STUDIES IN COMMUNICATION COMPLEXITY AND SEMIDEFINITE PROGRAMS

PENGHUI YAO

NATIONAL UNIVERSITY OF SINGAPORE

2013

STUDIES IN COMMUNICATION COMPLEXITY AND SEMIDEFINITE PROGRAMS

PENGHUI YAO

(B.Sc., ECNU)

CENTRE FOR QUANTUM TECHNOLOGIES

NATIONAL UNIVERSITY OF SINGAPORE

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

2013

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Penghui Yao

October 27, 2013

Acknowledgements

First and foremost I am deeply indebted to my supervisor Rahul Jain for giving me an opportunity to work with him and for giving me his guidance and support throughout my graduate career. He has played a fundamental role in my doctoral work. His enthusiasm and insistency have encouraged me to continue whenever I have faced difficulties. His insights have helped me greatly to proceed in my research. Rahul has shared with me much of his understandings and thoughts in computer science. All these will be the most valuable for my future research.

I am very grateful to my co-supervisor Miklos Santha. He encouraged me to apply to Centre for Quantum Technologies (CQT) to pursue my doctoral degree. He guided me in the early stages of my doctoral life and gave me freedom to pursue my research interests. He has created an intellectual group in CQT, where you don't feel research is a lonely job.

I would also like to thank my previous supervisor Angsheng Li, who had introduced me to computational complexity, an exciting and challenging area, and had supported my research for two years before I started my doctoral life in Singapore. I would like to express my gratitude to Hartmut Klauck, Troy Lee and Shengyu Zhang for their friendship. Many discussions with them have been instrumental in cleaning my doubts in research.

Colleagues and friends have given me various kinds of support over years. I would like to express my humble salutations to them. A very partial list includes Lin Chen, Thomas Decker, Donglin Deng, Raghav Kulkarni, Feng Mei, Attila Pereszlényi, Supartha Podder, Ved Prakash, Youming Qiao, Aarthi Sundaram, Weidong Tang, Sarvagya Upadhyay, Yibo Wang, Zhuo Wang, Jiabin You, Huangjun Zhu. I also wish to thank all the administrators of CQT for their excellent administrative support.

Finally, I would like to express the deepest thanks to my wife and my parents for their constant support in my endeavors. I dedicate this thesis to them.

Contents

Summary

This thesis contains two independent parts. The first part concerns fast parallel approximation algorithms for semidefinite programs. The second part concerns *strong direct product* results in communication complexity.

In the first part, we study fast parallel approximation algorithms for certain classes of semidefinite programs. Results are listed below.

- In Chapter 3, we present a fast parallel approximation algorithm for *positive semidefinite programs*. In positive semidefinite programs, all matrices involved in the specification of the problem are positive semidefinite and all scalars involved are non-negative. Our result generalizes the analogous result of Luby and Nisan [53] for positive linear programs.
- In Chapter 4, we present a fast parallel approximation algorithm for *mixed packing and covering semidefinite programs*. Mixed packing and covering semidefinite programs are natural generalizations of positive semidefinte programs. Our result generalizes the analogous result of Young [76] for linear mixed packing and covering programs.

In the second part, we are concerned with strong direct product theorems in communication complexity. A strong direct product theorem for a problem in a given model of computation states that, in order to compute k instances of the problem, if we provide resource which is less than k times the resource required for computing one instance of the problem, with constant success probability, then the probability of correctly computing all the k instances together, is exponentially small in k.

- In Chapter 6, we show a direct product theorem for any relation in the model of *two-party bounded-round public-coin communication complexity*. In particular, our result implies a strong direct product theorem for the *two-party constant-message public-coin communication complexity* of all relations.

- In Chapter 7, we show a strong direct product theorem for all relations in terms of the *smooth rectangle bound* in the model of *two-way public-coin communication complexity*. The smooth rectangle bound was introduced by Jain and Klauck [28] as a generic lower bound method for this model. Our result therefore implies a strong direct product theorem for all relations for which an (asymptotically) optimal lower bound can be provided using the smooth rectangle bound.

Chapter 1

Introduction

The thesis contains two independent parts. The first part concerns fast parallel approximation algorithms for semidefinite programs. The second part concerns strong direct product results in communication complexity. The first part is based on the following two papers.

- Rahul Jain and Penghui Yao. *A parallel approximation algorithm for positive semidefinite programming* [38]. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science*, FOCS'11, page 437-471, 2011.
- Rahul Jain and Penghui Yao. *A parallel approximation algorithm for mixed packing and covering semidefinite programs* [39]. CoRR, abs/1302.0275, 2012.

In this thesis, we concern fast parallel approximation algorithms for semidefinite programs. Fast parallel computation is captured by the complexity class NC. NC contains all the functions that can be computed by logarithmic space uniform Boolean circuits of poly-logarthmic depth. Many matrix operations can be implemented in NC circuits. We have further discussion on this class in Chapter 2. As computing an approximation solution to a semidefinite program, or even to a linear program is P-complete, not all semidefinite programs have fast parallel approximation algorithms under widely-believed assumption $\mathrm{P} \neq \mathrm{NC}$. Thus it is interesting to ask what subclasses of semidefinite programs have fast parallel approximation algorithms. Fast parallel approximation algorithms for approximating optimum solutions to different subclasses of semidefinite programs have been studied in several recent works (e.g. [3; 4; 26; 36; 37; 42]) leading to many interesting applications including the celebrated result QIP = PSPACE [26]. In this thesis, we concern two subclasses of semidefinite programs, positive semidefinite programs and mixed

packing and covering semidefinite programs. Positive semidefinite programs and mixed packing and covering semidefinite programs are two important subclasses of semidefinite programs. In positive semidefinite programs, all matrices involved in the specification of the problem are positive semidefinite and all scalars involved are non-negative. Mixed packing and covering semidefinite programs are natural generalizations of positive linear programs. In Chapter 2, we give the precise definitions of both subclasses of semidefinite programs and present some facts about parallel computation. In Chapter 3, we present a fast parallel approximation algorithm for positive semidefinite programs, which given an instance of a positive semidefinite program of size N and an approximation factor $\varepsilon > 0$, runs in parallel time $\text{poly}(\frac{1}{\varepsilon}) \cdot \text{polylog}(N)$, using $\text{poly}(N)$ processors, and outputs a value which is within multiplicative factor of $(1+\varepsilon)$ to the optimal. Our result generalizes the analogous result of Luby and Nisan [53] for positive linear programs and our algorithm is also inspired by their algorithm. In Chapter 4, we present a fast parallel approximation algorithm for a class of mixed packing and covering semidefinite programs. As a corollary we get a faster approximation algorithm for positive semidefinite programs with better dependence of the parallel running time on the approximation factor, as compared to the one in Chapter 3. Our algorithm and analysis is on similar lines as that of Young [76] who considered analogous linear programs. Although the result in Chapter 3 is improved and simplified, the techniques used in Chapter 3 are still interesting on its own.

The second part is based on the following two papers.

- Rahul Jain, Attila Pereszlényi and Penghui Yao. *A direct product theorem for bounded-round public-coin communication complexity* [30]. In Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science, FOCS '12, pages 167-176.

- Rahul Jain and Penghui Yao. *A strong direct product theorem in terms of the smooth rectangle bound* [40]. CoRR, abs/1209.0263, 2012.

A strong direct product theorem for a problem in a given model of computation states that, in order to compute k instances of the problem, if we provide resource which is less than k times the resource required for computing one instance of the problem with constant success probability, then the probability of correctly computing all the k instances together, is exponentially small in k.

Direct product questions and the weaker direct sum questions have been extensively investigated in different sub-models of communication complexity. A direct sum theorem

states that in order to compute k independent instances of a problem, if we provide resource less than k times the resource required to compute one instance of the problem with a constant success probability $p < 1$, then the success probability for computing all the k instances correctly is at most a constant $q < 1$. As far as we know, the first direct product theorem in communication complexity is Parnafes, Raz and Wigderson's [58] theorem for *forests* of communication protocols. Shaltiel's [66] showed a direct product theorem for the *discrepancy bound*, which is a powerful lower bound on the distributional communication complexity, under the uniform distribution. Later, it was extended to arbitrary distributions by Lee, Shraibman and Špalek [51]; to the multiparty case by Viola and Wigderson [71]; to the generalized discrepancy bound by Sherstov [67]. Klauck, Špalek, de Wolf's [48] showed a strong direct product theorem for the *quantum* communication complexity of the Set Disjointness problem, one of the most well-studied problems in communication complexity. Klauck's [46] extended it to the public-coin communication complexity (which was re-proven using very different arguments in Jain [25]). Other examples are Jain, Klauck and Nayak's [29] theorem for the *subdistribution bound*, Ben-Aroya, Regev, de Wolf's [10] theorem for the one-way quantum communication complexity of the Index function problem; Jain's [25] theorem for randomized one-way communication complexity and Jain's [25] theorem for *conditional relative min-entropy bound* (which is a lower bound on the public-coin communication complexity). Direct sum theorems have been shown in several models, like the public-coin one-way model [33], public-coin simultaneous message passing model [33], entanglement-assisted quantum one-way communication model [35], private-coin simultaneous message passing model [27] and constant-round public-coin two-way model [13]. Very recently, Braverman, Rao, Weinstein and Yehudayoff [14] have shown a direct product theorem for public-coin two-way communication models, which improves the analogous direct sum result in [8]. On the other hand, strong direct product conjectures have been shown to be false by Shaltiel [66] in some models of distributional communication complexity (and of *query complexity* and *circuit complexity*) under specific choices for the error parameter.

Examples of direct product theorems in others models of computation include Yao's *XOR lemma* [74], Raz's [61] theorem for two-prover games; Shaltiel's [66] theorem for *fair decision trees*; Nisan, Rudich and Saks' [56] theorem for *decision forests*; Drucker's [20] theorem for randomized query complexity; Sherstov's [67] theorem for *approximate polynomial degree* and Lee and Roland's [50] theorem for quantum query complexity. Besides their inherent importance, direct product theorems have had various important applications such as in *probabilistically checkable proofs* [61]; in circuit complexity [74] and in

showing time-space tradeoffs [2; 46; 48].

Some definitions and basic facts on communication complexity and information theory are given in Chapter 5. In Chapter 6, we consider the model of two-party bounded-round public-coin communication and show a direct product theorem for the communication complexity of any relation in this model. In particular, our result implies a strong direct product theorem for the two-party constant-message public-coin communication complexity of all relations. As an immediate application of our result, we get a strong direct product theorem for the Pointer Chasing problem. This problem has been well studied for understanding round v/s communication trade-offs in both classical and quantum communication protocols [32; 44; 47; 57; 60]. Our result generalizes the result of Jain [25] which can be regarded as the special case when $t = 1$. We show the result using information theoretic arguments. Our arguments and techniques build on the ones used in Jain [25]. One key tool used in our work and also in Jain [25] is a message compression technique due to Braverman and Rao [13], who used it to show a *direct sum* theorem in the same model of communication complexity as considered by us. Another important tool that we use is a correlated sampling protocol, which for example, has been used in Holenstein [23] for proving a parallel repetition theorem for two-prover games. In Chapter 7, we consider the model of two-way public-coin communication and show a strong direct product theorem for all relations in terms of the smooth rectangle bound, introduced by Jain and Klauck [28] as a generic lower bound method in this model. Our result therefore implies a strong direct product theorem for all relations for which an (asymptotically) optimal lower bound can be provided using the smooth rectangle bound. In fact we are not aware of any relation for which it is known that the smooth rectangle bound does not provide an optimal lower bound. This lower bound subsumes many of the other known lower bound methods, for example the *rectangle bound* (a.k.a the *corruption bound*) [5; 9; 45; 63; 75], the *smooth discrepancy bound* (a.k.a the γ_2 *bound* [52] which in turn subsumes the *discrepancy bound*), the *subdistribution bound* [29] and the *conditional min-entropy bound* [25]. As a consequence, our result reproves some of the known strong direct product results, for example for Inner Product [49] Greater-Than [70] and Set-Disjointness [25; 46]. Our result also shows new strong direct product result for Gap-Hamming Distance [17; 68] and also implies near optimal direct product results for several important functions and relations used to show exponential separations between classical and quantum communication complexity, for which near optimal lower bounds are provided using the rectangle bound, for example by Raz [62], Gavinsky [21] and Klartag and Regev [65]. Our proof is based on information theoretic argument. A key

tool we use is a sampling protocol due to Braverman [12], in fact a modification of it used by Kerenidis, Laplante, Lerays, Roland and Xiao [43].

Chapter 2

Semidefinite programs and parallel computation

As discussed in the previous chapter, several different subclasses of semidefinite programs are shown to admit fast parallel approximation algorithms e.g. [3; 4; 26; 36; 37; 42]. However for each of the algorithms used for example in [26; 36; 37], in order to produce a $(1+\varepsilon)$ approximation of the optimal value for a given semidefinite program of size N, in the corresponding subclass that they considered, the (parallel) running time was $\text{polylog}(N) \cdot \text{poly}(\kappa) \cdot \text{poly}(\frac{1}{\varepsilon})$, where κ was a *width parameter* that depended on the input semidefinite program (and was defined differently for each of the algorithms). For the specific instances of the semidefinite programs arising out of the applications considered in [26; 36; 37], it was separately argued that the corresponding width parameter κ is at most $\text{polylog}(N)$ and therefore the running time remained $\text{polylog}(N)$ (for constant ε). It is therefore desirable to remove the polynomial dependence on the width parameter and obtain a truly polylog running time algorithm, for a reasonably large subclass of semidefinite programs.

We will introduce parallel commputation, and then describe positive semidefinite programs and mixed packing and covering semidefinite programs in this chapter. And in the subsequent two chapters, we will present a fast parallel approximation algorithm for each of them.

2.1 Parallel computation

To design fast parallel approximation algorithms, we will make use of various facts concerning parallel computation. Note that the complexity class NC contains all the func-

tions that can be computed by logarithmic-space uniform Boolean circuits of polylogarithmic depth. Many matrix operations can be performed by NC algorithms. Here we make an assumption that the entries of all the matrices we consider have rational real and imaginary parts. First, the elementary matrix operations, such as addition, multiplication, inversion can be implemented by NC algorithm. We refer the readers to von zur Gathen's survey[72] for more details. Second, matrix exponentials and spectral decompositions can be approximated with high accuracy in NC. More precisely, the following two problems are in NC.

- Matrix exponentials. Given input an $n \times n$ matrix M, a rational number $\varepsilon > 0$ and an integer number k expressed in unary notation (i.e. 1^k) satisfying $\|M\| \leq k$, output an $n \times n$ matrix X such that $\|\exp(M) - X\| \leq \varepsilon$.
- Spectral decompositions. Given input an $n \times n$ matrix M and a rational number $\varepsilon > 0$, output an $n \times n$ unitary matrix U and an $n \times n$ diagonal matrix Γ such that

$$\|M - U\Gamma U^*\| \leq \varepsilon.$$

Readers can refer to [26; 36] for more discussion.

2.2 Positive semidefinite programs

A positive semidefinite program can be expressed in the following standard form (we use symbols $\geq, \leq$ to also represent Löwner order, where $A \geq B$ means $A - B$ is positive semidefinite).

Primal problem P		Dual problem D	
minimize:	$\operatorname{Tr} CX$	maximize:	$\sum_{i=1}^{m} b_i y_i$
subject to:	$\forall i \in [m] : \operatorname{Tr} A_i X \geq b_i,$	subject to:	$\sum_{i=1}^{m} y_i \cdot A_i \leq C,$
	$X \geq 0.$		$\forall i \in [m] : y_i \geq 0.$

Here $C, A_1, \ldots, A_m$ are $n \times n$ positive semidefinite matrices and $b_1, \ldots, b_m$ are non-negative reals (in a general semidefinite program $C, A_1, \ldots, A_m$ are Hermitian and $b_1, \ldots, b_m$ are

reals). Let us assume that the conditions for strong duality are satisfied and the optimum value for P, denoted $\text{opt}(P)$, equals the optimum value for D, denoted $\text{opt}(D)$. Assume w.l.o.g $m \geq n$ (by repeating the first constrain in P if necessary).

We will show that the problem can be transformed to the following special form in parallel polylog time.

Special form Primal problem $\hat{P}$

$$\begin{aligned} \text{minimize:} \quad & \operatorname{Tr} \hat{X} \\ \text{subject to:} \quad & \forall i \in [m] : \operatorname{Tr} \hat{A}_i \hat{X} \geq 1, \\ & \hat{X} \geq 0. \end{aligned}$$

Lemma 2.2.1. *Let $\hat{X}$ be a feasible solution to $\hat{P}$ such that $\operatorname{Tr} \hat{X} \leq (1+\varepsilon)\text{opt}(\hat{P})$. For any $\varepsilon > 0$, a feasible solution X to P can be derived from $\hat{X}$ such that $\operatorname{Tr} X \leq (1+\varepsilon)^2 \text{opt}(P)$. Furthermore, X can be obtained from $\hat{X}$ in parallel time $\text{polylog}(m)$.*

Given the positive semidefinite program (P, D) as above, we first show that without loss of generality (P, D) can be in the following special form.

Special form Primal problem P

$$\begin{aligned} \text{minimize:} \quad & \operatorname{Tr} X \\ \text{subject to:} \quad & \forall i \in [m] : \operatorname{Tr} A_i X \geq 1, \\ & X \geq 0. \end{aligned}$$

Special form Dual problem D

$$\begin{aligned} \text{maximize:} \quad & \sum_{i=1}^{m} y_i \\ \text{subject to:} \quad & \sum_{i=1}^{m} y_i \cdot A_i \leq I, \\ & \forall i \in [m] : y_i \geq 0. \end{aligned}$$

Here $A_1, \ldots, A_m$ are $n \times n$ positive semidefinite matrices and I represents the identity matrix. Furthermore, for all i, norm of A_i, denoted $\|A_i\|$, is at most 1 and the minimum non-zero eigenvalue of A_i is at least $\frac{1}{\gamma}$ where $\gamma = \frac{m^2}{\varepsilon^2}$.

We show how to transform the primal problem to the special form and a similar transformation can be applied to dual problem. First observe that if for some i, $b_i = 0$, the corresponding constraint in primal problem is trivial and can be removed. Similarly if for some i, the support of A_i is not contained in the support of C, then y_i must be 0 and can be removed. Therefore we can assume w.l.o.g. that for all $i, b_i > 0$ and the support of A_i is contained in the support of C. Hence w.l.o.g we can take the support of C as the

whole space, in other words, C is invertible. For all $i \in [m]$, define $A'_i \stackrel{\text{def}}{=} \frac{C^{-1/2} A_i C^{-1/2}}{b_i}$. Consider the normalized Primal problem.

Normalized Primal problem P'

$$\begin{aligned} \text{minimize:} \quad & \operatorname{Tr} X' \\ \text{subject to:} \quad & \forall i \in [m] : \operatorname{Tr} A'_i X' \geq 1, \\ & X' \geq 0. \end{aligned}$$

Hence, we have the following claim.

Claim 2.2.2. If X is a feasible solution to P, then $C^{1/2} X C^{1/2}$ is a feasible solution to P' with the same objective value. Similarly if X' is a feasible solution to P', then $C^{-1/2} X' C^{-1/2}$ is a feasible solution to P with the same objective value. Hence $\text{opt}(P) = \text{opt}(P')$.

The next step to transforming the problem is to limit the range of eigenvalues of A'_is. Let $\beta = \min_i \|A'_i\|$.

Claim 2.2.3. $\frac{1}{\beta} \leq \text{opt}(P') \leq \frac{m}{\beta}$.

Proof. Note that $\frac{1}{\beta} I$ is a feasible solution for P'. This implies $\text{opt}(P') \leq \frac{n}{\beta} \leq \frac{m}{\beta}$. Let X' be an optimal feasible solution for P'. Let j be such that $\|A'_j\| = \beta$. Then $\beta \operatorname{Tr} X' \geq \operatorname{Tr} A'_j X' \geq 1$, hence $\frac{1}{\beta} \leq \text{opt}(P')$. □

Let $A'_i = \sum_{j=1}^{n} a'_{ij} |v_{ij}\rangle\langle v_{ij}|$ be the spectral decomposition of A'_i. Define for all $i \in [m]$ and $j \in [n]$,

$$a''_{ij} \stackrel{\text{def}}{=} \begin{cases} \frac{\beta m}{\varepsilon} & \text{if } a'_{ij} > \frac{\beta m}{\varepsilon}, \\ 0 & \text{if } a'_{ij} < \frac{\varepsilon \beta}{m}, \\ a'_{ij} & \text{otherwise.} \end{cases} \tag{2.1}$$

Define $A''_i = \sum_{j=1}^{n} a''_{ij} |v_{ij}\rangle\langle v_{ij}|$. Consider the transformed Primal problem P''.

Transformed Primal problem P''

$$\begin{aligned} \text{minimize:} \quad & \operatorname{Tr} X'' \\ \text{subject to:} \quad & \forall i \in [m] : \operatorname{Tr} A''_i X'' \geq 1, \\ & X'' \geq 0. \end{aligned}$$

Lemma 2.2.4. *1. Any feasible solution to P'' is also a feasible solution to P'.*

2. $\mathrm{opt}(P') \leq \mathrm{opt}(P'') \leq \mathrm{opt}(P')(1+\varepsilon)$.

Proof. 1. Follows immediately from the fact that $A_i'' \leq A_i'$.

2. First inequality follows from 1. Let X' be an optimal solution to P' and let $\tau = \mathrm{Tr}(X')$. Let $X'' = X' + \frac{\varepsilon\tau}{m} I$. Then, since $m \geq n$, $\mathrm{Tr}\, X'' \leq (1+\varepsilon)\,\mathrm{Tr}\, X'$. Thus it suffices to show that X'' is feasible to P''.

   Fix $i \in [m]$. Assume that there exists $j \in [n]$ such that $a_{ij}' \geq \frac{\beta m}{\varepsilon}$. Then, from Claim 2.2.3

$$\mathrm{Tr}\, A_i'' X_i'' \geq \mathrm{Tr}\, \frac{\beta m}{\varepsilon} |v_{ij}\rangle\langle v_{ij}| \cdot \frac{\varepsilon\tau}{m} I = \beta\tau \geq 1.$$

   Now assume that for all $j \in [n]$, $a_{ij} \leq \frac{\beta m}{\varepsilon}$. By (2.1) and definition of β, $\|A_i''\| = \|A_i'\| \geq \beta$ and $A_i'' \geq A_i' - \frac{\varepsilon\beta}{m} I$. Therefore

$$\begin{aligned} \mathrm{Tr}\, A_i'' X_i'' &\geq \mathrm{Tr}\, A_i'' X' + \beta\frac{\varepsilon\tau}{m} \\ &\geq \mathrm{Tr}\, A_i' X' + \beta\frac{\varepsilon\tau}{m} - \mathrm{Tr}\, \frac{\varepsilon\beta}{m} X' = \mathrm{Tr}\, A_i' X' \geq 1. \end{aligned}$$

□

Note that for all $i \in [m]$, the ratio between the largest eigenvalue and the smallest nonzero eigenvalue of A_i'' is at most $\frac{m^2}{\varepsilon^2} = \gamma$.

Finally, we get the special form Primal problem $\hat{P}$ as follows. Let $t = \max_{i\in[m]} \|A_i''\|$ and for all $i \in [m]$ define $\hat{A}_i \stackrel{\mathrm{def}}{=} \frac{A_i''}{t}$. Consider,

Special form Primal problem $\hat{P}$

$$\begin{aligned} \text{minimize:} \quad & \mathrm{Tr}\, \hat{X} \\ \text{subject to:} \quad & \forall i \in [m] : \mathrm{Tr}\, \hat{A}_i \hat{X} \geq 1, \\ & \hat{X} \geq 0. \end{aligned}$$

It is easily seen that there is a one-to-one correspondence between the feasible solutions to P'' and $\hat{P}$ and $\mathrm{opt}(\hat{P}) = t \cdot \mathrm{opt}(P'')$. Furthermore, X can be obtained from $\hat{X}$ in parallel time $\mathrm{polylog}(m)$ since all the operations involved can be implemented in NC circuits and the number of operations is$\mathrm{polylog}(m)$. Therefore $\hat{P}$ satisfies all the properties that we want and cumulating all we have shown above, we get Lemma 2.2.1.

2.3 Mixed packing and covering

Mixed packing and covering is a more general optimization problem, which can be formalized as the following feasibility problem.

Q1: Given $n \times n$ positive semidefinite matrices $P_1, \ldots, P_m, P$ and non-negative diagonal matrices $C_1, \ldots, C_m, C$ and $\varepsilon \in (0, 1)$, find an vector $x \geq 0$ such that

$$\sum_{i=1}^{m} x_i P_i \leq (1+\varepsilon)P \quad \text{and} \quad \sum_{i=1}^{m} x_i C_i \geq C$$

or show that the following is infeasible

$$\sum_{i=1}^{m} x_i P_i \leq P \quad \text{and} \quad \sum_{i=1}^{m} x_i C_i \geq C \ .$$

Given a fast parallel approximation algorithm for **Q1**, we can obtain a fast parallel approximation algorithm for the following optimization problem by the standard binary search method.

Q2: Given $n \times n$ positive semidefinite matrices $P_1, \ldots, P_m, P$ and non-negative diagonal matrices $C_1, \ldots, C_m, C$,

$$\begin{aligned} \text{maximize:} \quad & \gamma \\ \text{subject to:} \quad & \sum_{i=1}^{m} x_i P_i \leq P \\ & \sum_{i=1}^{m} x_i C_i \geq \gamma C \\ & \forall i \in [m] : x_i \geq 0. \end{aligned}$$

The following special case of **Q2** is positive semidefinite programs.

Q3: Given $n \times n$ positive semidefinite matrices $P_1, \ldots, P_m, P$ and non-negative scalars $c_1, \ldots, c_m$,

$$\begin{aligned} \text{maximize:} \quad & \sum_{i=1}^{m} x_i c_i \\ \text{subject to:} \quad & \sum_{i=1}^{m} x_i P_i \leq P \\ & \forall i \in [m] : x_i \geq 0. \end{aligned}$$

Chapter 3

A parallel approximation algorithm for positive semidefinite programming

3.1 Introduction

In this chapter, we consider the class of positive semidefinite programs given in Chapter 2 Section 2.2. We present an algorithm, which given as input, $(C, A_1, \ldots, A_m, b_1, \ldots, b_m)$, and an error parameter $\varepsilon > 0$, outputs a $(1+\varepsilon)$ approximation to the optimum value of the program, and has running time $\text{polylog}(n) \cdot \text{polylog}(m) \cdot \text{poly}(\frac{1}{\varepsilon})$. As can be noted, there is no polynomial dependence on any 'width' parameter on the running time of our algorithm.

Our algorithm is inspired by the algorithm used by Luby and Nisan [53] to solve positive linear programs. Positive linear programs can be considered as a special case of positive semidefinite programs in which the matrices used in the description of the program are all pairwise commuting. Our algorithm (and the algorithm in [53]) is based on the *multiplicative weights update* (MWU) method. This is a powerful technique for *experts learning* and finds its origins in various fields including learning theory, game theory, and optimization. The algorithms used in [3; 4; 26; 36; 37; 42] are based on its matrix variant the *matrix multiplicative weights update* method.

The algorithm starts with feasible primal variable X and feasible dual variable $(y_1, \cdots, y_m)$. The algorithm proceeds in phases, where in each phase the large eigenvalues of $\sum_{i=1}^{m} y_i^t A_i$ (X^t, y_i^ts represent the candidate primal and dual variables at time t, respectively) are

sought to be brought below a threshold determined for that phase. The primal variable X^t at time step t is chosen to be the projection onto the large eigenvalues (above the threshold) eigenspace of $\sum_{i=1}^{m} y_i^t A_i$. Using the sum of the primal variables generated so far, the dual variables are updated using the MWU method. A suitable scaling parameter λ_t is chosen during this update, which is small enough so that the change of dual objective value $\sum_{i=1}^{m} y_i$ at each update is small. It ensures that the output of the algorithm is a good approximation solution if the program is feasible. At the same time, λ_t is large enough so that there is reasonable progress in bringing down the large eigenvalues of $\sum_{i=1}^{m} y_i^t A_i$. This guarantees that only polylog number of phases are needed.

Due to the non-commutative nature of the matrices involved in our case, our algorithm primarily deviates from that of [53] in how the threshold is determined inside each phase. The problem that is faced is roughly as follows. Since A_i's could be non-commuting, when y_i^ts are scaled down, the sum of the large eigenvalues of $\sum_{i=1}^{m} y_i^t A_i$ may not come down and this scaling may just move the large eigenvalues eigenspace. Therefore a suitable extra condition needs to be ensured while choosing the threshold. Due to this, our analysis also primarily deviates from [53] in bounding the number of time steps required in any phase and is significantly more involved. The analysis requires us to study the relationship between the large eigenvalues eigenspaces before and after scaling (say W_1 and W_2). For this purpose we consider the decomposition of the underlying space into one and two-dimensional subspaces which are invariant under the actions of both Π_1 and Π_2 (projections onto W_1 and W_2 respectively) and this helps the analysis significantly. Such decomposition has been quite useful in earlier works as well for example in quantum walk [1; 64; 69] and quantum complexity theory [54; 55]. The result is improved later by Jain and Yao in [38], which is given in Chapter 4. However, the techniques used here are interesting in their own right.

We present the algorithm in the next section and its analysis, both optimality and the running time, in the subsequent section.

3.2 Algorithm

By Lemma 2.2.1, We may start with the following special positive semidefinte programs.

Special form Primal problem P

$$\begin{aligned} \text{minimize:} \quad & \operatorname{Tr} X \\ \text{subject to:} \quad & \forall i \in [m] : \operatorname{Tr} A_i X \geq 1, \\ & X \geq 0. \end{aligned}$$

Special form Dual problem D

$$\begin{aligned} \text{maximize:} \quad & \sum_{i=1}^{m} y_i \\ \text{subject to:} \quad & \sum_{i=1}^{m} y_i \cdot A_i \leq I, \\ & \forall i \in [m] : y_i \geq 0. \end{aligned}$$

In order to compactly describe the algorithm, and also the subsequent analysis, we introduce some notation. Let $Y = \operatorname{Diag}(y_1, \ldots, y_m)$ ($m \times m$ diagonal matrix with $Y(i,i) = y_i$ for $i \in [m]$). Let Φ be the map (from $n \times n$ positive semidefinite matrices to $m \times m$ positive semidefinite diagonal matrices) defined by $\Phi(X) = \operatorname{Diag}(\operatorname{Tr} A_1 X, \ldots, \operatorname{Tr} A_m X)$. Then its adjoint map Φ^* acts as $\Phi^*(Y) = \sum_{i=1}^{m} Y(i,i) \cdot A_i$ (for all diagonal matrices $Y \geq 0$). We let I represent the identity matrix (in the appropriate dimensions clear from the context). For Hermitian matrix B and real number l, let $N_l(B)$ represent the sum of eigenvalues of B which are at least l. The algorithm is mentioned in Figure 3.1.

