at the places where Λ has entry $-\infty$ is now $p_1 q_3 = 0$ and $p_3 q_1 = 0$ leading to four cases: $p_1 = p_3 = 0$, $p_1 = q_1 = 0$, $p_3 = q_3 = 0$ and $q_1 = q_3 = 0$.

In the cases $p_1 = q_1 = 0$ and $p_3 = q_3 = 0$, the function f is only a function of p_2 and q_2, and has form

$$f(p_2, q_2) = \widetilde{\mathbb{H}}[p_2] + \widetilde{\mathbb{H}}[q_2] + c q_2 p_2 + a(1-p_2)(1-q_2) + \frac{a+c}{7}.$$

The Jacobian and Hessian of f for $0 < p_2, q_2 < 1$ is

$$J(f) = \left(\log\left(\tfrac{1}{p_2} - 1\right) - a + (c+a)q_2, \quad \log\left(\tfrac{1}{q_2} - 1\right) - a + (c+a)p_2\right),$$

$$H(f) = \begin{pmatrix} -\frac{\log e}{p_2(1-p_2)} & c+a \\ c+a & -\frac{\log e}{q_2(1-q_2)} \end{pmatrix}.$$

By Sylvester's criterion, the Hessian is negative definite for $|c+a| < 4\log e$ (which holds for the actual parameters) as the upper left entry is negative and the determinant is nonnegative:

$$\det H(f) = \frac{\log^2 e}{p_2(1-p_2)\cdot q_2(1-q_2)} - (c+a)^2 \geq \frac{\log^2 e}{4\cdot 4} - (c+a)^2 > 0.$$

It follows that f is strictly concave, and hence if it has a critical point in the interior of its domain, then it is its unique maximum. Numerically solving $J(f) = 0$ provides indeed a critical point in the interior, namely, $p_2 = q_2 \approx 0.377964$.

The remaining cases are $p_1 = p_3 = 0$ and $q_1 = q_3 = 0$. We consider only the second one, as the first one is analogous. Now $q = [0, 1, 0]^T$ is fixed, hence

$$f(p) = \widetilde{\mathbb{H}}\,[p_1, p_2, p_3] + cp_2 + \frac{a+c}{7}$$

subject to $p_1 + p_2 + p_3 = 1$. Note that f is a concave function, and as we will see, it has a (unique) critical point in its interior, and hence it is its unique maximum.

We use Lagrange multipliers to find critical points, i.e., we look for the zeros of the Jacobian of

$$\begin{aligned} &f(p) - (\lambda - 1)(p_1 + p_2 + p_3) \\ &\qquad = p_1 \log \frac{1}{p_1} + p_2 \log \frac{1}{p_2} + p_3 \log \frac{1}{p_3} + cp_2 + \frac{a+c}{7} - (\lambda - 1)(p_1 + p_2 + p_3), \end{aligned}$$

for which the equations are

$$\begin{aligned} \log \frac{1}{p_1} - \lambda &= 0, & \log \frac{1}{p_2} + c - \lambda &= 0, \\ \log \frac{1}{p_3} - \lambda &= 0. \end{aligned}$$

This can be solved in p_1, p_2, p_3:

$$p_1 = p_3 = 2^{-\lambda}, \qquad p_2 = 2^{c-\lambda}.$$

The value of λ is determined by the condition $p_1 + p_2 + p_3 = 1$:

$$2 \cdot 2^{-\lambda} + 2^{c-\lambda} = 1,$$

which simplifies to

$$\begin{aligned} 2^{1-c} + 1 &= 2^{\lambda - c}, \\ \lambda &= \log(2^{1-c} + 1) + c > 1. \end{aligned}$$

In particular, $p_1 = p_3 = 2^{-\lambda}$ lie strictly between 0 and $1/2$, ensuring that p_1, p_2, p_3 is an inner point of the domain of f. Hence the maximum value of f is

$$\begin{aligned} &\max_p f(p) = 2 \cdot 2^{-\lambda}\lambda + 2^{c-\lambda}(\lambda - c) + c2^{c-\lambda} + \frac{a+c}{7} \\ &\qquad = \underbrace{(2^{1-c} + 1)2^{c-\lambda}}_{1}\lambda = \log(2^{1-c} + 1) + c + \frac{a+c}{7}. \end{aligned}$$

Summarizing, the overall maximum of f is the maximum of the maxima of the cases, i.e., $\mathbb{W}\,[N] \le \max_{p,q} f(p, q) \approx 1.79115$. □

Information cost (defined over Chakrabarti et al. [2001], Bar-Yossef et al. [2004], Barak et al. [2010]) is an information-theoretic lower bound on communication complexity that has analogous properties to common information—it also obeys a direct sum theorem Bar-Yossef et al. [2004], and characterizes amortized communication complexity Braverman and Rao [2011]. For a boolean matrix M and distribution μ on rows and columns of M, the (internal) information cost of a randomized protocol Π is $\mathrm{IC}_\mu(\Pi) = \mathbb{I}[\Pi; A \mid B] + \mathbb{I}[\Pi; B \mid A]$. The information cost of M with respect to distribution μ and error ϵ is then the infimum over all protocols Π that compute M with error at most ϵ of $\mathrm{IC}_\mu(\Pi)$. The information cost of M with respect to any distribution is a lower bound on the randomized communication complexity of M.

Note that, as a protocol with error ϵ for M can be trivially transformed into a protocol with error ϵ for the negation of M, the information cost of M and $\overline{M} = \mathbb{1} - M$ are the same. We now see with the example of set intersection, the negation of disjointness, that this is not true for the common information.

Lemma 3.2 (Common information of set intersection and noninvarance under complements)**.** *Let M be the 2^n-by-2^n matrix where*

$$M(x,y) = \begin{cases} 0 & x \cap y = \emptyset \\ 1 & \textit{otherwise} \end{cases}.$$

Then $\mathbb{C}[M] = O(1)$, yet for $\overline{M} = \mathbb{1}_{2^n,2^n} - M$ we have $\mathbb{C}\left[\overline{M}\right] = \mathbb{C}\left[DISJ_n\right] = 2n/3$.

Proof. The complement of set intersection $\overline{M} = \mathrm{DISJ}_n = \mathrm{DISJ}_1^{\otimes n}$ is the disjointness matrix. It follows that $\mathbb{C}\left[\mathrm{DISJ}_n\right] = 2n/3$ as common information is additive under tensoring (Lemma 5.5) and $\mathbb{C}\left[\mathrm{DISJ}_1\right] = 2/3$ by Witsenhausen [1976].

We now establish an upper bound on the common information M, the set intersection matrix. Note that the number of ones in M is $m = 2^{2n} - 3^n = (1 - (3/4)^n)2^{2n}$. M has a covering of size n by the rectangles $R_i = \{(x,y) : x_i = y_i = 1\}$. We use this covering to define a partition of the ones of M inductively as follows. Let $S_1 = R_1$ and $S_i = \{(x,y) \in R_i : (x,y) \notin R_j, j < i\}$. In general S_i is not itself a rectangle, but can be partitioned into 3^{i-1} many rectangles, each of relative area 4^{-i}. Using this factorization, we can lower bound the private information by

$$\begin{aligned}
\mathbb{W}[M] &\geq \frac{2^{2n}}{m} \sum_{i=1}^{n} 3^{i-1} \frac{1}{4^i} \log \frac{2^{2n}}{4^i} \\
&= \frac{2 \cdot 3^{n-1}}{m} \sum_{i=1}^{n} \left(\frac{4}{3}\right)^{n-i} \cdot (n-i) \\
&= \frac{2 \cdot 3^{n-1}}{m} \cdot \frac{(4/3) \cdot ((n-1)(4/3)^n - n(4/3)^{n-1} + 1)}{(4/3-1)^2} \\
&= \frac{2^{2n}}{m}(2n - 8(1 - (3/4)^n)) = \frac{2n}{1-(3/4)^n} - 8
\end{aligned}$$

using the identity $\sum_{i=1}^{n}(n-i)x^i = x((n-1)x^n - nx^{n-1} + 1)/(x-1)^2$.

Thus we have

$$\mathbb{C}[M] \leq \log m - \frac{2^{2n}}{m}(2n - 8(1-(3/4)^n)) = 2n + \log\left(1 - \left(\frac{3}{4}\right)^n\right) - \frac{2n}{1-(3/4)^n} + 8 = 8 + o(1).$$

□

Another variant of information cost, known as external information cost has been defined in the literature. This definition is directly analogous to the definition of common information: the external information cost of a protocol Π with respect to a distribution $\mu \sim (A,B)$ is $\mathrm{IC}_\mu^0(\Pi) =$

$\mathbb{I}[A, B; \Pi]$. In the case of common information, however, this distinction is not important as for a factorization the external and internal information cost of a factorization are equivalent up to a constant.

Lemma 3.3 (External vs. internal common information). *Let M be a nonnegative matrix, $(A, B) \sim M$ and Π be a seed. Then*

$$\mathbb{I}[A, B; \Pi] = \mathbb{I}[A; \Pi \mid B] + \mathbb{I}[B; \Pi \mid A] + \mathbb{I}[A; B],$$

i.e., external information cost *and* internal information cost *differ by* $\mathbb{I}[A; B]$.

Proof. We consider

$$\begin{aligned}\mathbb{I}[A, B; \Pi] - \mathbb{I}[A; \Pi \mid B] - \mathbb{I}[B; \Pi \mid A] &= \mathbb{I}[B; \Pi] + \mathbb{I}[A; \Pi \mid B] - \mathbb{I}[A; \Pi \mid B] - \mathbb{I}[B; \Pi \mid A] \\ &= \mathbb{I}[B; \Pi] - \mathbb{I}[B; \Pi \mid A] = \mathbb{I}[B; A] - \underbrace{\mathbb{I}[B; A \mid \Pi]}_{=0} = \mathbb{I}[B; A].\end{aligned}$$

□

4 Common information as amortized log nonnegative rank

By Fact 2.5, the common information provides a lower bound on the logarithm of the nonnegative rank. This bound, however, can be arbitrarily far from the logarithm of the nonnegative rank, as can be seen in the next example.

Let $M_n \in \mathbb{R}^{n \times n}$ be the diagonal matrix given by $M_n(i, i) := 2^i / \sum_{j \in [n]} 2^j$ and $M_n(i, j) = 0$ whenever $i \neq j$. Clearly the nonnegative rank of M_n is n, however, the factorization Π given by the 1×1 rectangles arising from the elements on the main diagonal shows $\mathbb{C}[M_n] \leq \mathbb{H}[\Pi] = O(1)$.

We will now show, however, that common information does capture the *amortized log nonnegative rank*, under small ℓ_1 perturbations. This result is inspired by a one-shot statement in [Jain et al., 2013, Lemma 4.1], and the quantitative analysis is improved here for tensored matrices $M^{\otimes n}$.

For this theorem the following formula for the conditional mutual information between P and Q given R will be useful:

$$\mathbb{I}[P; Q \mid R] = \mathbb{E}_{x \sim P, y \sim Q, z \sim R}\left[\log \frac{q(x \mid y, z)}{q(x \mid z)}\right],$$

where $q(x \mid y, z) := \mathbb{P}[P = x \mid Q = y, R = z]$ is a probability vector and similarly for $q(x \mid z)$.

