Settling the complexity of the k-disjointness and the k-Hamming distance problems

Mert Sağlam

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Shayan Oveis Gharan, Chair

Paul Beame

Anup Rao

Program Authorized to Offer Degree:
Computer Science and Engineering

University of Washington

Abstract

Settling the complexity of the k-disjointness and the k-Hamming distance problems

Mert Sağlam

Chair of the Supervisory Committee:
Title of Chair Shayan Oveis Gharan
Computer Science and Engineering

Suppose that two parties, traditionally called Alice and Bob, are given respectively the inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ to a function $f\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and are required to compute $f(x, y)$. Since each party only has one part of the input, they can compute $f(x, y)$ only if some communication takes place between them. The communication complexity of a given function is the minimum amount of communication (in bits) needed to evaluate it on any input with high probability.

We study the communication complexity of two related problems, the k-Hamming distance and k-disjointness and give tight bounds to both of these problems: The r-round communication complexity of the k-disjointness problem is $\Theta(k \log^{(r)} k)$, whereas a tight $\Omega(k \log k)$ bound holds for the k-Hamming distance problem for any number of rounds.

The lower bound direction of our first result is obtained by proving a *super-sum* result on computing the OR of n equality problems, which is the first of its kind. Using our second bound, we settle the complexity of various property testing problems such as k-linearity, which was open since 2002 or earlier. Our lower bounds are obtained via information theoretic arguments and along the way we resolve a question conjectured by Erdős and Simonovits in 1982, which incidentally was studied even earlier by Blakley and Dixon in 1966.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

In communication complexity, one tries to understand the limitations of computation through communication bottlenecks. In computation, communication bottlenecks are everywhere; inside the same core, between the cores, inside the same die, package, between the memory and the processor, between software components and more. These bottlenecks are not specific to the current designs of computers; in fact, any physical computer has to deal with similar communication issues as dictated by laws of nature. Communication bottlenecks remain relevant and challenging even higher up in the abstraction hierarchy, including theoretical abstractions such as data structures and algorithms.

In 1979 to study such bottlenecks in a generic way, Yao introduced the following abstract model [87]. Two players, called Alice and Bob, are required to evaluate a known function f on inputs (x, y), where Alice has only x and Bob has only y. The players communicate by sending messages, that is, bit strings, to each other in turns until one of the players is able to deduce $f(x, y)$. The communication complexity of one such function f is taken to be the minimum number of bits of communication needed to be able to evaluate it on any input (x, y) with hight probability. Note here that the inputs x and y, for instance, may correspond to data stored in separate parts of a computational device and the communication complexity of the function would then model the data transfer between these parts.

One problem that proved its significance early on in the history of communication complexity is set disjointness. In the set disjointness problem, the players receive respectively subsets S, T of a ground set, say $\{1, 2, \ldots, n\}$, and are required to figure out if their sets S and T intersect. The first lower bound for this problem that applies even to randomized protocols was given by Babai, Frankl and Simon in [5]. Their lower bound shows that any randomized protocol for the set disjointness problem over the ground set $\{1, 2, \ldots, n\}$ requires at least $\Omega(\sqrt{n})$ bits of communication. On the flip side, the best known protocol was the trivial one round protocol wherein one player sends their entire input to the other, thereby communicating n bits and solving the problem deterministically. This gap between the upper bound and the lower bound was closed in a landmark paper of Kalyanasundaram and Schnitger from 1987, which showed a tight $\Omega(n)$ bits lower bound. This result was later simplified in influential papers of Razborov [77] and Bar-Yossef, Jayram, Kumar and Sivakumar [7], which led to a whole new understanding of communication through the lens of information theory.

The k-disjointness problem. To gather a more refined understanding of the set disjointness problem, in this work we study a variant of this problem where the set sizes are restricted to be at most k for some arbitrary parameter $k \leq n$. Needless to say, we can always take k large enough to recover the unrestricted version of the problem. This restriction may seem dull initially given the earlier remark, however it allows us to uncover an entirely new and surprising aspect of the set disjointness problem: that the complexity of the problem exhibits a rounds versus communication trade-off, that vanishes very quickly with increasing number of rounds.

In fact, we are not the first to study this restriction on the size of the sets. In 1990s Håstad and Wigderson studied this variant of the problem, henceforth called the k-disjointness problem, and gave a protocol which communicates only $O(k)$ bits over $O(\log k)$ rounds, solving the problem with constant probability [47, 73, 48]. In the first part of this thesis, we improve this protocol to run in just $\log^* k$ rounds, while we simultaneously reduce the error probability to exponentially small in k. In fact, for an arbitrary integer $r \leq \log^* k$, this protocol can be run in r rounds with communication cost $O(k \log^{(r)} k)$ bits in total, and error probability well below polynomial for $r \geq 2$. More importantly, we show that this improvement is final: we prove that any r-round set disjointness protocol with even constant error probability requires at least one message of size $\Omega(k \log^{(r)} k)$ bits. This proof works even when r is a function of k such as $r = \log^* k$, settling the complexity of the problem entirely.