3.3 Analysis

For all of this section, let $\varepsilon_1 = \frac{3\varepsilon}{\ln n}$. In the following we assume $\varepsilon < \frac{1}{1000}$ and $n > 1000$.

3.3.1 Optimality

In this section we present the analysis assuming that all the operations performed by the algorithm are perfect. Note that the algorithm only involves elementary matrix operations (addition, substraction and multiplication), matrix exponentials and matrix spectral decomposition. All those operation can be performed with high precision. And the number of operations is polylog to the size of inputs, which will be shown in the next subsection. We claim, without going into further details, that similar analysis can be performed while taking into account the accuracy loss due to the actual operations of the algorithm in the limited running time.

We start with following claims.

Claim 3.3.1. For all $t \leq t_f$, λ_t satisfies the conditions 1. and 2. in Step (3d) in the Algorithm.

Proof. Easily verified. □

Input : Positive semidefinite matrices $A_1, \ldots, A_m$ and error parameter $\varepsilon > 0$.

Output : X^* feasible for P and Y^* feasible for D.

1. Let $\varepsilon_0 = \frac{\varepsilon^2}{\ln^2 n}$, $t = 0, X_0 = 0$. Let k_s be the smallest positive number such that $(1+\varepsilon_0)^{k_s} \leq \|\Phi^*(I)\| < (1+\varepsilon_0)^{k_s+1}$. Let $k = k_s$.

2. Let $Y_t = \exp(-\Phi(X_t))$.

3. If $\operatorname{Tr} Y_t > \frac{1}{m^{1/\varepsilon}}$, do

   (a) If $\|\Phi^*(Y_t)\| < (1+\varepsilon_0)^k$, then set $k \leftarrow k-1$ and repeat this step.

   (b) Set $\mathsf{thr}' = k$.

   (c) If
   $$N_{(1+\varepsilon_0)^{\mathsf{thr}'-1}}(\Phi^*(Y_t)) \geq (1+\frac{2}{5}\varepsilon) N_{(1+\varepsilon_0)^{\mathsf{thr}'}}(\Phi^*(Y_t)).$$
   then $\mathsf{thr}' \leftarrow \mathsf{thr}' - 1$ and repeat this step. Else set $\mathsf{thr} = \mathsf{thr}'$.

   (d) Let Π_t be the projector on the eigenspace of $\Phi^*(Y_t)$ with eigenvalues at least $(1+\varepsilon_0)^{\mathsf{thr}}$. For $\lambda > 0$, let $P_\lambda^{\geq}$ be the projection onto eigenspace of $\Phi(\lambda\Pi_t)$ with eigenvalues at least $2\sqrt{\varepsilon}$. Let $P_\lambda^{\leq}$ be the projection onto eigenspace of $\Phi(\lambda\Pi_t)$ with eigenvalues at most $2\sqrt{\varepsilon}$. Find λ_t such that

   1. $\operatorname{Tr}(P_{\lambda_t}^{\geq} Y_t P_{\lambda_t}^{\geq})\Phi(\Pi_t) \geq \sqrt{\varepsilon} \operatorname{Tr} Y_t \Phi(\Pi_t)$ and,
   2. $\operatorname{Tr}(P_{\lambda_t}^{\leq} Y_t P_{\lambda_t}^{\leq})\Phi(\Pi_t) \geq (1-\sqrt{\varepsilon}) \operatorname{Tr} Y_t \Phi(\Pi_t)$ as follows.

   i. Sort $\{\operatorname{Tr} A_i \Pi_t\}_{i=1}^m$ in non-increasing order. Suppose $\operatorname{Tr} A_{j_1}\Pi_t \geq \operatorname{Tr} A_{j_2}\Pi_t \geq \cdots \geq \operatorname{Tr} A_{j_m}\Pi_t$.

   ii. Let y_j be the j-th diagonal entry of Y_j. Find index $r \in [m]$ satisfying
   $$\sum_{k=1}^{r} y_{j_k} \operatorname{Tr} A_{j_k}\Pi_t \geq \sqrt{\varepsilon} \sum_{k=1}^{m} y_{j_k} \operatorname{Tr} A_{j_k}\Pi_t, \text{ and}$$
   $$\sum_{k=r}^{m} y_{j_k} \operatorname{Tr} A_{j_k}\Pi_t \geq (1-\sqrt{\varepsilon}) \sum_{k=1}^{m} y_{j_k} \operatorname{Tr} A_{j_k}\Pi_t.$$

   iii. Let $\lambda_t = \frac{2\sqrt{\varepsilon}}{\operatorname{Tr} A_{j_r}\Pi_t}$.

   (e) Let $X_{t+1} = X_t + \lambda_t \Pi_t$. Set $t \leftarrow t+1$ and go to Step 2.

4. Let $t_f = t$, $k_f = k$. Let α be the minimum eigenvalue of $\Phi(X_{t_f})$. Output $X^* = X_{t_f}/\alpha$.

5. Let t' be such that $\operatorname{Tr} Y_{t'} / \|\Phi^*(Y_{t'})\|$ is the maximum among all time steps. Output $Y^* = Y_{t'} / \|\Phi^*(Y_{t'})\|$.

Figure 3.1: Algorithm

Claim 3.3.2. $\alpha > 0$.

Proof. Follows since $\frac{1}{m^{1/\varepsilon}} \geq \operatorname{Tr} Y_{t_f} = \operatorname{Tr} \exp(-\Phi(X_{t_f})) > \exp(-\alpha)$. □

Following lemma shows that for any time t, $\|\Phi^*(Y_t)\|$ is not much larger than $(1+\varepsilon_0)^{\mathsf{thr}}$.

Lemma 3.3.3. *For all $t \leq t_f$, $\|\Phi^*(Y_t)\| \leq (1+\varepsilon_0)^{\mathsf{thr}}(1+\varepsilon_1)$.*

Proof. Fix any $t \leq t_f$. As $\operatorname{Tr}(\Phi^*(Y_t)) \leq nN_{(1+\varepsilon_0)^k}(\Phi^*(Y_t))$, the loop at Step 3(c) runs at most $\frac{\ln n}{\ln(1+\frac{2\varepsilon}{5})}$ times. Hence

$$\begin{aligned}\|\Phi^*(Y_t)\| &\leq (1+\varepsilon_0)^{k+1} \leq (1+\varepsilon_0)^{\mathsf{thr}}(1+\varepsilon_0)^{\frac{\ln n}{\ln(1+\frac{2\varepsilon}{5})}+1} \\ &< (1+\varepsilon_0)^{\mathsf{thr}}(1+\frac{3\varepsilon}{\ln n}) = (1+\varepsilon_0)^{\mathsf{thr}}(1+\varepsilon_1).\end{aligned}$$

□

Following lemma shows that as t increases, there is a reduction in the trace of the dual variable in terms of the trace of the primal variable.

Lemma 3.3.4. *For all $t \leq t_f$ we have,* $\operatorname{Tr} Y_{t+1} \leq \operatorname{Tr} Y_t - \lambda_t \cdot (1-4\sqrt{\varepsilon}) \cdot \|\Phi^*(Y_t)\| \cdot (\operatorname{Tr} \Pi_t)$.

Proof. Fix any $t \leq t_f$. Let $B = P^{\leq}_{\lambda_t} \Phi(\lambda_t \Pi_t) P^{\leq}_{\lambda_t}$. Note that $B \leq \Phi(\lambda_t \Pi_t)$ and also $B \leq 2\sqrt{\varepsilon} I$. Second last inequality below follows from Lemma 3.3.3 which shows that all eigenvalues of $\Pi_t \Phi^*(Y_t) \Pi_t$ are at least $(1-\varepsilon_1)\|\Phi^*(Y_t)\|$.

$$\begin{aligned}\operatorname{Tr} Y_{t+1} &= \operatorname{Tr} \exp(-\Phi(X_t) - \Phi(\lambda_t \Pi_t)) \\ &\leq \operatorname{Tr} \exp(-\Phi(X_t) - B) \\ &= \operatorname{Tr} \exp(-\Phi(X_t)) \exp(-B) \\ &\leq \operatorname{Tr} \exp(-\Phi(X_t))(I - (1-2\sqrt{\varepsilon})B) \\ &= \operatorname{Tr} Y_t - (1-2\sqrt{\varepsilon}) \operatorname{Tr} Y_t B \\ &\leq \operatorname{Tr} Y_t - (1-\sqrt{\varepsilon})(1-2\sqrt{\varepsilon}) \operatorname{Tr} Y_t \Phi(\lambda_t \Pi_t) \\ &= \operatorname{Tr} Y_t - (1-\sqrt{\varepsilon})(1-2\sqrt{\varepsilon}) \operatorname{Tr} \Phi^*(Y_t) \lambda_t \Pi_t \\ &\leq \operatorname{Tr} Y_t - (1-\varepsilon_1)(1-\sqrt{\varepsilon})(1-2\sqrt{\varepsilon}) \lambda_t \|\Phi^*(Y_t)\| (\operatorname{Tr} \Pi_t) \\ &\leq \operatorname{Tr} Y_t - (1-4\sqrt{\varepsilon}) \lambda_t \|\Phi^*(Y_t)\| (\operatorname{Tr} \Pi_t).\end{aligned}$$

The first inequality holds because $A_1 \geq A_2$ implies $\operatorname{Tr} \exp(A_1) \geq \operatorname{Tr} \exp(A_2)$, the second equality because both B and $\Phi(X_t)$ are diagonal, the second inequality because $A \leq I$ implies $\exp(-\delta A) \leq I - \delta(1-\delta)A)$, and the third inequality is from step 3(d) part 1. □

Following lemma relates the trace of X_{t_f} with the trace of Y^* and Y_{t_f}.

Lemma 3.3.5. $\operatorname{Tr} X_{t_f} \leq \frac{1}{(1-4\sqrt{\varepsilon})} \cdot (\operatorname{Tr} Y^*) \cdot \ln(m/\operatorname{Tr} Y_{t_f})$.

Proof. Using Lemma 3.3.4 we have,

$$\begin{aligned}
\frac{\operatorname{Tr} Y_{t+1}}{\operatorname{Tr} Y_t} &\leq 1 - \frac{(1-4\sqrt{\varepsilon})\lambda_t \left\|\Phi^*(Y_t)\right\| (\operatorname{Tr} \Pi_t)}{\operatorname{Tr} Y_t} \\
&\leq \exp\left(-\frac{(1-4\sqrt{\varepsilon})\lambda_t \left\|\Phi^*(Y_t)\right\| (\operatorname{Tr} \Pi_t)}{\operatorname{Tr} Y_t}\right) \\
&\leq \exp\left(-\frac{(1-4\sqrt{\varepsilon})\lambda_t \operatorname{Tr} \Pi_t}{\operatorname{Tr} Y^*}\right) \\
&= \exp\left(-\frac{(1-4\sqrt{\varepsilon}) \operatorname{Tr}(X_{t+1} - X_t)}{\operatorname{Tr} Y^*}\right).
\end{aligned}$$

The second inequality holds because $\exp(-x) \geq 1 - x$, and second inequality is from property of Y^*. This implies,

$$\begin{aligned}
\operatorname{Tr} Y_{t_f} &\leq (\operatorname{Tr} Y_0) \exp\left(-\frac{(1-4\sqrt{\varepsilon}) \operatorname{Tr} X_{t_f}}{\operatorname{Tr} Y^*}\right) \\
\Rightarrow \operatorname{Tr} X_{t_f} &\leq \frac{(\operatorname{Tr} Y^*) \ln(m/(\operatorname{Tr} Y_{t_f}))}{(1-4\sqrt{\varepsilon})} \quad (\text{since } \operatorname{Tr} Y_0 = m).
\end{aligned}$$

□

We can now finally bound the trace of X^* in terms of the trace of Y^*.

Theorem 3.3.6. *X^* and Y^* are feasible for the P and D respectively and*

$$\operatorname{Tr} X^* \leq (1+5\sqrt{\varepsilon}) \operatorname{Tr} Y^* .$$

Therefore, since $\operatorname{opt}(P) = \operatorname{opt}(D)$,

$$\begin{aligned}
\operatorname{opt}(D) = \operatorname{opt}(P) &\leq \operatorname{Tr} X^* \leq (1+5\sqrt{\varepsilon}) \operatorname{Tr} Y^* \\
&\leq (1+5\sqrt{\varepsilon}) \operatorname{opt}(D) = (1+5\sqrt{\varepsilon}) \operatorname{opt}(P).
\end{aligned}$$

Proof. Note that $\Phi(X^*) = \Phi(X_{t_f})/\alpha \geq I$ and $\Phi^*(Y^*) = \Phi^*(Y_{t'})/\left\|\Phi^*(Y_{t'})\right\| \leq I$. X^* and Y^* are feasible for P and D respectively. From Lemma 3.3.5 we have,

$$\alpha \operatorname{Tr} X^* = \operatorname{Tr} X_{t_f} \leq \frac{1}{1-4\sqrt{\varepsilon}} \cdot (\operatorname{Tr} Y^*) \cdot \ln(m/\operatorname{Tr} Y_{t_f}) .$$

Since $Y_{t_f} = \exp(-\Phi(X_{t_f}))$ we have

$$\operatorname{Tr} Y_{t_f} \geq \left\|\exp(-\Phi(X_{t_f}))\right\| = \exp(-\alpha) \ .$$

Using above two equations we have,

$$\begin{aligned}
\operatorname{Tr} X^* &\leq \frac{1}{1-4\sqrt{\varepsilon}} \cdot (\operatorname{Tr} Y^*) \cdot \frac{\ln(m/\operatorname{Tr} Y_{t_f})}{\ln(1/\operatorname{Tr} Y_{t_f})} \\
&= \frac{1}{1-4\sqrt{\varepsilon}} \cdot (\operatorname{Tr} Y^*) \cdot \left(1 + \frac{\ln m}{\ln(1/\operatorname{Tr} Y_{t_f})}\right) \\
&\leq \frac{1+\varepsilon}{1-4\sqrt{\varepsilon}} \cdot (\operatorname{Tr} Y^*) \quad (\text{since } \operatorname{Tr} Y_{t_f} \leq \frac{1}{m^{1/\varepsilon}}) \\
&\leq (1+5\sqrt{\varepsilon}) \cdot \operatorname{Tr} Y^* \ .
\end{aligned}$$

□

3.3.2 Time complexity

In this section we are primarily interested in bounding the number of iterations of the algorithm, that is we will bound k_f and also the number of iterations for any given k. We claim, without going into further details, that the actions required by the algorithm in any given iteration can all be performed in time $\text{polylog}(n) \cdot \text{polylog}(m) \cdot \text{poly}(\frac{1}{\varepsilon})$, since operations for Hermitian matrices like eigenspace decomposition, exponentiation, and other operations like sorting and binary search for a list of real numbers etc. can be all be performed in polylog parallel time.

Let us first introduce some notation. Let A be a Hermitian matrix and l be a real number. Let

- Π_l^A denote the projector onto the space spanned by the eigenvectors of A with eigenvalues at least l. Let Π^A be shorthand for Π_1^A.
- $N_l(A)$ denote the sum of eigenvalues of A at least l. Thus $N_l(A) = \operatorname{Tr} \Pi_l^A A$. Let $N(A)$ be shorthand for $N_1(A)$.
- $\lambda_k(A)$ denote the k-th largest eigenvalue of A.
- $\lambda^{\downarrow}(A) \stackrel{\text{def}}{=} (\lambda_1(A), \cdots, \lambda_n(A))$.
- for any two vectors $u, v \in \mathcal{R}^n$ we say u majorizes v, denoted $u \succeq v$, iff $\sum_{i=1}^k u_i = \sum_{i=1}^k v_i$ and for any $j \in [n]$ we have, $\sum_{i=1}^j u_i \geq \sum_{i=1}^j v_i$.

We need the following facts.

Fact 3.3.7. [11] For $n \times n$ Hermitian matrices A and B, $A \geq B$ implies $\lambda_i(A) \geq \lambda_i(B)$ for all $1 \leq i \leq n$. Thus $N_l(A) \geq N_l(B)$ for any real number l.

Fact 3.3.8. [11] Let A be an $n \times n$ Hermitian matrix and $P_1, \cdots, P_r$ be a family of mutually orthogonal projections. Then $\lambda^{\downarrow}(A) \succeq \lambda^{\downarrow}(\sum_i P_i A P_i)$.

Fact 3.3.9. [41] For any two projectors Π and Δ, there exits an orthogonal decomposition of the underlying vector space into one dimensional and two dimensional subspaces that are invariant under both Π and Δ. Moreover, inside each two-dimensional subspace, Π and Δ are rank-one projectors.

Lemma 3.3.10. *Let k_f be the final value of k. Then $k_s - k_f = \mathcal{O}(\frac{\log m \log^2 n}{\varepsilon^3})$.*

Proof. Note that $\|\Phi^*(I)\| = \|\sum_{i=1}^m A_i\| \leq m$, since for each i, $\|A_i\| \leq 1$. Hence

$$k_s = \mathcal{O}((\log m)/\varepsilon_0) \ .$$

Let $Y_{t_f-1} = \mathrm{Diag}(y_1, \ldots y_m)$. We have (since $\operatorname{Tr} A_i \geq \frac{1}{\gamma} \geq \frac{\varepsilon^2}{m^2}$ for each i),

$$\begin{aligned} & m(1+\varepsilon_0)^{k_f+1} \geq m \left\|\Phi^*(Y_{t_f-1})\right\| \geq \operatorname{Tr} \Phi^*(Y_{t_f-1}) \\ =& \sum_{i=1}^m y_i \operatorname{Tr} A_i \geq \frac{\sum_{i=1}^m y_i}{\gamma} = \frac{\operatorname{Tr} Y_{t_f-1}}{\gamma} \\ \geq& \frac{1}{m^{1/\varepsilon}\gamma} \geq \frac{\varepsilon^2}{m^{2+1/\varepsilon}} \ . \end{aligned}$$

Hence $k_f \geq -\mathcal{O}(\frac{\log m}{\varepsilon\varepsilon_0})$. Therefore $k_s - k_f = \mathcal{O}(\frac{\log m}{\varepsilon\varepsilon_0}) = \mathcal{O}(\frac{\log m \log^2 n}{\varepsilon^3})$. □

Theorem 3.3.11. *For any fixed k, the number of iterations of the algorithm is at most $\mathcal{O}(\frac{\log^2 n}{\varepsilon_1^9 \varepsilon})$. Hence combined with Lemma 3.3.10, the total number of iterations of the algorithm is at most $\mathcal{O}(\frac{\log^{13} n \log m}{\varepsilon^{13}})$.*

Proof. Fix k. Assume that the Algorithm has reached step $3(d)$ for this fixed k , $\frac{6 \log^2 n}{\varepsilon_1^9 \varepsilon}$ times. As argued in the proof of Lemma 3.3.4, whenever Algorithm reaches step $3(d)$, $\mathsf{thr} \geq k - \frac{3 \ln n}{\varepsilon}$. Thus there exists a value s between k and $k - \frac{3 \ln n}{\varepsilon}$ such that $\mathsf{thr} = s$ at least $\frac{2 \log n}{\varepsilon_1^9}$ times.

From Lemma 3.3.3 we get that the sum of the eigenvalues above $(1+\varepsilon_0)^s$, is at most $n(1+\varepsilon_1)(1+\varepsilon_0)^s$ at the beginning of this phase. Whenever $\mathsf{thr} \neq s$ in this phase, using

Fact 3.3.7, we conclude that the eigenvalues of $\Phi^*(Y_t)$ above $(1+\varepsilon_0)^s$ do not increase. Whenever $\mathsf{thr} = s$ in this phase, using Lemma 3.3.12, we conclude that the eigenvalues of $\Phi^*(Y_t)$ above $(1+\varepsilon_0)^s$ reduce by a factor of $(1-\varepsilon_1^9)$. This can be seen by letting A in Lemma 3.3.12 to be $\frac{1-\exp(-2\sqrt{\varepsilon})}{(1+\varepsilon_0)^s} \cdot \Phi^*(P^{\geq}_{\lambda_t} Y_t P^{\geq}_{\lambda_t})$ and B to be $\frac{1}{(1+\varepsilon_0)^s}\Phi^*(Y^t) - A$. Now condition $3(d)(1.)$ of the Algorithm gives condition (2) of Lemma 3.3.12. Condition (1) of Lemma 3.3.12 can also be seen to be satisfied (using Lemma 3.3.3) and condition (4) of Lemma 3.3.12 is false due to condition $3(c)$ of the Algorithm. This implies condition (3) of Lemma 3.3.12 must also be false which gives us the desired conclusion.

Therefore the eigenvalues of $\Phi^*(Y_t)$ above $(1+\varepsilon_0)^s$ (in particular above $(1+\varepsilon_0)^k$) will vanish before $\mathsf{thr} = s$, $\frac{2\log n}{\varepsilon_1^9}$ times. Hence k must decrease before the Algorithm has reached step $3(d)$, $\frac{6\log^2 n}{\varepsilon_1^9 \varepsilon}$ times. □

Following is a key lemma. It states that for two positive semidefinite matrices A, B, if A has good weight in the large (above 1) eigenvalues space of $A+B$ and if the sum of large (above 1) eigenvalues of B is pretty much the same as for $A+B$, then the sum of eigenvalues of $A+B$, slightly below 1 should be a constant fraction larger than the sum above 1.

Lemma 3.3.12. *Let $\varepsilon' = \frac{\varepsilon_0}{1+\varepsilon_0}$. Let A, B be two $n \times n$ positive semidefinite matrices satisfying*

$$\|A+B\| \leq 1+\varepsilon_1 \text{ and } \|B\| \geq 1, \tag{3.1}$$

$$\operatorname{Tr} \Pi^{A+B} A \geq \varepsilon \operatorname{Tr} \Pi^{A+B}(A+B), \text{ and} \tag{3.2}$$

$$\operatorname{Tr} \Pi^{B} B \geq (1-\varepsilon_1^9) \operatorname{Tr} \Pi^{A+B}(A+B). \tag{3.3}$$

Then

$$N_{1-\varepsilon'}(A+B) > (1+\frac{2}{5}\varepsilon) N(A+B). \tag{3.4}$$

Proof. In order to prove this Lemma we need to first show a few other Lemmas. By Fact 3.3.9, Π^B and Π^{A+B} decompose the underlying space V as follows,

$$V = \left(\bigoplus_{i=1}^{k} V_i \right) \bigoplus W.$$

Above for each $i \in [k]$, V_i is either one-dimensional or two-dimensional subspace, invariant for both Π^B and Π^{A+B} and inside V_i at least one of Π^B and Π^{A+B} survives. W is the subspace where both Π^B and Π^{A+B} vanish. We identify the subspace V_i and the

projector onto itself. For any matrix M, define M_i to be $V_i M V_i$. We can see that both the projectors Π^B and Π^{A+B} are decomposed into the direct sum of one-dimensional projectors as follows.

$$\Pi^B = \bigoplus_{i=1}^{k} \Pi_i^B \quad \text{and} \quad \Pi^{A+B} = \bigoplus_{i=1}^{k} \Pi_i^{A+B}.$$

Lemma 3.3.13. *For any $i \in [k]$, $\Pi^{B_i} = \Pi_i^B$ and $\Pi_i^{A+B} = \Pi^{A_i+B_i}$. That is, the eigenspace of B_i with eigenvalues at least 1, is exactly the restriction of Π^B to V_i and similarly for $A_i + B_i$.*

Proof. We prove $\Pi^{B_i} = \Pi_i^B$ and the other equality follows similarly. If $\dim V_i = 1$, i.e. $V_i = \text{span}\{|v\rangle\}$, then either $\Pi^B|v\rangle = |v\rangle$ or $\Pi^B|v\rangle = 0$. For the first case, $\Pi_i^B = |v\rangle\langle v|$, and $B_i = \langle v|B|v\rangle|v\rangle\langle v|$ and $\langle v|B|v\rangle \geq 1$, which means $\Pi^{B_i} = |v\rangle\langle v|$. For the second case, $\Pi_i^B = 0$, $\langle v|B|v\rangle < 1$, i.e. $\Pi^{B_i} = 0$.

For the case $\dim V_i = 2$,

$$\begin{aligned} B_i &= V_i B V_i = V_i(\Pi^B B \Pi^B + (I - \Pi^B)B(I - \Pi^B))V_i \\ &= V_i(\bigoplus_j \Pi_j^B)B(\bigoplus_j \Pi_j^B)V_i + \\ &\quad V_i((W \oplus \bigoplus_j (V_j - \Pi_j^B))B(W \oplus \bigoplus_j (V_j - \Pi_j^B)))V_i \\ &= \Pi_i^B B \Pi_i^B + (V_i - \Pi_i^B)B(V_i - \Pi_i^B). \end{aligned}$$

Let $\Pi_i^B = |v_1\rangle\langle v_1|$ and $V_i - \Pi_i^B = |v_0\rangle\langle v_0|$, then

$$B_i = \langle v_1|B|v_1\rangle|v_1\rangle\langle v_1| + \langle v_0|B|v_0\rangle|v_0\rangle\langle v_0| \tag{3.5}$$

is the spectral decomposition of B_i. As $\Pi^B|v_1\rangle = \Pi_i^B|v_1\rangle = |v_1\rangle$ and $\Pi^B|v_0\rangle = \Pi_i^B|v_0\rangle = 0$, we have $\langle v_1|B|v_1\rangle \geq 1$ and $\langle v_0|B|v_0\rangle < 1$, and hence $\Pi^{B_i} = |v_1\rangle\langle v_1|$. □

Lemma 3.3.14.

$$\text{Tr}\, \Pi^B B = \sum_{i=1}^{k} \text{Tr}\, \Pi^{B_i} B_i, \tag{3.6}$$

$$\text{Tr}\, \Pi^{A+B} B = \sum_{i=1}^{k} \text{Tr}\, \Pi^{A_i+B_i} B_i, \text{ and} \tag{3.7}$$

$$\operatorname{Tr}\Pi^{A+B}(A+B)=\sum_{i=1}^{k}\operatorname{Tr}\Pi^{A_i+B_i}(A_i+B_i) \tag{3.8}$$

Then using Eq.(3.2) and Eq.(3.3) we get,

$$\sum_{i=1}^{k}\operatorname{Tr}\Pi^{A_i+B_i}B_i\leq(1-\varepsilon)\sum_{i=1}^{k}\operatorname{Tr}\Pi^{A_i+B_i}(A_i+B_i). \tag{3.9}$$

$$\sum_{i=1}^{k}\operatorname{Tr}\Pi^{B_i}B_i\geq(1-\varepsilon_1^9)\sum_{i=1}^{k}\operatorname{Tr}\Pi^{A_i+B_i}(A_i+B_i). \tag{3.10}$$

Proof. We prove (3.6) and (3.7) and (3.8) follow similarly.

$$\begin{aligned}\operatorname{Tr}\Pi^{B}B&=\sum_{i=1}^{k}\operatorname{Tr}\Pi_i^{B}B=\sum_{i=1}^{k}\operatorname{Tr}V_i\Pi_i^{B}V_iB\\ &=\sum_{i=1}^{k}\operatorname{Tr}\Pi_i^{B}V_iBV_i=\sum_{i=1}^{k}\operatorname{Tr}\Pi_i^{B}B_i=\sum_{i=1}^{k}\operatorname{Tr}\Pi^{B_i}B_i.\end{aligned}$$

□

Remarks:

1. In any one-dimensional subspace $V_i=\operatorname{span}\{|v\rangle\}$ in the decomposition of V as above, if $\Pi^{A+B}|v\rangle=0$, then $\langle v|(A+B)|v\rangle<1$, which implies $\langle v|B|v\rangle<1$, that is $\Pi^{B}|v\rangle=0$. But this contradicts the fact that at least one of Π^{B} and Π^{A+B} does not vanish in V_i. Thus Π^{A+B} never vanishes in any of V_i. Therefore for all $i\in[k]$ we have $\operatorname{Tr}\Pi^{A_i+B_i}(A_i+B_i)=\operatorname{Tr}\Pi_i^{A+B}(A_i+B_i)\geq 1$.

2. From (3.1), for all $i\in[k]$, $\operatorname{Tr}\Pi^{A_i+B_i}(A_i+B_i)\leq 1+\varepsilon_1$. Combined with (3.8), we have
$$k\leq N(A+B)\leq k(1+\varepsilon_1).$$

Lemma 3.3.15. *Let*

$$I=\{i:\operatorname{Tr}\Pi^{A_i+B_i}B_i\leq(1-\varepsilon^2)\operatorname{Tr}\Pi^{A_i+B_i}(A_i+B_i)\},$$

and

$$J=\{i:\operatorname{Tr}\Pi^{B_i}B_i\geq(1-\varepsilon_1^8)\operatorname{Tr}\Pi^{A_i+B_i}(A_i+B_i)\}.$$

Then

$$|I \cap J| > \frac{99}{100}\varepsilon k.$$

Proof. From (3.9),

$$\begin{aligned}
& (1-\varepsilon^2)\sum_{i\notin I} N(A_i+B_i) \le (1-\varepsilon)\sum_{i=1}^{k} N(A_i+B_i) \\
\Rightarrow\ & (\varepsilon-\varepsilon^2)\sum_{i\notin I} N(A_i+B_i) \le (1-\varepsilon)\sum_{i\in I} N(A_i+B_i) \\
\Rightarrow\ & \varepsilon(k-|I|) \le (1+\varepsilon_1)|I| \quad \text{(from Remarks 1. and 2.)} \\
\Rightarrow\ & |I| \ge \frac{\varepsilon}{1+\varepsilon_1+\varepsilon}k.
\end{aligned}$$

From (3.10) (since for all $i \in [k], N(A_i+B_i) \ge N(B_i)$),

$$\begin{aligned}
& \sum_{i\in J} N(A_i+B_i) + (1-\varepsilon_1^8)\sum_{i\notin J} N(A_i+B_i) \\
& \ge (1-\varepsilon_1^9)\sum_{i=1}^{k} N(A_i+B_i) \\
\Rightarrow\ & \varepsilon_1\sum_{i\in J} N(A_i+B_i) \ge (1-\varepsilon_1)\sum_{i\notin J} N(A_i+B_i) \\
\Rightarrow\ & \varepsilon_1(1+\varepsilon_1)|J| \ge (1-\varepsilon_1)(k-|J|) \\
\Rightarrow\ & |J| \ge \frac{1-\varepsilon_1}{1+\varepsilon_1^2}k.
\end{aligned}$$

The second last implication is from Remarks 1 and 2. Thus

$$|I \cap J| \ge \left(\frac{\varepsilon}{1+\varepsilon_1+\varepsilon} + \frac{1-\varepsilon_1}{1+\varepsilon_1^2} - 1\right)k > \frac{99}{100}\varepsilon k.$$

□

Remark:

3. Note that for any $i \in I \cap J$, $\dim V_i = 2$. Otherwise, either $\Pi^{A_i+B_i} = \Pi^{B_i}$ or $\Pi^{B_i} = 0$ and neither of these can happen in $I \cap J$ (from definitions of I and J).