We will also use the *cyclic property* of mutual information:

$$\mathbb{I}[P; Q] - \mathbb{I}[P; Q \mid R] = \mathbb{I}[Q; R] - \mathbb{I}[Q; R \mid P].$$

Theorem 4.1 (Common information = amortized log nonnegative rank). *Let $M \in \mathbb{R}_+^{m \times n}$ be a matrix with $w := \sum_{ij} M_{ij}$ and $k = \mathbb{C}[M]$. Then for any $\varepsilon > 0$ and $\delta \in (0, 1)$, for every multiplier $\ell \geq \max\{\Omega(\log^2(mn/\varepsilon)/\varepsilon^2 k^2) \cdot \ln(\delta^{-1}), \Omega(\delta/\varepsilon)\}$ there exists a nonnegative matrix $M_{\varepsilon,\delta,\ell} \in \mathbb{R}_+^{m^\ell \times n^\ell}$ with*

1. $\log \operatorname{rk}_+(M_{\varepsilon,\delta,\ell})/\ell \leq (1 + \varepsilon)\mathbb{C}[M] + O(\delta^3 \ln \delta^{-1})/\ell$,
2. $\left\|M^{\otimes \ell} - M_{\varepsilon,\delta,\ell}\right\|_1 \leq \delta w^\ell$.

In particular, we have

$$\lim_{\ell \to \infty, \delta \to 0, \varepsilon \to 0} \frac{\log \operatorname{rk}_+ M_{\varepsilon,\delta,\ell}}{\ell} = \mathbb{C}[M]$$

Proof. Without loss of generality we may assume $w = 1$, which allows us to identify matrices with probably distribution of row-column pairs. In particular, let q_0 be the probability distribution associated with M, and let $(A_0, B_0) \sim q_0$ be its random row-column pair. Let Π_0 be a seed of A_0, B_0 realizing $k = \mathbb{C}[M] = \mathbb{I}[A_0, B_0; \Pi_0]$, with size $r \leq mn$ which exists by Fact 2.5.

For a distribution p of a random variable X, let $p^{\otimes \ell}$ denote the distribution of ℓ independent copies of X, which is consistent with the notion of tensoring the associated matrix for p. In particular, the distribution of $M^{\otimes \ell}$ is $q_0^{\otimes \ell}$. We shall approximate the distributions q_0 and $q_0^{\otimes \ell}$ with better behaving distributions $q_1, q_2, \ldots$ and the same subscripts and superscripts will be used for the random variables, e.g., A_1, B_1, Π_1 will have distribution q_1. The first goal of the approximation is to bound the ratios $\log q_0(a, b|\pi)/q_0(a, b)$ appearing in the common information to allow us later to argue via concentration.

For a lower bound, we define q_1 by keeping the seed $\Pi_1 := \Pi_0$, and modifying only the distribution of A_0, B_0 conditioned on Π_0 to obtain A_1, B_1. The aim is to have $q_1(a|\pi), q_1(b|\pi) \geq \beta$ for a small positive parameter β chosen later. Therefore we introduce coins $C_A, C_B \in \{0, 1\}$ independently of $\Pi_1 = \Pi_0$, A_0, and B_0 with

$$\mathbb{P}[C_A = 1] = \beta m, \qquad \mathbb{P}[C_B = 1] = \beta n.$$

If $C_A = 1$ then we choose A_1 uniformly in the range of A_0 independently of Π_0, A_0, B_0. If $C_A = 0$ then we choose $A_1 = A_0$. We define B_1 similarly using C_B and B_0. Obviously, A_1 and B_1 are conditionally independent given Π_0. In other words, the conditional probabilities are

$$q_1(a|\pi) := (1 - \beta m) q_0(a|\pi) + \beta, \qquad q_1(b|\pi) := (1 - \beta n) q_0(b|\pi) + \beta.$$

In particular,

$$q_1(a, b|\pi) = q_1(a|\pi) \cdot q_1(b|\pi) \geq \beta^2.$$

For the mutual information we deduce the following bound

$$\begin{aligned} \mathbb{I}[A_1, B_1; \Pi_1] &\leq \mathbb{I}[A_1, B_1; \Pi_1 \,|\, C_A, C_B] \\ &= \mathbb{P}[C_A = 0, C_B = 0]\, \mathbb{I}[A_0, B_0; \Pi_0] + \mathbb{P}[C_A = 1, C_B = 0]\, \mathbb{I}[B_0; \Pi_0] \\ &+ \mathbb{P}[C_A = 0, C_B = 1]\, \mathbb{I}[A_0; \Pi_0] \leq \mathbb{I}[A_0, B_0; \Pi_0] = k. \end{aligned}$$

The first inequality holds because C_A, C_B are independent of Π_1 by construction using the cyclic property of $\mathbb{I}[A_1, B_1; \Pi_1 \,|\, C_A, C_B] - \mathbb{I}[A_1, B_1; \Pi_1]$. The equality is a special case of the law of total expectation, e.g., given $C_A = C_B = 0$ the distribution of A_1, B_1, Π_1 is that of A_0, B_0, Π_0. And given $C_A = 1, C_B = 0$ the distribution is that of B_0, Π_0 and an independent uniform random variable. The second inequality follows by upper bounding all mutual information terms by $\mathbb{I}[A_0, B_0; \Pi_0]$.

We estimate the total variation of $q_0^{\otimes \ell}$ and $q_1^{\otimes \ell}$. We start by comparing the conditional distributions of q_0 and q_1, using that for distributions p_1, p_2 we have $\|p_1 - p_2\|_1 = 2 \max_{X \text{ event}} (p_1(X) - p_2(X))$, and the maximizer can be explicitly given:

$$\begin{aligned} \sum_a |q_1(a|\pi) - q_0(a|\pi)| &= 2 \sum_{a \,:\, q_1(a|\pi) > q_0(a|\pi)} (q_1(a|\pi) - q_0(a|\pi)) \\ &= 2 \sum_{a \,:\, q_1(a|\pi) > q_0(a|\pi)} \beta(1 - m q_0(a|\pi)) \leq 2\beta(m - 1), \\ \sum_b |q_1(b|\pi) - q_0(b|\pi)| &\leq 2\beta(n - 1). \end{aligned}$$

Combining the estimates on rows and columns:

$$\left\|q_1(\cdot|\pi) - q_0(\cdot|\pi)\right\|_1 \leq 2\beta(m+n-2),$$

which remains valid by removing the conditioning on π via taking expectation:

$$\left\|q_1 - q_0\right\|_1 \leq 2\beta(m+n-2).$$

We can now estimate the total variation of $q_0^{\otimes\ell}$ and $q_1^{\otimes\ell}$ via

$$\left\|q_1^{\otimes\ell} - q_0^{\otimes\ell}\right\|_1 \leq \sum_{j=1}^{\ell} \left\|q_1^{\otimes j} \otimes q_0^{\otimes(\ell-j)} - q_1^{\otimes j-1} \otimes q_0^{\otimes(\ell-j+1)}\right\|_1 \leq 2\ell\beta(m+n-2).$$

Let q denote the conditional distribution q_1 given $q_1(\Pi_1) \geq \beta$. In particular, $q(a,b|\pi) = q_1(a,b|\pi)$ for all a, b, π. As there are r possible values of Π_1 we have

$$\mathbb{P}\left[q_1(\Pi_1) < \beta\right] \leq r\beta,$$

and hence, we estimate similarly as before:

$$\left\|q - q_1\right\|_1 \leq 2\,\mathbb{P}\left[q_1(\Pi_1) < \beta\right] \leq 2r\beta, \qquad \left\|q^{\otimes\ell} - q_1^{\otimes\ell}\right\|_1 \leq 2\ell r\beta.$$

As a result, we now have a distribution q close to q_0 such that whenever $q_1(\pi) \geq \beta$:

$$\frac{q(a,b|\pi)}{q(a,b)} \geq q_1(a,b|\pi) \geq \beta^2, \qquad \frac{q(a,b|\pi)}{q(a,b)} \leq \frac{1}{q(\pi)} \leq \frac{1}{q_1(\pi)} \leq \frac{1}{\beta}. \tag{2}$$

We check that the mutual information of q remains close to the common information of q_0. Let $\chi(X)$ denote the indicator of event X.

$$\begin{aligned}\mathbb{E}\left[\log\frac{q(A,B|\Pi)}{q(A,B)}\right] &= \mathbb{I}\left[A,B;\Pi\right] = \mathbb{I}\left[A_1,B_1;\Pi_1\,|\,q_1(\Pi_1) \geq \beta\right] \\ &\leq \frac{\mathbb{I}\left[A_1,B_1;\Pi_1\,|\,\chi(q_1(\Pi_1) \geq \beta)\right]}{\mathbb{P}\left[q_1(\Pi_1) \geq \beta\right]} = \frac{\mathbb{I}\left[A_1,B_1;\Pi_1\right] - \mathbb{I}\left[A_1,B_1;\chi(q_1(\Pi_1) \geq \beta)\right]}{\mathbb{P}\left[q_1(\Pi_1) \geq \beta\right]} \leq \frac{k}{1-r\beta},\end{aligned}$$

where the first inequality follows from the law of total expectation, and the following equality follows with the cyclic property of $\mathbb{I}\left[P;Q\right] - \mathbb{I}\left[P;Q\,|\,R\right]$.

From now on we will only work with q and $q_3 := q^{\otimes\ell}$. In order to ease notation we introduce independent copies $Z_1, \dots, Z_\ell$ of the pair (A,B) (we no longer need to handle the components of the pair separately), and independent copies $W_1, \dots, W_\ell$ of Π, so that the Z_i, W_i are mutually independent copies of $(A,B), \Pi$. Let $Z = (Z_1, \dots, Z_\ell), W = (W_1, \dots, W_\ell)$ denote the collection of the Z_i and W_i, respectively.