The k-Hamming distance problem. An alternative way of viewing the k-disjointness problem is through the characteristic vectors of sets S, T given to the players. We may think of each player as having received a k-sparse bit vector vector and the goal of the players is to understand whether the Hamming distance of their vectors is less than $2k$. Notice that the Hamming distance of these vectors is less than $2k$ if and only if the corresponding set S and T intersect. An important generalization of this problem is the *k-Hamming distance problem* wherein the players are given bit vectors, respectively x and y, within Hamming distance k and the goal of the players is to compute the Hamming distance between x and y exactly. Notice that, compared to the k-disjointness problem, here we removed the k-sparsity promise while keeping the closeness promise intact.

For the k-disjointness and the k-Hamming distance problems there is a simple 1-round $O(k \log k)$ bits protocol. As the work of Håstad and Wigderson and our work uncovered, for the k-disjointness problem the complexity goes to k very quickly as we allow more and more rounds. Interestingly, no such protocol could be found for the k-Hamming distance problem—not even a single improvement over the easy 1-round protocol. This lack of progress became all the more pressing when in 2011 Blais Brody and Matulef [15] showed a connection between the k-Hamming distance problem and a host of important property testing problems.

The property testing model. In the *property testing model* one is given black-box query access to an otherwise unknown function g, with the task of determining whether the function is inside a certain class or differs from any function inside the class in at least an ϵ-fraction of the possible outputs. A function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ is called k-linear if $f(x) = \langle w, x \rangle$ for some $w \in \mathbb{F}_2^n$ with $\|w\|_1 \leq k$. In other words, a k-linear function outputs the sum of at most k bits of the input in mod 2. In [13], Blais gave a $O(\log k)$ round tester with $O(k \log k)$ queries for k-linearity. This result was improved in [23] by Buhrman, Garcia-Soriano, Matsliah and de Wolf, who gave a nonadaptive $O(k \log k)$ bits tester. The progress stuck here however and the gap between the $O(k \log k)$ query upper bound and the $\Omega(k)$ lower bound remained unresolved for several years.

Multiple attempts have been made to show an $\Omega(k \log k)$ lower bound to the k-Hamming distance problem to close this gap. In [23], [32], and [75] an $\Omega(k \log k)$ lower bound was shown, but only for protocols having 1 round. However, these lower bounds do not rule out the possibility that the complexity of the k-Hamming distance problem decays to, say $O(k)$, with multiple rounds, just like the related k-disjointness problem. In fact, all three of these bounds apply to both the k-disjointness and the k-Hamming distance problems alike, and therefore are incapable of separating them. Some other attempts have been made to prove an $O(k)$ upper bound to the k-Hamming distance problem. One daunting challenge in proving a sharp lower bound the the k-Hamming distance problem lies in the fact that one of the most powerful proof techniques in communication complexity, the corruption bound, is *unable* prove any bound beyond $\Omega(k \log 1/\delta)$ where δ is the error probability of the protocol. The corruption bound fails due to the existence of large combinatorial rectangles in the communication matrix of the k-Hamming distance function.

In 2014, Blais, Brody and Ghazi [14] showed an $\Omega(k \log 1/\delta)$ lower bound for the k-Hamming distance problem using an information theoretic argument. While this may seem like a small improvement over the $\Omega(k)$ lower bound that follows from the work of [55], as one usually takes δ to be a constant, the importance of this result lies in that it is the first formal evidence that k-Hamming distance is harder than the k-disjointness problem. Recall that our upper bound [80] solves the k-disjointness problem with $O(k)$ bits of communication and exponentially small error probability in k, so no bound of the form $\Omega(k \log 1/\delta)$ can be true for k-disjointness. We remark that the lower bound [14] does not go further than $\Omega(k \log 1/\delta)$ not only due to the limitation of the corruption method alluded to in the earlier paragraph, but actually the problem they study, the $\mathrm{OR} \circ 1\mathrm{vs}3$ problem, to establish their k-Hamming distance lower bound admits an actual $O(k \log 1/\delta)$ upper bound. Notice that this is a much stronger statement than the existence of larger rectangles.

The usual decomposition technique employed in randomized communication complexity of considering a special set of possible inputs which can be interpreted as calculating a composed function, say $f \circ g^k$, with a simple f leads to the OR $\circ$ 1vs3 function of [14]. As mentioned earlier, this restriction of the problem admits an $O(k \log 1/\delta)$ bits upper bound, so it is of no help to prove an $\Omega(k \log k)$ lower bound. In the same work [14], the composed function MAJ $\circ$ 1vs3 was proposed as a candidate to obtain an $\Omega(k \log k)$ lower bound. However working with the majority function appears to be difficult and it is unclear if the decomposition bought us any comfort in proving the optimal lower bound at all.