The following lemma states that for each $i \in I \cap J$, the second eigenvalue of $A_i + B_i$ is close to 1.

Lemma 3.3.16. *Let P and Q be 2×2 positive semidefinite matrices satisfying*

$$\|Q\| \geq 1, \quad \|P+Q\| \leq 1+\varepsilon_1, \quad \lambda_2(P+Q) < 1, \tag{3.11}$$

$$\operatorname{Tr} \Pi^{P+Q} P \geq \varepsilon^2 \operatorname{Tr} \Pi^{P+Q}(P+Q) \quad \textit{and} \tag{3.12}$$

$$\operatorname{Tr} \Pi^{Q} Q \geq (1-\varepsilon_1^8) \operatorname{Tr} \Pi^{P+Q}(P+Q) \ . \tag{3.13}$$

Then $\lambda_2(P+Q) > 1 - \frac{1}{9}\varepsilon_1^3$.

Proof. We prove it by direct calculation. Let η be the maximum real number such that $P - \eta(I - \Pi^{P+Q}) \geq 0$. Set $P_1 = P - \eta(I - \Pi^{P+Q})$ and $Q_1 = Q + \eta(I - \Pi^{P+Q})$. P_1, Q_1 satisfy all the conditions in this Lemma and P_1 is a rank one matrix. Furthermore, set $P_2 = P_1/\|Q_1\|$ and $Q_2 = Q_1/\|Q_1\|$. Again all the conditions in this Lemma are still satisfied by P_2, Q_2 since $\Pi^{Q_2} = \Pi^{Q_1} = \Pi^{Q}$ and $\Pi^{P_2+Q_2} = \Pi^{P_1+Q_1} = \Pi^{P+Q}$. As $\lambda_2(P_2+Q_2) \leq \lambda_2(P_1+Q_1) = \lambda_2(P+Q)$, it suffices to prove that $\lambda_2(P_2+Q_2) > 1 - \frac{1}{9}\varepsilon_1^3$. Consider P_2, Q_2 in the diagonal bases of Q_2.

$$P_2 = \begin{pmatrix} |r|\cos^2\theta & r\sin\theta\cos\theta \\ r^*\sin\theta\cos\theta & |r|\sin^2\theta \end{pmatrix}, Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix}.$$

where $r \in \mathbb{C}$ and $0 \leq b < 1$. Set $\lambda = \|P_2 + Q_2\|$. Eq. (3.13) implies that

$$\lambda \leq \frac{1}{1-\varepsilon_1^8} < 1 + 2\varepsilon_1^8. \tag{3.14}$$

Since

$$\begin{aligned} \operatorname{Tr} \Pi^{Q_2} P_2 &= \operatorname{Tr} \Pi^{Q_2}(P_2+Q_2) - \operatorname{Tr} \Pi^{Q_2} Q_2 \\ &\leq \operatorname{Tr} \Pi^{P_2+Q_2}(P_2+Q_2) - \operatorname{Tr} \Pi^{Q_2} Q_2 \\ &\leq \varepsilon_1^8 \operatorname{Tr} \Pi^{P_2+Q_2}(P_2+Q_2) = \varepsilon_1^8 \lambda < 2\varepsilon_1^8, \end{aligned}$$

we have,

$$|r|\cos^2\theta < 2\varepsilon_1^8. \tag{3.15}$$

Observe that,

$$|v\rangle = \frac{1}{\sqrt{1 + \left(\frac{|r|\sin\theta\cos\theta}{\lambda - b - |r|\sin^2\theta}\right)^2}} \begin{pmatrix} 1 \\ \frac{r^*\sin\theta\cos\theta}{\lambda - b - |r|\sin^2\theta} \end{pmatrix},$$

is the eigenvector of $P_2 + Q_2$ with eigenvalue λ. Hence $\Pi^{P_2+Q_2} = |v\rangle\langle v|$. Note that $\lambda >$

$b + |r|\sin^2\theta$, because $\lambda_2(P_2 + Q_2) = 1 + |r| + b - \lambda < 1$. Consider

$$\begin{aligned}
& \mathrm{Tr}(\Pi^{P_2+Q_2} P_2) = \langle v|P_2|v\rangle \\
= & \frac{|r|\cos^2\theta + \frac{2|r|^2\sin^2\theta\cos^2\theta}{\lambda-b-|r|\sin^2\theta} + \frac{|r|^3\sin^4\theta\cos^2\theta}{(\lambda-b-|r|\sin^2\theta)^2}}{1 + \frac{|r|^2\sin^2\theta\cos^2\theta}{(\lambda-b-|r|\sin^2\theta)^2}} \\
= & \frac{|r|(\lambda-b)^2\cos^2\theta}{(\lambda-b-|r|\sin^2\theta)^2 + |r|^2\sin^2\theta\cos^2\theta} \\
\leq & \frac{|r|\cos^2\theta}{(1-\frac{|r|\sin^2\theta}{\lambda-b})^2} \\
< & \frac{2\varepsilon_1^8}{(1-\frac{|r|\sin^2\theta}{\lambda-b})^2}.
\end{aligned}$$

Combining with (3.12), we obtain

$$\begin{aligned}
& 2\varepsilon_1^8 \geq \varepsilon^2(1-\frac{|r|\sin^2\theta}{\lambda-b})^2 \\
& \Rightarrow (1-\frac{|r|\sin^2\theta}{\lambda-b})^2 < \frac{\varepsilon_1^6}{100} \\
& \Rightarrow |r|\sin^2\theta > (1-\frac{1}{10}\varepsilon_1^3)(\lambda-b) \\
& \Rightarrow |r|\sin^2\theta + (1-\frac{1}{10}\varepsilon_1^3)b > (1-\frac{1}{10}\varepsilon_1^3)\lambda \\
& \Rightarrow |r| + b > (1-\frac{1}{10}\varepsilon_1^3)\lambda > 1 - \frac{1}{10}\varepsilon_1^3 \ .
\end{aligned}$$

Hence

$$\begin{aligned}
\lambda_2(P_2+Q_2) = \mathrm{Tr}(P_2+Q_2) - \lambda &= 1 + |r| + b - \lambda \\
&> 2 - \frac{1}{10}\varepsilon_1^3 - (1+2\varepsilon_1^8) > 1 - \frac{1}{9}\varepsilon_1^3 \ .
\end{aligned}$$

□

We can finally prove Lemma 3.3.12. By Fact 3.3.7, $\lambda^{\downarrow}(A+B) \succeq \lambda^{\downarrow}(\sum_i A_i + B_i)$. Let $j_1 = \max\{j : \lambda_j(A+B) \geq 1\}$, $j_2 = \max\{j : \lambda_j(\sum_i (A_i+B_i)) \geq 1\}$, and $j_0 = j_1 + \frac{99}{100}\varepsilon k$. Then

$$\sum_{j\leq j_0} \lambda_j(A+B) \geq \sum_{j\leq j_0} \lambda_j \left(\sum_i (A_i+B_i)\right).$$

According to the decomposition in Fact 3.3.9, Lemma 3.3.14 and the remarks below

it, $j_1 = j_2 = k$ and

$$\sum_{j \le j_1} \lambda_j(A+B) = \operatorname{Tr} \Pi^{A+B}(A+B), \quad \text{and}$$

$$\sum_{j \le j_2} \lambda_j \left(\sum_i (A_i + B_i) \right) = \sum_i \operatorname{Tr} \Pi^{A_i+B_i}(A_i + B_i).$$

The RHS of both the equations are equal by Lemma 3.3.14. Therefore,

$$\sum_{k<j\le j_0} \lambda_j(A+B) \ge \sum_{k<j\le j_0} \lambda_j \left(\sum_i (A_i + B_i) \right).$$

By Lemma 3.3.15 and Lemma 3.3.16,

$$\sum_{k<j\le j_0} \lambda_j \left(\sum_i (A_i + B_i) \right) \ge \frac{99}{100}\varepsilon k \left(1 - \frac{1}{9}\varepsilon_1^3\right).$$

Let $x = N_{1-\varepsilon'}(A+B) - N(A+B)$, then

$$\sum_{k<j\le j_0} \lambda_j(A+B) \le x + \left(\frac{99}{100}\varepsilon k - x\right)(1-\varepsilon').$$

Therefore from previous three inequalities,

$$\frac{99}{100}\varepsilon k \left(1 - \frac{1}{9}\varepsilon_1^3\right) \le x + \left(\frac{99}{100}\varepsilon k - x\right)(1-\varepsilon'),$$

which implies

$$x \ge \frac{99}{100}\varepsilon k \left(1 - \frac{\varepsilon_1^3}{9\varepsilon'}\right).$$

Note that $\varepsilon_1^3 \ll \varepsilon'$, therefore from Remark 2.,

$$N_{1-\varepsilon'}(A+B) \ge k + \frac{99}{100}\varepsilon k \left(1 - \frac{\varepsilon_1^3}{9\varepsilon'}\right) > \left(1 + \frac{1}{2}\varepsilon\right) k$$
$$> \left(1 + \frac{2}{5}\varepsilon\right)(1+\varepsilon_1)k \ge \left(1 + \frac{2}{5}\varepsilon\right) N(A+B).$$

□

Chapter 4

A parallel approximation algorithm for mixed packing and covering semidefinite programs

4.1 Introduction

In this chapter, we continue investigating fast parallel approximation algorithms for semidefinite programs. We present an algorithm for **Q1** given in Chapter 2 Section 2.3 running in parallel time $\text{polylog}(n, m) \cdot \frac{1}{\varepsilon^4} \cdot \log \frac{1}{\varepsilon}$. Using this and standard binary search, a multiplicative $(1 - \varepsilon)$ approximate solution can be obtained for the optimization task **Q2** in parallel time $\text{polylog}(n, m, \frac{1}{\varepsilon})$.

Our algorithm for **Q1** and its analysis is on similar lines as the algorithm and analysis of Young [76] who had considered analogous questions for linear programs. As a corollary we get an algorithm for approximating positive semidefinite programs (**Q3**) with better dependence of the parallel running time on ε as compared that in the previous chapter (and arguably with simpler analysis). Very recently, in an independent work, Peng and Tangwongsan [59] also presented a fast parallel approximation algorithm for positive semidefinite programs. Their work is also inspired by Young [76].

4.2 Algorithm and analysis

Using standard arguments, the feasibility question **Q1** can be transformed, in parallel time $\text{polylog}(m, n)$, to the special case when P and C are identity matrices. (Similar

transformation is used in Chapter 2 Section 2.2 for positive semidefinite programs) Hence we consider the following special case from now on.

Q: Given $n \times n$ positive semidefinite matrices $P_1, \ldots, P_m, P$ and non-negative diagonal matrices $C_1, \ldots, C_m, C$ and $\varepsilon \in (0, 1)$, find a vector $x \geq 0$ such that

$$\sum_{i=1}^{m} x_i P_i \leq (1+\varepsilon) I \quad \text{and} \quad \sum_{i=1}^{m} x_i C_i \geq I$$

or show that the following is infeasible

$$\sum_{i=1}^{m} x_i P_i \leq I \quad \text{and} \quad \sum_{i=1}^{m} x_i C_i \geq I \ .$$

Our algorithm is presented in Figure 4.1 .

4.2.1 Idea of the algorithm

The algorithm starts with an initial value for x such that $\sum_{i=1}^{m} x_i P_i \leq I$. It makes increments to the vector x such that with each increment, the increase in $\|\sum_{i=1}^{m} x_i P_i\|$ is not more than $(1 + \mathcal{O}(\varepsilon))$ times the increase in the minimum eigenvalue of $\sum_{i=1}^{m} x_i C_i$. We argue that it is always possible to increment x in this manner if the input instance is feasible, hence the algorithm outputs infeasible if it cannot find such an increment to x. The algorithm stops when the minimum eigenvalue of $\sum_{i=1}^{m} x_i C_i$ has exceeded 1. Due to our condition on the increments, at the end of the algorithm we also have $\sum_{i=1}^{m} x_i P_i \leq (1 + \mathcal{O}(\varepsilon))I$. The change of the eigenvalues is generally hard to analyze directly. Using the idea from Young [76], We obtain handle on the largest and smallest eigenvalues of concerned matrices via their *soft* versions, which are more easily handled functions of those matrices (see definitions in the next section). Like the algorithm for positive semidefinite programs in Chapter 3, We set the changes in each step small enough to ensure the approximation. At the same time, they are large enough such that the algorithm terminates in polylog time.

4.2.2 Correctness analysis

We begin with the definitions of soft maximum and minimum eigenvalues of a positive semidefinite matrix A. They are inspired by analogous definitions made in Young [76] in

Input : $n \times n$ positive semidefinite matrices $P_1, \ldots, P_m$, non-negative diagonal matrices $C_1, \ldots, C_m$, and error parameter $\varepsilon \in (0, 1)$.

Output : Either infeasible, which means there is no x such that (I is the identity matrix),

$$\sum_{i=1}^{m} x_i P_i \leq I \quad \text{and} \quad \sum_{i=1}^{m} x_i C_i \geq I \ .$$

OR an $x^* \in \mathbb{R}^m$ such that

$$\sum_{i=1}^{m} x_i^* P_i \leq (1 + 9\varepsilon) I \quad \text{and} \quad \sum_{i=1}^{m} x_i^* C_i \geq I \ .$$

1. Set $x_j = \frac{1}{m\|P_j\|}$.
2. Set $N = \frac{1}{\varepsilon}\left(\|\sum_{i=1}^{m} x_i P_i\| + 2\ln n + \ln m\right)$.
3. While $\lambda_{\min}(\sum_{i=1}^{m} x_i C_i) < N$ ($\lambda_{\min}$ represents minimum eigenvalue), do
   (a) Set
   $$\text{local}_j(x) = \frac{\text{Tr}(\exp(\sum_{i=1}^{m} x_i P_i) \cdot P_j)}{\text{Tr}(\exp(-\sum_{i=1}^{m} x_i C_i) \cdot C_j)} \quad \text{and}$$
   $$\text{global}(x) = \frac{\text{Tr}\exp(\sum_{i=1}^{m} x_i P_i)}{\text{Tr}(\exp(-\sum_{i=1}^{m} x_i C_i))}.$$
   (b) If g is not yet set or $\min_j\{\text{local}_j(x)\} > g(1+\varepsilon)$, set $g = \text{global}(x)$.
   (c) If $\min_j\{\text{local}_j(x)\} > \text{global}(x)$, return infeasible.
   (d) For all $j \in [m]$, set $C_j = \Pi_j \cdot C_j \cdot \Pi_j$, where Π_j is the projection onto the eigenspace of $\sum_{i=1}^{m} x_i C_i$ with eigenvalues at most N.
   (e) Choose increment vector $\alpha \geq 0$ and scalar $\delta > 0$ such that
   $$\forall j: \ \alpha_j = x_j \delta \text{ if } \text{local}_j(x) \leq g(1+\varepsilon), \text{ else } \alpha_j = 0, \text{ and}$$
   $$\max\{\left\|\sum_{i=1}^{m} \alpha_i P_i\right\|, \left\|\sum_{i=1}^{m} \alpha_i C_i\right\|\} = \varepsilon.$$
   (f) Set $x = x + \alpha$.
4. Return $x^* = x/N$.

Figure 4.1: Algorithm

the context of vectors.

Definition 4.2.1. For positive semidefinite matrix A, define

$$\mathrm{Imax}(A) \stackrel{\mathrm{def}}{=} \ln \mathrm{Tr}\exp(A),$$

and

$$\mathrm{Imin}(A) \stackrel{\mathrm{def}}{=} -\ln \mathrm{Tr}\exp(-A).$$

Note that $\mathrm{Imax}(A) \geq \|A\|$ and $\mathrm{Imin}(A) \leq \lambda_{\min}(A)$, where $\lambda_{\min}(A)$ is the minimum eigenvalue of A.

The following lemma shows that if a small increment is made in the vector x, then changes in $\mathrm{Imax}(\sum_{j=1}^m x_j A_j)$ and $\mathrm{Imin}(\sum_{j=1}^m x_j A_j)$ can be bounded appropriately.

Lemma 4.2.2. *Let $A_1, \ldots, A_m$ be positive semidefinite matrices and let $x \geq 0, \alpha \geq 0$ be vectors in $\mathbb{R}^m$. If $\|\sum_{i=1}^m \alpha_i A_i\| \leq \varepsilon \leq 1$, then*

$$\mathrm{Imax}(\sum_{j=1}^m (x_j+\alpha_j)A_j) - \mathrm{Imax}(\sum_{j=1}^m x_j A_j) \leq \frac{(1+\varepsilon)}{\mathrm{Tr}(\exp(\sum_{i=1}^m x_i A_i))} \sum_{j=1}^m \alpha_j \,\mathrm{Tr}(\exp(\sum_{i=1}^m x_i A_i)A_j),$$

and

$$\mathrm{Imin}(\sum_{j=1}^m (x_j+\alpha_j)A_j) - \mathrm{Imin}(\sum_{j=1}^m x_j A_j) \geq \frac{(1-\varepsilon/2)}{\mathrm{Tr}(\exp(-\sum_{i=1}^m x_i A_i))} \sum_{j=1}^m \alpha_j \,\mathrm{Tr}(\exp(-\sum_{i=1}^m x_i A_i)A_j).$$

Proof. We will use the following Golden-Thompson inequality.

Fact 4.2.3. For Hermitian matrices A, B : $\mathrm{Tr}(\exp(A+B)) \leq \mathrm{Tr}\exp(A)\exp(B)$.

We will also need the following fact.

Fact 4.2.4. Let A be positive semidefinite with $\|A\| \leq \varepsilon \leq 1$. Then,

$$\exp(A) \leq I + (1+\varepsilon)A \quad \text{and} \quad \exp(-A) \leq I - (1-\varepsilon/2)A.$$

Consider,

$$\begin{aligned}
&\mathrm{Imax}(\sum_{j=1}^{m}(x_j+\alpha_j)A_j) - \mathrm{Imax}(\sum_{j=1}^{m} x_jA_j)\\
&= \ln\left(\frac{\mathrm{Tr}\exp(\sum_{i=1}^{m}(x_i+\alpha_i)A_i)}{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)}\right)\\
&\leq \ln\left(\frac{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)\exp(\sum_{j=1}^{m}\alpha_jA_j)}{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)}\right) \quad \text{(from Fact 4.2.3)}\\
&= \ln\left(\frac{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)(I+(1+\varepsilon)(\sum_{j=1}^{m}\alpha_jA_j))}{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)}\right) \quad \text{(from Fact 4.2.4)}
\end{aligned}$$

$$\begin{aligned}
&= \ln\left(1+\frac{(1+\varepsilon)\,\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)(\sum_{j=1}^{m}\alpha_jA_j)}{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)}\right)\\
&\leq \frac{(1+\varepsilon)\,\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)(\sum_{j=1}^{m}\alpha_jA_j)}{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_iA_i)} \quad \text{(since } \ln(1+a)\leq a \text{ for all real } a\text{)}
\end{aligned}$$

The desired bound on $\mathrm{Imin}(\sum_{j=1}^{m}(x_j+\alpha_j)A_j) - \mathrm{Imin}(\sum_{j=1}^{m} x_jA_j)$ follows by analogous calculations. □

The next two lemmas show that the increment of $\mathrm{Imax}(\sum_{i=1}^{m} x_iP_i)$ is bounded by the increment of $\mathrm{Imin}(\sum_{i=1}^{m})$ from above, as expected.

Lemma 4.2.5. *At step 3(e) of the algorithm, for any j with $\alpha_j > 0$ we have,*

$$\frac{\mathrm{Tr}(\exp(\sum_{i=1}^{m} x_iP_i)\cdot P_j)}{\mathrm{Tr}(\exp(\sum_{i=1}^{m} x_iP_i))} \leq (1+\varepsilon)\frac{\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_iC_i)\cdot C_j)}{\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_iC_i))}.$$

Proof. Consider any execution of step 3(e) of the algorithm. Fix j such $\alpha_j > 0$. Note that,

$$\frac{\mathrm{local}_j(x)}{\mathrm{global}(x)} = \frac{\mathrm{Tr}(\exp(\sum_{i=1}^{m} x_iP_i)\cdot P_j)\cdot\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_iC_i))}{\mathrm{Tr}(\exp(\sum_{i=1}^{m} x_iP_i))\cdot\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_iC_i)\cdot C_j)}.$$

We will show that $\mathrm{global}(x) \geq g$ throughout the algorithm and this will show the desired since that $\mathrm{local}_j(x) \leq (1+\varepsilon)g \leq (1+\varepsilon)\mathrm{global}(x)$.

At step 3(b) of the algorithm, g can be equal to $\mathrm{global}(x)$. Since x never decreases during the algorithm, at step 3(a), $\mathrm{global}(x)$ can only increase. At step 3(d), the modification of C_js only decreases $\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_iC_i))$ and hence again $\mathrm{global}(x)$ can only increase. □

Lemma 4.2.6. *For each increment of x at step 3(f) of the algorithm,*

$$\mathrm{Imax}(\sum_{j=1}^{m}(x_j+\alpha_j)P_j)-\mathrm{Imax}(\sum_{j=1}^{m}x_jP_j) \le (1+\varepsilon)^3\left(\mathrm{Imin}(\sum_{j=1}^{m}(x_j+\alpha_j)C_j)-\mathrm{Imin}(\sum_{j=1}^{m}x_jC_j)\right).$$

Proof. Consider,

$$\begin{aligned}
&\mathrm{Imax}(\sum_{j=1}^{m}(x_j+\alpha_j)P_j)-\mathrm{Imax}(\sum_{j=1}^{m}x_jP_j)\\
&\le \frac{(1+\varepsilon)}{\mathrm{Tr}(\exp(\sum_{i=1}^{m}x_iP_i))}\sum_{j=1}^{m}\alpha_j\,\mathrm{Tr}(\exp(\sum_{i=1}^{m}x_iP_i)P_j) \quad \text{(from Lemma 4.2.2)}\\
&\le \frac{(1+\varepsilon)^2}{\mathrm{Tr}(\exp(-\sum_{i=1}^{m}x_iC_i))}\sum_{j=1}^{m}\alpha_j\,\mathrm{Tr}(\exp(-\sum_{i=1}^{m}x_iC_i)C_j)\\
&\qquad\qquad \text{(from Lemma 4.2.5 and step 3(e) of the algorithm)}
\end{aligned}$$

$$\le \frac{(1+\varepsilon)^2}{1-\varepsilon/2}\left(\mathrm{Imin}(\sum_{j=1}^{m}(x_j+\alpha_j)C_j)-\mathrm{Imin}(\sum_{j=1}^{m}x_jC_j)\right) \quad \text{(from Lemma 4.2.2).}$$

This shows the desired. □

The following lemma shows that such j in step 3 (c) always exists if the program is feasible.

Lemma 4.2.7. *If the input instance $P_1,\ldots,P_m,C_1,\ldots,C_m$ is feasible, that is there exists vector $y\in\mathbb{R}^m$ such that*

$$\sum_{i=1}^{m}y_iP_i \le I \quad \text{and} \quad \sum_{i=1}^{m}y_iC_i \ge I \ ,$$

then always at step 3(c) of the algorithm, $\min_j\{\mathrm{local}_j(x)\}\le \mathrm{global}(x)$. Hence the algorithm will return some x^*.

If the algorithm outputs infeasible*, then the input instance is not feasible.*

Proof. Consider some execution of step 3(c) of the algorithm. Let $C'_1,\ldots,C'_m$ be the current values of $C_1,\ldots,C_m$. Note that if the input is feasible with vector y, then we will

also have

$$\frac{\mathrm{Tr}(\exp(\sum_{i=1}^m x_i P_i)(\sum_{j=1}^m y_j P_j))}{\mathrm{Tr}(\exp(\sum_{i=1}^m x_i P_i))} \leq 1 \leq \frac{\mathrm{Tr}(\exp(-\sum_{i=1}^m x_i C_i')(\sum_{j=1}^m y_j C_j'))}{\mathrm{Tr}(\exp(-\sum_{i=1}^m x_i C_i'))}.$$

Therefore there exists $j \in [m]$ such that

$$\frac{\mathrm{Tr}(\exp(\sum_{i=1}^m x_i P_i) P_j)}{\mathrm{Tr}(\exp(\sum_{i=1}^m x_i P_i))} \leq \frac{\mathrm{Tr}(\exp(-\sum_{i=1}^m x_i C_i') C_j')}{\mathrm{Tr}(\exp(-\sum_{i=1}^m x_i C_i'))},$$

and hence $\mathrm{local}_j(x) \leq \mathrm{global}(x)$.

If the algorithm outputs infeasible, then at that point $\min_j\{\mathrm{local}_j(x)\} > \mathrm{global}(x)$ and hence from the argument above $P_1, \ldots, P_m, C_1', \ldots, C_m'$ is infeasible which in turn implies that $P_1, \ldots, P_m, C_1, \ldots, C_m$ is infeasible. □

Finally, we are able to show that the algorithm outputs a good approximation solution.

Lemma 4.2.8. *If the algorithm returns some x^*, then*

$$\sum_{i=1}^m x_i^* P_i \leq (1+9\varepsilon) I \quad \text{and} \quad \sum_{i=1}^m x_i^* C_i \geq I.$$

Proof. Because of the condition of the while loop, it is clear that $\sum_{i=1}^m x_i^* C_i \geq I$.

For $x \in \mathbb{R}^m$, define

$$\Phi(x) \stackrel{\mathrm{def}}{=} \mathrm{Imax}(\sum_{j=1}^m x_j P_j) - (1+\varepsilon)^3 \cdot \mathrm{Imin}(\sum_{j=1}^m x_j C_j).$$

Note that the update of C_j's at step 3(d) only increase $\mathrm{Imin}(\sum_{j=1}^m x_j C_j)$. Hence using Lemma 4.2.6, we conclude that $\Phi(x)$ is non-decreasing during the algorithm. At step 1 of the algorithm,

$$\begin{aligned}\Phi(x) &\leq \mathrm{Imax}(\sum_{j=1}^m x_j P_j) = \ln \mathrm{Tr}(\exp(\sum_{i=1}^m x_i P_i)) \\ &\leq \ln(n \exp(\left\|\sum_{i=1}^m x_i P_i\right\|)) \leq \ln(n \exp(\sum_{i=1}^m \|x_i P_i\|)) = \ln n + 1.\end{aligned}$$

Hence just before the last increment,

$$\begin{aligned}\left\|\sum_{i=1}^{m} x_i P_i\right\| \le \mathrm{Imax}(\sum_{j=1}^{m} x_j P_j) &\le \Phi(x) + (1+\varepsilon)^3 \cdot \mathrm{Imin}(\sum_{j=1}^{m} x_j C_j) \\ &\le \ln n + 1 + (1+\varepsilon)^3 \cdot \mathrm{Imin}(\sum_{j=1}^{m} x_j C_j) \\ &\le \ln n + 1 + (1+\varepsilon)^3 \cdot \lambda_{\min}(\sum_{j=1}^{m} x_j C_j) \\ &\le \ln n + 1 + (1+\varepsilon)^3 N \le (1+8\varepsilon)N \ .\end{aligned}$$

In the last increment, because of the condition on step 3(e) of the algorithm, $\|\sum_{i=1}^{m} x_i P_i\|$ increase by at most ε. Hence $\sum_{i=1}^{m} x_i^* P_i \le (1+9\varepsilon)I$. □

4.2.3 Running time analysis

Note that each individual step in the algorithm can be performed in parallel time polylog(mn). Please refer to Chapter 2, Section 2.1. We show that the wile loop is executed polylog times, then show that only polylog iterations are required in each loop.

Lemma 4.2.9. *Assume that the algorithm does not return* infeasible *for some input instance. The number of times g is increased at step 3(b) of the algorithm is $\mathcal{O}(N/\varepsilon)$.*

Proof. At the beginning of the algorithm $\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_i C_i)) \le n$ since each eigenvalue of $\exp(-\sum_{i=1}^{m} x_i C_i)$ is at most 1. Also $\mathrm{Tr}\exp(\sum_{i=1}^{m} x_i P_i) \ge 1$. Hence

$$g = \mathrm{global}(x) = \frac{\mathrm{Tr}\exp(\sum_{i=1}^{m} x_i P_i)}{\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_i C_i))} \ge \frac{1}{n} \ge \frac{1}{\exp(N)}.$$

At the end of the algorithm $\lambda_{\min}(\sum_{i=1}^{m} x_i C_i) \le N + \varepsilon \le 2N$. Hence

$$\mathrm{Tr}(\exp(-\sum_{i=1}^{m} x_i C_i)) \ge \left\|\exp(-\sum_{i=1}^{m} x_i C_i)\right\| = \exp(-\lambda_{\min}(\sum_{i=1}^{m} x_i C_i)) \ge \exp(-2N).$$

Also (using Lemma 4.2.8)

$$\mathrm{Tr}(\exp(\sum_{i=1}^{m} x_i P_i)) \le n \left\|\exp(\sum_{i=1}^{m} x_i P_i)\right\| \le n\exp((1+9\varepsilon)N) \le \exp(11N).$$

Hence $g \le \mathrm{global}(x) \le \exp(13N)$.

Whenever g is updated at step 3(b) of the algorithm, we have

$$\text{global}(x) \geq \min_j \{\text{local}_j(x)\} > (1+\varepsilon)g$$

just before the update and $\text{global}(x) = g$ just after the update. Thus g increases by at least $(1+\varepsilon)$ multiplicative factor. Hence the number of times g increases is $\mathcal{O}(N/\varepsilon)$. □

Lemma 4.2.10. *Assume that the algorithm does not return* infeasible *for some input instance. The number of iterations of the while loop in the algorithm for a fixed value of g is $\mathcal{O}(N \log(mN)/\varepsilon)$.*

Proof. From Lemma 4.2.8 and step 3(d) of the algorithm we have

$$\max\{\left\|\sum_{i=1}^{m} x_i P_i\right\|, \left\|\sum_{i=1}^{m} x_i C_i\right\|\} = \mathcal{O}(N)$$

throughout the algorithm. On the other hand we have $\max\{\|\sum_{i=1}^{m} \delta x_i P_i\|, \|\sum_{i=1}^{m} \delta x_i C_i\|\} = \varepsilon$ at step 3(e). Hence $\delta = \Omega(\varepsilon/N)$ throughout the algorithm.