In a first step we show that the encoding length of the ratios of the tensored distribution strongly concentrates around the common information via Hoeffding's inequality. Note that $\beta^2 \leq \frac{q(Z_i|W_i)}{q(Z_i)} \leq \frac{1}{\beta}$ by (2) as $q_1(W_i) \geq \beta$ holds almost surely because $W_i \sim q$. Observe that

$$\begin{aligned}\mathbb{P}\left[\log\frac{q_3(Z|W)}{q_3(Z)} > (1+\varepsilon)k\ell\right] &= \mathbb{P}\left[\frac{1}{\ell}\sum_{i=1}^{\ell}\log\frac{q(Z_i|W_i)}{q(Z_i)} > (1+\varepsilon)k\right] \\ &\leq \mathbb{P}\left[\frac{1}{\ell}\sum_{i=1}^{\ell}\log\frac{q(Z_i|W_i)}{q(Z_i)} - \mathbb{E}\left[\frac{1}{\ell}\sum_{i=1}^{\ell}\log\frac{q(Z_i|W_i)}{q(Z_i)}\right] > (1+\varepsilon)k - \frac{k}{1-r\beta}\right]\end{aligned}$$

Note that $(1+\varepsilon)k - k/(1-r\beta) = (\varepsilon - r\beta/(1-r\beta))k$. We apply Hoeffding's inequality, so that the following inequality chain holds:

$$\mathbb{P}\left[\log \frac{q_3(Z|W)}{q_3(Z)} > (1+\varepsilon)k\ell\right] \leq \mathbb{P}\left[\frac{1}{\ell}\sum_{i=1}^{\ell} \log \frac{q(Z_i|W_i)}{q(Z_i)} - \mathbb{E}\left[\frac{1}{\ell}\sum_{i=1}^{\ell} \log \frac{q(Z_i|W_i)}{q(Z_i)}\right] > \left(\varepsilon - \frac{r\beta}{1-r\beta}\right)k\right]$$
$$\leq \exp\left(-\frac{2\ell k^2}{9\log^2 1/\beta}\left(\varepsilon - \frac{r\beta}{1-r\beta}\right)^2\right) =: \delta_1.$$

Therefore with high probability the conditional distribution does not deviate much from the unconditional one, i.e., the set

$$G_1 := \left\{(z,w) \in [m^\ell n^\ell] \times [r^\ell] \,\middle|\, q_3(z|w) \leq 2^{(1+\varepsilon)k\ell} q_3(z)\right\}$$

has high probability

$$q_3(G_1) \geq 1 - \delta_1.$$

We are ready to introduce the matrix $M_{\varepsilon,\delta,\ell}$ by means of the associated distribution. We sample τ independent copies $W_1, \dots, W_\tau$ of W according to the distribution q_3; in particular several of the W_i may coincide. Let $J \in [\tau]$ be chosen uniformly, and $\widetilde{W} := W_J$ will be the seed for the random row and column. We define the conditional distribution of $\widetilde{Z}|W_J = w$ to coincide with $Z|W = w$. This uniquely defines the distribution of $\widetilde{Z}, \widetilde{W}$, and we let $M_{\varepsilon,\delta,\ell}$ be the matrix of $\widetilde{Z}$ given $W_1, \dots, W_\tau$:

$$M_{\varepsilon,\delta,\ell}(z) = \widetilde{q_3}(z) = \frac{\sum_{i\in[\tau]} q_3(z|w_i)}{\tau}.$$

Thus $M_{\varepsilon,\delta,\ell}$ is a random matrix with $\operatorname{rk}_+ M_{\varepsilon,\delta,\ell} \leq \tau$. We show that with high probability, it is close to $M^{\otimes\ell}$, i.e.,

$$\mathbb{E}\left[\left\|M^{\otimes\ell} - M_{\varepsilon,\delta,\ell}\right\|_1\right] \leq \delta.$$

We will need the set

$$G_2 := \left\{z : \sum_{w:\ (z,w)\in G_1} q_3(z,w) \geq \delta_2 q_3(z)\right\},$$

that contains all row-columns pairs that are within a δ_2-ratio, with δ_2 chosen later. We approximate q_3 by a measure q_4 defined via

$$q_4(z,w) := \begin{cases} q_3(z,w) & \text{if } (z,w) \in G_1 \text{ and } z \in G_2, \\ 0 & \text{otherwise.} \end{cases}$$

Note that q_4 need not be a probability distribution, however it is close to q_3:

$$\|q_3 - q_4\|_1 = \sum_{z,w} |q_4(z,w) - q_3(z,w)| = \sum_{(z,w)\notin G_1} q_3(z,w) + \sum_{\substack{(z,w)\in G_1\\ z\notin G_2}} q_3(z,w)$$
$$\leq 1 - q_3(G_1) + \sum_{z\notin G_2} \delta_2 \cdot q_3(z) = 1 - q_3(G_1) + \delta_2(1 - q_3(G_2)) \leq \delta_1 + \delta_2. \quad (3)$$

We define $q_4(z|w) := q_4(z,w)/q_3(w)$ and $q_4(z) := \sum_w q_4(z,w) = \sum_w q_3(w) q_4(z|w)$. As an approximation for $\widetilde{q_3}$, we use

$$\widetilde{q_4}(z) := \frac{\sum_{i\in[\tau]} q_4(z|w_i)}{\tau}.$$

Therefore $\mathbb{E}\left[q_4(z|w_i)\right]_{w_i\sim W} = q_4(z)$. Moreover, for $(z,w)\in G_1$ and $z\in G_2$, we have by definition of the sets

$$q_4(z|w) = q_3(z|w) \overset{(z,w)\in G_1}{\leq} 2^{(1+\varepsilon)k\ell} q_3(z) \overset{z\in G_2}{\leq} \frac{2^{(1+\varepsilon)k\ell}}{\delta_2} q_4(z),$$

and $q_4(z|w) \leq \frac{2^{(1+\varepsilon)k\ell}}{\delta_2} q_4(z)$ trivially holds if $(z,w)\notin G_1$ or $z\notin G_2$.

With the bounds on the ratios, we will now invoke Chernoff's bound to estimate the error arising from the sample set $\{w_i \mid i\in[\tau]\}$. Whenever $q_4(z)\neq 0$, we have

$$\begin{aligned}\mathbb{P}_{w_i\sim W}\left[|\widetilde{q_4}(z) - q_4(z)| > \delta_2 q_4(z)\right] &= \mathbb{P}_{w_i\sim W}\left[\left|\frac{\sum_{i\in[\tau]} q_4(z|w_i)}{\tau} - q_4(z)\right| > \delta_2 q_4(z)\right] \\ &\leq 2\exp\left(-\frac{\delta_2^3 \tau}{3\cdot 2^{(1+\varepsilon)k\ell}}\right) =: \delta_4.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}\left[|\widetilde{q_4}(z) - q_4(z)|\right] \leq\ & \mathbb{P}\left[|\widetilde{q_4}(z) - q_4(z)| \leq \delta_2 q_4(z)\right]\cdot \delta_2 q_4(z) \\ & + \mathbb{P}\left[|\widetilde{q_4}(z) - q_4(z)| > \delta_2 q_4(z)\right]\left(\widetilde{q_4}(z) + q_4(z)\right) \\ =\ & \delta_2 q_4(z) + \mathbb{P}\left[|\widetilde{q_4}(z) - q_4(z)| > \delta_2 q_4(z)\right]\left(\widetilde{q_4}(z) + (1-\delta_2) q_4(z)\right) \\ \leq\ & \delta_2 q_4(z) + (\widetilde{q_4}(z) + (1-\delta_2) q_4(z))\delta_4.\end{aligned}$$

This obviously holds for $q_4(z) = 0$ as well. Summing up for all z we obtain

$$\mathbb{E}\left[\|\widetilde{q_4} - q_4\|_1\right] \leq \delta_2 + (2-\delta_2)\delta_4 < \delta_2 + 2\delta_4.$$

We can easily estimate the distance between the approximations $\widetilde{q_3}$ and $\widetilde{q_4}$:

$$\begin{aligned}\mathbb{E}\left[\|\widetilde{q_3} - \widetilde{q_4}\|_1\right] &= \mathbb{E}_{w_i}\left[\left\|\frac{\sum_{i\in[\tau]} q_3(\cdot|w_i)}{\tau} - \frac{\sum_{i\in[\tau]} q_4(\cdot|w_i)}{\tau}\right\|_1\right] \\ &\leq \sum_{i\in[\tau]} \frac{\mathbb{E}_{w_i}\left[\|q_3(\cdot|w_i) - q_4(\cdot|w_i)\|_1\right]}{\tau} = \|q_3 - q_4\|_1 \overset{(3)}{\leq} \delta_1 + \delta_2.\end{aligned}$$

Finally, the total variation of q_3 and $\widetilde{q_3}$ can be bounded:

$$\mathbb{E}\left[\|q_3 - \widetilde{q_3}\|_1\right] \leq \mathbb{E}\left[\|q_3 - q_4\|_1\right] + \mathbb{E}\left[\|q_4 - \widetilde{q_4}\|_1\right] + \mathbb{E}\left[\|\widetilde{q_4} - \widetilde{q_3}\|_1\right] \leq 2\delta_1 + 3\delta_2 + 2\delta_4.$$

At last, we combine the various bounds above to bound the distance of $M^{\otimes\ell}$ and $M_{\varepsilon,\delta,\ell}$:

$$\begin{aligned}\mathbb{E}\left[\left\|M^{\otimes\ell} - M_{\varepsilon,\delta,\ell}\right\|_1\right] = \mathbb{E}\left[\left\|q_0^{\otimes\ell} - \widetilde{q_3}\right\|_1\right] &\leq \left\|q_0^{\otimes\ell} - q_1^{\otimes\ell}\right\|_1 + \left\|q_1^{\otimes\ell} - q_3\right\|_1 + \mathbb{E}\left[\|q_3 - \widetilde{q_3}\|_1\right] \\ &\leq 2\ell\beta(m+n+r-2) + 2\delta_1 + 3\delta_2 + 2\delta_4.\end{aligned}$$

Now we choose the free parameters β, δ_2 to make this bound smaller than δ, in particular,

$$\ell\beta(m+n+r-2) = \frac{\delta}{8}, \qquad \delta_2 = \frac{\delta}{12},$$

$$\delta_1 \leq \frac{\delta}{8}, \qquad \delta_4 = 2\exp\left(-\frac{\delta_2^3\tau}{3\cdot 2^{(1+\varepsilon)k\ell}}\right) \leq \frac{\delta}{8}.$$

The last inequality holds provided

$$\tau \geq \frac{5184 \cdot 2^{(1+\varepsilon)k\ell}}{\delta^3} \ln\left(\frac{16}{\delta}\right).$$

To ease the estimation of δ_1, we require $\varepsilon - r\beta/(1-r\beta) \geq \varepsilon/2$, i.e., $\beta \leq \varepsilon/r(2+\varepsilon)$, which means

$$\ell \geq \frac{\delta r(2+\varepsilon)}{8(m+n+r-2)\varepsilon}.$$

Thus

$$\delta_1 = \exp\left(-\frac{2\ell k^2}{9\log^2 1/\beta}\left(\varepsilon - \frac{r\beta}{1-r\beta}\right)^2\right) \leq \exp\left(-\frac{\ell\varepsilon^2 k^2}{18\log^2(r(2+\varepsilon)/\varepsilon)}\right) \leq \delta/8$$

if

$$\ell \geq \frac{18\log^2(r(2+\varepsilon)/\varepsilon)}{\varepsilon^2 k^2} \ln(8/\delta).$$

□

As a corollary, we obtain that common information is the best bound in a natural class.

Corollary 4.2 (Common information as limit superior)**.** *Let X be a real-valued function with domain the set of nonnegative matrices, satisfying the following continuity condition: For every nonnegative matrix M and $\varepsilon > 0$, there is a constant $c > 0$ such that for every positive integer n and nonnegative matrix N*

$$X(N) \geq X(M^{\otimes n}) - n\varepsilon - nc \left\|N - M^{\otimes n}\right\|_1.$$

If for all nonnegative matrices M we have $X(M) \leq \log \operatorname{rk}_+ M$ then

$$\limsup_{n\to\infty} \frac{X(M^{\otimes n})}{n} \leq \mathbb{C}\left[M\right].$$

If additionally for all nonnegative matrices M we have $\mathbb{C}\left[M\right] \leq X(M)$, then $\lim_{n\to\infty} \frac{X(M^{\otimes n})}{n} = \mathbb{C}\left[M\right]$.