If one does not go through the decomposition route, the only standard technique which appears to be able to provide a global analysis for the k-Hamming distance problem is the spectral norm bound. While it is possible to show an $\Omega(k \log \frac{\log 1/\delta}{k} + k)$ lower bound for the log-spectral-norm of the corresponding k-threshold function via a duality argument, anything beyond this bound proved difficult. Through some experimentation with linear program solvers, we believe that this lower bound is tight for the log-spectral-norm of the k-threshold function. In fact, in an upcoming work with de Wolf and Hoyer, we show that there are quantum communication protocols for the k-Hamming distance problem with just $O(k)$ qubits of communication and exponentially small error probability. This implies that the log-spectral-norm of k-threshold function is $O(k)$, even though this upper bound does not match the log-spectral-norm lower bound in the δ parameter.

Given that various powerful lower bound techniques in communication complexity are all stuck at $\Omega(k)$, and all upper bound attempts are stuck at $O(k \log k)$, it was far from clear what the correct bound for the communication complexity of the k-Hamming distance should be. In our work (recently published in [79]), we manage to find a connection between the k-Hamming distance problem and a very natural statement about how heat behaves in time. It turns out, similar questions were considered earlier in the literature. In 1966, Blakley and Dixon [18] studied a special case of our main inequality, defined and proven in Chapter 4 of this thesis. In 1982, Erdős and Simonovits [34] studied a slightly more restrictive case of the inequality of [18] and posed it as a conjecture. In Chapter 4 we prove that the inequality of Blakley and Dixon holds true even in a more general form, which in turn affirmatively answers the conjecture of Erdős and Simonovits. Furthermore in the second part of Chapter 4, we prove a substantial refinement of this inequality, which in essence argues that the probability mass contained in a certain region of a Markov chain must be nearly convex as a function of time. In Chapter 5 we prove that the aforementioned inequality about Markov chains is precisely what is needed to get an $\Omega(k \log k)$ lower bound to the k-Hamming distance problem, thereby establishing the complexity of k-Hamming distance problem and the tight bounds for the property testing of k-linearity.

1.1 Our techniques.

An important challenge in proving the inequality of Blakley and Dixon [18] is the following. Our generalized version of the inequality states that for vectors u, v with nonnegative entries and a symmetric matrix S with nonnegative entries, we have $\left\langle v, S^k u \right\rangle^{1/k} \geq \langle v, S^t u \rangle^{1/t}$ for $k > t$ integers of the same parity. We remark that the nonnegativity requirement on u, v is crucial and the inequality fails for general u, v. In fact, one way to deduce this is to take our recent $O(k)$ qubits quantum communication protocol for the k-Hamming distance problem and obtain u, v from the log-spectral-norm relaxation of the protocol. The key technical challenge in proving this inequality seems to be devising a proof that is sensitive to u, v being in the the positive orthant. In [17], Blakley and Roy showed that the special case $\left\langle v, S^k u \right\rangle^{1/k} \geq \langle v, Su \rangle$ is true by taking a geometric view and working with the positive orthant directly. For $t > 1$, working with the positive orthant directly appears to be very difficult. Instead, we break apart from the geometric view and into a probabilistic perspective to make positivity an inherent part of the technique. We cast this inequality as a probabilistic statement and analyze it through the lens of information theory by viewing it as the success probability of discrete time stochastic process.

1.2 Future directions

A surprising part of our $\Omega(k \log k)$ lower bound for the k-Hamming distance problem is that the proof does not use at all the tensor structure of the hypercube. The only two properties of the hypercube that we use are that 1) it is an undirected graph 2) if we perform the standard discrete time random walk from a vertex u, at time $t \ll \sqrt{n}$, we end up with a vertex v satisfying $\mathrm{Ham}(u, v) = t$ with probability at least, say 0.9.

We believe a similar proof the Gap Hamming Distance problem may be possible. Current lower bound proofs for the Gap Hamming Distance depend heavily on the product structure of the hypercube through concentration of measure phenomenon. We discuss this possibility in Conjecture 5.7.1.

1.3 Outline

In this thesis we study the communication complexity of the k-disjointness and k-Hamming distance problems. In Chapter 5, we give our k-disjointness protocol, and show a matching lower bound after we take a brief detour into combinatorics to develop our isoperimetric inequality used in the lower bound proof.

Our lower bound for the k-Hamming distance, given in Chapter 5, requires us to understand how heat behaves in discrete time. We develop this theory in Chapter 4. The notation we use and some formal preliminaries are given in Chapter 2.

Chapter 2
PRELIMINARIES

In this chapter we list the notation we use, conventions we adopt and formally and define formally and in more detail the computational models that we work with in this thesis.