Let x_j be increased in the last iteration of the while loop for a fixed value of g. Note that x_j is initially $1/(m\|P_j\|)$ and at the end x_j is at most $10N/\|P_j\|$ (since, using Lemma 4.2.8, $\|x_j P_j\| \leq \|\sum_{i=1}^{m} x_j P_j\| \leq 10N$). Hence the algorithm makes at most $\mathcal{O}(\log(mN)/\delta) = \mathcal{O}(N \log(mN)/\varepsilon)$ increments for each x_j.

Note that $\text{local}_j(x)$ only increases throughout the algorithm by steps 3(d) and 3(e) of the algorithm. Hence since the last iteration of the while loop (for this fixed g) increases x_j, it must be that each iteration of the while loop increases x_j. Hence, the number of iterations of the while loop (for this fixed g) is $\mathcal{O}(N \log(mN)/\varepsilon)$. □

Hence combining the above lemmas and using $N = \mathcal{O}(\frac{\ln(mn)}{\varepsilon})$, we get

Corollary 4.2.11. *The parallel running time of the algorithm is upper bounded by* $\text{polylog}(mn) \cdot \frac{1}{\varepsilon^4} \cdot \log \frac{1}{\varepsilon}$.

Chapter 5

Information theory and communication complexity

In this chapter, we give some definitions and facts on information theory and communication complexity, which will be used in the subsequent chapters.

5.1 Information theory

For integer $n \geq 1$, let $[n]$ represent the set $\{1, 2, \ldots, n\}$. Let $\mathcal{X}$, $\mathcal{Y}$ be finite sets and k be a natural number. Let $\mathcal{X}^k$ be the set $\mathcal{X} \times \cdots \times \mathcal{X}$, the cross product of $\mathcal{X}$, k times. Let μ be a (probability) distribution on $\mathcal{X}$. Let $\mu(x)$ represent the probability of $x \in \mathcal{X}$ according to μ. Let X be a random variable distributed according to μ, which we denote by $X \sim \mu$. We use the same symbol to represent a random variable and its distribution whenever it is clear from the context. The expectation value of function f on $\mathcal{X}$ is denoted as $\mathbb{E}_{x \leftarrow X}[f(x)] \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot f(x)$. The entropy of X is defined as $\mathrm{H}(X) \stackrel{\text{def}}{=} -\sum_x \mu(x) \cdot \log \mu(x)$. For two distributions μ, λ on $\mathcal{X}$, the distribution $\mu \otimes \lambda$ is defined as $(\mu \otimes \lambda)(x_1, x_2) \stackrel{\text{def}}{=} \mu(x_1) \cdot \lambda(x_2)$. XY represents the joint distribution of X and Y. Let $\mu^k \stackrel{\text{def}}{=} \mu \otimes \cdots \otimes \mu$, k times. The ℓ_1 distance between μ and λ is defined to be half of the ℓ_1 norm of $\mu - \lambda$; that is, $\|\lambda - \mu\|_1 \stackrel{\text{def}}{=} \frac{1}{2} \sum_x |\lambda(x) - \mu(x)| = \max_{S \subseteq \mathcal{X}} |\lambda_S - \mu_S|$, where $\lambda_S \stackrel{\text{def}}{=} \sum_{x \in S} \lambda(x)$. We say that λ is ε-close to μ if $\|\lambda - \mu\|_1 \leq \varepsilon$.

The relative entropy between distributions X and Y on $\mathcal{X}$ is defined as $\mathrm{S}(X\|Y) \stackrel{\text{def}}{=} \mathbb{E}_{x \leftarrow X}\left[\log \frac{\Pr[X=x]}{\Pr[Y=x]}\right]$. Here we assume $0 \log \frac{0}{0} = 0$. The relative min-entropy between them is defined as $\mathrm{S}_\infty(X\|Y) \stackrel{\text{def}}{=} \max_{x \in \mathcal{X}} \left\{\log \frac{\Pr[X=x]}{\Pr[Y=x]}\right\}$. $\mathrm{S}_\infty(X\|Y) \stackrel{\text{def}}{=} \infty$ if there exists x such that $\Pr[X = x] > 0$ and $\Pr[Y = x] = 0$. It is easy to see that $\mathrm{S}(X\|Y) \leq \mathrm{S}_\infty(X\|Y)$.

Let X, Y, Z be jointly distributed random variables. Let Y_x denote the distribution of Y conditioned on $X = x$. The conditional entropy of Y conditioned on X is defined as $\mathrm{H}(Y|X) \stackrel{\text{def}}{=} \mathbb{E}_{x \leftarrow X}[\mathrm{H}(Y_x)] = \mathrm{H}(XY) - \mathrm{H}(X)$. The mutual information between X and Y is defined as: $\mathrm{I}(X:Y) \stackrel{\text{def}}{=} \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(XY) = \mathbb{E}_{y \leftarrow Y}[\mathrm{S}(X_y \| X)] = \mathbb{E}_{x \leftarrow X}[\mathrm{S}(Y_x \| Y)]$. It can be checked from the definition that $\mathrm{I}(X:Y) = \mathrm{S}(XY \| X \otimes Y)$. We say that X and Y are independent iff $\mathrm{I}(X:Y) = 0$, or equivalently, $XY = X \otimes Y$. The conditional mutual information between X and Y, conditioned on Z, is defined as: $\mathrm{I}(X:Y|Z) \stackrel{\text{def}}{=} \mathbb{E}_{z \leftarrow Z}[\mathrm{I}(X:Y|Z=z)] = \mathrm{H}(X|Z) + \mathrm{H}(Y|Z) - \mathrm{H}(XY|Z)$. The following is the *chain rule* for mutual information : $\mathrm{I}(X:YZ) = \mathrm{I}(X:Z) + \mathrm{I}(X:Y|Z)$.

Let X, X', Y, Z be jointly distributed random variables. We define the joint distribution of $(X'Z)(Y|X)$ by: $\Pr[(X'Z)(Y|X) = x, z, y] \stackrel{\text{def}}{=} \Pr[X' = x, Z = z] \cdot \Pr[Y = y | X = x]$. We say that X, Y, Z is a Markov chain iff $XYZ = (XY)(Z|Y)$ and we denote it by $X \leftrightarrow Y \leftrightarrow Z$. Suppose Alice is given $x \sim X$ and Bob is given $y \sim Y$, then Bob can sample distribution Z_{xy} without knowing x if and only if $X \leftrightarrow Y \leftrightarrow Z$. It is easy to see that X, Y, Z is a Markov chain if and only if $\mathrm{I}(X:Z|Y) = 0$. Ibinson, Linden and Winter [24] showed that if $\mathrm{I}(X:Z|Y)$ is small then XYZ is close to being a Markov chain.

Lemma 5.1.1 ([24])**.** *For any random variables X, Y and Z, it holds that*

$$\mathrm{I}(X:Z|Y) = \min\left\{\mathrm{S}(XYZ \| X'Y'Z') : X' \leftrightarrow Y' \leftrightarrow Z'\right\}.$$

The minimum is achieved by distribution $X'Y'Z' = (XY)(Z|Y)$.

We need the following basic facts. A very good text for reference on information theory is [19].

Fact 5.1.2. Relative entropy is jointly convex in its arguments. That is, for distributions $\mu, \mu^1, \lambda, \lambda^1 \in \mathcal{X}$ and $p \in [0,1]$: $\mathrm{S}(p\mu + (1-p)\mu^1 \| \lambda + (1-p)\lambda^1) \leq p \cdot \mathrm{S}(\mu \| \lambda) + (1-p) \cdot \mathrm{S}(\mu^1 \| \lambda^1)$.

Fact 5.1.3. Relative entropy satisfies the following chain rule. Let XY and X^1Y^1 be random variables on $\mathcal{X} \times \mathcal{Y}$. It holds that: $\mathrm{S}(X^1Y^1 \| XY) = \mathrm{S}(X^1 \| X) + \mathbb{E}_{x \leftarrow X^1}[\mathrm{S}(Y^1_x \| Y_x)]$. In particular, using Fact 5.1.2: $\mathrm{S}(X^1Y^1 \| X \otimes Y) = \mathrm{S}(X^1 \| X) + \mathbb{E}_{x \leftarrow X^1}[\mathrm{S}(Y^1_x \| Y)] \geq \mathrm{S}(X^1 \| X) + \mathrm{S}(Y^1 \| Y)$.

Fact 5.1.4. Let XY and X^1Y^1 be random variables on $\mathcal{X} \times \mathcal{Y}$. It holds that

$$\mathrm{S}\left(X^1Y^1 \middle\| X \otimes Y\right) \geq \mathrm{S}\left(X^1Y^1 \middle\| X^1 \otimes Y^1\right) = \mathrm{I}\left(X^1 : Y^1\right).$$

Fact 5.1.5. For distributions λ and μ: $\quad 0 \leq \|\lambda - \mu\|_1 \leq \sqrt{\mathrm{S}(\lambda\|\mu)}$.

Fact 5.1.6. Let λ and μ be distributions on $\mathcal{X}$. For any subset $\mathcal{S} \subseteq \mathcal{X}$, it holds that: $\sum_{x\in\mathcal{S}} \lambda(x) \cdot \log \frac{\lambda(x)}{\mu(x)} \geq -1$.

Fact 5.1.7. The ℓ_1 distance and relative entropy are monotone non-increasing when subsystems are considered. Let XY and X^1Y^1 be random variables on $\mathcal{X} \times \mathcal{Y}$, then

$$\begin{aligned} &\left\|XY - X^1Y^1\right\|_1 \geq \left\|X - X^1\right\|_1 \quad \text{and} \\ &\mathrm{S}\left(XY\middle\|X^1Y^1\right) \geq \mathrm{S}\left(X\middle\|X^1\right). \end{aligned}$$

Fact 5.1.8. For function $f : \mathcal{X} \times \mathcal{R} \to \mathcal{Y}$ and random variables X, X_1 on $\mathcal{X}$ and R on $\mathcal{R}$, such that R is independent of (XX_1), it holds that: $\|Xf(X,R) - X_1 f(X_1,R)\|_1 = \|X - X_1\|_1$.

Fact 5.1.9. (**Classical substate theorem** [31]) Let X, X' be two distributions on $\mathcal{X}$. For any $\delta \in (0,1)$, it holds that

$$\Pr_{x\leftarrow X'}\left[\frac{\Pr[X'=x]}{\Pr[X=x]} \leq 2^{(\mathrm{S}(X'\|X)+1)/\delta}\right] \geq 1-\delta.$$

Lemma 5.1.10. *Given random variables A, A' and $\varepsilon > 0$, if $\|A - A'\|_1 \leq \varepsilon$, then for any $r \in (0,1)$,*

$$\begin{aligned} &\Pr_{a\leftarrow A}\left[\left|1 - \tfrac{\Pr[A'=a]}{\Pr[A=a]}\right| \leq \tfrac{\varepsilon}{r}\right] \geq 1-2r; \text{ and} \\ &\Pr_{a\leftarrow A'}\left[\left|1 - \tfrac{\Pr[A'=a]}{\Pr[A=a]}\right| \leq \tfrac{\varepsilon}{r}\right] \geq 1-2r-\varepsilon. \end{aligned}$$

Proof. Let $G = \left\{a : \left|1 - \frac{\Pr[A'=a]}{\Pr[A=a]}\right| \leq \frac{\varepsilon}{r}\right\}$, then

$$\begin{aligned} 2\varepsilon &\geq \sum_a \left|\Pr[A=a] - \Pr[A'=a]\right| \\ &\geq \sum_{a\notin G} \left|\Pr[A=a] - \Pr[A'=a]\right| \\ &= \sum_{a\notin G} \Pr[A=a]\left|1 - \frac{\Pr[A'=a]}{\Pr[A=a]}\right| \geq \Pr_{a\leftarrow A}[a\notin G]\cdot\frac{\varepsilon}{r}. \end{aligned}$$

Thus $\Pr_{a\leftarrow A}[a \in G] \geq 1-2r$. The second inequality follows immediately. □

The following definition was introduced by Holenstein [23]. It plays a critical role in his proof of a parallel repetition theorem for two-prover games.

Definition 5.1.11 ([23])**.** For two distributions (X_0Y_0) and (X_1SY_1T), we say that (X_0, Y_0) is $(1-\varepsilon)$-embeddable in (X_1S, Y_1T) if there exists a probability distribution R over a set $\mathcal{R}$, which is independent of X_0Y_0 and functions $f_A : \mathcal{X}\times\mathcal{R} \to \mathcal{S}$, $f_B : \mathcal{Y}\times\mathcal{R} \to \mathcal{T}$, such that

$$\|X_0Y_0f_A(X_0, R)f_B(Y_0, R) - X_1Y_1ST\|_1 \leq \varepsilon.$$

The following lemma was shown by Holenstein [23] using a correlated sampling protocol.

Lemma 5.1.12 ([23])**.** *For random variables S, X and Y, if*

$$\begin{aligned}\|XYS - (XY)(S|X)\|_1 &\leq \varepsilon \quad \textit{and}\\ \|XYS - (XY)(S|Y)\|_1 &\leq \varepsilon,\end{aligned}$$

then (X, Y) is $(1-4\varepsilon)$-embeddable in (XS, YS).

We need the following generalization of the previous lemma.

Lemma 5.1.13. *For joint random variables (A', B', C') and (A, B), satisfying*

$$\mathrm{S}(A'B'\|AB) \leq \varepsilon. \tag{5.1}$$

$$\mathbb{E}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B_a\right)\right] \leq \varepsilon, \tag{5.2}$$

$$\mathbb{E}_{(b,c)\leftarrow B',C'}\left[\mathrm{S}\left(A'_{b,c}\middle\|A_b\right)\right] \leq \varepsilon, \tag{5.3}$$

it holds that (A, B) is $(1-5\sqrt{\varepsilon})$-embeddable in $(A'C', B'C')$.

Proof. Using the definition of the relative entropy, we have the following.

$$\begin{aligned}&\mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B_a\right)\right] - \mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B'_a\right)\right] = \mathop{\mathbb{E}}_{(a,b,c)\leftarrow A',B',C'}\left[\log\frac{\Pr[B'=b|A'=a]}{\Pr[B=b|A=a]}\right]\\ &= \mathop{\mathbb{E}}_{a\leftarrow A'}[\mathrm{S}(B'_a\|B_a)] \quad \geq \quad 0.\end{aligned}$$

This means that

$$\mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B'_a\right)\right] \leq \mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B_a\right)\right] \leq \varepsilon. \tag{5.4}$$

Then

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B'_a\right)\right] \\
&= \mathrm{S}(A'C'B'\|(A'C')\,(B'|A')) && (5.5)\\
&= \mathrm{S}(A'B'C'\|(A'B')\,(C'|A')) && (5.6)\\
&\geq \|A'B'C' - (A'B')\,(C'|A')\|_1^2\,. && (5.7)
\end{aligned}$$

Above, Eq. (5.5) follows from the chain rule for the relative entropy, Eq. (5.6) follows because $(A'C')\,(B'|A')$ and $(A'B')\,(C'|A')$ are identically distributed, and Eq. (5.7) follows from Fact 5.1.5. Now from Equations (5.7) and (5.4) we get

$$\|A'B'C' - (A'B')\,(C'|A')\|_1 \leq \sqrt{\varepsilon}.$$

By similar arguments we get

$$\|A'B'C' - (A'B')\,(C'|B')\|_1 \leq \sqrt{\varepsilon}.$$

The inequalities above and Lemma 5.1.12 imply that (A', B') is $(1 - 4\sqrt{\varepsilon})$-embeddable in $(A'C', B'C')$. Furthermore from Fact 5.1.5 and $\mathrm{S}(A'B'\|AB) \leq \varepsilon$ we get

$$\|A'B' - AB\|_1 \leq \sqrt{\varepsilon}.$$

Finally using the inequality above, Fact 5.1.8 and the triangle inequality for the ℓ_1 norm, we get that (A, B) is $(1 - 5\sqrt{\varepsilon})$-embeddable in $(A'C', B'C')$.

□

5.2 Communication complexity

In this work, we are concerned with the model of communication complexity which was introduced by Yao [73]. In this model there are different parties who wish to compute a joint relation of their inputs. They do local computation, use public/private coins, and communicate between them to achieve this task. The player receiving the last message outputs the answer. The resource that is counted is the number of bits communicated. The text by Kushilevitz and Nisan [49] is an excellent reference for this model.

Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $k \geq 1$ be an integer and $\varepsilon \in (0, 1)$. And let

$f^k \subseteq \mathcal{X}^k \times \mathcal{Y}^k \times \mathcal{Z}^k$ be defined to be cross product of f with itself k times. In a protocol for computing f^k, Alice will receive input in $\mathcal{X}^k$, Bob will receive input in $\mathcal{Y}^k$ and the output of the protocol will be in $\mathcal{Z}^k$.

Two-way public-coin communication complexity. In a two-way public-coin communication protocol, Alice is given $x \in \mathcal{X}$, and Bob is given $y \in \mathcal{Y}$. They are supposed to output $z \in \mathcal{Z}$ such that $(x, y, z) \in f$ via exchanging messages and doing local computations. They may share public coins before the inputs are revealed to them. The *transcript* of a protocol is the concatenation of the public coins and all messages exchanged between Alice and Bob. Let $\mathrm{R}_{\varepsilon}^{\mathrm{pub}}(f)$ represent the two-way public-coin randomized communication complexity of f with the worst case error ε, that is the communication of the best two-way public-coin for f with error for each input (x, y) being at most ε. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Let $\mathrm{D}_{\varepsilon}^{\mu}(f)$ represent the two-way distributional communication complexity of f under distribution μ with distributional error ε, that is the communication of the best two-way deterministic protocol for f, with average error over the distribution of the inputs drawn from μ, at most ε. Following is Yao's min-max principle which connects the worst case error and the distributional error settings, see. e.g., [49, Theorem 3.20, page 36].

Fact 5.2.1 (Yao's principle, [73])**.** $\mathrm{R}_{\varepsilon}^{\mathrm{pub}}(f) = \max_{\mu} \mathrm{D}_{\varepsilon}^{\mu}(f)$.

Two-party bounded-round public-coin communication complexity. In a two-party t-message public-coin model of communication, Alice with input $x \in \mathcal{X}$ and Bob with input $y \in \mathcal{Y}$, do local computation using public coins shared between them and exchange t messages, with Alice sending the first message. At the end of their protocol the party receiving the t-th message outputs some $z \in \mathcal{Z}$. The output is declared correct if $(x, y, z) \in f$ and wrong otherwise. Let $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f)$ represent the two-party t-message public-coin communication complexity of f with worst case error ε, i.e., the communication of the best two-party t-message public-coin protocol for f with error for each input (x, y) being at most ε. We similarly consider two-party t-message deterministic protocols where there are no public coins used by Alice and Bob. Let $\mu \in \mathcal{X} \times \mathcal{Y}$ be a distribution. We let $\mathrm{D}_{\varepsilon}^{(t),\mu}(f)$ represent the two-party t-message distributional communication complexity of f under μ with expected error ε, i.e., the communication of the best two-party t-message deterministic protocol for f, with distributional error (average error over the inputs) at most ε under μ. We have similar Yao's principle for this model.

Lemma 5.2.2 (Yao's principle, [73])**.** $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f) = \max_{\mu} \mathrm{D}_{\varepsilon}^{(t),\mu}(f)$.

The following fact about communication protocols can be verified using the rectangle property of communication protocols.

Fact 5.2.3. Let there be t messages $M_1, \ldots, M_t$ in a deterministic communication protocol between Alice and Bob with inputs X, Y respectively where X and Y are independent. Then for any $s \in [t]$, X and Y are independent even conditioned on $M_1, \ldots, M_s$.

5.2.1 Smooth rectangle bounds

Besides showing direct product results, another major focus in communication complexity has been to investigate generic lower bound methods, that apply to all functions (and possibly to all relations). In the model we are concerned with, various generic lower bound methods are known, for example the *partition bound* [28], the *information complexity* [18], the smooth rectangle bound [28] (which in turn subsumes the rectangle bound a.k.a the corruption bound) [5; 9; 45; 63; 75], the smooth discrepancy bound a.k.a the γ_2 bound [52] (which in turn subsumes the discrepancy bound), the subdistribution bound [29] and the conditional min-entropy bound [25]. Proving strong direct product results in terms of these lower bound methods is a reasonable approach to attacking the general question. Indeed, many lower bounds have been shown to satisfy strong direct product theorems, example the discrepancy bound [51], the subdistribution bound under product distributions [29], the smooth discrepancy bound [67] and the conditional min-entropy bound [25].

Smooth rectangle bound was introduced by Jain and Klauck in [28], which generalizes the rectangle bound (a.k.a. the corruption bound) [5; 9; 45; 63; 75]. Roughly speaking, the rectangle bound for relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ under a distribution μ, with respect to an element $z \in \mathcal{Z}$, and error ε, tries to capture the size (under μ) of a largest rectangle for which z is a right answer for $1 - \varepsilon$ fraction of inputs inside the rectangle. It is not hard to argue that the rectangle bound forms a lower bound on the distributional communication complexity of f under μ. The smooth rectangle bound for f further captures the maximum, over all relations g that are close to f under μ, of the rectangle bound of g under μ. The distributional error setting can eventually be related to the worst case error setting via the well known Yao's principle [75].

Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\varepsilon, \delta \geq 0$. With a slight abuse of notation, we write $f(x, y) \stackrel{\text{def}}{=} \{z \in \mathcal{Z} | \ (x, y, z) \in f\}$, and $f^{-1}(z) \stackrel{\text{def}}{=} \{(x, y) : (x, y, z) \in f\}$.

Definition 5.2.4. (**Smooth-rectangle bound** [28]) The (ε, δ)-smooth rectangle bound

of f, denoted by $\widetilde{\mathrm{srec}}_{\varepsilon,\delta}(f)$, is defined as follows:

$$\widetilde{\mathrm{srec}}_{\varepsilon,\delta}(f) \stackrel{\text{def}}{=} \max\{\widetilde{\mathrm{srec}}^{\lambda}_{\varepsilon,\delta}(f) \mid \lambda \text{ a distribution over } \mathcal{X}\times\mathcal{Y}\};$$
$$\widetilde{\mathrm{srec}}^{\lambda}_{\varepsilon,\delta}(f) \stackrel{\text{def}}{=} \max\{\widetilde{\mathrm{srec}}^{z,\lambda}_{\varepsilon,\delta}(f) \mid z \in \mathcal{Z}\};$$
$$\widetilde{\mathrm{srec}}^{z,\lambda}_{\varepsilon,\delta}(f) \stackrel{\text{def}}{=} \max\{\widetilde{\mathrm{rec}}^{z,\lambda}_{\varepsilon}(g) \mid g \subseteq \mathcal{X}\times\mathcal{Y}\times\mathcal{Z};\ \Pr_{(x,y)\leftarrow\lambda}[f(x,y)\neq g(x,y)] \le \delta\};$$
$$\widetilde{\mathrm{rec}}^{z,\lambda}_{\varepsilon}(g) \stackrel{\text{def}}{=} \min\{\mathrm{S}_{\infty}(\lambda_R\|\lambda) \mid R \text{ is a rectangle in } \mathcal{X}\times\mathcal{Y},\ \lambda(g^{-1}(z)\cap R) \ge (1-\varepsilon)\lambda(R)\}.$$

When $\delta = 0$, the smooth rectangle bound equals the rectangle bound (a.k.a. the corruption bound) [5; 9; 45; 63; 75]. Definition 5.2.4 is a generalization of the one in [28], where it is only defined for boolean functions.

Jain and Klauck showed that the smooth rectangle bound is stronger than every lower bound method we mentioned above except the partition bound and the information complexity. Jain and Klauck showed that the partition bound subsumes the smooth rectangle bound and in a recent work Kerenidis, Laplante, Lerays, Roland and Xiao [43] showed that the information complexity subsumes the smooth rectangle bound (building on the work of Braverman and Weinstein [16] who showed that the information complexity subsumes the discrepancy bound). New lower bounds for specific functions have been discovered using the smooth rectangle bound, for example Chakrabarti and Regev's [17] optimal lower bound for the Gap-Hamming Distance partial function. Klauck [46] used the smooth rectangle bound to show a strong direct product result for the Set-Disjointness function, via exhibiting a lower bound on a related function. On the other hand, as far as we know, no function (or relation) is known for which its smooth rectangle bound is (asymptotically) strictly smaller than its two-way public-coin communication complexity. Hence establishing whether or not the smooth rectangle bound is a tight lower bound for all functions and relations in this model is an important open question.

In [28], Jain and Klauck provide an alternate definition of smooth-rectangle bound for boolean functions.

Definition 5.2.5. For function $f : \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$, the ϵ- smooth rectangle bound of f denoted $\mathrm{srec}_{\epsilon}(f)$ is defined to be $\max\{\mathrm{srec}^{z}_{\epsilon}(f) : z \in \mathcal{Z}\}$, where $\mathrm{srec}^{z}_{\epsilon}(f)$ is given by the optimal value of the following linear program. ($\mathcal{W}$ represents the set of all rectangles.)

Primal

$$
\begin{aligned}
&\text{min:} \quad \sum_{W \in \mathcal{W}} v_W \\
&\forall (x,y) \in f^{-1}(z) : \sum_{W:(x,y)\in W} v_W \geq 1-\epsilon, \\
&\forall (x,y) \in f^{-1}(z) : \sum_{W:(x,y)\in W} v_W \leq 1, \\
&\forall (x,y) \in f^{-1} - f^{-1}(z) : \sum_{W:(x,y)\in W} v_W \leq \epsilon, \\
&\forall W : v_W \geq 0 \ .
\end{aligned}
$$

Dual

$$
\begin{aligned}
&\text{max:} \quad \sum_{(x,y)\in f^{-1}(z)} ((1-\epsilon)\lambda_{x,y} - \phi_{x,y}) - \sum_{(x,y)\notin f^{-1}(z)} \epsilon \cdot \lambda_{x,y} \\
&\forall W : \sum_{(x,y)\in f^{-1}(z)\cap W} (\lambda_{x,y} - \phi_{x,y}) - \sum_{(x,y)\in (W \cap f^{-1} - f^{-1}(z))} \lambda_{x,y} \leq 1, \\
&\forall (x,y) : \lambda_{x,y} \geq 0; \phi_{x,y} \geq 0 \ .
\end{aligned}
$$

The following lemma lower bounds the natural definition in terms of the linear programming definition of smooth rectangle bound. A similar, but weaker, relationship was shown in [28].

Lemma 5.2.6. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Let $z \in Z$ and $\varepsilon > 0$. There exists a distribution $\mu \in \mathcal{X} \times \mathcal{Y}$ and $\delta, \beta > 0$ such that*

$$
\widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon^2)\frac{\delta}{\beta},\delta}(f) \geq \log(\mathrm{srec}^z_\varepsilon(f)) + 3\log\varepsilon.
$$

Proof. See Appendix A. □

The smooth rectangle bound is a lower bound on the two-way public-coin communication complexity. It is first proved by Jain and Klauck in [28]. The proof is contained in Appendix A for completeness.

Lemma 5.2.7. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\lambda \in \mathcal{X} \times \mathcal{Y}$ be a distribution and let $z \in \mathcal{Z}$. Let $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow\lambda}[f(x,y) = \{z\}]$. Let $\varepsilon, \varepsilon', \delta > 0$ be such that $\frac{\delta+\varepsilon}{\beta-2\varepsilon} < (1+\varepsilon')\frac{\delta}{\beta}$. Then,*

$$
\mathrm{R}_\varepsilon(f) \geq \mathrm{D}^\lambda_\varepsilon(f) \geq \widetilde{\mathrm{srec}}^{z,\lambda}_{(1+\varepsilon')\delta/\beta,\delta}(f) - \log\frac{4}{\varepsilon}.
$$

Chapter 6

A direct product theorem for two-party bounded-round public-coin communication complexity

6.1 Introduction

In this chapter, we show a direct product theorem for the two-party bounded-round public-coin communication complexity. In this model, for computing a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ ($\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are finite sets), one party, say Alice, is given an input $x \in \mathcal{X}$ and the other party, say Bob, is given an input $y \in Y$. They are supposed to do local computations using public coins shared between them, communicate a fixed number of messages between them and at the end, output an element $z \in Z$. They are said to succeed if $(x, y, z) \in f$. In this chapter we only consider *complete* relations, that is for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, there is some $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. Using the notations introduced in Section 5.2, we show that

Theorem 6.1.1. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ a relation, $\varepsilon > 0$ and $k, t \geq 1$ be integers. There exists a constant $\kappa \geq 0$ such that,*

$$\mathrm{R}^{(t),\mathrm{pub}}_{1-(1-\varepsilon/2)^{\Omega(k\varepsilon^2/t^2)}}(f^k) = \Omega\left(\frac{\varepsilon \cdot k}{t} \cdot \left(\mathrm{R}^{(t),\mathrm{pub}}_{\varepsilon}(f) - \frac{\kappa t^2}{\varepsilon}\right)\right).$$

In particular, it implies a strong direct product theorem for the two-party constant-

message public-coin communication complexity of all relations f.[1] Our result generalizes the result of Jain [25] which can be regarded as the special case when $t = 1$. Our result can be considered as an important progress towards settling the strong direct product conjecture for the two-party public-coin communication complexity, a major open question in this area.

As a direct consequence of our result we get a direct product theorem for the Pointer Chasing problem defined as follows. Let $n, t \geq 1$ be integers. Alice and Bob are given functions $F_A : [n] \to [n]$ and $F_B : [n] \to [n]$, respectively. Let F^t represent alternate composition of F_A and F_B done t times, starting with F_A. The parties are supposed to communicate and determine $F^t(1)$. In the bit version of the problem, the players are supposed to output the least significant bit of $F^t(1)$. We refer to the t-pointer chasing problem as FP_t and the bit version as BP_t. The pointer chasing problem naturally captures the trade-off between number of messages exchanged and the communication used. There is a straightforward t-message deterministic protocol with $t \cdot \log n$ bits of communication for both FP_t and BP_t. However if only $t - 1$ messages are allowed to be exchanged between the parties, exponentially more communication is required. The communication complexity of this problem has been very well studied both in the classical and quantum models of communication complexity [32; 44; 47; 57; 60]. Some tight lower bounds that we know so far are as follows.