Proof. Let M be a nonnegative matrix and $\varepsilon > 0$ fixed. Let c be the constant depending on M and ε from the continuity condition. By Theorem 4.1 for every large enough nonnegative integer n there is an approximation $\widetilde{M}$ of M satisfying $\log \operatorname{rk}_+ \widetilde{M} \leq n(1+\varepsilon)\mathbb{C}\left[M\right]$ and $\left\|M^{\otimes n} - \widetilde{M}\right\|_1 \leq \varepsilon/c$.

$$X(M^{\otimes n}) \leq X(\widetilde{M}) + n\varepsilon + nc\left\|M^{\otimes n} - \widetilde{M}\right\|_1 \leq \log \operatorname{rk}_+ \widetilde{M} + 2n\varepsilon \leq (1+\varepsilon)n\mathbb{C}\left[M\right] + 2n\varepsilon,$$

i.e.,

$$\frac{X(M^{\otimes n})}{n} \leq \mathbb{C}\left[M\right] + (2 + \mathbb{C}\left[M\right])\varepsilon.$$

It follows that

$$\limsup_{n\to\infty} \frac{X(M^{\otimes n})}{n} \leq \inf_{\varepsilon>0}(\mathbb{C}\left[M\right] + (2+\mathbb{C}\left[M\right])\varepsilon) = \mathbb{C}\left[M\right]$$

as claimed. □

Remark 4.3. Lemma 5.7 together with Proposition 5.6 shows that common information satisfies the conditions for X.

5 A dual approach to common information

In this section we extend the dual characterization of common information from [Witsenhausen, 1976, §4], and use it to establish continuity and additivity under tensoring. As UDISJ, the main example we consider is only a partial matrix, we also generalize common information to partial matrices, and extend the characterization to obtain lower bounds.

5.1 Common information of partial matrices

We first extend the lower bound in [Witsenhausen, 1976, Theorem 2] from full matrices to partial ones. The obtained lower bounds may no longer be tight due to inherent discontinuity of common information of partial matrices, as exhibited in Example 5.4.

Definition 5.1. The *common information* (*private information*) of a partial matrix M is the infimum (supremum) of common information (private information) over all its nonnegative extensions

$$\mathbb{C}\left[M\right] \coloneqq \inf_{\widetilde{M} \supseteq M} \mathbb{C}\left[\widetilde{M} \middle| Z\right] \qquad \text{and} \qquad \mathbb{W}\left[M\right] \coloneqq \sup_{\widetilde{M} \supseteq M} \mathbb{W}\left[\widetilde{M} \middle| Z\right].$$

where Z is the event of being in the domain of definition of M.

Clearly $\mathbb{C}\left[M\right] = \mathbb{H}\left[M\right] - \mathbb{W}\left[M\right]$, where $\mathbb{H}\left[M\right]$ is the entropy of M restricted to its domain. We are ready to formulate the lower bound on common information.

Proposition 5.2 (Common information via rank-1 factors)**.** *Let M be a partial nonnegative $m \times n$ matrix. Then its common information is lower bounded by*

$$\mathbb{C}\left[M\right] \geq \sup_{\Lambda \in \mathbb{R}^Z} \inf_{\substack{p,q \geq 0 \\ \|p\|_1 = \|q\|_1 = 1}} \mathbb{H}\left[M\right] - \mathbb{H}\left[p,q \,|\, Z\right] + \frac{q^T \Lambda p}{\sum_{a,b \in Z} p_a q_b} - \operatorname{Tr}\left[\Lambda \frac{M}{\|M\|_1}\right], \tag{4}$$

where Z is the event of being in the domain of M. Similarly, the private information $\mathbb{W}\left[M\right]$ is upper bounded by

$$\mathbb{W}\left[M\right] \leq \inf_{\Lambda \in \mathbb{R}^{n \times m}} \max_{\substack{p,q \geq 0 \\ \|p\|_1 = \|q\|_1 = 1}} \mathbb{H}\left[p,q \,|\, Z\right] - \frac{q^T \Lambda p}{\sum_{a,b \in Z} p_a q_b} + \operatorname{Tr}\left[\Lambda \frac{M}{\|M\|_1}\right]. \tag{5}$$

If M is a full matrix, then equality holds for both quantities above.

Proof. Equality in the case of full matrices is [Witsenhausen, 1976, Theorem 2]. The proof of the inequality follows by a direct calculation. Without loss of generality, we assume $\|M\|_1 = 1$. Let $\Lambda \in \mathbb{R}^Z$ and

$$\alpha \coloneqq \inf_{\substack{p,q \geq 0 \\ \|p\|_1 = \|q\|_1 = 1}} \mathbb{H}\left[M\right] - \mathbb{H}\left[p,q \,|\, Z\right] + \frac{q^T \Lambda p}{\sum_{a,b \in Z} p_a q_b} - \operatorname{Tr}[\Lambda M].$$

Furthermore, let $\widetilde{M} = \sum_i \lambda_i p_i q_i^T$ be an extension of M with a rank-1 factorization coming from a seed Π. We need to show that $\mathbb{I}\left[\widetilde{M}; \Pi \middle| Z\right] \geq \alpha$; note that $\mathbb{H}\left[\widetilde{M} \middle| Z\right] = \mathbb{H}\left[M\right]$.

Therefore we restrict the factorization to the domain of M, omit factors which are 0 on the whole domain, and rescale the entries to be probability distributions possibly changing the coefficients λ_i: $M = \sum_i \mu_i (p_i q_i^T \mid Z)$. In particular, by summing up all the entries, we obtain $\sum_i \mu_i = 1$. Now an easy calculation establishes the claim:

$$\begin{aligned} \mathbb{I}\left[\widetilde{M}; \Pi \middle| Z\right] &= \mathbb{H}\left[\widetilde{M} \middle| Z\right] - \mathbb{H}\left[\widetilde{M} \middle| Z, \Pi\right] = \mathbb{H}\left[M \,|\, Z\right] - \sum_i \mu_i \mathbb{H}\left[p_i, q_i \,|\, Z\right] \\ &= \sum_i \mu_i \left(\mathbb{H}\left[M\right] - \mathbb{H}\left[p_i, q_i \,|\, Z\right]\right) \geq \sum_i \mu_i \left(\alpha + \operatorname{Tr}[\Lambda(M - (p_i q_i^T \mid Z))]\right) = \alpha \end{aligned}$$

as $\sum_i \mu_i(\mathrm{Tr}[\Lambda(M - (p_i q_i^T \mid Z))] = 0$. This proves the lower bound on $\mathbb{C}[M]$. The upper bound on $\mathbb{W}[M]$ follows via $\mathbb{W}[M] = \mathbb{H}[M] - \mathbb{C}[M]$. □

Note that (4) and (5) are invariant under additive shifts of Λ of the form $\Lambda + \rho \cdot \mathbb{1}$, but not under rescalings.

The supremum in (4) cannot be replaced by maximum even for full matrices. We see this in the next example with the 2×2 DISJ matrix.

Example 5.3 (Common information of DISJ via (4))**.** We consider the matrix $D := \left(\begin{smallmatrix} a & b \\ c & 0 \end{smallmatrix}\right)$ where $a + b + c = 1$. The common information for this matrix has been established to be $\mathbb{C}[D] = (b + c)\log(b + c) - b \log b - c \log c = \mathbb{H}[D] - \widetilde{\mathbb{H}}[a]$ in Witsenhausen [1976]. We will now show that $\mathbb{C}[D]$ can only be reached in the limit and for every single instance of Λ we have that

$$\inf_{\substack{p,q \geq 0 \\ \|p\|_1 = \|q\|_1 = 1}} \mathbb{H}[D] - \mathbb{H}[p] - \mathbb{H}[q] + q^T \Lambda p - \mathrm{Tr}[\Lambda D] < \mathbb{H}[D] - \widetilde{\mathbb{H}}[a],$$

or equivalently $K := \sup_{p,q} \mathbb{H}[p] - \mathbb{H}[q] + q^T \Lambda p - \mathrm{Tr}[\Lambda D] > \widetilde{\mathbb{H}}[a]$. Here and below we drop the conditions on p, q for readability.

First let us show that $\mathbb{W}[D] \leq \widetilde{\mathbb{H}}[a]$ cannot be obtained via a single supporting hyperplane, i.e., $K > \widetilde{\mathbb{H}}[a]$. Recall that

$$\widetilde{\mathbb{H}}[p] + \widetilde{\mathbb{H}}[q] \leq K - q^T \Lambda p + \mathrm{Tr}(\Lambda D)$$

for all $0 \leq p, q \leq 1$. We examine this for the pairs p, q where the bound is supposed to be tight, i.e., for the pairs appearing in the best factorization: $([a, b + c], [1, 0])$ and $([1, 0], [a, b + c])$. Actually, we also consider nearby pairs $p = [a, b + c]$ and $q = [1 - x, x]$ for which we obtain

$$\widetilde{\mathbb{H}}[a] + \widetilde{\mathbb{H}}[x] \leq K - a\Lambda_{11} - (b + c)\Lambda_{21} + [a(\Lambda_{11} - \Lambda_{12}) + (b + c)(\Lambda_{21} - \Lambda_{22})]x + a\Lambda_{11} + b\Lambda_{12} + c\Lambda_{21}$$

for all $0 \leq x \leq 1$, therefore

$$\widetilde{\mathbb{H}}[a] < K - a\Lambda_{11} - (b + c)\Lambda_{21} + a\Lambda_{11} + b\Lambda_{12} + c\Lambda_{21} = K + b(\Lambda_{12} - \Lambda_{21}).$$

Thus $K > \widetilde{\mathbb{H}}[a]$ if $\Lambda_{12} \leq \Lambda_{21}$. A similar argument applies when $\Lambda_{21} \leq \Lambda_{12}$ finishing the proof of $K > \widetilde{\mathbb{H}}[a]$.

We will now show that for an arbitrary $\varepsilon > 0$ there exists Λ so that

$$\sup_{p,q} q^T \Lambda p + \widetilde{\mathbb{H}}[p] + \widetilde{\mathbb{H}}[q] - \mathrm{Tr}(\Lambda D) \leq \widetilde{\mathbb{H}}[a] + \varepsilon$$

if $0 < a < 1/2$. Actually, we will choose a Λ of the form

$$\Lambda = \begin{bmatrix} -\widetilde{\mathbb{H}}'[a] & 0 \\ 0 & -C \end{bmatrix}$$

where $C > 0$ is a large constant to be chosen later. Observe that $\mathrm{Tr}(\Lambda D) = -\widetilde{\mathbb{H}}'[a]\,a$. Let us introduce the shorthand

$$\psi(p, q) := \widetilde{\mathbb{H}}[p] + \widetilde{\mathbb{H}}[q] - \widetilde{\mathbb{H}}'[a]\,pq - C(1 - p)(1 - q) + a\widetilde{\mathbb{H}}'[a].$$

Let us choose $0 < \delta < 1/2$ such that $\widetilde{\mathbb{H}}'[a]\,\delta + \widetilde{\mathbb{H}}[\delta] \leq \varepsilon$ and let $C = (2 + \widetilde{\mathbb{H}}'[a]\,a)/\delta^2$. First suppose that both $p, q \leq 1 - \delta$. In this case $\psi(p, q) \leq 2 - C\delta^2 + \widetilde{\mathbb{H}}'[a]\,a \leq 0$ and the claim holds.