2.1 Notation

We denote by $[n]$ the set $\{1, 2, \ldots, n\}$. Throughout this thesis, we take exp and log functions to the base 2. For exponentials and logarithms in other bases such as b, we write $\exp_b$ and $\log_b$. We adopt the convention $\ln := \log_e$ where $e = 2.718\ldots$ is the limiting value of $(1+1/n)^n$ as $n \to \infty$. We will also use the iterated versions of these functions:

$$\begin{aligned} &\log^{(0)} x := x, && \exp^{(0)} x := x, \\ &\log^{(r)} x := \log\left(\log^{(r-1)} x\right), && \exp^{(r)} x := \exp\left(\exp^{(r-1)} x\right) \quad \text{for } r \geq 1. \end{aligned}$$

Moreover we define the $\log^* x$ to be the smallest integer r for which $\log^{(r)} x < 2$. For instance, we have $\log^* 16 = 3$ and $\log^* 2^{16} = 4$. The function $\log^*$ is conventionally called the *iterated logarithm*, which we adopt. To differentiate, we call the function $\log^{(r)}$ the *r-iterated logarithm*.

2.2 Random variables and distributions

Let Ω be a countable set. For a function $\mu\colon \Omega \to \mathbb{R}_+$ and a set $\Psi \subseteq \Omega$, we use the shorthand

$$\mu(\Psi) := \sum_{x \in \Psi} \mu(x).$$

A function $\mu\colon \Omega \to \mathbb{R}_+$ is said to be a distribution on Ω if $\mu(\Omega) = 1$ and a subdistribution if $\mu(\Omega) \leq 1$. For a function μ on Ω, we define

$$\operatorname{supp}(\mu) := \{x \in \Omega \mid \mu(x) > 0\}.$$

For two distributions $\mu\colon \Omega_1 \to \mathbb{R}_+$ and $\nu\colon \Omega_2 \to \mathbb{R}_+$, let us denote by $\mu\nu$ the distribution on $\Omega_1 \times \Omega_2$ given by $(\mu\nu)(x_1, x_2) = \mu(x_1)\nu(x_2)$.

For a discrete random variable X, we denote by $\operatorname{dist}(X)$ the distribution function of X and we define $\operatorname{supp}(X) := \operatorname{supp}(\operatorname{dist}(X))$. If X is so that $\operatorname{dist}(X)\colon \Omega \to \mathbb{R}_+$, then we say that X has sample space Ω. Two random variables X and Y are said to be independent if $\operatorname{dist}(XY) = \operatorname{dist}(X)\operatorname{dist}(Y)$.

Lemma 2.2.1 (Jensen [54], Formula (5))**.** *Let X be a real-valued random variable and f be a convex function. We have $\mathbb{E}\left[f(X)\right] \geq f(\mathbb{E}[X])$. When f is strictly convex, the inequality holds with equality if and only if X is constant with probability 1.*

2.3 Facts from information theory

In this section we review the definitions and facts we use from information theory. Let μ and ν be two nonnegative functions on Ω. The Kullback-Leibler divergence [85, 59] of μ from ν, denoted $\mathbf{D}(\mu \,\|\, \nu)$, is defined by

$$\mathbf{D}(\mu \,\|\, \nu) := \sum_{x \in \Omega} \mu(x) \log \frac{\mu(x)}{\nu(x)}\,. \tag{2.3.1}$$

Here, if $\mu(x) = 0$ for some x, then its contribution to the summation is taken as 0, even when $\nu(x) = 0$. The divergence is undefined if there is an $x \in \Omega$ such that $\mu(x) > 0$ and $\nu(x) = 0$. It can be shown that if the related series converges for the right hand side of Eq. (2.3.1), it converges absolutely, which justifies leaving the summation order unspecified. A fundamental property of $\mathbf{D}(\cdot \,\|\, \cdot)$ is that the divergence of a distribution from a subdistribution is always nonnegative.

Lemma 2.3.1 (Gibbs [38], Theorem VIII)**.** *Let $\mu, \nu\colon \Omega \to \mathbb{R}$ be such that μ is a distribution and ν is a subdistribution. We have $\mathbf{D}(\mu \,\|\, \nu) \geq 0$ with equality if and only if $\mu = \nu$.*

Lemma 2.3.2 (Kullback and Leibler [59], Lemma 3.2)**.** *Let $\mu, \nu\colon \Omega \to \mathbb{R}_+$ be so that μ is a distribution on Ω and* $\operatorname{supp}(\mu) = \Psi \subseteq \Omega$*. We have*

$$\mathbf{D}(\mu \,\|\, \nu) \geq -\log \nu(\Psi)$$

with equality if and only if $\mu(x) = \nu(x)/\nu(\Psi)$ *for* $x \in \Psi$ *and* $\mu(x) = 0$ *for* $x \notin \Psi$*.*

Proof. By Eq. (2.3.1) we write

$$\mathbf{D}(\mu \,\|\, \nu) = -\sum_{x \in \Psi} \mu(x) \log \frac{\nu(x)}{\mu(x)} \geq -\log \sum_{x \in \Psi} \nu(x) = -\log \nu(\Psi)\,,$$

where the inequality follows from Lemma 2.2.1 and concavity of $z \mapsto \log z$ on $\mathbb{R}_+$. If $\mu(x) = \nu(x)/\nu(\Psi)$ for $x \in \Psi$, we have $\mathbf{D}(\mu \,\|\, \nu) = -\log \nu(\Psi)$ by direct computation. Otherwise $\mathbf{D}(\mu \,\|\, \nu) > -\log \nu(\Psi)$ by strict concavity of $z \mapsto \log z$. □