Theorem 6.1.2. *For integer $t \geq 1$,*

1. *[60]* $\mathrm{R}^{(t-1),\mathrm{pub}}_{1/3}(\mathrm{FP}_t) \geq \Omega(n \log^{(t-1)} n)$;
   $\mathrm{R}^{(t-1),\mathrm{pub}}_{1/3}(\mathrm{BP}_t) \geq \Omega(n)$.

As a consequence of Theorem 6.1.1 we get strong direct product results for this problem. Note that in the descriptions of FP_t and BP_t, t is a fixed constant, not dependent on the input size.

Corollary 6.1.3. *For integers $t, k \geq 1$,*

1. $\mathrm{R}^{(t-1),\mathrm{pub}}_{1-2^{-\Omega(k/t^2)}}(\mathrm{FP}^k_t) \geq \Omega\left(\frac{k}{t} \cdot n \log^{(t-1)} n\right)$;
2. $\mathrm{R}^{(t-1),\mathrm{pub}}_{1-2^{-\Omega(k/t^2)}}(\mathrm{BP}^k_t) \geq \Omega\left(\frac{k}{t} \cdot n\right)$.

[1] When $\mathrm{R}^{(t),\mathrm{pub}}_{\varepsilon}(f)$ is a constant, a direct product result can be shown via direct arguments as for example in [25].

6.1.1 Our techniques

We prove our direct product result using information theoretic arguments. Information theory is a versatile tool in communication complexity, especially in proving lower bounds and direct sum and direct product theorems [6; 8; 13; 18; 25; 27; 33; 34; 35]. The broad argument that we use is as follows. For a given relation f, let the communication required for computing one instance with t messages and constant success be c. Let us consider a protocol for computing f^k with t messages and communication cost $o(kc)$. Let us condition on success on a set C of coordinates. If the overall success in coordinates in $C \subseteq [k]$ is already as small as we want then we are done and stop. Otherwise we exhibit another coordinate j outside C such that the success in the j-th coordinate, even conditioned on the success in the l coordinates, is bounded away from 1. This way the overall success keeps going down and becomes exponentially small (in k) eventually. More concretely, the distribution of inputs X_jY_j (conditioning on the success of the coordinates in C), in the j-th coordinate is quite close to μ and the joint distribution X_jY_jM (where M is the message transcript of $\mathcal{P}$) can be approximated very well by Alice and Bob using a t message protocol for f, when they are given input according to μ, using communication less than c. This shows that success in the j-th coordinate must be bounded away from one. We do this argument in the distributional setting where one is concerned with average error over the inputs coming from a specified distribution rather than the worst case error over all inputs. The distributional setting is then related to the worst case setting by the well known Yao's principle [73].

To simulate the transcript, we adopt the message compression protocol due to Braverman and Rao [13], where they used the protocol to show a direct sum theorem for the same communication model we are considering. Informally, the protocol can be stated as follows.

Braverman-Rao protocol (informal). *Given a Markov chain $Y \leftrightarrow X \leftrightarrow M$, there exists a public-coin protocol between Alice and Bob, with input X, Y, respectively, with a single message from Alice to Bob of $\mathcal{O}(\mathrm{I}(X : M|Y))$ bits, such that at the end of the protocol, Alice and Bob both possess a random variable M', close to M in ℓ_1 distance.*

Consider the situation after conditioning on the success in the set C as above, and let X_jY_j represent the input in the jth coordinate. The Braverman-Rao compression protocol cannot be directly applied at this stage. Take the first message M_1 sent by Alice, for instance. $Y_jX_jM_1$ is not necessarily a Markov chain even if the initial distribution is product. However, we are able to show that $Y_jX_jM_1$ is 'close' to being a Markov

chain by further conditioning on appropriate sub-events. We then use a more 'robust' Braverman-Rao compression protocol (along the lines of the original), where by being 'robust', we mean that the communication cost and the error do not vary much even for XYM which is close to being a Markov chain (similar arguments were used in Jain [25]). We then apply such a robust message compression protocol to each successive message. We accumulate some errors for each of these messages. Thus in order to keep the overall error bounded, we are able to make our argument for protocols with a bounded number of message exchanges.

Another difficulty that is faced in this argument is that since μ may be a non-product distribution, Alice and Bob may obtain information about each other's input in the j-th coordinate via their inputs in other coordinates. This is overcome by splitting the distribution μ into a convex combination of several product distributions. This idea of splitting a non-product distribution into convex combination of product distributions has been used in several previous works to handle non-product distributions in different settings [6; 8; 13; 23; 25; 61; 63]. This splitting of non-product distribution leads us to use another important tool namely the *correlated sampling* protocol, that was also used for example by Holenstein [23] while arguing a strong direct product result for the two-prover one-round games.

As mentioned previously, we build on the arguments used in Jain [25]. Jain shows a new characterization of the two-party one-way public-coin communication complexity and uses it to show a strong direct product result for all relations in this model. We are unable to arrive at such a characterization for protocols with more than one message and use a more direct approach, as outlined above, to arrive at our direct product result.

6.2 Proof of Theorem 6.1.1

We start by showing a few lemmas which are helpful in the proof of the main result. The following lemma was shown in Jain [25] and follows primarily from a message compression argument due to Braverman and Rao [13].

Theorem 6.2.1. *Let $\delta > 0, c \geq 0$. Let X', Y', N be random variables for which $Y' \leftrightarrow X' \leftrightarrow N$ is a Markov chain and the following holds,*

$$\Pr_{(x,y,m)\leftarrow X',Y',N}\left[\log \frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} > c\right] \leq \delta. \tag{6.1}$$

There exists a public-coin protocol between Alice and Bob, with inputs X', Y' respec-

tively, with a single message from Alice to Bob of $c + \mathcal{O}(\log(1/\delta))$ *bits, such that at the end of the protocol, both Alice and Bob possess a random variable* M *satisfying* $\|X'Y'N - X'Y'M\|_1 \leq 2\delta$.

Remark 6.2.2. In [13], the condition $\mathrm{I}(X' : N|Y') \leq c$ is used instead of (6.1). It is changed to the current one in Jain [25]. By Markov's inequality, $\mathrm{I}(X' : N|Y') \leq c$ implies

$$\Pr_{(x,y,m)\leftarrow X',Y',N}\left[\log \frac{\Pr[N = m|X' = x]}{\Pr[N = m|Y' = y]} > \frac{c+1}{\delta}\right] \leq \delta.$$

This modification is essential in our arguments since the condition (6.1) is robust when the underlying joint distribution is perturbed slightly, while $\mathrm{I}(X' : N|Y')$ may change a lot with such a perturbation.

As mentioned in Subsection 6.1.1, we have to work with approximate Markov chains in our arguments for the direct product. The following lemma makes Theorem 6.1.1 more robust to deal with approximate Markov chains.

Lemma 6.2.3. *Let* $c \geq 0, 1 > \varepsilon > 0, \varepsilon' > 0$. *Let* X', Y', M' *be random variables for which the following holds,*

$$\mathrm{I}(X' : M'|Y') \leq c \text{ and } \mathrm{I}(Y' : M'|X') \leq \varepsilon.$$

There exists a public-coin protocol between Alice and Bob, with inputs X', Y' *respectively, with a single message from Alice to Bob of* $\frac{c+5}{\varepsilon'} + \mathcal{O}(\log \frac{1}{\varepsilon'})$ *bits, such that at the end of the protocol, both Alice and Bob possess a random variable* M *satisfying* $\|X'Y'M' - X'Y'M\|_1 \leq 3\sqrt{\varepsilon} + 6\varepsilon'$.

Proof. Let us introduce a new random variable N with joint distribution $X'Y'N \stackrel{\text{def}}{=} (X'Y')(M'|X')$. Note that $Y' \leftrightarrow X' \leftrightarrow N$ is a Markov chain. Using Lemma 5.1.1, we have

$$\mathrm{S}(X'Y'M'\|X'Y'N) = \mathrm{I}(Y' : M'|X') \leq \varepsilon. \tag{6.2}$$

Applying Fact 5.1.5, we get $\|X'Y'M' - X'Y'N\|_1 \leq \sqrt{\varepsilon}$. Theorem 6.2.1 and the following claim together imply that there exists a public-coin protocol between Alice and Bob, with input X', Y', respectively, with a single message from Alice to Bob of $\frac{c+5}{\varepsilon'} + \mathcal{O}(\log \frac{1}{\varepsilon'})$ bits, at the end of which both Alice and Bob possess a random variable N' satisfying $\|X'Y'N' - X'Y'N\|_1 \leq 2\sqrt{\varepsilon} + 6\varepsilon'$. Finally using the triangle inequality for the ℓ_1 norm we conclude the desired. □

Claim 6.2.4.

$$\Pr_{(m,x,y)\leftarrow N,X',Y'}\left[\log\frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]}\geq\frac{c+5}{\varepsilon'}\right]\leq 3\varepsilon'+\sqrt{\varepsilon}.$$

As mentioned in Remark 6.2.2, although mutual information is not robust, an upper bound on the mutual information implies an upper bound on the majority of the logarithm of a ratio, which turns out to be robust. We can also apply this trick to the bounds on other information theoretic quantities. The following claim is a robust version of the inequality $\mathrm{S}(M'X'Y'\|NXY)\geq 0$.

Claim 6.2.5. Given $0<\varepsilon,\varepsilon'<1$, we have

$$\Pr_{(m,x,y)\leftarrow M',X',Y'}\left[\log\frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}<\frac{\varepsilon+1}{\varepsilon'}\right]>1-\varepsilon'.$$

Proof. We prove it by applying Markov inequality to $\mathrm{S}(M'X'Y'\|NXY)\geq 0$. let us define the set

$$G_1\stackrel{\text{def}}{=}\left\{(m,x,y):\log\frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}<\frac{\varepsilon+1}{\varepsilon'}\right\}.$$

Consider,

$$\begin{aligned}
0\geq -\mathrm{S}(M'X'Y'\|NXY)&\geq -\mathop{\mathbb{E}}_{(x,y)\leftarrow X',Y'}\left[\mathrm{S}\left(M'_{xy}\middle\|N_{xy}\right)\right]\\
&=\mathop{\mathbb{E}}_{(m,x,y)\leftarrow M',X',Y'}\left[\log\frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\right] &(6.3)\\
&=\sum_{(m,x,y)\in G_1}\left(\Pr[M'=m,X'=x,Y'=y]\cdot\log\frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\right)\\
&\quad+\sum_{(m,x,y)\notin G_1}\left(\Pr[M'=m,X'=x,Y'=y]\cdot\log\frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\right)\\
&\geq\sum_{(m,x,y)\in G_1}\left(\Pr[M'=m,X'=x,Y'=y]\cdot\log\frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\right)\\
&\quad+\Pr[(M',X',Y')\notin G_1]\cdot\frac{\varepsilon+1}{\varepsilon'} &(6.4)\\
&=\sum_{(m,x,y)\notin G_1}\left(\Pr[M'=m,X'=x,Y'=y]\cdot\log\frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[N=m|X'=x,Y'=y]}\right)\\
&\quad-\mathrm{S}(M'X'Y'\|NX'Y')+\Pr[(M',X',Y')\notin G_1]\cdot\frac{\varepsilon+1}{\varepsilon'} &(6.5)\\
&\geq -1-\varepsilon+\Pr[(M',X',Y')\notin G_1]\cdot\frac{\varepsilon+1}{\varepsilon'}. &(6.6)
\end{aligned}$$

Above, Eq. (6.3) and Eq. (6.5) follow from the definition of the relative entropy, and Eq. (6.4) follows from the definition of G_1. To get Eq. (6.6), we use Fact 5.1.6 and Eq. (6.2). Eq. (6.6) implies that $\Pr[(M', X', Y') \notin G_1] \leq \varepsilon'$. □

Applying Markov inequality to the condition $\mathrm{I}(M' : X'|Y') \leq c$, we can get the following claim.

Claim 6.2.6. $\Pr_{(m,x,y)\leftarrow M',X',Y'}\left[\log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]} < \frac{c+1}{\varepsilon'}\right] \geq 1-\varepsilon'$.

Proof. Let us define

$$G_2 \stackrel{\text{def}}{=} \left\{(m,x,y) : \log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]} < \frac{c+1}{\varepsilon'}\right\}.$$

Consider,

$$\begin{aligned}
c &\geq \mathrm{I}(M' : X'|Y') && (6.7)\\
&= \mathop{\mathbb{E}}_{(m,x,y)\leftarrow M',X',Y'}\left[\log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]}\right] && (6.8)\\
&= \sum_{(m,x,y)\in G_2}\left(\Pr[M'=m,X'=x,Y'=y]\cdot \log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]}\right)\\
&\quad + \sum_{(m,x,y)\notin G_2}\left(\Pr[M'=m,X'=x,Y'=y]\cdot \log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]}\right)\\
&\geq \frac{c+1}{\varepsilon'}\cdot \Pr[(M',X',Y') \notin G_2] - 1. && (6.9)
\end{aligned}$$

Above Eq. (6.7) is one of the assumptions in the lemma; Eq. (6.8) follows from the definition of the conditional mutual information; Eq. (6.9) follows from the definition of G_2 and Fact 5.1.6. Eq. (6.9) implies that $\Pr[(M', X', Y') \notin G_2] \leq \varepsilon'$. □

Applying Markov inequality to (6.2), we have the following claim.

Claim 6.2.7. $\Pr_{(m,x,y)\leftarrow M',X',Y'}\left[\log \frac{\Pr[M'=m,Y'=y]}{\Pr[N=m,Y'=y]} < \frac{\varepsilon+1}{\varepsilon'}\right] \geq 1-\varepsilon'$.

Proof. Define

$$G_3 \stackrel{\text{def}}{=} \left\{(m,x,y) : \log \frac{\Pr[M'=m,Y'=y]}{\Pr[N=m,Y'=y]} < \frac{\varepsilon+1}{\varepsilon'}\right\}.$$

Consider,

$$\begin{aligned}
\varepsilon &\geq \mathrm{S}(X'Y'M' \| X'Y'N) \\
&\geq \mathrm{S}(Y'M' \| Y'N) \qquad (6.10) \\
&= \mathop{\mathbb{E}}_{(m,x,y)\leftarrow M',X',Y'} \left[\log \frac{\Pr[M'=m, Y'=y]}{\Pr[N=m, Y'=y]} \right] \\
&= \sum_{(m,x,y)\in G_3} \left(\Pr[M'=m, X'=x, Y'=y] \cdot \log \frac{\Pr[M'=m, Y'=y]}{\Pr[N=m, Y'=y]} \right) \\
&\quad + \sum_{(m,x,y)\notin G_3} \left(\Pr[M'=m, X'=x, Y'=y] \cdot \log \frac{\Pr[M'=m, Y'=y]}{\Pr[N=m, Y'=y]} \right) \\
&\geq -1 + \Pr[(M', X', Y') \notin G_3] \cdot \frac{\varepsilon+1}{\varepsilon'}. \qquad (6.11)
\end{aligned}$$

Above Eq. (6.10) follows from Fact 5.1.7 and Eq. (6.11) follows from definition of G_3. This implies $\Pr[(M', X', Y') \notin G_3] \leq \varepsilon'$. □

With those claims above we can prove Claim 6.2.4.

Proof of Claim 6.2.4:.

$$\begin{aligned}
&\log \frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} = \log \frac{\Pr[N=m|X'=x, Y'=y]}{\Pr[N=m|Y'=y]} \\
&= \log \frac{\Pr[N=m|X'=x, Y'=y]}{\Pr[M'=m|X'=x, Y'=y]} + \log \frac{\Pr[M'=m|X'=x, Y'=y]}{\Pr[M'=m|Y'=y]} \\
&\quad + \log \frac{\Pr[M'=m, Y'=y]}{\Pr[N=m, Y'=y]}. \qquad (6.12)
\end{aligned}$$

From union bound and above we get (recall $1 > \varepsilon > 0$),

$$\begin{aligned}
&\Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}} \left[\log \frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} \geq \frac{c+5}{\varepsilon'} \right] \\
&= \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}} \left[\log \frac{\Pr[N=m|X'=x, Y'=y]}{\Pr[N=m|Y'=y]} \geq \frac{c+5}{\varepsilon'} \right] \\
&\leq \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}} \left[\log \frac{\Pr[N=m|X'=x, Y'=y]}{\Pr[M'=m|X'=x, Y'=y]} \geq \frac{\varepsilon+1}{\varepsilon'} \right] \\
&+ \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}} \left[\log \frac{\Pr[M'=m|X'=x, Y'=y]}{\Pr[M'=m|Y'=y]} \geq \frac{c+1}{\varepsilon'} \right]
\end{aligned}$$

$$+ \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}}\left[\log\frac{\Pr[M'=m,Y'=y]}{\Pr[N=m,Y'=y]} \geq \frac{\varepsilon+1}{\varepsilon'}\right]. \tag{6.13}$$

By Claim 6.2.5, 6.2.6 and 6.2.7, each term is bounded from above by ε'. Combining the bounds for the three terms we get

$$\Pr_{(m,x,y)\leftarrow M',X',Y'}\left[\log\frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} \geq \frac{c+5}{\varepsilon'}\right] \leq 3\varepsilon'.$$

Using $\|X'Y'M' - X'Y'N\|_1 \leq \sqrt{\varepsilon}$ (as was shown previously), we finally have,

$$\Pr_{(m,x,y)\leftarrow N,X',Y'}\left[\log\frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} \geq \frac{c+5}{\varepsilon'}\right] \leq 3\varepsilon' + \sqrt{\varepsilon}.$$

□

The following lemma generalizes the lemma above to deal with multiple messages, as needed for our purposes.

Lemma 6.2.8. *Let $t \geq 1$ be an integer. Let $\varepsilon' > 0$, $c_s \geq 0, 1 > \varepsilon_s > 0$ for each $1 \leq s \leq t$. Let $R', X', Y', M'_1, \ldots, M'_t$, be random variables for which the following holds (below $M'_{<s} \stackrel{\text{def}}{=} M'_1 \cdots M'_{s-1}$),*

$$\mathrm{I}(X' : M'_s | Y'R'M'_{<s}) \leq c_s, \mathrm{I}(Y' : M'_s | X'R'M'_{<s}) \leq \varepsilon_s, \tag{6.14}$$

for odd s, and

$$\mathrm{I}(Y' : M'_s | X'R'M'_{<s}) \leq c_s, \mathrm{I}(X' : M'_s | Y'R'M'_{<s}) \leq \varepsilon_s,$$

for even s.

There exists a public-coin t-message protocol $\mathcal{P}_t$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message. The total communication of $\mathcal{P}_t$ is $\frac{\sum_{s=1}^t c_s + 5t}{\varepsilon'} + \mathcal{O}\left(t \log \frac{1}{\varepsilon'}\right)$, and at end of the protocol, both Alice and Bob possess random variables $M_1, \ldots, M_t$, satisfying: $\|R'X'Y'M_1 \cdots M_t - R'X'Y'M'_1 \cdots M'_t\|_1 \leq 3\sum_{s=1}^t \sqrt{\varepsilon_s} + 6\varepsilon' t$.

Proof. We prove the lemma by induction on t. For the base case $t = 1$, note that

$$\mathrm{I}(X'R' : M'_1 | Y'R') = \mathrm{I}(X' : M'_1 | Y'R') \leq c_1$$

and

$$\mathrm{I}(Y'R' : M_1'|X'R') = \mathrm{I}(Y' : M_1'|X'R') \leq \varepsilon_1.$$

Lemma 6.2.3 implies (by taking $X', Y,' M'$ in Lemma 6.2.3 to be $X'R', Y'R', M_1'$ respectively) that Alice, with input $X'R'$, and Bob, with input $Y'R'$, can run a public-coin protocol with a single message from Alice to Bob of

$$\frac{c_1 + 5}{\varepsilon'} + \mathcal{O}(\log \frac{1}{\varepsilon'})$$

bits and generate a new random variable M_1 satisfying

$$\|R'X'Y'M_1' - R'X'Y'M_1\|_1 \leq 3\sqrt{\varepsilon_1} + 6\varepsilon'.$$

Now let $t > 1$. Assume t is odd, for even t a similar argument follows. From the induction hypothesis there exists a public-coin $t-1$ message protocol $\mathcal{P}_{t-1}$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message, and total communication

$$\frac{\sum_{s=1}^{t-1} c_s + 5(t-1)}{\varepsilon'} + \mathcal{O}\left((t-1)\log\frac{1}{\varepsilon'}\right), \tag{6.15}$$

such that at the end both Alice and Bob possess random variables $M_1, \ldots, M_{t-1}$ satisfying

$$\|R'X'Y'M_1 \cdots M_{t-1} - R'X'Y'M_1' \cdots M_{t-1}'\|_1 \leq 3\sum_{s=1}^{t-1}\sqrt{\varepsilon_s} + 6\varepsilon'(t-1). \tag{6.16}$$

Note that

$$\mathrm{I}(Y'R'M_{<t}' : M_t'|X'R'M_{<t}') = \mathrm{I}(Y' : M_t'|X'R'M_{<t}') \leq c_t$$

and

$$\mathrm{I}(X'R'M_{<t}' : M_t'|Y'R'M_{<t}') = \mathrm{I}(X' : M_t'|Y'R'M_{<t}') \leq \varepsilon_t.$$

Therefore Lemma 6.2.3 implies (by taking $X', Y,' M'$ in Lemma 6.2.3 to be $X'R'M_{<t}', Y'R'M_{<t}', M_t'$ respectively) that Alice, with input $X'R'M_{<t}'$, and Bob, with input $Y'R'M_{<t}'$, can run a public coin protocol $\mathcal{P}$ with a single message from Alice to Bob of

$$\frac{c_t + 5}{\varepsilon'} + \mathcal{O}\left(\log\frac{1}{\varepsilon'}\right) \tag{6.17}$$

bits and generate a new random variable M_t'' satisfying

$$\left\|R'X'Y'M_1'\cdots M_{t-1}'M_t' - R'X'Y'M_1'\cdots M_{t-1}'M_t''\right\|_1 \leq 3\sqrt{\varepsilon_t} + 6\varepsilon'. \tag{6.18}$$

Fact 5.1.8 and Eq. (6.16) imply that Alice, on input $X'R'M_{<t}$ and Bob on input $Y'R'M_{<t}$, on running the same protocol $\mathcal{P}$ will generate a new random variable M_t satisfying

$$\begin{aligned}
&\|R'X'Y'M_1\cdots M_{t-1}M_t - R'X'Y'M_1'\cdots M_{t-1}'M_t''\|_1 \\
&= \|R'X'Y'M_1\cdots M_{t-1} - R'X'Y'M_1'\cdots M_{t-1}'\|_1 \\
&\leq 3\sum_{s=1}^{t-1}\sqrt{\varepsilon_s} + 6\varepsilon'(t-1).
\end{aligned} \tag{6.19}$$

Therefore by composing protocol $\mathcal{P}_{t-1}$ and protocol $\mathcal{P}$, using Equations (6.15), (6.17), (6.18), (6.19) and the triangle inequality for the ℓ_1 norm, we get a public-coin t-message protocol $\mathcal{P}_t$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message, and total communication

$$\frac{\sum_{s=1}^{t} c_s + 5t}{\varepsilon'} + \mathcal{O}\left(t\log\frac{1}{\varepsilon'}\right),$$

such that at the end Alice and Bob both possess random variables $M_1,\ldots,M_t$ satisfying

$$\|R'X'Y'M_1\cdots M_t - R'X'Y'M_1'\cdots M_t'\|_1 \leq 3\sum_{s=1}^{t}\sqrt{\varepsilon_s} + 6\varepsilon' t. \qquad \square$$

In the lemma above, Alice and Bob shared an input R' (potentially correlated with $X'Y'$). Eventually we need Alice and Bob to generate this shared part themselves using correlated sampling. The following lemma, obtained from the lemma above, is the one that we finally use in the proof of our main result.

Lemma 6.2.9. *Let random variables $R', X', Y', M_1', \ldots, M_t'$ and numbers $\varepsilon', c_s, \varepsilon_s$ satisfy all the conditions in Lemma 6.2.8. Let $\tau > 0$ and let random variables (X, Y) be $(1-\tau)$-embeddable in $(X'R', Y'R')$. There exists a public-coin t-message protocol $\mathcal{Q}_t$ between Alice, with input X, and Bob, with input Y, with Alice sending the first message, and total communication $\frac{\sum_{s=1}^{t} c_s + 5t}{\varepsilon'} + \mathcal{O}\left(t\log\frac{1}{\varepsilon'}\right)$ bits, such that at the end Alice possesses $R_A M_1\cdots M_t$ and Bob possesses $R_B M_1\cdots M_t$, such that: $\|XYR_AR_BM_1\cdots M_t - X'Y'R'R'M_1'\cdots M_t'\|_1 \leq \tau + 3\sum_{s=1}^{t}\sqrt{\varepsilon_s} + 6\varepsilon' t$.*

Proof. In $\mathcal{Q}_t$, Alice and Bob, using public coins and no communication first generate

R_A, R_B such that $\|XYR_AR_B - X'Y'R'R'\|_1 \leq \tau$. They can do this from the Definition 5.1.11 of embedding. Now they will run protocol $\mathcal{P}_t$ (as in Lemma 6.2.8) with Alice's input being XR_A and Bob's input being YR_B and at the end both possess $M_1, \ldots, M_t$. From Lemma 6.2.8, the communication of $\mathcal{Q}_t$ is as desired. Now from Fact 5.1.8, Lemma 6.2.8 and the triangle inequality for the ℓ_1 norm,

$$\|XYR_AR_BM_1 \cdots M_t - X'Y'R'R'M'_1 \cdots M'_t\|_1 \leq \tau + 3\sum_{s=1}^{t} \sqrt{\varepsilon_s} + 6\varepsilon' t. \qquad \square$$

We are now ready to prove our main result, Theorem 6.1.1. We restate it here for convenience.

Theorem 6.2.10. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ a relation, $\varepsilon > 0$ and $k, t \geq 1$ be integers. There exists a constant $\kappa \geq 0$ such that,*

$$\mathrm{R}^{(t),\mathrm{pub}}_{1-(1-\varepsilon/2)^{\Omega(k\varepsilon^2/t^2)}}(f^k) = \Omega\left(\frac{\varepsilon \cdot k}{t} \cdot \left(\mathrm{R}^{(t),\mathrm{pub}}_{\varepsilon}(f) - \frac{\kappa t^2}{\varepsilon}\right)\right).$$

Proof of Theorem 6.1.1: Let $\delta \stackrel{\mathrm{def}}{=} \frac{\varepsilon^2}{7500t^2}$ and $\delta_1 = \frac{\varepsilon}{3000t}$. From Yao's principle, Lemma 5.2.2, it suffices to prove that for any distribution μ on $\mathcal{X} \times \mathcal{Y}$, $\mathrm{D}^{(t),\mu^k}_{1-(1-\varepsilon/2)^{\lfloor \delta k \rfloor}}(f^k) \geq \delta_1 kc$, where $c \stackrel{\mathrm{def}}{=} \mathrm{D}^{(t),\mu}_{\varepsilon}(f) - \frac{\kappa t^2}{\varepsilon}$, for constant κ to be chosen later. Let $XY \sim \mu^k$. Let $\mathcal{Q}$ be a t-message deterministic protocol between Alice, with input X, and Bob, with input Y, that computes f^k, with Alice sending the first message and total communication $\delta_1 kc$ bits. We assume t is odd for the rest of the argument and Bob makes the final output (the case when t is even follows similarly). The following Claim 6.2.11 implies that the success of $\mathcal{Q}$ is at most $(1-\varepsilon/2)^{\lfloor \delta k \rfloor}$ and this shows the desired. $\square$

Claim 6.2.11. For each $i \in [k]$, define a binary random variable $T_i \in \{0, 1\}$, which represents the success of $\mathcal{Q}$ (that is Bob's output being correct) on the i-th instance. That is, $T_i = 1$ if the protocol $\mathcal{Q}$ computes the i-th instance of f correctly, and $T_i = 0$ otherwise. Let $k' \stackrel{\mathrm{def}}{=} \lfloor \delta k \rfloor$. There exist k' coordinates $\{i_1, \ldots, i_{k'}\}$ such that for each $1 \leq r \leq k' - 1$,

$$\text{either} \quad \Pr\left[T^{(r)} = 1\right] \leq (1-\varepsilon/2)^{k'} \quad \text{or}$$

$$\Pr\left[T_{i_{r+1}} = 1 \middle| T^{(r)} = 1\right] \leq 1 - \varepsilon/2,$$

where $T^{(r)} \stackrel{\mathrm{def}}{=} \prod_{j=1}^{r} T_{i_j}$.

Proof of Claim 6.2.11: For $s \in [t]$, denote the s-th message of Ω by M_s. Define $M \stackrel{\text{def}}{=} M_1 \cdots M_t$. In the following we assume $1 \leq r < k'$, however same arguments also work when $r = 0$, that is for identifying the first coordinate, which we skip for the sake of avoiding repetition. Suppose we have already identified r coordinates $i_1, \ldots, i_r$ satisfying that $\Pr[T_{i_1} = 1] \leq 1 - \varepsilon/2$ and $\Pr[T_{i_{j+1}} = 1 | T^{(j)} = 1] \leq 1 - \varepsilon/2$ for $1 \leq j \leq r-1$. If $\Pr\left[T^{(r)} = 1\right] \leq (1-\varepsilon/2)^{k'}$, we are done. So from now on, assume $\Pr\left[T^{(r)} = 1\right] > (1-\varepsilon/2)^{k'} \geq 2^{-\delta k}$.

Let D be a random variable uniformly distributed in $\{0,1\}^k$ and independent of XY. Let $U_i = X_i$ if $D_i = 0$, and $U_i = Y_i$ if $D_i = 1$. For any random variable L, let us introduce the notation: $L^1 \stackrel{\text{def}}{=} (L | T^{(r)} = 1)$. For example, $X^1Y^1 = (XY | T^{(r)} = 1)$. If $L = L_1 \cdots L_k$, define $L_{-i} \stackrel{\text{def}}{=} L_1 \cdots L_{i-1} L_{i+1} \cdots L_k$, and $L_{<i} \stackrel{\text{def}}{=} L_1 \cdots L_{i-1}$. Random variable $L_{\leq i}$ is defined analogously. Let $C \stackrel{\text{def}}{=} \{i_1, \ldots, i_r\}$. Define $R_i \stackrel{\text{def}}{=} D_{-i} U_{-i} X_{C \cup [i-1]} Y_{C \cup [i-1]}$ for $i \in [k]$. We denote an element from the range of R_i by r_i. [1]

To prove the claim, we show that there exists a coordinate $j \notin C$ such that,

1. $(X_j Y_j)$ can be embedded well in $(X_j^1 R_j^1, Y_j^1 R_j^1)$ (with appropriate parameters as required in Lemma 5.1.13.)