Now consider the case that at least one of p, q is at least $1-\delta$. As Λ is symmetric we may suppose without loss of generality that $q \geq 1-\delta$. Then we can upper bound $\psi(p,q)$ as follows, using the concavity of entropy

$$\psi(p,q) \leq \widetilde{\mathbb{H}}[a] + \widetilde{\mathbb{H}}'[a](p-a) + \widetilde{\mathbb{H}}[\delta] - \widetilde{\mathbb{H}}'[a](p-\delta) + a\widetilde{\mathbb{H}}'[a]$$
$$= \widetilde{\mathbb{H}}[a] + \widetilde{\mathbb{H}}[\delta] + \widetilde{\mathbb{H}}'[a]\,\delta \leq \widetilde{\mathbb{H}}[a] + \varepsilon$$

as claimed.

We will see later in Lemma 5.7 that common information is a continuous quantity for *full* matrices, with a proof based on the tightness of the dual characterization. The next example, however, shows that common information of partial matrices can be discontinuous, ruling out the tightness of the lower bound for partial matrices in general.

Example 5.4 (Discontinuity of common information of a partial matrix)**.** Despite continuity for full matrices, common information is not continuous for partial matrices, as the following examples shows:

$$\mathbb{C}\left[\begin{pmatrix} \varepsilon & 1 \\ 1 & * \end{pmatrix}\right] = \begin{cases} 0, & \varepsilon > 0, \\ 1, & \varepsilon = 0. \end{cases}$$

Here $*$ denotes an undefined nonnegative entry. Note that for $\varepsilon > 0$ the matrix has a rank-1 extension, while for $\varepsilon = 0$ no factor can have both its entries in the antidiagonal non-zero, i.e., it must reveal the exact entry of $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & * \end{smallmatrix}\right)$. Therefore $\mathbb{C}[M] = \mathbb{H}[M \mid Z] = 1$.

5.2 Continuity and tensoring for common information

The dual characterization of common information has several applications. We first see how the supporting hyperplanes of the information set naturally tensor, leading to a simplified form of the dual formulation for a matrix which is a tensor product. Then we see how the dual characterization implies that common information is robust under small ℓ_1 perturbations.

We prove that common information is additive under tensoring of matrices. The core of the proof is a direct sum property of mutual information (see [Cover and Thomas, 2006, Theorem 2.5.2]): for arbitrary random variables A, B, C

$$\mathbb{I}[A_1, A_2; B] = \mathbb{I}[A_1; B] + \mathbb{I}[A_2; B \mid A_1].$$

In particular, $\mathbb{I}[A_1, A_2; B] \geq \mathbb{I}[A_1; B] + \mathbb{I}[A_2; B]$ if A_1 and A_2 are independent.

Lemma 5.5 (Common information and tensoring)**.** *Let M, N be arbitrary nonnegative matrices. Then $\mathbb{C}[M \otimes N] = \mathbb{C}[M] + \mathbb{C}[N]$. In particular $\mathbb{C}[M^{\otimes n}] = n\mathbb{C}[M]$ for all $n \in \mathbb{N}$.*

Proof. First we identify the distribution induced by $M \otimes N$. Let $(A_M, B_M) \sim M$ and $(A_N, B_N) \sim N$ be independent pairs of random variables with distribution induced by M and N, respectively. Then the distribution of $(A_M, A_N; B_M, B_N)$ is induced by $M \otimes N$.

Now let Π be a seed for $M \otimes N$. We have

$$\mathbb{I}[M \otimes N; \Pi] = \mathbb{I}[A_M, B_M, A_N, B_N; \Pi] \geq \mathbb{I}[A_M, B_M; \Pi] + \mathbb{I}[A_N, B_N; \Pi] = \mathbb{I}[M; \Pi] + \mathbb{I}[N; \Pi],$$

where the latter inequality follows from the direct sum property and the independence of (A_M, B_M) and (A_N, B_N). It suffices to observe that Π is a seed both for (A_M, B_M) and (A_N, B_N) so that when taking the infimum over all seeds Π for $M \otimes N$ we have

$$\begin{aligned}
\mathbb{C}\left[M \otimes N\right] &= \inf_{\Pi \text{ seed for } (A_M, A_N),\,(B_M, B_N)} \mathbb{I}\left[M \otimes N; \Pi\right] \\
&\geq \inf_{\Pi \text{ seed for } (A_M, A_N),\,(B_M, B_N)} \left(\mathbb{I}\left[M; \Pi\right] + \mathbb{I}\left[N; \Pi\right]\right) \\
&\geq \inf_{\Pi \text{ seed for } A_M,\, B_M} \mathbb{I}\left[M; \Pi\right] + \inf_{\Pi \text{ seed for } A_N,\, B_N} \mathbb{I}\left[N; \Pi\right] = \mathbb{C}\left[M\right] + \mathbb{C}\left[N\right].
\end{aligned}$$

We will now show that the inequality is tight. For this let Π_M be any seed for M and Π_N be any seed for N with Π_M and Π_N being conditionally independent given A_M, A_N, B_M, B_N. Clearly, Π_M, Π_N is a seed for $M \otimes N$. By the chain rule we have

$$\mathbb{I}\left[A_M, B_M, A_N, B_N; \Pi_M, \Pi_N\right] = \mathbb{I}\left[A_M, B_M; \Pi_M, \Pi_N\right] + \mathbb{I}\left[A_N, B_N; \Pi_M, \Pi_N \,|\, A_M, B_M\right]$$

We further have, (again using the chain rule)

$$\mathbb{I}\left[A_M, B_M; \Pi_M, \Pi_N\right] = \underbrace{\mathbb{I}\left[A_M, B_M; \Pi_N\right]}_{=0\text{, by independence}} + \underbrace{\mathbb{I}\left[A_M, B_M; \Pi_M \,|\, \Pi_N\right]}_{=\mathbb{I}\left[A_M, B_M; \Pi_M\right]}$$

and similarly

$$\mathbb{I}\left[A_N, B_N; \Pi_M, \Pi_N \,|\, A_M, B_M\right] = \underbrace{\mathbb{I}\left[A_N, B_N; \Pi_M \,|\, A_M, B_M\right]}_{=0} + \underbrace{\mathbb{I}\left[A_N, B_N; \Pi_N \,|\, A_M, B_M, \Pi_M\right]}_{=\mathbb{I}\left[A_N, B_N; \Pi_N\right]},$$

so that

$$\mathbb{I}\left[A_M, B_M, A_N, B_N; \Pi_M, \Pi_N\right] = \mathbb{I}\left[A_M, B_M; \Pi_M\right] + \mathbb{I}\left[A_N, B_N; \Pi_N\right].$$

Taking the infimum over all seeds Π_M for M and Π_N for N we obtain

$$\begin{aligned}
\mathbb{C}\left[M \otimes N\right] &= \inf_{\Pi \text{ seed for } (A_M, A_N),\,(B_M, B_N)} \mathbb{I}\left[A_M, B_M, A_N, B_N; \Pi\right] \\
&\leq \inf_{\substack{\Pi_M \text{ seed for } A_M,\, B_M \\ \Pi_N \text{ seed for } A_M,\, B_M}} \mathbb{I}\left[A_M, B_M, A_N, B_N; \Pi_M, \Pi_N\right] \\
&= \inf_{\Pi_M \text{ seed for } A_M,\, B_M} \mathbb{I}\left[A_M, B_M; \Pi_M\right] + \inf_{\Pi_N \text{ seed for } A_M,\, B_M} \mathbb{I}\left[A_N, B_N; \Pi_N\right] \\
&= \mathbb{C}\left[M\right] + \mathbb{C}\left[N\right].
\end{aligned}$$

□

As an application of Lemma 5.5, in the dual formulation for a tensor product $M_1 \otimes \cdots \otimes M_n$, we can restrict the parameter Λ in the minimax formula (4) from Proposition 5.2 to be a tensor sum of matrices corresponding to the components M_i.

Proposition 5.6. *Let $M_i \in \mathbb{R}_+^{m_i \times n_i}$ be nonnegative matrices with $i \in [\ell]$. Then*

$$\mathbb{C}\left[M_1 \otimes \cdots \otimes M_\ell\right] = \sup_{\Lambda_i \in \mathbb{R}^{n_i \times m_i} : i = 1, \ldots, \ell} \; \inf_{\substack{p, q \geq 0 \\ \|p\|_1 = \|q\|_1 = 1}} \left(\mathbb{H}\left[M\right] - \mathbb{H}\left[p\right] - \mathbb{H}\left[q\right] + q^T \Lambda p - \operatorname{Tr}[\Lambda M]\right), \quad (6)$$

where $\Lambda := \Lambda_1 \oplus \cdots \oplus \Lambda_\ell$.

Proof. Adding up (4) from Proposition 5.2 for $M_1, \dots, M_\ell$ together with Lemma 5.5 provides

$$\begin{aligned}
\mathbb{C}\left[M_1 \otimes \cdots \otimes M_\ell\right] &= \mathbb{C}\left[M_1\right] + \cdots + \mathbb{C}\left[M_\ell\right] \\
&= \sum_{i=1}^{\ell} \sup_{\Lambda_i \in \mathbb{R}^{n_i \times m_i}} \inf_{\substack{p_i, q_i \geq 0 \\ \|p_i\|_1 = \|q_i\|_1 = 1}} \mathbb{H}\left[M_i\right] - \mathbb{H}\left[p_i\right] - \mathbb{H}\left[q_i\right] + q_i^T \Lambda_i p_i - \operatorname{Tr}[\Lambda_i M_i] \\
&= \sup_{\Lambda_i \in \mathbb{R}^{n_i \times m_i} : i=1,\dots,\ell} \inf_{\substack{p_i, q_i \geq 0 \\ \|p_i\|_1 = \|q_i\|_1 = 1 \\ i=1,\dots,\ell}} \sum_{i=1}^{\ell} \left(\mathbb{H}\left[M_i\right] - \mathbb{H}\left[p_i\right] - \mathbb{H}\left[q_i\right] + q_i^T \Lambda_i p_i - \operatorname{Tr}[\Lambda_i M_i] \right).
\end{aligned}$$

Note that the last formula is obtained from the right-hand side of (6) by restricting p and q to product distributions $p = p_1 \times \cdots \times p_\ell$ and $q = q_1 \times \cdots \times q_\ell$.

To finish the proof, we show that the minimum of the inner formula is not enlarged by allowing arbitrary distributions p and q. Indeed, the following computation establishes that the inner formula decreases by replacing p and q with the products $p_1 \times \cdots \times p_\ell$ and $q_1 \times \cdots \times q_\ell$ of their marginal distribution (omitting terms not depending on p and q):

$$-\mathbb{H}\left[p\right] - \mathbb{H}\left[q\right] + q^T \Lambda p = -\mathbb{H}\left[p\right] - \mathbb{H}\left[q\right] + \sum_{i=1}^{\ell} q_i^T \Lambda_i p_i \geq \sum_{i=1}^{\ell} \left(-\mathbb{H}\left[p_i\right] - \mathbb{H}\left[q_i\right] + q_i^T \Lambda_i p_i \right). \quad \square$$

We now show that the common information of close by matrices cannot discontinuously increase.