We extend the divergence notation $\mathbf{D}(\cdot \,\|\, \cdot)$ to apply to random variables as follows. Let X, Y be discrete random variables on the same sample space Ω. Define

$$\mathbf{D}(X \,\|\, Y) := \mathbf{D}(\mathrm{dist}(X) \,\|\, \mathrm{dist}(Y))\,. \tag{2.3.2}$$

With this notation in hand, we are ready to define the conditional divergence. Let X_1X_2 and Y_1Y_2 be random variables defined on the sample space $\Omega_1 \times \Omega_2$. The divergence of $X_1 \,|\, X_2$ from $Y_1 \,|\, Y_2$ is defined by

$$\mathbf{D}(X_1 \,|\, X_2 \,\|\, Y_1 \,|\, Y_2) := \mathop{\mathbb{E}}_{x_2 \sim X_2} \mathbf{D}(X_1 \,|\, X_2 = x_2 \,\|\, Y_1 \,|\, Y_2 = x_2)\,. \tag{2.3.3}$$

Here, for each $x_2 \in \mathrm{supp}(X_2)$, $X_1 \,|\, X_2 = x_2$ and $Y_1 \,|\, Y_2 = x_2$ are random variables on the sample space Ω_1 obtained from, respectively X_1X_2 and Y_1Y_2, by conditioning on the second coordinate equaling x_2.

Lemma 2.3.3 (e.g., [31])**.** *Let* X_1X_2 *and* Y_1Y_2 *be random variables, both on the sample space* $\Omega_1 \times \Omega_2$*. We have*

$$\mathbf{D}(X_1X_2 \,\|\, Y_1Y_2) = \mathbf{D}(X_1 \,\|\, Y_1) + \mathbf{D}(X_2 \,|\, X_1 \,\|\, Y_2 \,|\, Y_1)\,.$$

Proof. Let $\mu, \nu \colon \Omega_1 \times \Omega_2 \to \mathbb{R}_+$ be the distributions of respectively X_1X_2 and Y_1Y_2. Using the shorthands $\mu(\Omega_1, x_2) := \sum_{x_1 \in \Omega_1} \mu(x_1, x_2)$ and $\nu(\Omega_1, x_2) := \sum_{x_1 \in \Omega_1} \nu(x_1, x_2)$, we write

$$\begin{aligned}
\mathbf{D}(X_1 \,|\, X_2 \,\|\, Y_1 \,|\, Y_2) &= \sum_{x_2 \in \Omega_2} \mu(\Omega_1, x_2) \sum_{x_1 \in \Omega_1} \frac{\mu(x_1, x_2)}{\mu(\Omega_1, x_2)} \log \frac{\mu(x_1, x_2)\nu(\Omega_1, x_2)}{\nu(x_1, x_2)\mu(\Omega_1, x_2)} \\
&= \sum_{x_1, x_2} \mu(x_1, x_2) \log \frac{\mu(x_1, x_2)}{\nu(x_1, x_2)} + \sum_{x_2} \mu(\Omega_1, x_2) \log \frac{\nu(\Omega_1, x_2)}{\mu(\Omega_1, x_2)}
\end{aligned}$$

by splitting the terms inside the logarithm. Using Eq. (2.3.1) together with Eq. (2.3.2) we conclude

$$\mathbf{D}(X_1 \,|\, X_2 \,\|\, Y_1 \,|\, Y_2) = \mathbf{D}(X_1X_2 \,\|\, Y_1Y_2) - \mathbf{D}(X_2 \,\|\, Y_2)\,.$$

Rearranging, we obtain the statement of the lemma. □

The next lemma establishes that the Kullback-Leibler divergence is jointly convex in its parameters. This fact is also called the data processing inequality for Kullback-Leibler divergence.

Lemma 2.3.4. *Let $\mu_1, \mu_2 \colon \Omega_1 \to \mathbb{R}_+$ be distributions supported on Ω_1 and let $\nu_1, \nu_2 \colon \Omega_2 \to \mathbb{R}_+$ be distributions supported on Ω_2. For any $a \in [0,1]$, we have*

$$\mathbf{D}(a\mu_1 + (1-a)\mu_2 \,\|\, a\nu_1 + (1-a)\nu_2) \le a\mathbf{D}(\mu_1 \,\|\, \nu_1) + (1-a)\mathbf{D}(\mu_2 \,\|\, \nu_2)$$

Further for two reals $p, q \in [0,1]$, we use the shorthand notation $\mathbf{D}_2(p \,\|\, q)$ to denote the divergence of the random variable P from Q where P, Q are Bernoulli random variables with expectation respectively p and q.