2. Random variables $R_j^1, X_j^1, Y_j^1, M_1^1, \ldots, M_t^1$ satisfy the conditions of Lemma 6.2.8 with appropriate parameters.

Applying Markov inequality to Claim 6.2.12, we can get a coordinate $j \notin C$ such that

$$\mathrm{S}\left(X_j^1 Y_j^1 \middle\| X_j Y_j\right) \leq 12\delta, \tag{6.20}$$

$$\mathop{\mathbb{E}}_{(r_j, x_j) \leftarrow R_j^1, X_j^1} \left[\mathrm{S}\left(\left(Y_j^1\right)_{r_j, x_j} \middle\| \left(Y_j\right)_{x_j}\right)\right] \leq 12\delta, \tag{6.21}$$

$$\mathop{\mathbb{E}}_{(r_j, y_j) \leftarrow R_j^1, Y_j^1} \left[\mathrm{S}\left(\left(X_j^1\right)_{r_j, y_j} \middle\| \left(X_j\right)_{y_j}\right)\right] \leq 12\delta, \tag{6.22}$$

$$\sum_{s \text{ odd}} \mathrm{I}\left(X_j^1 : M_s^1 \middle| R_j^1 Y_j^1 M_{<s}^1\right) + \sum_{s \text{ even}} \mathrm{I}\left(Y_j^1 : M_s^1 \middle| R_j^1 X_j^1 M_{<s}^1\right) \leq 12\delta_1 c, \tag{6.23}$$

[1] We justify here the composition of R_i. Random variables $D_{-i}U_{-i}$ are useful since conditioning on them makes the distribution of inputs product across Alice and Bob (for fixed values of X_iY_i) and is helpful in our arguments later. Random variables X_CY_C are helpful since conditioning on them ensures that the inputs become product even conditioned on success on C. Random variables $X_{[i-1]}Y_{[i-1]}$ are helpful since the following chain rule is used to draw a new coordinate outside C with low information content:

$$\mathrm{I}(XY : M) = \sum_i \mathrm{I}\left(X_iY_i : M \middle| X_{[i-1]}Y_{[i-1]}\right).$$

$$\sum_{s \text{ odd}} \mathrm{I}\left(Y_j^1 : M_s^1 \middle| R_j^1 X_j^1 M_{<s}^1\right) + \sum_{s \text{ even}} \mathrm{I}\left(X_j^1 : M_s^1 \middle| R_j^1 Y_j^1 M_{<s}^1\right) \leq 12\delta t. \tag{6.24}$$

Set $\varepsilon' \stackrel{\text{def}}{=} \frac{\varepsilon}{125t}$, and

$$\varepsilon_s \stackrel{\text{def}}{=} \begin{cases} \mathrm{I}\left(Y_j^1 : M_s^1 \middle| R_j^1 X_j^1 M_{<s}^1\right) & s \in [t] \text{ odd}, \\ \mathrm{I}\left(X_j^1 : M_s^1 \middle| R_j^1 Y_j^1 M_{<s}^1\right) & s \in [t] \text{ even}. \end{cases};$$

$$c_s \stackrel{\text{def}}{=} \begin{cases} \mathrm{I}\left(Y_j^1 : M_s^1 \middle| R_j^1 X_j^1 M_{<s}^1\right) & s \in [t] \text{ even}, \\ \mathrm{I}\left(X_j^1 : M_s^1 \middle| R_j^1 Y_j^1 M_{<s}^1\right) & s \in [t] \text{ odd}. \end{cases}$$

By (6.24), $\sum_{s=1}^{t} \sqrt{\varepsilon_s} \leq \sqrt{12\delta t}$. From Equations (6.20)(6.21)(6.22) and Lemma 5.1.13 we can infer that $(X_j Y_j)$ is $(1 - 10\sqrt{3\delta})$-embeddable in $(X_j^1 R_j^1; Y_j^1 R_j^1)$. This, combined with Equations (6.23)(6.24) and Lemma 6.2.9 (take $\varepsilon', \varepsilon_s, c_s$ in the lemma to be as defined above and take $XYX'Y'R'M_1' \cdots M_t'$ in the lemma to be $X_j Y_j X_j^1 Y_j^1 R_j^1 M_1^1 \cdots M_t^1$) imply the following (for appropriate constant κ). There exists a public-coin t-message protocol $\mathcal{Q}^1$ between Alice, with input X_j, and Bob, with input Y_j, with Alice sending the first message and total communication, $\frac{12\delta_1 c + 5t}{\varepsilon'} + \mathcal{O}(t \log \frac{1}{\varepsilon'}) < \mathrm{D}_{\varepsilon}^{(t),\mu}(f)$, such that at the end Alice possesses $R_A M_1 \cdots M_t$ and Bob possesses $R_B M_1 \cdots M_t$, satisfying

$$\left\| X_j Y_j R_A R_B M_1 \cdots M_t - X_j^1 Y_j^1 R_j^1 R_j^1 M_1^1 \cdots M_t^1 \right\|_1 \leq 10\sqrt{3\delta} + 3\sqrt{12\delta t} + 6\varepsilon' t < \varepsilon/2.$$

Assume for contradiction that $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] > 1 - \varepsilon/2$. Consider a protocol $\mathcal{Q}^2$ (with no communication) for f between Alice, with input $X_j^1 R_j^1 M_1^1 \cdots M_t^1$, and Bob, with input $Y_j^1 R_j^1 M_1^1 \cdots M_t^1$, as follows. Bob generates the rest of the random variables present in Y^1 (not present in his input) himself since, conditioned on his input, those other random variables are independent of Alice's input (here we use Fact 5.2.3). Bob then generates the output for the j-th coordinate in $\mathcal{Q}$, and makes it the output of $\mathcal{Q}^2$. This ensures that the success probability of Bob in $\mathcal{Q}^2$ is $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] > 1 - \varepsilon/2$. Now consider protocol $\mathcal{Q}^3$ for f, with Alice's input X_j and Bob's input Y_j, which is a composition of $\mathcal{Q}^1$ followed by $\mathcal{Q}^2$. This ensures, using Fact 5.1.8, that success probability of Bob (averaged over public coins and the inputs $X_j Y_j$) in $\mathcal{Q}^3$ is larger than $1 - \varepsilon$. Finally by fixing the public coins of $\mathcal{Q}^3$, we get a deterministic protocol $\mathcal{Q}^4$ for f with Alice's input X_j and Bob's input Y_j such that the communication of $\mathcal{Q}^4$ is less than $\mathrm{D}_{\varepsilon}^{(t),\mu}(f)$ and Bob's success probability (averaged over the inputs $X_j Y_j$) in $\mathcal{Q}^4$ is larger than $1 - \varepsilon$. This is a contradiction to the definition of $\mathrm{D}_{\varepsilon}^{(t),\mu}(f)$ (recall that $X_j Y_j$ are distributed according

to μ). Hence it must be that $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] \leq 1 - \varepsilon/2$. The claim now follows by setting $i_{r+1} = j$. □

Claim 6.2.12. It holds that

1. $\sum_{i \notin C} \mathrm{S}(X_i^1 Y_i^1 \| X_i Y_i) < \delta k$.
2. $\frac{1}{2} \sum_{i \notin C} \mathbb{E}_{(r_i, x_i) \leftarrow R_i^1, X_i^1} \left[\mathrm{S}\left((Y_i^1)_{r_i, x_i} \middle\| (Y_i)_{x_i} \right) \right] + \frac{1}{2} \sum_{i \notin C} \mathbb{E}_{(r_i, y_i) \leftarrow R_i^1, Y_i^1} \left[\mathrm{S}\left((X_i^1)_{r_i, y_i} \middle\| (X_i)_{y_i} \right) \right] < \delta k$
3. $\frac{1}{2} \sum_{i \notin C} \left(\sum_{s \text{ odd}} \mathrm{I}(X_i^1 : M_s^1 | R_i^1 Y_i^1 M_{<s}^1) + \sum_{s \text{even}} \mathrm{I}(Y_i^1 : M_s^1 | R_i^1 X_i^1 M_{<s}^1) \right) < \delta k$.
4. $\sum_{i \notin C} \left(\sum_{s \text{ odd}} \mathrm{I}(Y_i^1 : M_s^1 | R_i^1 X_i^1 M_{<s}^1) + \sum_{s \text{ even}} \mathrm{I}(X_i^1 : M_s^1 | R_i^1 Y_i^1 M_{<s}^1) \right) \leq 2\delta k t$.

Proof. 1.

$$\delta k > \mathrm{S}_\infty\left(X^1 Y^1 \middle\| XY\right) \geq \mathrm{S}\left(X^1 Y^1 \middle\| XY\right) \geq \sum_{i \notin C} \mathrm{S}\left(X_i^1 Y_i^1 \middle\| X_i Y_i\right), \tag{6.25}$$

where first inequality follows from the assumption that $\Pr\left[T^{(r)} = 1\right] > 2^{-\delta k}$, and the last inequality follows from Fact 5.1.3. The following calculations are helpful for achieving conditions (5.2) and (5.3) of Lemma 5.1.13.

2.

$$\begin{aligned}
&\delta k > \mathrm{S}_\infty\left(X^1 Y^1 D^1 U^1 \middle\| XY DU\right) \\
&\geq \mathrm{S}\left(X^1 Y^1 D^1 U^1 \middle\| XY DU\right) \\
&\geq \mathop{\mathbb{E}}_{\substack{(d,u,x_C,y_C) \\ \leftarrow D^1, U^1, X_C^1, Y_C^1}} \left[\mathrm{S}\left(\left(X^1 Y^1\right)_{d,u,x_C,y_C} \middle\| (XY)_{d,u,x_C,y_C} \right) \right] && (6.26) \\
&= \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(d,u,x_{C\cup[i-1]},y_{C\cup[i-1]}) \\ \leftarrow D^1, U^1, X_{C\cup[i-1]}^1, Y_{C\cup[i-1]}^1}} \left[\mathrm{S}\left(\left(X_i^1 Y_i^1\right)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]}}} \middle\| (X_i Y_i)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]}}} \right) \right] && (6.27) \\
&= \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(d_i,u_i,r_i) \\ \leftarrow D_i^1, U_i^1, R_i^1}} \left[\mathrm{S}\left((X_i^1 Y_i^1)_{d_i,u_i,r_i} \middle\| (X_i Y_i)_{d_i,u_i,r_i} \right) \right] && (6.28) \\
&= \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{(r_i,x_i) \leftarrow R_i^1, X_i^1} \left[\mathrm{S}\left(\left(Y_i^1\right)_{r_i,x_i} \middle\| (Y_i)_{x_i} \right) \right] + \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{(r_i,y_i) \leftarrow R_i^1, Y_i^1} \left[\mathrm{S}\left(\left(X_i^1\right)_{r_i,y_i} \middle\| (X_i)_{y_i} \right) \right]. && (6.29)
\end{aligned}$$

Above, Eq. (6.26) and Eq. (6.27) follow from Fact 5.1.3; Eq. (6.28) is from the definition of R_i. Eq. (6.29) follows since D_i^1 is independent of R_i^1 and with probability half D_i^1 is 0, in which case $U_i^1 = X_i^1$ and with probability half D_i^1 is 1 in which case $U_i^1 = Y_i^1$.

3.

$$\begin{aligned}
\delta_1 ck &\geq \left|M^1\right| \geq \mathrm{I}\left(X^1Y^1 : M^1 \middle| D^1U^1X_C^1Y_C^1\right) \\
&= \sum_{i\notin C} \mathrm{I}\left(X_i^1Y_i^1 : M^1 \middle| D^1U^1X_{C\cup[i-1]}^1Y_{C\cup[i-1]}^1\right) \\
&= \sum_{i\notin C}\sum_{s=1}^{t} \mathrm{I}\left(X_i^1Y_i^1 : M_s^1 \middle| D^1U^1X_{C\cup[i-1]}^1Y_{C\cup[i-1]}^1M_{<s}^1\right) \\
&= \sum_{i\notin C}\sum_{s=1}^{t} \mathrm{I}\left(X_i^1Y_i^1 : M_s^1 \middle| D_i^1U_i^1R_i^1M_{<s}^1\right) \\
&= \sum_{i\notin C}\left(\left(\sum_{s \text{ odd}} + \sum_{s \text{ even}}\right) \mathrm{I}\left(X_i^1Y_i^1 : M_s^1 \middle| D_i^1U_i^1R_i^1M_{<s}^1\right)\right) \\
&\geq \frac{1}{2}\sum_{i\notin C}\Big(\sum_{s \text{ odd}} \mathrm{I}\left(X_i^1 : M_s^1 \middle| R_i^1Y_i^1M_{<s}^1\right) + \sum_{s\text{even}} \mathrm{I}\left(Y_i^1 : M_s^1 \middle| R_i^1X_i^1M_{<s}^1\right)\Big). \qquad (6.30)
\end{aligned}$$

Above we have used the chain rule for mutual information several times. Last inequality follows since D_i^1 is independent of $(X_i^1Y_i^1R_i^1M^1)$ and with probability half D_i^1 is 0, in which case $U_i^1 = X_i^1$ and with probability half D_i^1 is 1 in which case $U_i^1 = Y_i^1$.

4.

$$\begin{aligned}
\delta k &\geq \mathrm{S}_\infty\left(D^1U^1X^1Y^1M_{\leq s}^1 \middle\| DUXYM_{\leq s}\right) \\
&\geq \mathrm{S}\left(D^1U^1X^1Y^1M_{\leq s}^1 \middle\| DUXYM_{\leq s}\right) \\
&\geq \mathbb{E}_{(d,u,x_C,y_C,m_{\leq s})\leftarrow D^1,U^1,X_C^1,Y_C^1,M_{\leq s}^1}\left[\mathrm{S}\left((X^1Y^1)_{d,u,x_C,y_C,m_{\leq s}} \middle\| (XY)_{d,u,x_C,y_C,m_{\leq s}}\right)\right] \\
&= \sum_{i\notin C} \mathop{\mathbb{E}}_{\substack{(d,u,x_{C\cup[i-1]},y_{C\cup[i-1]},m_{\leq s}) \\ \leftarrow D^1,U^1,X_{C\cup[i-1]}^1,Y_{C\cup[i-1]}^1,M_{\leq s}^1}} \left[\mathrm{S}\left((X_i^1Y_i^1)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]},m_{\leq s}}} \middle\| (X_iY_i)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]},m_{\leq s}}}\right)\right] \\
&= \sum_{i\notin C} \mathop{\mathbb{E}}_{\substack{(d_i,u_i,r_i,m_{\leq s}) \\ \leftarrow D_i^1,U_i^1,R_i^1,M_{\leq s}^1}} \left[\mathrm{S}\left((X_i^1Y_i^1)_{d_i,u_i,r_i,m_{\leq s}} \middle\| (X_iY_i)_{d_i,u_i,r_i,m_{\leq s}}\right)\right] \qquad (6.31)
\end{aligned}$$

$$\geq \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{\substack{(x_i,r_i,m_{\leq s})\\ \leftarrow X_i^1,R_i^1,M_{\leq s}^1}} \left[\mathrm{S}\left((Y_i^1)_{x_i,r_i,m_{\leq s}}\middle\|(Y_i)_{x_i,r_i,m_{\leq s}}\right)\right]$$

$$= \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{\substack{(x_i,r_i,m_{\leq s})\\ \leftarrow X_i^1,R_i^1,M_{\leq s}^1}} \left[\mathrm{S}\left((Y_i^1)_{x_i,r_i,m_{\leq s}}\middle\|(Y_i)_{x_i,r_i,m_{<s}}\right)\right] \tag{6.32}$$

$$= \frac{1}{2}\sum_{i\notin C} \mathbb{E}_{(x_i,r_i,m_{<s})\leftarrow X_i^1,R_i^1,M_{<s}^1}\left[\mathrm{S}\left((Y_i^1M_s^1)_{x_i,r_i,m_{<s}}\middle\|(Y_i)_{x_i,r_i,m_{<s}}\otimes(M_s^1)_{x_i,r_i,m_{<s}}\right)\right]$$

$$\geq \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{\substack{(x_i,r_i,m_{<s})\\ \leftarrow X_i^1,R_i^1,M_{<s}^1}} \left[\mathrm{I}\left((Y_i^1)_{x_i,r_i,m_{<s}} : (M_s^1)_{x_i,r_i,m_{<s}}\right)\right] \tag{6.33}$$

$$= \frac{1}{2}\sum_{i\notin C} \mathrm{I}\left(Y_i^1 : M_s^1\middle|X_i^1R_i^1M_{<s}^1\right). \tag{6.34}$$

Above we have used Fact 5.1.3 several times. Eq. (6.31) follows from the definition of R_i; Eq. (6.32) follows from the fact that $Y \leftrightarrow X_iR_iM_{<s} \leftrightarrow M_s$ for any i, whenever s is odd; Eq. (6.33) follows from Fact 5.1.4. From a symmetric argument, we can show that when $s \in [t]$ is even, $\frac{1}{2}\sum_{i\notin C}\mathrm{I}\left(X_i^1 : M_s^1|Y_i^1R_i^1M_{<s}^1\right) \leq \delta k$. This and Eq. (6.34) together imply

$$\sum_{i\notin C}\Big(\sum_{s \text{ odd}}\mathrm{I}\left(Y_i^1 : M_s^1\middle|R_i^1X_i^1M_{<s}^1\right) + \sum_{s \text{ even}}\mathrm{I}\left(X_i^1 : M_s^1\middle|R_i^1Y_i^1M_{<s}^1\right)\Big) \leq 2\delta kt. \tag{6.35}$$

□

Chapter 7

A strong direct product theorem in terms of the smooth rectangle bound

7.1 Introduction

In this chapter, we investigate direct product problems for the model two-way public-coin communication (Please refer to Section 5.2). We assume that the last $\lceil \log |\mathcal{Z}| \rceil$ bits of the transcript of a protocol is the output. For most of interesting functions (relations), the lengths of the outputs are much smaller than the communication cost in this model.

7.1.1 Result

In this chapter, we show a strong direct product theorem in terms of the smooth rectangle bound (please refer to Definition 5.2.4). Using the notations introduced in Section 5.2, we show that

Theorem 7.1.1. *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Let $z \in \mathcal{Z}$ and $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow\mu}[f(x,y) = \{z\}]$. Let $\varepsilon', \delta > 0$. There exists a small enough $\varepsilon > 0$ such that the following holds. For all integers $t \geq 1$,*

$$\mathrm{R}^{\mathrm{pub}}_{1-(1-\varepsilon)^{\lfloor \varepsilon^2 t/32 \rfloor}}(f^t) \geq \frac{\varepsilon^2}{32} \cdot t \cdot \left(11\varepsilon \cdot \widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}(f) - 2\right).$$

Our result implies a strong direct product theorem for all relations for which an (asymptotically) optimal lower bound can be provided using the smooth rectangle bound. Combining Theorem 7.1.1 with Lemma 5.2.6, we get the following result.

Theorem 7.1.2. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a (partial) function. For every $\epsilon \in (0,1)$, there exists small enough $\eta \in (0, 1/3)$ such that the following holds. For all integers $t \geq 1$,*

$$\mathrm{R}^{\mathrm{pub}}_{1-(1-\eta)\lfloor \eta^2 t/32 \rfloor}(f^{(t)}) \geq \frac{\eta^2}{32} \cdot t \cdot \left(11\eta \cdot \log \mathrm{srec}_{\epsilon}(f) - 3 \log \frac{1}{\epsilon} - 2 \right).$$

As a consequence, our results reprove some of the known strong direct product results, for example for Inner Product [49] and Set-Disjointness [25; 46]. Recently smooth rectangle bound has been used to provide new tight lower bounds for several functions, for example for the Gap-Hamming Distance [17; 68] partial function and the Greater-Than function [70]. These results, along with our result, imply strong direct product for these functions. Smooth rectangle bound has also been used to provide near optimal lower bounds for several important functions and relations used to show exponential separations between classical and quantum communication complexity for example Vector in Subspace by Raz [62] and Klartag and Regev [65], and Hidden Matching by Gavinsky [21]. These results combined with our result imply near optimal strong direct product results for these functions and relations.

In a recent work, Harsha and Jain [22] have shown that the smooth-rectangle bound provides an optimal lower bound of $\Omega(n)$ for the Tribes function. For this function all other weaker lower bound methods mentioned before like the rectangle bound, the sub-distribution bound, the smooth discrepancy bound, the conditional min-entropy bound etc. fail to provide an optimal lower bound since they are all $O(\sqrt{n})$. Earlier Jayram, Kumar and Sivakumar [7] had shown a lower bound of $\Omega(n)$ using information complexity. The result of [22] along with Theorem 7.1.2 implies a strong direct product result for the Tribes function. This adds to the growing list of functions for which a strong direct product result can be shown via Theorem 7.1.2.

In [43], Kerenidis et. al. introduced the *relaxed partition bound* (a weaker version of the partition bound [28]) and showed it to be stronger than the smooth rectangle bound. For boolean functions, or more generally for the functions with constant-size output, the smooth rectangle bound and the relaxed partition bound are in-fact equivalent, which can be checked by by comparing the corresponding linear-programs. Thus our result also implies a strong direct product theorem in terms of the relaxed partition bound for boolean functions (and more generally when the size of output set is a constant).

7.1.2 Our techniques

The broad argument of the proof is similar to the one in Chapter 6. We show our result in the distributional error setting and translate it to the worst case error setting using Yao's principle Fact 5.2.1. Let f be a relation, μ be a distribution on $\mathcal{X} \times \mathcal{Y}$, and c be the smooth rectangle bound of f under the distribution μ with output $z \in \mathcal{Z}$. Consider a protocol Π which computes f^k with inputs drawn from distribution μ^k and communication $o(c \cdot k)$ bits. Let C be a subset of the coordinates $\{1, 2, \ldots, k\}$. If the probability that Π computes all the instances in C correctly is as small as desired, then we are done. Otherwise, we exhibit a new coordinate $j \notin C$, such that the probability, conditioned on success in C, of the protocol Π answering correctly in the j-th coordinate is bounded away from 1. Same as proving Theorem 6.1.1, we introduce a new random variable R_j, such that conditioned on it and X_jY_j (input in the jth coordinate), Alice and Bob's inputs in the other coordinates become independent when the distribution of the input μ is non-product. Let the random variables $X_j^1 Y_j^1 R_j^1 M^1$ represent the inputs in the jth coordinate, the new variable R_j and the message transcript of Π, conditioned on the success on C. The first useful property that we observe is that the joint distribution of $X_j^1 Y_j^1 R_j^1 M^1$ can be written as,

$$\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = xyr_j m\right] = \frac{1}{q}\mu(x,y)u_x(r_j,m)u_y(r_j,m),$$

where u_x, u_y are functions and q is a positive real number. The marginal distribution of $X_j^1 Y_j^1$ is no longer μ though. However using the same arguments as in [25] and in the previous chapter, one can show that the distribution of $X_j^1 Y_j^1$ is close, in ℓ_1 distance, to μ and $\mathrm{I}\left(X_j^1 : R_j^1 M^1 \middle| Y_j^1\right) + \mathrm{I}\left(Y_j^1 : R_j^1 M^1 \middle| X_j^1\right) \leq o(c)$, where $\mathrm{I}(:)$ represents the mutual information (please refer to Section 5.1 for precise definitions) .

Now, assume for contradiction that the success in the jth coordinate in Π is large, like 0.99, conditioned on success in C. Using the conditions obtained in the previous paragraph, we argue that there exists a zero-communication public-coin protocol Π', between Alice and Bob, with inputs drawn from μ. In Π' Alice and Bob are allowed to abort the protocol or output an element in $\mathcal{Z}$. We show that the probability of non-abort for this protocol is large, like 2^{-c}, and conditioned on non-abort, the probability that Alice and Bob output a correct answer for their inputs is also large, like 0.99. This allows us to exhibit (by fixing the public coins of Π' appropriately), a large rectangle (with weight under μ like 2^{-c}) such that z is a correct answer for a large fraction (like 0.99) of the inputs inside the rectangle. This shows that the rectangle bound of f, under μ

with output z, is smaller than c. With careful analysis we are also able to show that the smooth rectangle bound of f under μ, with output z, is smaller than c, reaching a contradiction to the definition of c.

The sampling protocol that we use to obtain the public-coin zero-communication protocol, is the same as that in Kerenidis et al. [43], which in turn is a modification of a protocol due to Braverman [12][1] (a variation of which also appears in [16]). However our analysis of the protocol's correctness deviates significantly in parts from the earlier works [12; 16; 43] due to the fact that for us the marginal distribution of X^1Y^1 need not be the same as that of μ, in fact for some inputs (x, y), the probability under the two distributions can be significantly different.

There is another important original contribution of our work, not present in the previous works [12; 16; 43]. We observe a crucial property of the protocol Π' which turns out to be very important in our arguments. The property is that the bad inputs (x, y) for which the distribution of Π''s sample for $R_j^1 M^1$, conditioned on non-abort, deviates a lot from the desired $R_j^1 M^1 |\ (X^1Y^1 = xy)$, their probability is nicely reduced (as compared to $\Pr[X^1Y^1 = xy]$) in the final distribution of Π', conditioned on non-abort. This helps us to argue that the distribution of inputs and outputs in Π', conditioned on non-abort, is close in ℓ_1 distance to $X_j^1 Y_j^1 R_j^1 M^1$, implying good success in Π', conditioned on non-abort.

7.2 Proof

The following lemma builds a connection between the zero-communication protocols and the smooth rectangle bound.

Lemma 7.2.1. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, $X'Y' \in \mathcal{X} \times \mathcal{Y}$ be a distribution and $z \in \mathcal{Z}$. Let $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow X'Y'}[f(x, y) = \{z\}]$. Let $c \geq 1$. Let $\varepsilon, \varepsilon', \delta > 0$ be such that $(\delta+2\varepsilon)/(\beta-3\varepsilon) < (1 + \varepsilon')\delta/\beta$. Let Π be a zero-communication public-coin protocol with input $X'Y'$, public coin R, Alice's output $A \in \mathcal{Z}\cup\{\perp\}$, and Bob's output $B \in \mathcal{Z}\cup\{\perp\}$. Let $X^1Y^1A^1B^1R^1 \stackrel{\text{def}}{=} (X'Y'ABR|\ A = B \neq \perp)$. Let*

1. $\Pr[A = B \neq \perp] \geq 2^{-c}$; 2. $\|X^1Y^1 - X'Y'\| \leq \varepsilon$.

3. $\Pr[(X^1, Y^1, A^1) \in f] \geq 1 - \varepsilon$.

Then $\widetilde{\text{srec}}^{z,X'Y'}_{(1+\varepsilon')\delta/\beta,\delta}(f) < \frac{c}{\varepsilon}$.

[1] A protocol, achieving similar task, however working only for product distributions on inputs was first shown by Jain, Radhakrishnan and Sen [35].

Proof. Let $g \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, satisfy $\Pr_{(x,y)\leftarrow X'Y'}[f(x,y) \neq g(x,y)] \leq \delta$. It suffices to show that $\widetilde{\mathrm{rec}}^{z,X'Y'}_{(1+\varepsilon')\delta/\beta}(g) \leq \frac{c}{\varepsilon}$. Since $\Pr[A = B \neq \perp] \geq 2^{-c}$,

$$\begin{aligned} c &\geq \mathrm{S}_\infty\left(X^1Y^1R^1A^1B^1 \middle\| X'Y'RAB\right) \\ &\geq \mathrm{S}\left(X^1Y^1R^1A^1B^1 \middle\| X'Y'RAB\right) \\ &\geq \mathop{\mathbb{E}}_{r\leftarrow R^1, a\leftarrow A^1}\left[\mathrm{S}\left((X^1Y^1)_{r,a} \middle\| X'Y'\right)\right] \quad \text{(from Fact 5.1.3).} \end{aligned} \tag{7.1}$$

Since $\|X^1Y^1 - X'Y'\| \leq \varepsilon$,

$$\Pr_{xyr\leftarrow X^1Y^1R^1}[f(x,y) = \{z\}] \geq \Pr_{xy\leftarrow X'Y'}[f(x,y) = \{z\}] - \varepsilon \geq \beta - \varepsilon. \tag{7.2}$$

Since $\Pr[(X^1, Y^1, A^1) \in f] \geq 1 - \varepsilon$, we have $\Pr[A^1 = B^1 = z] \geq \beta - 2\varepsilon$. Since

$$\Pr_{(x,y)\leftarrow X'Y'}[f(x,y) \neq g(x,y)] \leq \delta,$$

by item 2 of this lemma, we have

$$\Pr_{xyra\leftarrow X^1Y^1R^1A^1}[(x,y,a) \in g] \geq \Pr_{xyra\leftarrow X^1Y^1R^1A^1}[(x,y,a) \in f] - \delta - \varepsilon \geq 1 - 2\varepsilon - \delta. \tag{7.3}$$

By standard application of Markov's inequality on equations (7.1), (7.2), (7.3), we get an r_0, such that

$$\begin{aligned} \mathrm{S}\left((X^1Y^1)_{r_0,z} \middle\| X'Y'\right) &\leq \frac{c}{\varepsilon}, \\ \Pr_{xy\leftarrow(X^1Y^1)_{r_0,z}}[z \notin g(x,y)] &\leq (\delta + 2\varepsilon)/(\beta - 3\varepsilon) \leq (1+\varepsilon')\delta/\beta. \end{aligned}$$

Here, $(X^1Y^1)_{r_0,z} = (X^1Y^1|(R^1 = r_0, A^1 = z)$. Note that the distribution of $(X^1Y^1)_{r_0,z}$ is the distribution of $X'Y'$ restricted to some rectangle and then rescaled to make a distribution. Hence

$$\mathrm{S}\left((X^1Y^1)_{r_0,z} \middle\| X'Y'\right) = \mathrm{S}_\infty\left((X^1Y^1)_{r_0,z} \middle\| X'Y'\right).$$

Thus $\widetilde{\mathrm{rec}}^{z,X'Y'}_{(1+\varepsilon')\delta/\beta}(g) < \frac{c}{\varepsilon}$. □

The following is our main lemma. A key tool that we use here is a sampling protocol that appears in [43] (protocol Π' as shown in Figure 7.1), which is a variant of a sampling protocol that appears in [16], which in turn is a variant of a sampling protocol that

appears in [12]. Naturally similar arguments and calculations, as in this lemma, are made in previous works [12; 16; 43], however with a key difference. In their setting $\sum_m u_x(m)u_y(m) = 1$ for all (x, y). However in our setting this number could be much smaller than one for different (x, y). Hence our arguments and calculations deviate from previous works at several places significantly. Another important original contribution of our work is Claim 7.2.6 which is used in the proof of the main lemma. We highlight its importance later just before its proof.