Lemma 5.7 (Continuity of common information)**.** *Let $N, M \in \mathbb{R}_+^{m \times n}$ be nonnegative matrices with $\|M\|_1 = \|N\|_1 = 1$ and let $\varepsilon > 0$. Then*

$$\mathbb{C}\left[M\right] \leq \mathbb{C}\left[N\right] + \|M - N\|_1 \log \frac{\|M - N\|_1}{mn} + \|\Lambda\|_\infty \|M - N\|_1 + \varepsilon,$$

where Λ is an ε-realizer of the common information of M, i.e., for all $p, q \geq 0$

$$\mathbb{C}\left[M\right] - \varepsilon \leq \mathbb{H}\left[M\right] - \mathbb{H}\left[p\right] - \mathbb{H}\left[q\right] + q^T \Lambda p - \operatorname{Tr}[\Lambda M].$$

Proof. The statement follows directly from the characterization of the common information in Proposition 5.2:

$$\begin{aligned}
\mathbb{C}\left[N\right] &\geq \min_{p,q} \mathbb{H}\left[N\right] - \mathbb{H}\left[p\right] - \mathbb{H}\left[q\right] + q^T \Lambda p - \operatorname{Tr}[\Lambda N] \\
&\geq \mathbb{C}\left[M\right] - \varepsilon + \mathbb{H}\left[N\right] - \mathbb{H}\left[M\right] + \operatorname{Tr} \Lambda (M - N) \\
&\geq \mathbb{C}\left[M\right] - \varepsilon - \|M - N\|_1 \log \frac{\|M - N\|_1}{mn} - \|\Lambda\|_\infty \|M - N\|_1 .
\end{aligned}$$

$\square$

6 Consequences for (U)DISJ

We will now use the dual approach to derive lower bounds on the DISJ as well as the UDISJ (partial) matrices under any type of small perturbation.

As a start, we will establish a stronger lower bound on the common information of the UDISJ (partial) matrix than in Braun and Pokutta [2013]. This improvement is based on the result from Kaibel and Weltge [2013] that every combinatorial rectangle with no uniquely intersecting pairs of subsets can have at most 2^n disjoint pairs of subsets. We give an alternative proof of this fact using a compression argument.

Lemma 6.1 (Recoding disjoint sets). *Let $A, B \in \{0,1\}^n$ be two independent random strings satisfying $\mathbb{P}[|A \cap B| = 1] = 0$. Let $S = \{(a,b) \in \{0,1\}^n \mid a \cap b = \emptyset \wedge \mathbb{P}[A = a, B = b] > 0\}$. Then*

1. *there exists a nonsingular binary code for S (depending on the distribution of A, B) of length n, i.e., we can encode each of the elements in S with at most n bits. In particular, $|S| \leq 2^n$.*
2. *$\mathbb{H}[A, B \mid A \cap B = \emptyset] \leq n$.*

Proof.

Encoding step: We encode the pair a, b. We choose $C_1, \ldots, C_n \in \{0,1\}$ inductively. Suppose that C_j with $j < i$ has been chosen. For readability let

$$p_i := \mathbb{P}_{a\sim A}\left[a_i = 1 \,\middle|\, \substack{a_j=0 \text{ for } j<i \text{ with } C_j=0 \\ a_j=A_j \text{ for } j>i}\right] \qquad \text{and} \qquad q_i := \mathbb{P}_{b\sim B}\left[b_i = 1 \,\middle|\, \substack{b_j=0 \text{ for } j<i \text{ with } C_j=1 \\ b_j=B_j \text{ for } j>i}\right].$$

Note that p_i is a function of $C_1, \ldots, C_{i-1}$ and $A_{i+1}, \ldots, A_n$; similarly for q_i. By independence and $\mathbb{P}[|A \cap B| = 1] = 0$ we have $p_i q_i = 0$.

$$\text{If } p_i = 0, \quad C_i := \begin{cases} 1 & \text{if } B_i = 0 \\ 0 & \text{if } B_i = 1; \end{cases} \qquad \text{if } q_i = 0, \quad C_i := \begin{cases} 0 & \text{if } A_i = 0 \\ 1 & \text{if } A_i = 1. \end{cases}$$

If both $p_i = q_i = 0$, then choose the C_i arbitrarily.

Decoding step: We will now show that we can exactly decode A, B from C. This will in particular imply that $|S| \leq 2^n$ and hence $\mathbb{H}[A, B \mid A \cap B = \emptyset] = \mathbb{H}[C \mid A \cap B = \emptyset] \leq n$. We inductively decode A, B from C, however in reverse direction. Suppose that $(A_n, B_n), \ldots, (A_{i+1}, B_{i+1})$ have been decoded. We decode

$$A_i := \begin{cases} 0, & \text{if } C_i = 0 \text{ or } p_i = 0 \\ 1, & \text{otherwise.} \end{cases} \qquad B_i := \begin{cases} 1, & \text{if } C_i = 1 \text{ or } q_i = 0 \\ 0, & \text{otherwise.} \end{cases}$$

It remains to show that the decoding is exact. Observe that $\mathbb{P}[A_i = 1 \wedge p_i = 0] = 0$ and therefore the decoding is exact if $p_i = 0$. If $p_i \neq 0$, then $q_i = 0$ and $C_i = A_i$ by definition of C_i. Similarly B_i is decoded correctly. □

Theorem 6.2 (Lower bound on common information of UDISJ). *Let M be the UDISJ (partial) matrix of strings of length n. Then*

$$\mathbb{C}[M] \geq n \log 3/2 \approx 0.585 \cdot n.$$

Proof. Let Π be a seed of an extension $\widetilde{M}$ of M. Given an arbitrary $\Pi = \pi$, the variables A, B become independent, and hence the set of pairs (a, b) with non-zero probability is a rectangle. Obviously, there is no (a, b) with $|a \cap b| = 1$ and non-zero probability. By Kaibel and Weltge [2013], every rectangle for UDISJ contains at most 2^n pairs (a, b) with a, b being disjoint, hence $\mathbb{H}[\widetilde{M} \mid A \cap B = \emptyset, \Pi = \pi] \leq n$. Taking expectation, $\mathbb{H}[\widetilde{M} \mid A \cap B = \emptyset, \Pi] \leq n$, i.e.,

$$\mathbb{C}[\widetilde{M} \mid A \cap B = \emptyset] = \mathbb{H}[\widetilde{M} \mid A \cap B = \emptyset] - \mathbb{C}[\widetilde{M} \mid A \cap B = \emptyset, \Pi] \geq n \log 3 - n = n \log 3/2. \quad \square$$

6.1 Approximate direct sum lower bound

We revisit approaches from Braverman and Moitra [2012], Braun and Pokutta [2013] to obtain a tight lower bound on the conditional common information of approximate UDISJ. We use the same conditional as in several previous works including Bar-Yossef et al. [2004], Braverman and Moitra [2012], Braun and Pokutta [2013]. This conditional is a variant of the disjointness $A \cap B = \emptyset$ with the remarkable feature that it preserves the independence of A and B under any seed. As demonstrated by Theorem 6.2, breaking independence can lead to improved lower bounds, however it is not clear how to handle the perturbed case using this conditional.

Let $C = (C_1, \dots, C_n)$ be n fair coins with sides labelled with $\mathbb{A}$ and $\mathbb{B}$. The coins are independent of A, B and any seed Π. We define $D = (D_1, \dots, D_n)$ via

$$D_i := \begin{cases} A_i & \text{if } C_i = \mathbb{A}, \\ B_i & \text{if } C_i = \mathbb{B}. \end{cases}$$

The exact condition is the event $D = 0$ together with the random variable C.

We shall use D_{-i} and C_{-i} to denote the collections $(D_j : j \neq i)$ and $(C_j : j \neq i)$, respectively.

Theorem 6.3. *Let M be a nonnegative square matrix with rows and columns indexed by subsets of $[n]$. Then for all $0 < \varepsilon < (2/5) \log 3/2 \approx 0.234$*

$$\mathbb{W}\left[M \,|\, D = 0, C\right] \leq (1 - \alpha + \varepsilon)n - \frac{s(\alpha - \varepsilon) + t(\beta - \varepsilon + 2\log\varepsilon)}{r}, \tag{7}$$

where

$$r := \sum_{\substack{a,b \subseteq [n] \\ a \cap b = \emptyset}} 2^{-|a|-|b|} M(a,b), \quad t := \sum_{\substack{a,b \subseteq [n] \\ |a \cap b| = 1}} 2^{-|a|-|b|+2} M(a,b), \qquad s := \sum_{\substack{a,b \subseteq [n] \\ a \cap b = \emptyset}} (|a| + |b|) 2^{-|a|-|b|} M(a,b),$$

$$\alpha := 1 - \frac{\log 3}{2} \approx 0.208, \qquad \beta := 6 - 3\log 3 - 2\log(5\ln 2) \approx -2.341.$$

In particular,

$$\mathbb{W}\left[M \,|\, D = 0, C\right] \leq (1 - \alpha)n - \frac{s\alpha + t(\gamma + 2\log(t/(t + rn + s)))}{r}, \tag{8}$$

where $\gamma := \beta + 2\log(2e^{-1}\log e) \approx -2.169$ provided $t/(t + rn + s) < \frac{\log 3/2}{5 \ln 2} \approx 0.169$.

As the upper bound is invariant under scalings of M, the parameters r, u, t are not normalized.

Example 6.4. For the modified partial UDISJ matrix M

$$M(a,b) := \begin{cases} 1, & a \cap b = \emptyset \\ \delta, & |a \cap b| = 1 \end{cases}$$

we have $r = 2^n$, $t = \delta r n$, $s = rn/2$ leading to the lower bound

$$\mathbb{C}\left[M \,|\, D = 0, C\right] \geq \left(\frac{3}{2}\alpha + \delta\left(\gamma - 2\log\left(1 + \frac{2}{\delta}\right)\right)\right)n$$

for $\delta < 3/(2(\frac{5\ln 2}{\log(3/2)} - 1)) \approx 0.305$. The lower bound is exact for $\delta = 0$ and the above formula is a continuous extension. It complements the lower bound $(1-\delta)n/8$ for all $0 < \delta < 1$ from [Braun and Pokutta, 2013, Theorem 4.1]; see also Braverman and Moitra [2012] for a slightly weaker bound.