2.3.1 Mutual information and Shannon entropy

Let X and Y be jointly distributed random variables. The mutual information of X and Y, denoted $\mathbf{I}(X : Y)$, is defined as

$$\mathbf{I}(X : Y) := \mathbf{D}(\mathrm{dist}(X,Y) \,\|\, \mathrm{dist}(X)\,\mathrm{dist}(Y))\,. \tag{2.3.4}$$

The mutual information of a random variable with itself, i.e., the quantity $\mathbf{I}(X : X)$ is called the Shannon entropy of X and denoted by $\mathrm{H}(X)$. If $X \in [t]^n$ and $L \subseteq [n]$, then the projection of X to the coordinates in L is denoted by X_L. Namely, X_L is obtained from $X = (X_1, \ldots, X_n)$ by keeping only the coordinates X_i with $i \in L$. The following lemma of Chung et al. [28] relates the entropy of a variable to the entropy of its projections.

Lemma 2.3.5 (Chung et al. [28])**.** *Let $\mathrm{supp}(X) \subseteq [t]^n$. We have $\frac{l}{n}\mathrm{H}(X) \le \mathbb{E}_L[\mathrm{H}(X_L)]$, where the expectation is taken for a uniform random l-subset L of $[n]$.*

2.4 Concentration bounds

Let $X_1, \ldots, X_n$ be random variables such that $\mathbb{E}[X_i] = \epsilon$ for some $0 \leq \epsilon \leq 1$. Define $X = X_1 + \cdots + X_n$. By linearity of expectation we have $\mathbb{E}[X] = \epsilon n$. If each X_i is chosen independently, due to the concentration of measure phenomenon, it is well understood that X takes values close to its expectation with very high probability. The classical results of Chernoff [27] and Hoeffding [45] give quantitative and intuitive bounds on the deviation probability.

Theorem 2.4.1 (Chernoff [27])**.** *Let $X = X_1 + \cdots + X_n$, where X_i for $i \in [n]$ are independent binary random variables with expectation ϵ. Then for any $\epsilon \leq \gamma \leq 1$ we have*

$$\Pr\left[X \geq \gamma n\right] \leq \exp\left(-n\mathbf{D}_2(\gamma \,\|\, \epsilon)\right).$$

Proof. Let E be the event that $X \geq \gamma n$. Let μ be the Bernoulli distribution that equals 1 with probability ϵ. We have

$$\begin{aligned} -\log \Pr[E] &= \mathbf{D}(\mathrm{dist}(X_1 \ldots X_n \,|\, E) \,\|\, \mu^n) && \text{(by Lemma 2.3.2)} \\ &\geq \sum_{i=1}^{n} \mathbf{D}(\mathrm{dist}(X_i \,|\, E) \,\|\, \mu) && \text{(by Lemma 2.3.3)} \\ &\geq n\mathbf{D}_2(\gamma \,\|\, \epsilon) && \text{(as } \mathbf{D}_2(\delta \,\|\, \epsilon) \geq \mathbf{D}_2(\gamma \,\|\, \epsilon) \text{ for } \epsilon \leq \gamma \leq \delta) \end{aligned}$$

Hence, $\Pr[E] \leq \exp\left(-n\,\mathbf{D}_2(\gamma \,\|\, \epsilon)\right)$ as required. □

2.5 Communication complexity and protocols

The model. In the two party communication complexity model we have two parties called respectively Alice and Bob who are required to evaluate a function $f \colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ (known to both of them) on some input (x, y) where x is revealed to Alice only and y is revealed to Bob only.

In the randomized variant of this model, which is what we solely study in this thesis, the players also have access to a shared random source. Without loss of generality, the random source can be taken as an infinite read-only string of bits, chosen uniformly at random. The players, having received their inputs x and y respectively, and with access to the shared random source, engage in a dialogue by alternately sending each other messages in rounds, in order to compute $f(x, y)$. Here, a message is a bit string of arbitrary length.

Protocols. A *communication protocol* specifies, for each round, which player's turn it is to speak and what message should be sent and whether the protocol terminates with some answer. The protocol specifies the message to be sent through a function mapping the random string, the current players input and all the messages the current player received so far to a bit string which is to be sent to the other player. This ensures that the message sent by the players, say Alice for illustration, can depend on only information known to her at the time: her own input, the shared random source and the message she received in the previous rounds. Some message are marked as answers; instead of being sent to the other player, these message are to be announced as the output of the protocol, after which the protocol terminates. We say that a communication protocol for a function $f\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ has at most δ-error if for any input pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$ with probability at least $1 - \delta$ the output of the protocol equals $f(x, y)$, where the probability is over the random choices that come from the shared random source.

In an r-round protocol there are at most r messages (excluding the output message) sent for any input and any configuration of the shared random source. To illustrate the way the number of messages is counted, consider the following protocol. Alice sends a single message to Bob and in return Bob replies with another message after which Alice announces the answer of the protocol. This protocol has two rounds. The complexity of a protocol is defined to be the the maximum, over all input pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and all choices of the random source, of the total number of bits sent by the two parties.