Lemma 7.2.2. ***(Main Lemma)*** *Let $c \geq 1$. Let p be a distribution over $\mathcal{X} \times \mathcal{Y}$ and $z \in \mathcal{Z}$. Let $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow p}[f(x, y) = \{z\}]$. Let $0 < \varepsilon < 1/3$ and $\delta, \varepsilon' > 0$ be such that $\frac{\delta+22\varepsilon}{\beta-33\varepsilon} < (1+\varepsilon')\frac{\delta}{\beta}$. Let XYM be random variables jointly distributed over the set $\mathcal{X}\times\mathcal{Y}\times\mathcal{M}$ such that the last $\lceil \log |\mathcal{Z}| \rceil$ bits of M represents an element in $\mathcal{Z}$. Let $u_x : \mathcal{M} \to [0, 1]$, $u_y : \mathcal{M} \to [0, 1]$ be functions for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. If it holds that,*

1. *For all $(x, y, m) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{M}$,*

$$\Pr[XYM = xym] = \frac{1}{q}p(x, y)u_x(m)u_y(m),$$

   where $q \stackrel{\text{def}}{=} \sum_{xym} p(x, y)u_x(m)u_y(m)$;

2. *$\mathrm{S}(XY\|p) \leq \varepsilon^2/4$;*

3. *$\mathrm{I}(X : M|Y) + \mathrm{I}(Y : M|X) \leq c$;*

4. *$\mathrm{err}_f(XYM) \leq \varepsilon$, where $\mathrm{err}_f(XYM) \stackrel{\text{def}}{=} \Pr_{xym\leftarrow XYM}[(x, y, \tilde{m}) \notin f]$, and $\tilde{m}$ represents the last $\lceil \log |\mathcal{Z}| \rceil$ bits of m;*

then $\widetilde{\mathrm{srec}}^{z,p}_{(1+\varepsilon')\delta/\beta,\delta}(f) < \frac{2c}{11\varepsilon^3}$. □

Note by direct calculations,

$$\Pr[XY = xy] = \frac{1}{q}p(x, y)\alpha_{xy}, \quad \text{where } \alpha_{xy} \stackrel{\text{def}}{=} \sum_m u_x(m)u_y(m); \tag{7.4}$$

$$\Pr[X = x] = \frac{1}{q}p(x)\alpha_x, \quad \text{where } \alpha_x \stackrel{\text{def}}{=} \sum_y p(y|x)\alpha_{xy}; \tag{7.5}$$

$$\Pr[Y = y] = \frac{1}{q}p(y)\alpha_y, \quad \text{where } \alpha_y \stackrel{\text{def}}{=} \sum_x p(x|y)\alpha_{xy}; \tag{7.6}$$

$$\Pr[X_y = x] = \frac{p(x|y)\alpha_{xy}}{\alpha_x}, \quad \Pr[Y_x = y] = \frac{p(y|x)\alpha_{xy}}{\alpha_y}; \tag{7.7}$$

$$\Pr[M_{xy} = m] = u_x(m)u_y(m)/\alpha_{xy}; \tag{7.8}$$

$$\Pr[M_x = m] = \frac{u_x(m)v_x(m)}{\alpha_x}, \quad \text{where } v_x(m) \stackrel{\text{def}}{=} \sum_y p(y|x)u_y(m); \tag{7.9}$$

$$\Pr[M_y = m] = \frac{u_y(m)v_y(m)}{\alpha_y}, \quad \text{where } v_y(m) \stackrel{\text{def}}{=} \sum_x p(x|y)u_x(m). \tag{7.10}$$

Like in Chapter 6, we apply Markov inequality to Item 2 and Item 3 of Lemma 7.2.2 to show most of (x, y) have nice properties. Let us define the sets of good (x, y).

$$G_1 \stackrel{\text{def}}{=} \{(x,y) : \left|1 - \frac{\alpha_{xy}}{q}\right| \leq \frac{1}{2} \text{ and } \left|1 - \frac{\alpha_x}{q}\right| \leq \frac{1}{2} \text{ and } \left|1 - \frac{\alpha_{xy}}{q}\right| \leq \frac{1}{2}\}; \tag{7.11}$$

$$G_2 \stackrel{\text{def}}{=} \{(x,y) : \mathrm{S}(M_{xy}\|M_x) + \mathrm{S}(M_{xy}\|M_y) \leq c/\varepsilon\}; \tag{7.12}$$

$$G \stackrel{\text{def}}{=} \{(x,y) : \Pr_{m \leftarrow M_{xy}}\left[\frac{u_y(m)}{v_x(m)} \leq 2^{\Delta} \text{ and } \frac{u_x(m)}{v_y(m)} \leq 2^{\Delta}\right] \geq 1 - 2\varepsilon\}. \tag{7.13}$$

We begin by showing that $G_1 \cap G_2$ is a large set and also $G_1 \cap G_2 \subseteq G$.

Claim 7.2.3. 1. $\Pr_{(x,y)\leftarrow p}[(x,y) \in G_1] > 1 - 6\varepsilon$,

2. $\Pr_{(x,y)\leftarrow p}[(x,y) \in G_2] \geq 1 - 3\varepsilon/2$,

3. $\Pr_{(x,y)\leftarrow p}[(x,y) \in G_1 \cap G_2] \geq 1 - 15\varepsilon/2$,

4. $G_1 \cap G_2 \subseteq G$.

Proof. Note item 1. and item 2. imply item 3. Now we show 1. Note that (using item 2. of Lemma 7.2.2 and Fact 5.1.5) $\|XY - p\|_1 \leq \varepsilon/2$. From Lemma 5.1.10 and (7.4), we have

$$\Pr_{(x,y)\leftarrow p}\left[\left|1 - \frac{\alpha_{xy}}{q}\right| \leq 1/2\right] \geq 1 - 2\varepsilon.$$

By the monotonicity of ℓ_1-norm, we have $\|X - p_{\mathcal{X}}\|_1 \leq \frac{\varepsilon}{2}$ and $\|Y - p_{\mathcal{Y}}\|_1 \leq \frac{\varepsilon}{2}$. Similarly, from (7.5) and (7.6) we have

$$\Pr_{(x,y)\leftarrow p}\left[\left|1 - \frac{\alpha_x}{q}\right| \leq 1/2\right] \geq 1 - 2\varepsilon, \quad \text{and} \quad \Pr_{(x,y)\leftarrow p}\left[\left|1 - \frac{\alpha_y}{q}\right| \leq 1/2\right] \geq 1 - 2\varepsilon.$$

By the union bound, item 1. follows.

Next we show 2. From item 3. of Lemma 7.2.2,

$$\mathop{\mathbb{E}}_{(x,y)\leftarrow XY}[\mathrm{S}(M_{xy}\|M_x) + \mathrm{S}(M_{xy}\|M_y)] = \mathrm{I}(X : M|Y) + \mathrm{I}(Y : M|X) \leq c.$$

Markov's inequality implies $\Pr_{(x,y)\leftarrow XY}[(x,y) \in G_2] \geq 1-\varepsilon$. Then item 2. follows from the fact that XY and p are $\varepsilon/2$-close.

Finally we show 4. For any $(x,y) \in G_1 \cap G_2$,

$$\begin{aligned}
&\mathrm{S}(M_{xy}\|M_x) \leq c/\varepsilon \\
&\Rightarrow \Pr_{m\leftarrow M_{xy}}\left[\frac{\Pr[M_{xy}=m]}{\Pr[M_x=m]} \leq 2^{\frac{c/\varepsilon+1}{\varepsilon}}\right] \geq 1-\varepsilon \quad \text{(from Fact 5.1.9)} \\
&\Rightarrow \Pr_{m\leftarrow M_{xy}}\left[\frac{u_y(m)\alpha_x}{v_x(m)\alpha_{xy}} \leq 2^{\frac{c/\varepsilon+1}{\varepsilon}}\right] \geq 1-\varepsilon \quad \text{(from (7.8) and (7.9))} \\
&\Rightarrow \Pr_{m\leftarrow M_{xy}}\left[\frac{u_y(m)}{v_x(m)} \leq 2^{\Delta}\right] \geq 1-\varepsilon. \quad ((x,y) \in G_1 \text{ and the choice of } \Delta)
\end{aligned}$$

Similarly, $\Pr_{m\leftarrow M_{xy}}\left[\frac{u_x(m)}{v_y(m)} \leq 2^{\Delta}\right] \geq 1-\varepsilon$. By the union bound,

$$\Pr_{m\leftarrow M_{xy}}\left[\frac{u_y(m)}{v_x(m)} \leq 2^{\Delta} \text{ and } \frac{u_x(m)}{v_y(m)} \leq 2^{\Delta}\right] \geq 1-2\varepsilon,$$

which implies $(x,y) \in G$. Hence $G_1 \cap G_2 \subseteq G$. □

Following few claims establish the desired properties of protocol Π' (Figure 7.1).

Definition 7.2.4. Define the following events.

- E occurs if the smallest $i \in \mathcal{A}$ satisfies $\mathbf{h}(\mathbf{m_i}) = \mathbf{r}$ and $i \in \mathcal{B}$. Note that E implies $\mathcal{A} \neq \emptyset$.
- B_c (subevent of E) occurs if E occurs and there exist $j \in \mathcal{B}$ such that $\mathbf{h}(\mathbf{m_j}) = \mathbf{r}$ and $\mathbf{m_i} \neq \mathbf{m_j}$, where i is the smallest element in $\mathcal{A}$.
- $H \stackrel{\text{def}}{=} E - B_c$.

Below we use conditioning on (x,y) as shorthand for "Alice's input is x and Bob's input is y".

Claim 7.2.5. For any $(x,y) \in G_1 \cap G_2$, we have

1. for all $i \in [T]$,

$$\frac{1}{2}\cdot\frac{q}{|\mathcal{M}|2^{\Delta}} \leq \Pr_{\mathbf{r}_{\Pi'}}[\text{Alice accepts } \mathbf{m_i}|\ (x,y)] \leq \frac{3}{2}\cdot\frac{q}{|\mathcal{M}|2^{\Delta}},$$

Alice's input is x. Bob's input is y. Common input is $c, \varepsilon, q, \mathcal{M}$.

1. Alice and Bob both set $\Delta \stackrel{\text{def}}{=} \frac{c/\varepsilon+1}{\varepsilon} + 2, T \stackrel{\text{def}}{=} \frac{2}{q}|\mathcal{M}|2^{\Delta} \ln \frac{1}{\varepsilon}$ and $k \stackrel{\text{def}}{=} \log(\frac{3}{\varepsilon}(\ln \frac{1}{\varepsilon}))$.
2. For $i = 1, \cdots, T$:
   (a) Alice and Bob, using public coins, jointly sample $\mathbf{m_i} \leftarrow \mathcal{M}, \boldsymbol{\alpha_i}, \boldsymbol{\beta_i} \leftarrow [0, 2^{\Delta}]$, uniformly.
   (b) Alice accepts $\mathbf{m_i}$ if $\boldsymbol{\alpha_i} \leq u_x(\mathbf{m_i})$, and $\boldsymbol{\beta_i} \leq 2^{\Delta} v_x(\mathbf{m_i})$.
   (c) Bob accepts $\mathbf{m_i}$ if $\boldsymbol{\alpha_i} \leq 2^{\Delta} v_y(\mathbf{m_i})$, and $\boldsymbol{\beta_i} \leq u_y(\mathbf{m_i})$.
3. Let $\mathcal{A} \stackrel{\text{def}}{=} \{i \in [T] : \text{Alice accepts } \mathbf{m_i}\}$ and $\mathcal{B} \stackrel{\text{def}}{=} \{i \in [T] : \text{Bob accepts } \mathbf{m_i}\}$.
4. Alice and Bob, using public coins, choose a uniformly random function $\mathbf{h} : \mathcal{M} \to \{0,1\}^k$ and a uniformly random string $\mathbf{r} \in \{0,1\}^k$.
   (a) Alice outputs $\perp$ if either $\mathcal{A}$ is empty or $\mathbf{h}(\mathbf{m_i}) \neq \mathbf{r}$ (where i is the smallest element in non-empty $\mathcal{A}$). Otherwise, she outputs the element in $\mathcal{Z}$, represented by the last $\lceil \log |\mathcal{Z}| \rceil$ bits of $\mathbf{m_i}$.
   (b) Bob finds the smallest $j \in \mathcal{B}$ such that $\mathbf{h}(\mathbf{m_j}) = \mathbf{r}$. If no such j exists, he outputs $\perp$. Otherwise, he outputs the element in $\mathcal{Z}$, represented by the last $\lceil \log |\mathcal{Z}| \rceil$ bits of $\mathbf{m_j}$.

Figure 7.1: Protocol Π'

and

$$\frac{1}{2} \cdot \frac{q}{|\mathcal{M}|2^{\Delta}} \leq \Pr_{\mathbf{r}_{\Pi'}}[\text{Bob accepts } \mathbf{m_i} | \ (x, y)] \leq \frac{3}{2} \cdot \frac{q}{|\mathcal{M}|2^{\Delta}},$$

where $\mathbf{r}_{\Pi'}$ is the internal randomness of protocol Π';

2. $\Pr_{\mathbf{r}_{\Pi'}}[B_c | \ (x, y), E] \leq \varepsilon$;
3. $\Pr_{\mathbf{r}_{\Pi'}}[H | \ (x, y)] \geq (1 - 4\varepsilon) \cdot 2^{-k-\Delta-2}$.

The proof requires long but direct calculation. Similar arguments and calculation are made in [43]. We defer the proof to the end of this chapter.

The following claim is an important original contribution of this work (not present in the previous works [12; 16; 43].) The claim helps us establish a crucial property of Π'. The property is that the bad inputs (x, y) for which the distribution of Π''s sample for

M, conditioned on non-abort, deviates a lot from the desired, their probability is nicely reduced in the final distribution of Π', conditioned on non-abort. This helps us to argue that the joint distribution of inputs and the transcript in Π', conditioned on non-abort, is still close in ℓ_1 distance to XYM.

Claim 7.2.6. Let AB and $A'B'$ be random variables over $\mathcal{A}_1 \times \mathcal{B}_1$ and $h : \mathcal{A}_1 \to [0, +\infty)$ be a function. Suppose for any $a \in \mathcal{A}_1$, there exist functions $f_a, g_a : \mathcal{B}_1 \to [0, +\infty)$, such that

1. $\sum_{a,b} h(a) f_a(b) = 1$, and $\Pr[AB = ab] = h(a) f_a(b)$;
2. $f_a(b) \geq g_a(b)$, for all $(a, b) \in \mathcal{A}_1 \times \mathcal{B}_1$;
3. $\Pr[A'B' = ab] = h(a) g_a(b)/C$, where $C = \sum_{a,b} h(a) g_a(b)$;
4. $\Pr_{a \leftarrow A}[\Pr_{b \leftarrow B_a}[f_a(b) = g_a(b)] \geq 1 - \delta_1] \geq 1 - \delta_2$, for $\delta_1 \in [0, 1), \delta_2 \in [0, 1)$.

Then $\|AB - A'B'\|_1 \leq \delta_1 + \delta_2$.

Proof. Set $G \stackrel{\text{def}}{=} \{(a, b) : f_a(b) = g_a(b)\}$. By condition 4, $\Pr_{(a,b) \leftarrow AB}[(a, b) \in G] \geq 1 - \delta_1 - \delta_2$. Then

$$C = \sum_{a,b} h(a) g_a(b) \geq \sum_{a,b:(a,b) \in G} h(a) f_a(b) = \Pr_{(a,b) \leftarrow AB}[(a, b) \in G] \geq 1 - \delta_1 - \delta_2. \tag{7.14}$$

We have

$$\begin{aligned}
& \|AB - A'B'\|_1 = \frac{1}{2} \sum_{a,b} |h(a) f_a(b) - \frac{1}{C} h(a) g_a(b)| \\
\leq\ & \frac{1}{2} \sum_{a,b} \left(|h(a) f_a(b) - h(a) g_a(b)| + |h(a) g_a(b) - \frac{1}{C} h(a) g_a(b)| \right) \\
\leq\ & \frac{1}{2} \left(\sum_{a,b} (h(a) f_a(b) - h(a) g_a(b)) + \frac{1 - C}{C} \sum_{a,b} h(a) g_a(b) \right) \quad \text{(using item 2. of this claim)} \\
\leq\ & \frac{1}{2} \left(\sum_{a,b:(a,b) \notin G} h(a) f_a(b) + 1 - C \right) \\
=\ & \frac{1}{2} \left(\Pr_{(a,b) \leftarrow AB}[(a, b) \notin G] + 1 - C \right) \leq \delta_1 + \delta_2 \quad \text{(from (7.14))}
\end{aligned}$$

□

Claim 7.2.7. $\Pr_{p,\mathbf{r}_{\Pi'}}[H] \geq (1 - \frac{23}{2}\varepsilon) \cdot 2^{-k-\Delta-2}$.

Proof. By the definition of H, we have

$$\begin{aligned}
\Pr_{p,\mathbf{r}_{\Pi'}}[H] &\geq \sum_{(x,y)\in G_1\cap G_2} p(x,y) \Pr_{\mathbf{r}_{\Pi'}}[H|\ (x,y)] \\
&\geq (1-4\varepsilon)\cdot 2^{-k-\Delta-2} \sum_{(x,y)\in G_1\cap G_2} p(x,y) \\
&\geq (1-\frac{23}{2}\varepsilon)\cdot 2^{-k-\Delta-2}.
\end{aligned}$$

The second inequality is by Claim 7.2.5, item 3, and the last inequality is by Claim 7.2.3 item 3. □

With the previous claim, we are able to show that the protocol Π' nicely simulate the distribution XYM.

Claim 7.2.8. Let the input of protocol Π' be drawn according to p. Let $X^1Y^1M^1$ represent the input and the transcript (the part of the public coins drawn from $\mathcal{M}$) conditioned on H. Then we have $\|XYM - X^1Y^1M^1\|_1 \leq 10\varepsilon$. Note that this implies that $\left\|X^1Y^1A^1B^1 - XY\tilde{M}\tilde{M}\right\|_1 \leq 10\varepsilon$, where $\tilde{M}$ represents the last $\lceil \log|\mathcal{Z}| \rceil$ bits of M and A^1, B^1 represent outputs of Alice and Bob respectively, conditioned on H.

Proof. For any (x,y), define

$$w_{xy}(m) \stackrel{\text{def}}{=} \min\left\{u_x(m), 2^{\Delta}v_y(m)\right\} \cdot \min\left\{u_y(m), 2^{\Delta}v_x(m)\right\}.$$

From step 2 (a),(b),(c), of protocol Π', $\Pr[M^1X^1Y^1 = mxy] = \frac{1}{C}p(x,y)w_{xy}(m)$, where $C = \sum_{xym} p(x,y)w_{xy}(m)$. Now,

$$\begin{aligned}
&\Pr_{(x,y)\leftarrow XY}\left[\Pr_{m\leftarrow M_{xy}}[w_{xy}(m) = u_x(m)u_y(m)] \geq 1-2\varepsilon\right] \\
&\quad = \Pr_{(x,y)\leftarrow XY}[(x,y)\in G] \geq 1-8\varepsilon.
\end{aligned}$$

The last inequality above follows using items 3. and 4. of Claim 7.2.3 and the fact that XY and p are $\varepsilon/2$-close.

Finally using Claim 7.2.6, (by substituting $\delta_1 \leftarrow 2\varepsilon, \delta_2 \leftarrow 8\varepsilon, A \leftarrow XY, B \leftarrow M, A' \leftarrow X^1Y^1, B' \leftarrow M^1, h \leftarrow \frac{p}{q}, f_{(x,y)}(m) \leftarrow u_x(m)u_y(m)$ and $g_{(x,y)}(m) \leftarrow w_{xy}(m)$), we get that $\|X^1Y^1M^1 - XYM\|_1 \leq 10\varepsilon$. □

We are now ready to finish the proof of Lemma 7.2.2.

Proof of Lemma 7.2.2: Consider the protocol Π'. We claim that it satisfies Lemma 7.2.1 by taking the correspondence between quantities in Lemma 7.2.1 and Lemma 7.2.2 as follows : $c \leftarrow (c/\varepsilon^2 + 3/\varepsilon), \varepsilon \leftarrow 11\varepsilon, \beta \leftarrow \beta, \delta \leftarrow \delta, z \leftarrow z, X'Y' \leftarrow p$.

Item 1. of Lemma 7.2.1 is implied by Claim 7.2.7 since $(1-\frac{23}{2}\varepsilon)\cdot 2^{-k-\Delta-2} \geq 2^{-(c/\varepsilon^2+3/\varepsilon)}$, from choice of parameters.

Item 2. of Lemma 7.2.1 is implied since $\|X^1Y^1 - p\|_1 \leq \|X^1Y^1 - XY\|_1 + \|XY - p\|_1 \leq \frac{21}{2}\varepsilon$, using item 2. of Lemma 7.2.2, Fact 5.1.5 and Claim 7.2.8.

Item 3. of Lemma 7.2.1 is implied since

$$\mathrm{err}_f\left(X^1Y^1M^1\right) \leq \mathrm{err}_f(XYM) + \left\|X^1Y^1M^1 - XYM\right\|_1 \leq 11\varepsilon,$$

using item 4. in Lemma 7.2.2 and Claim 7.2.8.

This implies

$$\widetilde{\mathrm{srec}}^{z,p}_{(1+\varepsilon')\delta/\beta,\delta}(f) < \frac{c/\varepsilon^2 + 3/\varepsilon}{11\varepsilon} \leq \frac{2c}{11\varepsilon^3}.$$

□

We can now prove our main result.

Theorem 7.2.9. *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, and $t > 1$ be an integer. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Let $z \in \mathcal{Z}$ and $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow\mu}[f(x,y) = \{z\}]$. Let $0 < \varepsilon < 1/3$ and $\varepsilon', \delta > 0$ be such that $\frac{\delta+22\varepsilon}{\beta-33\varepsilon} < (1+\varepsilon')\frac{\delta}{\beta}$. It holds that,*

$$\mathrm{R}^{\mathrm{pub}}_{1-(1-\varepsilon)^{\lfloor\varepsilon^2 t/32\rfloor}}(f^t) \geq \frac{\varepsilon^2}{32} \cdot t \cdot \left(11\varepsilon \cdot \widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}(f) - 2\right).$$

Proof. Set $\delta_1 \stackrel{\text{def}}{=} \varepsilon^2/32$. define

$$c \stackrel{\text{def}}{=} 11\varepsilon \cdot \widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}(f) - 2$$

and $XY \sim \mu^k$. By Fact 5.2.1, it suffices to show

$$\mathrm{D}^{\mu^t}_{1-(1-\varepsilon)^{\lfloor\varepsilon^2 t/32\rfloor}}(f^t) \geq \delta_1 tc.$$

Let Π be a deterministic two-way communication protocol, that computes f^t, with total communication $\delta_1 ct$ bits. The following claim implies that the success of Π is at most $(1-\varepsilon)^{\lfloor\delta_1 t\rfloor}$, and this shows the desired. □

Claim 7.2.10. For each $i \in [t]$, define a binary random variable $T_i \in \{0,1\}$, which

represents the success of Π on the i-th instance. That is, $T_i = 1$ if the protocol computes the i-th instance of f correctly, and $T_i = 0$ otherwise. Let $t' \stackrel{\text{def}}{=} \lfloor \delta_1 t \rfloor$. There exists t' coordinates $\{i_1, \cdots, i_{t'}\}$ such that for each $1 \leq r \leq t' - 1$,

1. either $\Pr\left[T^{(r)} = 1\right] \leq (1-\varepsilon)^{t'}$ or
2. $\Pr\left[T_{i_{r+1}} = 1 |\ T^{(r)} = 1\right] \leq 1 - \varepsilon$, where $T^{(r)} \stackrel{\text{def}}{=} \prod_{j=1}^{r} T_{i_j}$.

Proof. Suppose we have already identified r coordinates, $i_1, \cdots, i_r$ satisfying that $\Pr[T_{i_1}] \leq 1 - \varepsilon$ and $\Pr\left[T_{i_{j+1}} = 1 |\ T^{(j)} = 1\right] \leq 1 - \varepsilon$ for $1 \leq j \leq r-1$. If $\Pr\left[T^{(r)} = 1\right] \leq (1-\varepsilon)^{t'}$, then we are done. So from now on we assume $\Pr\left[T^{(r)} = 1\right] > (1-\varepsilon)^{t'} \geq 2^{-\delta_1 t}$. Here we assume $r \geq 1$. Similar arguments also work when $r = 0$, that is for identifying the first coordinate, which we skip for the sake of avoiding repetition.

Let D be a random variable uniformly distributed in $\{0,1\}^t$ and independent of XY. Let $U_i = X_i$ if $D_i = 0$, and $U_i = Y_i$ if $D_i = 1$. For any random variable L, define $L^1 \stackrel{\text{def}}{=} (L | T^{(r)} = 1)$. If $L = L_1 \cdots L_t$, define $L_{-i} \stackrel{\text{def}}{=} L_1 \cdots L_{i-1} L_{i+1} \cdots L_t$. Let $C \stackrel{\text{def}}{=} \{i_1, \cdots, i_r\}$. Define $R_i \stackrel{\text{def}}{=} D_{-i} U_{-i} X_{C \cup [i-1]} Y_{C \cup [i-1]}$.

Now let us apply Lemma 7.2.2 by substituting $XY \leftarrow X_j^1 Y_j^1, M \leftarrow R_j^1 M^1, p \leftarrow X_j Y_j, z \leftarrow z, \varepsilon \leftarrow \varepsilon, \delta \leftarrow \delta, \beta \leftarrow \beta, \varepsilon' \leftarrow \varepsilon'$ and $c \leftarrow 16\delta_1(c+1)$. Condition 1. in Lemma 7.2.2 is implied by Claim 7.2.12. Conditions 2. and 3. are implied by Claim 7.2.13. Also we have $\widetilde{\text{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}(f) > \frac{32\delta_1(c+1)}{11\varepsilon^3}$, by our choice of c. Hence condition 4. must be false and hence $\text{err}_f\left(X_j^1 Y_j^1 M^1\right) = \text{err}_f\left(X_j^1 Y_j^1 R_j^1 M^1\right) > \varepsilon$. This shows condition 2. of this Claim. □

The following fact can be easily verified by induction on the number of message exchanges in a private-coin protocol (please refer for example to [12] for an explicit proof). It is also implicit in the *cut and paste* property of private-coins protocol used in Bar-Yossef, Jayram, Kumar and Sivakumar [6] and in Jain, Radhakrishnan and Sen [35].

Lemma 7.2.11. *For any private-coin two-way communication protocol, with input $XY \sim \mu$ and transcript $M \in \mathcal{M}$, the joint distribution can be written as*

$$\Pr[XYM = xym] = \mu(x,y) u_x(m) u_y(m),$$

where $u_x : \mathcal{M} \to [0,1]$ and $u_y : \mathcal{M} \to [0,1]$, for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$.

Claim 7.2.12. Let $\mathcal{R}$ denote the space of R_j. There exist functions $u_{x_j}, u_{y_j} : \mathcal{R} \times \mathcal{M} \to [0,1]$ for all $(x_j, y_j) \in \mathcal{X} \times \mathcal{Y}$ and a real number $q > 0$ such that

$$\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = x_j y_j r_j m\right] = \frac{1}{q}\mu(x_j, y_j) u_{x_j}(r_j, m) u_{y_j}(r_j, m).$$

Proof. Note that $X_j Y_j$ is independent of R_j. Now consider a private-coin two-way protocol Π_1 with input $X_j Y_j$ as follows. Let Alice generate R_j and send to Bob. Alice and Bob then generate $(X_{-j})_{x_j r_j}$ and $(Y_{-j})_{y_j r_j}$, respectively. Then they run the protocol Π. Thus, from Lemma 7.2.11,

$$\Pr[X_j Y_j R_j M = x y_j r m] = \mu(x_j, y_j) \cdot v_{x_j}(r_j, m) \cdot v_{y_j}(r_j, m),$$

where $v_{x_j}, v_{y_j} : \mathcal{R} \times \mathcal{M} \to [0,1]$, for all $(x_j, y_j) \in \mathcal{X} \times \mathcal{Y}$.

Note that conditioning on $T^{(r)} = 1$ corresponds to choosing a subset, say S, of $\mathcal{R} \times \mathcal{M}$. Let

$$q \stackrel{\text{def}}{=} \sum_{x_j y_j r_j m : (r_j, m) \in S} \mu(x_j, y_j) v_{x_j}(r_j, m) v_{y_j}(r_j, m) \ .$$

Then

$$\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = x_j y_j r_j m\right] = \frac{1}{q}\mu(x_j, y_j) v_{x_j}(r_j, m) v_{y_j}(r_j, m),$$

for $(r_j, m) \in S$ and $\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = x_j y_j r_j m\right] = 0$ otherwise.

Now define

$$u_{x_j}(r_j, m) \stackrel{\text{def}}{=} v_{x_j}(r_j, m), \text{ and } u_{y_j}(r_j, m) \stackrel{\text{def}}{=} v_{y_j}(r_j, m),$$

for $(r_j, m) \in S$ and define them to be 0 otherwise. The claim follows. □

Claim 7.2.13. If $\Pr\left[T^{(r)} = 1\right] > 2^{-\delta_1 t}$, then there exists a coordinate $j \notin C$ such that

$$\mathrm{S}\left(X_j^1 Y_j^1 \middle\| X_j Y_j\right) \leq 8\delta_1 = \tfrac{\varepsilon^2}{4}, \tag{7.15}$$

and

$$\mathrm{I}\left(X_j^1 : M^1 R_j^1 \middle| Y_j^1\right) + \mathrm{I}\left(Y_j^1 : M^1 R_j^1 \middle| X_j^1\right) \leq 16\delta_1(c+1). \tag{7.16}$$

Proof. This follows using Claim 6.2.11. □

Now let us prove Claim 7.2.5.

Proof of Claim 7.2.5:

1. We do the argument for Alice. Similar argument follows for Bob. Note that $u_x(m), v_x(m) \in [0,1]$. Then for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\Pr_{\mathbf{r}_{\Pi'}}[\text{Alice accepts } \mathbf{m_i}|\ (x,y)] = \frac{1}{|\mathcal{M}|}\sum_m \frac{u_x(m)v_x(m)}{2^\Delta} = \frac{\alpha_x}{|\mathcal{M}|2^\Delta}.$$

Item 1 follows by the fact that $(x,y) \in G_1$.