The core of the proof is a bound on the conditional private information which arises from a fusion of Braverman and Moitra [2012] and Braun and Pokutta [2013]. We reuse the form which appeared as part of the *advantage* estimation in Braverman and Moitra [2012]:

Lemma 6.5. *For all* $0 \leq p, q \leq 1$, $0 < \varepsilon < (2/5) \log 3/2 \approx 0.234$ *and* $\alpha := 1 - \frac{\log 3}{2}$ *we have*

$$p\widetilde{\mathbb{H}}\left[q\right] + q\widetilde{\mathbb{H}}\left[p\right] \leq p + q - 2(\alpha - \varepsilon) + 2(\beta - \alpha - 2\log\varepsilon)(1-p)(1-q). \tag{9}$$

Proof. For convenience, first we prove a reparametrized version of the bound: let $\delta := 2^{-5\varepsilon/2}$, therefore $\varepsilon = -(2/5)\log\delta$, $2/3 \leq \delta < 1$ and $\widetilde{\mathbb{H}}'\left[\delta\right] \leq -1$. We claim

$$p(1 - \widetilde{\mathbb{H}}\left[q\right]) + q(1 - \widetilde{\mathbb{H}}\left[p\right]) \geq 2\alpha + \frac{4}{5}\log\delta - \left(\frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - 4\widetilde{\mathbb{H}}'\left[\delta\right]\right)(1-p)(1-q). \tag{10}$$

We start with the case $q \leq 4/5$, where the main estimation arises from. As the binary entropy function is convex, we can estimate its value by its gradient:

$$\begin{aligned}
\widetilde{\mathbb{H}}\left[p\right] &\leq \widetilde{\mathbb{H}}\left[\delta\right] + \widetilde{\mathbb{H}}'\left[\delta\right](p - \delta), \\
\widetilde{\mathbb{H}}\left[q\right] &\leq \widetilde{\mathbb{H}}\left[\frac{1}{3}\right] + \widetilde{\mathbb{H}}'\left[\frac{1}{3}\right]\left(q - \frac{1}{3}\right) = \widetilde{\mathbb{H}}\left[\frac{1}{3}\right] + q - \frac{1}{3},
\end{aligned}$$

leading to

$$\begin{aligned}
p(1 - \widetilde{\mathbb{H}}\left[q\right]) + q(1 - \widetilde{\mathbb{H}}\left[p\right]) &\geq p\left(1 - \widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - q + \frac{1}{3}\right) + q(1 - \widetilde{\mathbb{H}}\left[\delta\right] - \widetilde{\mathbb{H}}'\left[\delta\right](p - \delta)) \\
= \underbrace{\frac{4}{3} - \widetilde{\mathbb{H}}\left[\frac{1}{3}\right]}_{=2\alpha} - q\left(\underbrace{\widetilde{\mathbb{H}}\left[\delta\right] + \widetilde{\mathbb{H}}'\left[\delta\right](1-\delta)}_{-\log\delta}\right) &+ \left(\underbrace{\frac{1}{3} - \widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - \widetilde{\mathbb{H}}'\left[\delta\right]}_{\geq 4/3 - \widetilde{\mathbb{H}}[1/3] > 0} + (1 + \widetilde{\mathbb{H}}'\left[\delta\right])(1-q)\right)(p-1) \\
&\geq 2\alpha + \frac{4}{5}\log\delta - \left(\frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - 4\widetilde{\mathbb{H}}'\left[\delta\right]\right)(1-p)(1-q),
\end{aligned}$$

where the last inequality uses $5(1-q) \geq 1$. This finishes the case $q \leq 4/5$. The case $p \leq 4/5$ is analogous and therefore omitted. The remaining case is $p, q \geq 4/5$, where a simple estimation suffices:

$$\begin{aligned}
p(1 - \widetilde{\mathbb{H}}\left[q\right]) + q(1 - \widetilde{\mathbb{H}}\left[p\right]) &\geq 2 \cdot \frac{4}{5} \cdot \left(1 - \widetilde{\mathbb{H}}\left[\frac{4}{5}\right]\right) > 2\alpha \\
&> 2\alpha - \frac{4}{5}\left(\widetilde{\mathbb{H}}\left[\delta\right] + \widetilde{\mathbb{H}}'\left[\delta\right](1-\delta)\right) - (1-p)(1-q)\left(\frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - 4\widetilde{\mathbb{H}}'\left[\delta\right]\right).
\end{aligned}$$

Note that $\frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - 4\widetilde{\mathbb{H}}'\left[\delta\right] \geq \frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] + 4 > 0$.

Finally, we bring (10) into the form (9). This is fairly straightforward, the only estimation needed is the coefficient of $(1-p)(1-q)$:

$$\begin{aligned}
\frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - 4\widetilde{\mathbb{H}}'\left[\delta\right] = \frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - 4\log(2^{5\varepsilon/2} - 1) &< \frac{8}{3} - 5\widetilde{\mathbb{H}}\left[\frac{1}{3}\right] - 4\log\left(\frac{5\varepsilon\ln 2}{2}\right) \\
&= 10 - 5\log 3 - 4\log(5\ln 2) - 4\log\varepsilon = 2(\beta - \alpha - 2\log\varepsilon).
\end{aligned}$$

□

We are now ready to prove the bound for approximations of UDISJ.

Proof of Theorem 6.3. The second inequality follows from the first one by substituting the optimal value $\varepsilon := 2(t/(t+rn+s))\log e$. Therefore we prove only the first inequality.

First we express r, s, t as probabilities. Without loss of generality, we may assume $\sum_{a,b} M(a,b) = 1$, leading to $r = \mathbb{P}[D=0]$ and $t = \sum_{i=1}^{n} \mathbb{P}[D_{-i}=0, A_i=1, B_i=1]$. Instead of s it will be more convenient to use $u := \sum_{i=1}^{n} \mathbb{P}[D_{-i}=0] = t + rn + s$.

We start with the case $n=1$, and make the identifications $D = D_1, C = C_1, A = A_1, B = B_1$ for simplicity. Hence the parameters are $r = \mathbb{P}[D=0] = \frac{\mathbb{P}[A=0]+\mathbb{P}[B=0]}{2}$, $u = 1$, and $t = \mathbb{P}[A=1, B=1]$. Let Π be a seed, then Lemma 6.5 applies to upper bound the conditional entropy. We shall use the suggestive notation $p(\pi) := \mathbb{P}[A=0 \mid \Pi=\pi]$ and $q(\pi) := \mathbb{P}[B=0 \mid \Pi=\pi]$.

$$
\begin{aligned}
\mathbb{H}[A,B \mid D=0, C, \Pi] = {} & \quad \mathbb{P}[C=\mathcal{A} \mid D=0]\,\mathbb{H}[A,B \mid D=0, C=\mathcal{A}, \Pi] \\
& + \mathbb{P}[C=\mathcal{B} \mid D=0]\,\mathbb{H}[A,B \mid D=0, C=\mathcal{B}, \Pi] \\
&= \frac{\mathbb{P}[A=0]}{2\,\mathbb{P}[D=0]}\mathbb{H}[B \mid A=0, \Pi] + \frac{\mathbb{P}[B=0]}{2\,\mathbb{P}[D=0]}\mathbb{H}[A \mid B=0, \Pi] \\
&= \frac{\mathbb{E}_{\pi\sim\Pi}\left[p(\pi)\widetilde{\mathbb{H}}[q(\pi)] + q(\pi)\widetilde{\mathbb{H}}[p(\pi)]\right]}{2\,\mathbb{P}[D=0]} \\
&\leq \frac{\mathbb{E}_{\pi\sim\Pi}\left[p(\pi)+q(\pi)-2(\alpha-\varepsilon)+2(\beta-\alpha-2\log\varepsilon)(1-p(\pi))(1-q(\pi))\right]}{2\,\mathbb{P}[D=0]} \\
&= \frac{\mathbb{P}[A=0]+\mathbb{P}[B=0]-2(\alpha-\varepsilon)+2(\beta-\alpha-2\log\varepsilon)\,\mathbb{P}[A=1,B=1]}{2\,\mathbb{P}[D=0]} \\
&= 1 - \frac{\alpha-\varepsilon+t(\beta-\alpha-2\log\varepsilon)}{r}.
\end{aligned}
$$

We now turn to the case of general n. To simplify formulas, we introduce shorthand notations:

$$
u_i := \mathbb{P}[D_{-i}=0], \qquad\qquad t_i := \mathbb{P}[D_{-i}=0, A_i=1, B_i=1],
$$

leading to $u = \sum_{i=1}^{n} u_i$ and $t = \sum_{i=1}^{n} t_i$.

For any $1 \leq i \leq n$ we apply the $n=1$ case to A_i, B_i with distribution conditioned on $D_{-i}=0$, and use the seed Π, C_{-i}. In (8) we need to replace r, t by $r/u_i, t_i/u_i$, respectively:

$$
\mathbb{H}[A_i,B_i \mid D=0, C, \Pi] \leq 1 - \frac{\alpha-\varepsilon+(t_i/u_i)(\beta-\alpha-2\log\varepsilon)}{r/u_i} = 1 - \frac{u_i(\alpha-\varepsilon)+t_i(\beta-\alpha-2\log\varepsilon)}{r}.
$$

We sum up these inequalities to obtain as claimed:

$$
\begin{aligned}
\mathbb{H}[A,B \mid D=0, C, \Pi] &\leq \sum_{i=1}^{n} \mathbb{H}[A_i,B_i \mid D=0, C, \Pi] \\
&\leq n - \sum_{i=1}^{n} \frac{u_i(\alpha-\varepsilon)+t_i(\beta-\alpha-2\log\varepsilon)}{r} = n - \frac{u(\alpha-\varepsilon)+t(\beta-\alpha-2\log\varepsilon)}{r} \\
&= (1-\alpha+\varepsilon)n - \frac{s(\alpha-\varepsilon)+t(\beta-\varepsilon+2\log\varepsilon)}{r}.
\end{aligned}
$$

□

It is interesting to note that the conditional common information under $D=0, C$ is not maximized by $\left(\begin{smallmatrix}1/3 & 1/3\\ 1/3 & 0\end{smallmatrix}\right)$ as in the unconditional case, but by $\left(\begin{smallmatrix}2/8 & 3/8\\ 3/8 & 0\end{smallmatrix}\right)$.

6.2 Lower bound for perturbed UDISJ matrices

We use Theorem 6.3 to lower bound the common information of perturbed UDISJ matrices in terms of the size of the perturbation. For measuring the size of perturbation, a natural choice is the ℓ_1-norm of the conditional distribution $M \mid D = 0$, where a disjoint pair of subsets a, b have probability proportional to $2^{-|a|-|b|} M(a, b)$, however, this considers only disjoint a, b. Therefore we also use an analogous norm for $|a \cap b| = 1$. (Note that we do not condition on C rather we condition M on the event $D = 0$.) All in all, we introduce the norms

$$\|M\|_{\emptyset} := \sum_{a,b: a \cap b = \emptyset} 2^{-|a|-|b|} |M(a, b)|, \qquad \|M\|_{\{\cdot\}} := \frac{1}{n} \sum_{a,b: |a \cap b| = 1} 2^{-|a|-|b|} |M(a, b)|$$

for all (not necessarily nonnegative) matrices M. The purpose of the division by n in $\|M\|_{\{\cdot\}}$ is to scale it to the same range as $\|M\|_{\emptyset}$, e.g., $\|\mathbb{1}\|_{\emptyset} = 2^n$ and $\|\mathbb{1}\|_{\{\cdot\}} = 2^{n-3}$.