Let P be a protocol for a function $f\colon \mathcal{X} \times \mathcal{Y}$. We denote by $P(x, y, r)$ the output of the protocol on input $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and when the shared random string is fixed to r. We denote by $P(x, y)$ the random variable denoting the output of the protocol on inputs (x, y). The transcript of the protocol P, denoted $\Pi_P(x, y, r)$, is the concatenation of all the messages sent by the players (excluding the answer of the protocol) on inputs $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and when the shared random string is r. Likewise, $\Pi_P(x, y)$ denotes the random variable entailing the communication took place the two players when the random source is not fixed. We denote by $|\Pi_P|$ the length of the transcript, in bits.

Communication complexity The δ-error randomized communication complexity of a function $f \colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, denoted $R_\delta(f)$, is the minimum over all δ-error protocols for f, of the complexity of the protocol. We also define the r-round δ-error randomized communication complexity of a function f, denoted $R^r_\delta(f)$, wherein the minimization this time is over all r-round protocols for f with at most δ-error. Let us summarize the two key definitions of this section in notation.

For a function $f \colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$,

$$R_\delta(f) := \min_P \max_{x\in\mathcal{X}, y\in\mathcal{Y}, r} |\Pi_P(x, y, r)| \tag{2.5.1}$$

$$R^r_\delta(f) := \min_P \max_{x\in\mathcal{X}, y\in\mathcal{Y}, r} |\Pi_P(x, y, r)| \tag{2.5.2}$$

where P ranges over all δ-error protocols for f in Eq. (2.5.1) and P ranges over all δ-error r-round protocols for f in Eq. (2.5.2).

2.5.1 The k-disjointness problem

In the k-disjointness problem (also known as the sparse set disjointness or small set disjointness), each of the players receives a subset of $[n]$ of size at most k. Call the set Alice receives S and the set Bob receives T. The goal of the players is to determine whether their sets S and T intersect. We denote this problem by Disj^n_k.

One useful way to think about Disj^n_k problem is to associate the sets players receive with their characteristic vectors. This way, Alice and Bob each receive an n bit string with at most k ones and their goal is to determine if there is a coordinate in which both of their strings is one. Notice that the characteristic vectors have a common 1 if and only if their Hamming distance is at most $2k-2$.

2.5.2 The k-Hamming distance problem

In the k-Hamming distance problem the players get n bit strings, say $x, y \in \{0,1\}^n$, respectively with the promise that $\mathrm{Ham}(x,y) \leq k$. Their goal is to determine whether $\mathrm{Ham}(x,y)$ under this promise. Note that by the previous paragraph, the k-disjointness problem is a special case of the $2k$-Hamming distance problem, therefore up to constant factors, the communication complexity of Disj_k^n is upper bounded by Ham_k^n.

Chapter 3

THE K-DISJOINTNESS PROBLEM

In the set disjointness problem two players, traditionally called Alice and Bob, receive a subset of $[n] := \{1, \ldots, n\}$ each and the goal of the players is to determine whether their sets intersects or not. Since each player knows only one of the sets, this goal is achievable only if some communication takes place between the players. The communication complexity of the set disjointness problem is then defined to be the minimum amount of communication needed by the players to determine whether their sets intersect, with high probability, for any possible sets they may receive.

Set disjointness is perhaps the most studied problem in communication complexity. In the most common version the players receive subsets of $[n]$ with no restriction on the sizes of the sets. The primary question is whether the players can significantly improve on the trivial deterministic protocol, wherein the first player sends the entire input to the other player, thereby communicating n bits. The first lower bound on the randomized complexity of this problem was given in [5] by Babai, Frankl and Simon, who showed that any ϵ-error protocol for disjointness must communicate at least $\Omega(\sqrt{n})$ bits. The tight bound of $\Omega(n)$-bits was first given by Kalyanasundaram and Schnitger [55] and was later simplified by Razborov [77] and Bar-Yossef, Jayram, Kumar and Sivakumar [7].

In the sparse set disjointness problem Disj_k^n, the sets given to the players are guaranteed to have at most k elements. The deterministic communication complexity of this problem is well understood. The trivial protocol, where Alice sends her entire input to Bob solves the problem in one round using $O(k\log(n/k) + k)$ bits. On the other hand, an $\Omega(k\log(n/k) + k)$ bit total communication lower bound can be shown even for protocols with an arbitrary number of rounds, say using the rank method; see [60], page 175.