2. Define E_i (subevent of E) when i is the smallest element of $\mathcal{A}$. For all $(x,y) \in G_1 \cap G_2$, we have :

$$\begin{aligned}
&\Pr_{\mathbf{r}_{\Pi'}}[B_c|\ (x,y), E_i] \\
&= \Pr_{\mathbf{r}_{\Pi'}}[\exists j\ :\ j \in \mathcal{B} \text{ and } \mathbf{h}(\mathbf{m_j}) = \mathbf{r} \text{ and } \mathbf{m_j} \neq \mathbf{m_i}|\ (x,y), E_i] \\
&\leq \sum_{j\in[T], j\neq i} \Pr_{\mathbf{r}_{\Pi'}}[j \in \mathcal{B} \text{ and } \mathbf{h}(\mathbf{m_j}) = \mathbf{r} \text{ and } \mathbf{m_j} \neq \mathbf{m_i}|\ (x,y), E_i] \quad \text{(from the union bound)} \\
&\leq \sum_{j\in[T], j\neq i} \Pr_{\mathbf{r}_{\Pi'}}[j \in \mathcal{B}|\ (x,y), E_i] \cdot \Pr_{\mathbf{r}_{\Pi'}}[\mathbf{h}(\mathbf{m_j}) = \mathbf{r}|\ (x,y), E_i, j \in \mathcal{B}, \mathbf{m_j} \neq \mathbf{m_i}] \\
&\leq T \cdot \frac{3q}{|\mathcal{M}|2^{\Delta+1}} \cdot \frac{1}{2^k} \quad \text{(two-wise independence of } \mathbf{h} \text{ and item 1. of this Claim)} \\
&\leq \varepsilon. \quad \text{(from choice of parameters)}
\end{aligned}$$

Since above holds for every i, it implies $\Pr_{\mathbf{r}_{\Pi'}}[B_c|\ (x,y), E] \leq \varepsilon$.

3. Consider,

$$\begin{aligned}
&\Pr_{\mathbf{r}_{\Pi'}}[E|\ (x,y)] = \Pr_{\mathbf{r}_{\Pi'}}[\mathcal{A} \neq \varnothing|\ (x,y)] \cdot \Pr_{\mathbf{r}_{\Pi'}}[E|\ \mathcal{A} \neq \varnothing, (x,y)] \\
&\geq \left(1 - \left(1 - \frac{1}{2} \cdot \frac{q}{|\mathcal{M}|2^\Delta}\right)^T\right) \cdot \Pr_{\mathbf{r}_{\Pi'}}[E|\ \mathcal{A} \neq \varnothing, (x,y)] \quad \text{(using item 1. of this claim)} \\
&\geq (1-\varepsilon) \cdot \Pr_{\mathbf{r}_{\Pi'}}[E|\ \mathcal{A} \neq \varnothing, (x,y)] \ \text{(from choice of parameters)} \\
&= (1-\varepsilon) \cdot \Pr_{\mathbf{r}_{\Pi'}}[\mathbf{h}(\mathbf{m_i}) = \mathbf{r}|\ \mathcal{A} \neq \varnothing, (x,y)] \cdot \Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{B}|\ i \in \mathcal{A},\ \mathbf{h}(\mathbf{m_i}) = \mathbf{r}, (x,y)]
\end{aligned}$$

(from here on we condition on i being the first element of $\mathcal{A}$)

$$\begin{aligned}
&= (1-\varepsilon) \cdot 2^{-k} \cdot \Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{B}|\ i \in \mathcal{A}, (x,y)] \\
&= (1-\varepsilon) \cdot 2^{-k} \cdot \frac{\Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{B} \text{ and } i \in \mathcal{A}|\ (x,y)]}{\Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{A}|\ (x,y)]}
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{2}{3q}(1-\varepsilon)\cdot 2^{-k}\cdot|\mathcal{M}|2^{\Delta}\cdot \Pr_{\mathbf{r}_{\Pi'}}[i\in\mathcal{B} \text{ and } i\in\mathcal{A}|\ (x,y)] \quad \text{(using item 1. of this claim)}\\
&= \frac{2}{3q}(1-\varepsilon)\cdot 2^{-k}\cdot|\mathcal{M}|2^{\Delta}\cdot \sum_{m\in\mathcal{M}}\frac{1}{|\mathcal{M}|2^{2\Delta}}\min\left\{u_x(m),2^{\Delta}v_y(m)\right\}\cdot\min\left\{u_y(m),2^{\Delta}v_x(m)\right\}\\
&\qquad\qquad \text{(from construction of protocol } \Pi')\\
&\geq \frac{2}{3q}(1-\varepsilon)\cdot 2^{-k}\cdot|\mathcal{M}|2^{\Delta}\cdot \sum_{m\in G_{xy}}\frac{u_x(m)u_y(m)}{|\mathcal{M}|2^{2\Delta}}\\
&\qquad\qquad (G_{xy}\stackrel{\text{def}}{=}\{m: u_x(m)\leq 2^{\Delta}v_y(m) \text{ and } u_y(m)\leq 2^{\Delta}v_x(m)\})\\
&= \frac{2}{3q}(1-\varepsilon)\cdot 2^{-k}\cdot|\mathcal{M}|2^{\Delta}\cdot \frac{\alpha_{xy}}{|\mathcal{M}|2^{2\Delta}}\sum_{m\in G_{xy}}\frac{u_x(m)u_y(m)}{\alpha_{xy}}\\
&\geq \frac{1}{3}(1-\varepsilon)\cdot 2^{-k-\Delta}\cdot \Pr_{m\leftarrow M_{xy}}[m\in G_{xy}] \quad \text{(since } (x,y)\in G_1 \text{ and (7.8))}\\
&\geq \frac{1}{3}(1-\varepsilon)\cdot 2^{-k-\Delta}\cdot(1-2\varepsilon) \quad \text{(since } (x,y)\in G, \text{ using item 4. of Claim 7.2.3)}\\
&\geq (1-3\varepsilon)\cdot 2^{-k-\Delta-2}.
\end{aligned}
$$

Finally, using item 2. of this Claim.

$$
\Pr_{\mathbf{r}_{\Pi'}}[H|\ (x,y)] = \Pr_{\mathbf{r}_{\Pi'}}[E|\ (x,y)]\left(1-\Pr_{\mathbf{r}_{\Pi'}}[B_c|\ (x,y),E]\right)\geq(1-4\varepsilon)\cdot 2^{-k-\Delta-2}.
$$

□

Chapter 8

Conclusions and open problems

In this thesis, we have studied two independent topics. The first topic is concerned with fast parallel approximation algorithms for semidefinite programs. The second topic is concerned with strong direct product results in communication complexity. In this chapter, we briefly recall our main results and list some related open problems for further study.

8.1 Fast parallel approximation algorithms for semidefinite programs

In Chapter 3, we presented a fast parallel approximation algorithm for positive semidefinite programs. Our result generalizes the algorithm of Luby and Nisan [53]. To generalize their algorithm, the difficulty we faced was the non-commutative nature of the matrices involved. To handle it, we introduced new techniques, which are independently interesting and may have other applications. In Chapter 4, we presented a fast parallel approximation algorithm for mixed packing and covering problem, which strengthened the result in Chapter 3. Some related open problems are listed below.

8.1.1 Open problems

1. The programs we considered in Chapter 4 are not the most general mixed packing and covering programs since the covering constraints in the programs are linear. A natural question that arises is as follows. Can we get a fast parallel approximation algorithm for the following more general mixed packing and covering program?

Given $n \times n$ positive semidefinite matrices $P_1, \ldots, P_m, P, C_1, \ldots, C_m, C$,

$$\begin{aligned} \text{maximize:} \quad & \gamma \\ \text{subject to:} \quad & \sum_{i=1}^{m} x_i P_i \leq P \\ & \sum_{i=1}^{m} x_i C_i \geq \gamma C \\ & \forall i \in [m] : x_i \geq 0. \end{aligned}$$

2. Can we find interesting applications of the fast parallel approximation algorithms exhibited in this thesis ?

8.2 Strong direct product problems

In Chapter 6, we proved a direct product theorem for bounded-round public-coin communication complexity. As an application, we showed the strong direct product theorem for the Pointer Chasing. Very recently, our result is improved by Braverman, Rao, Weinstein and Yehudayoff [15] with better dependence on the number of rounds in the direct product result using a new sampling technique introduced in [14]. In Chapter 7, we provided a strong direct product result for the two-way public-coin communication complexity in terms of an important and widely used lower bound method, the smooth rectangle bound.

8.2.1 Open problems

As we mentioned in Chapter 5, strong direct product problems are central problems in complexity theory. They have been studied in various models for several years. In communication complexity, much progress has been made in the last decade. Some natural questions that arise from this work are:

1. In quantum communication complexity, strong direct product quesions are widely open. Can the techniques in Chapter 6 be extended to show direct product theorems for bounded-round quantum communication complexity?

2. Is the smooth rectangle bound a tight lower bound for two-way public-coin communication complexity for all relations? If yes, this would imply a strong direct product result for the two-way public-coin communication complexity for all relations, settling a major open question in this area. To start with, we can ask: is the smooth rectangle bound polynomially tight for the two-way public-coin communication complexity for all relations?

3. Or on the other hand, can we exhibit a relation for which the smooth rectangle bound is (asymptotically) strictly smaller than its two-way public-coin communication complexity?

4. Can we show similar direct product results in terms of possibly stronger lower bound methods like the partition bound and the information complexity?

5. It will be interesting to obtain new optimal lower bounds for the functions and relations using the smooth rectangle bound, implying strong direct product results for them.

Appendix A

Smooth rectangle bound

A.1 Proof of Lemma 5.2.6

Let $(\lambda'_{x,y}, \phi'_{x,y})$ be an optimal solution to the Dual. For $(x,y) \in f^{-1}(z)$, if $\lambda'_{x,y} > \phi'_{x,y}$ define $\lambda = \lambda'_{x,y} - \phi'_{x,y}$ and $\phi_{x,y} = 0$. Otherwise define $\lambda = 0$ and $\phi_{x,y} = \phi'_{x,y} - \lambda'_{x,y}$. For $(x,y) \notin f^{-1}(z)$ define $\phi_{x,y} = 0$. We note that $(\lambda_{x,y}, \phi_{x,y})$ is an optimal solution to the Dual with potentially higher objective value. Hence $(\lambda_{x,y}, \phi_{x,y})$ is also an optimal solution to the Dual.

Let us define three sets

$$U_1 \stackrel{\text{def}}{=} \{(x,y)|\ f(x,y) = z, \lambda_{x,y} > 0\},$$

$$U_2 \stackrel{\text{def}}{=} \{(x,y)|\ f(x,y) = z, \phi_{x,y} > 0\},$$

$$U_0 \stackrel{\text{def}}{=} \{(x,y)|\ f(x,y) \neq z, \lambda_{x,y} > 0\}.$$

Define,

$$\forall (x,y) \in U_1 : \mu'(x,y) \stackrel{\text{def}}{=} \lambda_{x,y},$$

$$\forall (x,y) \in U_2 : \mu'(x,y) \stackrel{\text{def}}{=} \varepsilon \phi_{x,y},$$

$$\forall (x,y) \in U_0 : \mu'(x,y) \stackrel{\text{def}}{=} \varepsilon \lambda_{x,y}.$$

Define $r \stackrel{\text{def}}{=} \sum_{x,y} \mu'(x,y)$ and define probability distribution $\mu \stackrel{\text{def}}{=} \mu'/r$. Let $\text{srec}^z_\epsilon(f) = 2^c$.

Define function g such that $g(x,y) = z$ for $(x,y) \in U_1$; $g(x,y) = f(x,y)$ for $(x,y) \in U_0$ and $g(x,y) = z'$ (for some $z' \neq z$) for $(x,y) \in U_2$. Then,

$$\begin{aligned} 2^c &= \sum_{(x,y)\in f^{-1}(z)} ((1-\epsilon)\lambda_{x,y} - \phi_{x,y}) - \sum_{(x,y)\notin f^{-1}(z)} \epsilon \cdot \lambda_{x,y} \\ &= (1-\epsilon)\mu'(U_1) - \frac{1}{\varepsilon}\mu'(U_2) - \mu'(U_0) \end{aligned}$$

This implies $r \geq \mu'(U_1) \geq 2^c$. Consider rectangle W.

$$\begin{aligned} &\sum_{(x,y)\in f^{-1}(z)\cap W} (\lambda_{x,y} - \phi_{x,y}) - \sum_{(x,y)\in(W - f^{-1}(z))} \lambda_{x,y} \leq 1 \\ &\Rightarrow \sum_{(x,y)\in U_1\cap W} \mu_{x,y} - \frac{1}{\varepsilon} \sum_{(x,y)\in U_2\cap W} \mu_{x,y} - \sum_{(x,y)\in U_0\cap W} \frac{1}{\varepsilon}\mu_{x,y} \leq \frac{1}{r} \\ &\Rightarrow \varepsilon \left(\sum_{(x,y)\in U_1\cap W} \mu_{x,y} - \frac{1}{r} \right) \leq \sum_{(x,y)\in U_2\cap W} \mu_{x,y} + \sum_{(x,y)\in U_0\cap W} \mu_{x,y} \\ &\Rightarrow \varepsilon \left(\sum_{(x,y)\in g^{-1}(z)\cap W} \mu_{x,y} - \frac{1}{r} \right) \leq \sum_{(x,y)\in W - g^{-1}(z)} \mu_{x,y} \\ &\Rightarrow \varepsilon \left(\sum_{(x,y)\in W} \mu_{x,y} - \frac{1}{r} \right) \leq (1+\varepsilon) \cdot \sum_{(x,y)\in W - g^{-1}(z)} \mu_{x,y} \\ &\Rightarrow \varepsilon \left(\sum_{(x,y)\in W} \mu_{x,y} - 2^{-c} \right) \leq (1+\varepsilon) \cdot \sum_{(x,y)\in W - g^{-1}(z)} \mu_{x,y}. \end{aligned}$$

Now consider a W with $\mu(W) \geq 2^{-c}/\varepsilon^3$. We have $\mu(W - g^{-1}(z)) \geq \frac{(1-\varepsilon^3)\varepsilon}{1+\varepsilon}\mu(W)$. Define $\beta \stackrel{\text{def}}{=} \mu(U_1 \cup U_2)$, $\delta \stackrel{\text{def}}{=} \mu(U_2)$. Now,

$$(1-\varepsilon)r\beta \geq (1-\varepsilon)\mu'(U_1) \geq \frac{1}{\varepsilon}\mu'(U_2) = \frac{1}{\varepsilon} r\delta.$$

Hence we have

$$\mu(W - g^{-1}(z)) \geq \frac{(1-\varepsilon^3)\delta}{(1-\varepsilon^2)\beta}\mu(W) \geq (1+\varepsilon^2)\frac{\delta}{\beta}\mu(W).$$

This implies $\widetilde{\mathrm{rec}}^{z,\mu}_{(1+\varepsilon^2)\delta/\beta}(g) \geq c + 3\log\varepsilon$. This implies that

$$\widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon^2)\frac{\delta}{\beta},\delta}(f) \geq c + 3\log\varepsilon = \log(\mathrm{srec}^{z}_{\varepsilon}(f)) + 3\log\varepsilon.$$

A.2 Smooth lower bound vs. communication complexity

Jain and Klauck show that the smooth rectangle bound is a lower bound on public-coin two-way communication complexity, as stated in Lemma 5.2.7. We contain the proof here for completeness.

Proof of Lemma 5.2.7: Let $c \stackrel{\text{def}}{=} \widetilde{\mathrm{srec}}^{z,\lambda}_{(1+\varepsilon')\frac{\delta}{\beta},\delta}(f)$. Let g be such that $\widetilde{\mathrm{rec}}^{z,\lambda}_{(1+\varepsilon')\frac{\delta}{\beta}}(g) = c$ and

$$\Pr_{(x,y)\leftarrow\lambda}[f(x,y) \neq g(x,y)] \leq \delta.$$

If $\mathrm{D}^{\lambda}_{\varepsilon}(f) \geq c - \log(4/\varepsilon)$ then we are done using Fact 5.2.1.

So lets assume for contradiction that $\mathrm{D}^{\lambda}_{\varepsilon}(f) < c - \log(4/\varepsilon)$. This implies that there exists a deterministic protocol Π for f with communication $c - \log(4/\varepsilon)$ and distributional error under λ bounded by ε. Since

$$\Pr_{(x,y)\leftarrow\lambda}[f(x,y) \neq g(x,y)] \leq \delta,$$

the protocol Π will have distributional error at most $\varepsilon + \delta$ for g. Let M represent the message transcript of Π and let O represent protocol's output. We assume that the last $\lceil \log|Z| \rceil$ bits of M contain O. We have,

1. $\Pr_{m\leftarrow M}[\Pr[M = m] \leq 2^{-c}] \leq \varepsilon/4$, since the total number of message transcripts in Π is at most $2^{c-\log(4/\varepsilon)}$.

2. $\Pr_{m\leftarrow M}[O = z|\ M = m] > \beta - \varepsilon$,
since $\Pr_{(x,y)\leftarrow\lambda}[f(x,y) = \{z\}] = \beta$ and distributional error of Π under λ is bounded by ε for f.

3. $\Pr_{m\leftarrow M}\left[\Pr_{(x,y)\leftarrow(XY)_m}[(x,y,O) \notin g|\ M = m] \geq \frac{\varepsilon+\delta}{\beta-2\varepsilon}\right] \leq \beta - 2\varepsilon$, since distributional error of Π under λ is bounded by $\varepsilon + \delta$ for g.

Using all of above we obtain a message transcript m such that $\Pr[M = m] > 2^{-c}$ and $(O = z|\ M = m)$ and

$$\begin{aligned}\Pr_{(x,y)\leftarrow(XY|M=m)}[(x,y,O) \notin g|\ M = m] &\leq \frac{\varepsilon+\delta}{\beta-2\varepsilon} \\ &< (1+\varepsilon')\frac{\delta}{\beta}.\end{aligned}$$

This and the fact that the support of $(XY|\ M = m)$ is a rectangle, implies that $\widetilde{\mathrm{rec}}^{z,\lambda}_{(1+\varepsilon')\frac{\delta}{\beta}}(g) < c$, contradicting the definition of c. Hence it must be that $\mathrm{D}^{\lambda}_{\varepsilon}(f) \geq c - \log(4/\varepsilon)$, which using Fact 5.2.1 shows the desired. □

Bibliography

[1] Andris Ambainis, Andrew M. Childs, Ben W. Reichardt, Robert Spalek, and Shengyu Zhang. Any AND-OR formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '07, pages 363–372, Washington, DC, USA, 2007. IEEE Computer Society. 13

[2] Andris Ambainis, Robert Spalek, and Ronald de Wolf. A new quantum lower bound method, with applications to direct product theorems and time-space tradeoffs. *Algorithmica*, 55:422–461, 2009. 4

[3] Sanjeev Arora, Elad Hazan, and Satyen Kale. Fast algorithms for approximate semide.nite programming using the multiplicative weights update method. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '05, pages 339–348, Washington, DC, USA, 2005. IEEE Computer Society. 1, 6, 12

[4] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing*, STOC '07, pages 227–236, New York, NY, USA, 2007. ACM. 1, 6, 12

[5] Laszlo Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th Annual Symposium on Foundations of Computer Science*, SFCS '86, pages 337–347, Washington, DC, USA, 1986. IEEE Computer Society. 4, 42, 43

[6] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 209–218, Washington, DC, USA, 2002. IEEE Computer Society. 47, 48, 74

[7] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *Proceedings of the 17th IEEE Annual Conference on Computational Complexity*, CCC '02, page 93, Washington, DC, USA, 2002. IEEE Computer Society. 63

[8] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, pages 67–76, New York, NY, USA, 2010. ACM. 3, 47, 48

[9] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Comput. Complex.*, 15(4):391–432, December 2006. 4, 42, 43

[10] Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and ldcs. In *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '08, pages 477–486, Washington, DC, USA, 2008. IEEE Computer Society. 3

[11] Rajendra Bhatia. *Matrix Analysis*. Springer, New York, USA, 1996. 19

[12] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th annual ACM Symposium on Theory of Computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM. 5, 65, 67, 70, 74

[13] Mark Braverman and Anup Rao. Information equals amortized communication. In *Proceedings of the 52nd Symposium on Foundations of Computer Science*, FOCS '11, pages 748–757, Washington, DC, USA, 2011. IEEE Computer Society. 3, 4, 47, 48, 49

[14] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:143, 2012. 3, 79

[15] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. In *Proceedings of the 40th international conference on Automata, languages and programming*, ICALP'13, to appear, Berlin, Heidelberg, 2013. Springer-Verlag. 79

[16] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, Lecture Notes in Computer Science, pages 459–470. Springer Berlin Heidelberg, 2012. 43, 65, 66, 67, 70

[17] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. In *Proceedings of the 43rd annual ACM Symposium on Theory of Computing*, STOC '11, pages 51–60, New York, NY, USA, 2011. ACM. 4, 43, 63

[18] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, FOCS '01, page 270, Washington, DC, USA, 2001. IEEE Computer Society. 42, 47

[19] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, New York, NY, USA, 1991. 37

[20] Andrew Drucker. Improved direct product theorems for randomized query complexity. In *Proceedings of the 2011 IEEE 26th Annual Conference on Computational Complexity*, CCC '11, pages 1–11, Washington, DC, USA, 2011. IEEE Computer Society. 3

[21] Dmitry Gavinsky. Classical interaction cannot replace a quantum message. In *Proceedings of the 40th annual ACM Symposium on Theory of Computing*, STOC '08, pages 95–102, New York, NY, USA, 2008. ACM. 4, 63

[22] Prahladh Harsha and Rahul Jain. A strong direct product theorem for the Tribes function via the smooth-rectangle bound. *CoRR*, abs/1201.6090, 2013. 63

[23] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing*, STOC '07, pages 411–419, New York, NY, USA, 2007. ACM. 4, 39, 48

[24] Ben Ibinson, Noah Linden, and Andreas Winter. Robustness of quantum markov chains. *Communications in Mathematical Physics*, 277:289–304, 2008. 37

[25] Rahul Jain. New strong direct product results in communication complexity. *Journal of the ACM*, to appear, 2013. 3, 4, 42, 46, 47, 48, 49, 63, 64

[26] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. QIP = PSPACE. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, pages 573–582, New York, NY, USA, 2010. ACM. 1, 6, 7, 12

[27] Rahul Jain and Hartmut Klauck. New results in the simultaneous message passing model via information theoretic techniques. In *Proceedings of the 2009 24th Annual IEEE Conference on Computational Complexity*, CCC '09, pages 369–378, Washington, DC, USA, 2009. IEEE Computer Society. 3, 47

[28] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, CCC '10, pages 247–258, Washington, DC, USA, 2010. IEEE Computer Society. vi, 4, 42, 43, 44, 63

[29] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds: extended abstract. In *Proceedings of the 40th annual ACM Symposium on Theory of Computing*, STOC '08, pages 599–608, New York, NY, USA, 2008. ACM. 3, 4, 42

[30] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, FOCS '12, pages 167–176, Washington, DC, USA, 2012. IEEE Computer Society. 2

[31] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 429–438, Washington, DC, USA, 2002. IEEE Computer Society. 38

[32] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. The quantum communication complexity of the pointer chasing problem: The bit version. In *Proceedings of the 22nd Conference on Foundations of Software Technology and Theoretical Computer Science*, FSTTCS '02, pages 218–229, London, UK, 2002. Springer-Verlag. 4, 46

[33] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Proceedings of the 30th in-*

ternational conference on Automata, languages and programming, ICALP'03, pages 300–315, Berlin, Heidelberg, 2003. Springer-Verlag. 3, 47

[34] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of set disjointness. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 220–229, Washington, DC, USA, 2003. IEEE Computer Society. 47

[35] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 285–296, Washington, DC, USA, 2005. IEEE Computer Society. 3, 47, 65, 74

[36] Rahul Jain, Sarvagya Upadhyay, and John Watrous. Two-message quantum interactive proofs are in PSPACE. In *Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '09, pages 534–543, Washington, DC, USA, 2009. IEEE Computer Society. 1, 6, 7, 12

[37] Rahul Jain and John Watrous. Parallel approximation of non-interactive zero-sum quantum games. In *Proceedings of the 2009 24th Annual IEEE Conference on Computational Complexity*, CCC '09, pages 243–253, Washington, DC, USA, 2009. IEEE Computer Society. 1, 6, 12

[38] Rahul Jain and Penghui Yao. A parallel approximation algorithm for positive semidefinite programming. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science*, FOCS '11, pages 437–471, Washington, DC, USA, 2011. IEEE Computer Society. 1, 13

[39] Rahul Jain and Penghui Yao. A parallel approximation algorithm for mixed packing and covering semidefinite programs. *CoRR*, abs/1302.0275, 2012. 1

[40] Rahul Jain and Penghui Yao. A strong direct product theorem in terms of the smooth rectangle bound. *CoRR*, abs/1209.0263, 2012. 2

[41] Camille Jordan. Essai sur la géométrie à n dimensions. *Bulletin de la S. M. F*, 3:103–174, 1875. 19

[42] S. Kale. *Efficient algorithms using the multiplicative weights update method*. PhD thesis, Princeton University, 2007. 1, 6, 12

[43] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jeremie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, FOCS '12, pages 500–509, Washington, DC, USA, 2012. IEEE Computer Society. 5, 43, 63, 65, 66, 67, 70

[44] Hartmut Klauck. On quantum and probabilistic communication: Las vegas and one-way protocols. In *Proceedings of the thirty-second annual ACM Symposium on Theory of Computing*, STOC '00, pages 644–651, New York, NY, USA, 2000. ACM. 4, 46

[45] Hartmut Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proceedings of the 2003 IEEE 18th Annual Conference on Computational Complexity*, CCC '18, pages 118–134, Washington, DC, USA, 2003. IEEE Computer Society. 4, 42, 43

[46] Hartmut Klauck. A strong direct product theorem for disjointness. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, pages 77–86, New York, NY, USA, 2010. ACM. 3, 4, 43, 63

[47] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proceedings of the thirty-third annual ACM Symposium on Theory of Computing*, STOC '01, pages 124–133, New York, NY, USA, 2001. ACM. 4, 46

[48] Hartmut Klauck, Robert Spalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '04, pages 12–21, Washington, DC, USA, 2004. IEEE Computer Society. 3, 4

[49] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996. 4, 40, 41, 63

[50] Troy Lee and Jeremie Roland. A strong direct product theorem for quantum query complexity. In *Proceedings of the 2012 IEEE Conference on Computational Complexity*, CCC '12, pages 236–246, Washington, DC, USA, 2012. IEEE Computer Society. 3

[51] Troy Lee, Adi Shraibman, and Robert Spalek. A direct product theorem for discrepancy. In *Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*, CCC '08, pages 71–80, Washington, DC, USA, 2008. IEEE Computer Society. 3, 42

[52] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing*, STOC '07, pages 699–708, New York, NY, USA, 2007. ACM. 4, 42

[53] Michael Luby and Noam Nisan. A parallel approximation algorithm for positive linear programming. In *Proceedings of the twenty-fifth annual ACM Symposium on Theory of Computing*, STOC '93, pages 448–457, New York, NY, USA, 1993. ACM. v, 2, 12, 13, 78

[54] Chris Marriott and John Watrous. Quantum Arthur—Merlin games. *Comput. Complex.*, 14:122–152, June 2005. 13

[55] Daniel Nagaj, Pawel Wocjan, and Yong Zhang. Fast amplification of QMA. *Quantum Information and Computation*, 9(11–12):1053–1068, 2009. 13

[56] Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. *SIAM J. Comput.*, 28:1035–1050, February 1999. 3

[57] Noam Nisan and Avi Widgerson. Rounds in communication complexity revisited. In *Proceedings of the twenty-third annual ACM Symposium on Theory of Computing*, STOC '91, pages 419–429, New York, NY, USA, 1991. ACM. 4, 46

[58] Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the gcd problem, in old and new communication models. In *Proceedings of the twenty-ninth annual ACM Symposium on Theory of Computing*, STOC '97, pages 363–372, New York, NY, USA, 1997. ACM. 3

[59] Richard Peng and Kanat Tangwongsan. Faster and simpler width-independent parallel algorithms for positive semidefinite programming. In *Proceedinbgs of the 24th ACM symposium on Parallelism in algorithms and architectures*, SPAA '12, pages 101–108, New York, NY, USA, 2012. ACM. 27

[60] Stephen J. Ponzio, Jaikumar Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing. *J. Comput. Syst. Sci.*, 62:323–355, March 2001. 4, 46

[61] Ran Raz. A parallel repetition theorem. In *Proceedings of the twenty-seventh annual ACM Symposium on Theory of Computing*, STOC '95, pages 447–456, New York, NY, USA, 1995. ACM. 3, 48

[62] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the thirty-first annual ACM Symposium on Theory of Computing*, STOC '99, pages 358–367, New York, NY, USA, 1999. ACM. 4, 63

[63] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, December 1992. 4, 42, 43, 48

[64] Oded Regev. Witness-preserving QMA amplification, quantum computation, lecture notes, 2006. 13

[65] Oded Regev and Bo'az Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In *Proceedings of the 43rd annual ACM Symposium on Theory of Computing*, STOC '11, pages 31–40, New York, NY, USA, 2011. ACM. 4, 63

[66] Ronen Shaltiel. Towards proving strong direct product theorems. *Comput. Complex.*, 12:1–22, July 2004. 3

[67] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In *Proceedings of the 43rd annual ACM Symposium on Theory of Computing*, STOC '11, pages 41–50, New York, NY, USA, 2011. ACM. 3, 42

[68] Alexander A. Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(8):197–208, 2012. 4, 63

[69] Mario Szegedy. Quantum speed-up of markov chain based algorithms. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 32–41, Washington, DC, USA, 2004. IEEE Computer Society. 13

[70] Emanuele Viola. The communication complexity of addition. In *Proceedings of the Twenty-fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '13, pages 632–651. SIAM, 2013. 4, 63

[71] Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008. 3

[72] J. von zur Gathen. Parallel linear algebra. In editor In J. Reif, editor, *Synthesis of Parallel Algorithms*, chapter 13. Morgan Kaufmann Publishers, Inc., 1993. 7

[73] Andrew C. Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM Symposium on Theory of Computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM. 40, 41, 47

[74] Andrew C. Yao. Theory and applications of trapdoor functions. In *Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science*, pages 80–91, 1982. 3

[75] Andrew C. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 24th Annual Symposium on Foundations of Computer Science*, SFCS '83, pages 420–428, Washington, DC, USA, 1983. IEEE Computer Society. 4, 42, 43

[76] Neal. Young. Sequential and parallel algorithms for mixed packing and covering. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, FOCS '01, pages 538–, Washington, DC, USA, 2001. IEEE Computer Society. v, 2, 27, 28