We put the matrix in subscript for the expressions r, s, t in Theorem 6.3. Obviously,

$$|t_M - t_N| \leq 4n \|M - N\|_{\{\cdot\}}, \qquad |s_M - s_N| \leq n \|M - N\|_{\emptyset}.$$

We are ready to formulate our lower bound for perturbed partial UDISJ matrices:

Corollary 6.6. *Let M be the unique disjointness matrix and N be a partial matrix defined on the same domain with $r_M = r_N = 1$ and $\|N - M\|_{\emptyset} < 1/4$ and $\|N - M\|_{\{\cdot\}} < (4 \cdot ((8 \log e)/\alpha - 4))^{-1} \approx 0.005$. Then*

$$\begin{aligned} \mathbb{C}\left[N \,|\, D = 0, C\right] \geq & \left(\frac{6 - 3\log 3}{4} - a\|N - M\|_{\emptyset} - b\|N - M\|_{\{\cdot\}} + 8\|N - M\|_{\{\cdot\}} \log\|N - M\|_{\{\cdot\}} \right) n \\ & + \|N - M\|_{\emptyset} \log\|N - M\|_{\emptyset} \end{aligned}$$

where $a = 1 + \frac{\log 3}{2} \approx 1.792$ and $b = 8 \log\left(\frac{4 \cdot \delta + 1/2}{8e^{-1} \log e}\right) - 4\beta \approx -14.909.$.

Proof of Corollary 6.6. We apply Theorem 6.3. Note that (7) holds with equality for M with $\varepsilon = 0$ and $t_M = 0$. For N we shall use a $\varepsilon > 0$ specified later.

$$\begin{aligned} \mathbb{C}\left[M \,|\, D = 0, C\right] &= \mathbb{H}\left[M \,|\, D = 0, C\right] - (1 - \alpha)n + s_M \alpha + t_M \beta, \\ \mathbb{C}\left[N \,|\, D = 0, C\right] &\geq \mathbb{H}\left[N \,|\, D = 0, C\right] - (1 - \alpha + \varepsilon)n + s_N(\alpha - \varepsilon) + t_N(\beta - \varepsilon + 2\log\varepsilon). \end{aligned}$$

We estimate the difference of entropies via [Cover and Thomas, 2006, Theorem 17.3.3] using that $r_M = r_N = 1$:

$$|\mathbb{H}\left[M \,|\, D = 0, C\right] - \mathbb{H}\left[N \,|\, D = 0, C\right]| \leq -\|N - M\|_{\emptyset} \log \frac{\|N - M\|_{\emptyset}}{3^n}.$$

Therefore

$$\begin{aligned} \mathbb{C}\left[N \,|\, D = 0, C\right] - \mathbb{C}\left[M \,|\, D = 0, C\right] \geq\ & \mathbb{H}\left[N \,|\, D = 0, C\right] - \mathbb{H}\left[M \,|\, D = 0, C\right] + (s_N - s_M)(\alpha - \varepsilon) \\ & + (t_N - t_M)(\beta - \varepsilon + 2\log\varepsilon) - s_M \varepsilon - t_M(\varepsilon - 2\log\varepsilon) \\ \geq\ & \|N - M\|_{\emptyset} \log \frac{\|N - M\|_{\emptyset}}{3^n} - (\alpha - \varepsilon)n\|N - M\|_{\emptyset} \\ & + 4n(\beta - \varepsilon + 2\log\varepsilon)\|N - M\|_{\{\cdot\}} - \frac{\varepsilon n}{2}. \end{aligned}$$

We choose ε to maximize this quantity:

$$\varepsilon := \frac{8\log e}{4 + \frac{1/2 - \|N-M\|_\emptyset}{\|N-M\|_{\{\cdot\}}}}.$$

The upper bounds $\|N - M\|_\emptyset < 1/4$ and $\|N - M\|_\emptyset < (4 \cdot ((8\log e)/\alpha - 4))^{-1} =: \delta$ on the norms in the hypothesis ensure $\varepsilon < \alpha < (2/5)\log(3/2)$ required for Theorem 6.3 and the above estimation. Also note that

$$\varepsilon = \frac{8\|N-M\|_{\{\cdot\}}\log e}{4\|N-M\|_{\{\cdot\}} + 1/2 - \|N-M\|_\emptyset} > \frac{8\|N-M\|_{\{\cdot\}}\log e}{4\cdot\delta + 1/2} = 2^{-(b+4\beta)/8} e\|N-M\|_{\{\cdot\}}.$$

We plug the value of ε into the estimation on $\mathbb{C}[B \mid D = 0, C]$:

$$\begin{aligned}
&\mathbb{C}[N \mid D = 0, C] - \mathbb{C}[M \mid D = 0, C] \\
&\quad\geq \left(-(\alpha + \log 3)\|N-M\|_\emptyset + 4\beta\|N-M\|_{\{\cdot\}} + 8\|N-M\|_{\{\cdot\}}\log\left(\frac{\varepsilon}{e}\right)\right) n \\
&\qquad + \|N-M\|_\emptyset \log\|N-M\|_\emptyset \\
&\quad> \left(-(\alpha + \log 3)\|N-M\|_\emptyset + 4\beta\|N-M\|_{\{\cdot\}} + 8\|N-M\|_{\{\cdot\}}\log 2^{-(b+4\beta)/8}\|N-M\|_{\{\cdot\}}\right) n \\
&\qquad + \|N-M\|_\emptyset \log\|N-M\|_\emptyset \\
&\quad= \left(-a\|N-M\|_\emptyset - b\|N-M\|_{\{\cdot\}} + 8\|N-M\|_{\{\cdot\}}\log\|N-M\|_{\{\cdot\}}\right) n \\
&\qquad + \|N-M\|_\emptyset \log\|N-M\|_\emptyset.
\end{aligned}$$

□

6.3 Lower bound for perturbed DISJ matrices

Similar to Section 6.2, we will now use Lemma 5.7 in order to lower bound the common information of unstructured perturbations of the DISJ matrix.

Lemma 6.7. *Let $M_n \in \mathbb{R}_+^{2^n \times 2^n}$ be the n-dimensional DISJ matrix. Then there exists a constant $1 > C > 0$ so that for any nonnegative matrix $N \in \mathbb{R}_+^{2^n \times 2^n}$ with $\left\|M_n - \frac{3^n}{\|N\|_1}N\right\|_1 \leq C \cdot 3^n$ we have $\mathbb{C}[N] = \Omega(n)$.*

Proof. Pick $\varepsilon > 0$ small enough and consider M_1. By Example 5.3 we know that the for any $\varepsilon > 0$, there exists Λ_1, so that $\mathbb{C}[M_1] - \varepsilon \leq \mathbb{H}[M_1] - \mathbb{H}[p] - \mathbb{H}[q] + q^T\Lambda_1 p - \operatorname{Tr}[\Lambda M_1]$. Let the largest absolute entry in Λ_1 be K. By Proposition 5.6, this Λ_1 can be extended to a Λ for M_n, so that

$$\mathbb{C}[M_n] - \varepsilon n \leq \mathbb{H}[M_n] - \mathbb{H}[p] - \mathbb{H}[q] + q^T\Lambda p - \operatorname{Tr}[\Lambda M_n],$$

where the largest absolute entry of Λ is bounded by nK. By Lemma 5.7 we have

$$\mathbb{C}[M_n] \leq \mathbb{C}[N] - L\log\frac{L}{4^n} + KnL + \varepsilon n,$$

where $L := \left\|M_n / \|M_n\|_1 - N / \|N\|_1\right\|_1$. The above can be rewritten as

$$\begin{aligned}
&\mathbb{C}[N] - L\log\frac{L}{4^n} + KnL + \varepsilon n = \mathbb{C}[N] + L\log\frac{4^n}{L} + KnL + \varepsilon n \\
&= \mathbb{C}[N] + 2Ln - L\log L + KnL + \varepsilon n = \mathbb{C}[N] + Ln(2 - \frac{1}{n}\log L + K) + \varepsilon n \\
&\leq \mathbb{C}[N] + Ln(2 + K) + \varepsilon n.
\end{aligned}$$

Let $\frac{2}{3} - \varepsilon > \delta > 0$ and put $L := \frac{2/3-\varepsilon-\delta}{2+K}$. Using the fact that $\mathbb{C}[M_n] = \frac{2}{3}n$ we obtain

$$\frac{2}{3}n \leq \mathbb{C}[N] + Ln(2+K) + \varepsilon n = \mathbb{C}[N] + \left(\frac{2}{3} - \delta\right)n,$$

so that $\delta n \leq \mathbb{C}[N]$ follows. The result follows by observing that $\|M_n\|_1 = 3^n$. □

We immediately obtain the following corollary

Corollary 6.8. *Let $M_n \in \mathbb{R}_+^{2^n \times 2^n}$ be the n-dimensional DISJ matrix. Then there exists a constant $1 > C > 0$ so that for any deformation N of M_n such that $\|N\|_1 = \|M_n\|_1$ and $\|M_n - N\|_1 \leq C \cdot 3^n$ we have $\mathbb{C}[N] = \Omega(n)$.*

In particular we can exchange up to $C/2$ entries 1 by 0 and vice versa and the resulting matrix will have linear common information and hence an exponential nonnegative rank.

Acknowledgements

The authors would like to thank Samuel Fiorini for valuable discussions on Lemma 6.1. Part of this work was conducted at the *ELC Workshop on Polyhedral Approaches: Extension complexity and pivoting lower bounds*[1] and the authors would like to thank the organizers for providing such a stimulating environment. Research reported in this paper was partially supported by NSF grant CMMI-1300144.

[1] `http://cgm.cs.mcgill.ca/~avis/Kyoto/workshop/workshop.html`

References

Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004. doi: 10.1016/j.jcss.2003.11.006. URL `http://www.sciencedirect.com/science/article/pii/S0022000003001855`.

B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. In *42nd ACM Symposium on Theory of Computing*, pages 67–76, 2010.

G. Braun and S. Pokutta. Common information and unique disjointness. *Proceedings of FOCS*, 2013.

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

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

M. Braverman and A. Rao. Information equals amortized communication. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 748–757. IEEE, 2011.

A. Chakrabarti, Y. Shi, A. Wirth, and A. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *42nd IEEE Symp. on Foundations of Computer Science (FOCS 2001)*, pages 270–278, 2001.

T. Cover and J. Thomas. *Elements of information theory*. Wiley-interscience, 2006.

H. Fawzi and P. Parrilo. New lower bounds on nonnegative rank using conic programming. Technical Report arXiv:1210.6970, arXiv, 2012.

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

R. Jain, Y. Shi, Z. Wei, and S. Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *Proceedings of SODA 2013*, 2013.

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

M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. *SIAM J. Discrete Math.*, 8:76–92, 1995. doi: 10.1137/S0895480192238482.

L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383–390, 1975. ISSN 0012-365X. doi: 10.1016/0012-365X(75)90058-8.

L. Lovász. Communication complexity: A survey. In B. Korte, L. Lovász, H. Prömel, and A. Schrijver, editors, *Paths, flows, and VLSI-layout*, pages 235–265. Springer, 1990.

A. Moitra. A singly-exponential time algorithm for computing nonnegative rank. *arXiv preprint arXiv:1205.0044*, 2012.

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

S. A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM J. Optim.*, 20(3):1364–1377, 2009.

H. S. Witsenhausen. Values and bounds for the common information of two discrete random variables. *SIAM Journal on Applied Mathematics*, 31(2):313–333, 1976.

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

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

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de