The randomized complexity of the problem is far more subtle. The result of Kalyanasundaram and Schnitger cited above immediately imply an $\Omega(k)$ lower bound for this version of the problem. The folklore 1-round protocol solves the problem using $O(k \log k)$ bits, wherein Alice sends $O(\log k)$-bit hashes for each element of her set. Håstad and Wigderson [48] gave a protocol that matches the $\Omega(k)$ lower bound mentioned above. Their $O(k)$-bit randomized protocol runs in $O(\log k)$-rounds and errs with a small constant probability. In **??**, we improve this protocol to run in $\log^* k$ rounds, still with $O(k)$ total communication, but with exponentially small error in k. We also present an r-round protocol for any $r < \log^* k$ with total communication $O(k \log^{(r)} k)$ and error probability well below $1/k$; see Theorem 5.1.1. (Here $\log^{(r)}$ denotes the iterated logarithm function, see Chapter 2.) As the exists-equal problem with parameters t and n (see below) is a special case of Disj_n^{tn}, our lower bounds for the exists-equal problem (see below) show that complexity of this algorithm is optimal for any number $r \leq \log^* k$ of rounds, even if we allow the much larger error probability of $1/3$. Buhrman et al. [23] and Woodruff [86] (as presented in [75]) show an $\Omega(k \log k)$ lower bound for 1-round complexity of Disj_k^n by a reduction from the indexing problem (this reduction was also mentioned in [**?**]). We note that these lower bounds do not apply to the exists-equal problem, as the input distribution they use generates instances inherently specific to the disjointness problem; furthermore this distribution admits a $O(\log k)$ protocol in two rounds.

In the equality problem Alice and Bob receive elements x and y of a universe $[t]$ and they have to decide whether $x = y$. We define the two player communication game exists-equal with parameters t and n as follows. Each player is given an n-dimensional vector from $[t]^n$, namely x and y. The value of the game is one if there exists a coordinate $i \in [n]$ such that $x_i = y_i$, zero otherwise. Clearly, this problem is the OR of n independent instances of the equality problem.

The direct sum problem in communication complexity is the study of whether n instances of a problem can be solved using less than n times the communication required for a single instance of the problem. This question has been studied extensively for specific communication problems as well as some class of problems [25, 50, 51, 10, 36, 49, 42, 8]. The so called direct sum approach is a very powerful tool to show lower bounds for communication games. In this approach, one expresses the problem at hand, say as the OR of n instances of a simpler function and the lower bound is obtained by combining a lower bound for the simpler problem with a direct sum argument. For instance, the two-player and multi-player disjointness bounds of [7], the lopsided set disjointness bounds [76], and the lower bounds for several communication problems that arise from streaming algorithms [52, 63] are a few examples of results that follow this approach.

Exists-equal with parameters t and n is a special case of Disj_n^{tn}, so our protocols in Section 3.4 solve exists-equal. We show that when $t = \Omega(n)$ these protocols are optimal, namely every r-round randomized protocol ($r \leq \log^* n$) with at most $1/3$ error error probability needs to send at least one message of size $\Omega(n \log^{(r)} n)$ bits. See Theorem 4.0.4. Our result shows that computing the OR of n instances of the equality problem requires *strictly more* than n times the communication required to solve a single instance of the equality problem when the number of rounds is smaller than $\log^* n - O(1)$. Recall that the equality problem admits an ϵ-error $\log(1/\epsilon)$-bit one-round protocol in the common random source model.

For $r = 1$, our result implies that to compute the OR of n instances of the equality problem with *constant probability*, no protocol can do better than solving each instance of the equality problem with *high probability* so that the union bound can be applied when taking the OR of the computed results. The single round case of our lower bound also generalizes the $\Omega(n \log n)$ lower bound of Molinaro et al. [67] for the one round communication problem, where the players have to find all the answers of n equality problems, outputting an n bit string.

3.1 Lower bound techniques

We obtain our general lower bound via a round elimination argument. In such an argument one assumes the existence of a protocol P that solves a communication problem, say f, in r rounds. By suitably modifying the inner working of P, one obtains another protocol P' with $r - 1$ rounds, which typically solves smaller instances of f or has larger error than P.

Iterating this process, one obtains a protocol with zero rounds. If the protocol we obtain solves non-trivial instances of f with good probability, we conclude that we have arrived at a contradiction, therefore the protocol we started with, P, cannot exist. Although round elimination arguments have been used for a long time, our round elimination lemma is the first to prove a *super-linear* communication lower bound in the number of primitive problems involved, obtaining which requires new and interesting ideas.

At the heart of the general round elimination lemma is a new isoperimetric inequality on the discrete cube $[t]^n$ equipped with the Hamming distance. We present this result, Theorem 3.6.5, in Section 3.6. The first isoperimetric inequality on this metric space was proven by Lindsey in [61], where the subsets of $[t]^n$ of a certain size with the so called minimum induced-edge number were characterized. This result was rediscovered in [58] and [29] as well. See [4] for a generalization of this inequality to universes which are n-dimensional boxes with arbitrary side lengths. In [19], Bollobás et al. study isoperimetric inequalities on $[t]^n$ endowed with the ℓ_1 distance. For the purposes of our proof we need to find sets S that minimize a substantially more complicated measure. This measure also captures how spread out S is and can be described roughly as the average over points $x \in [t]^n$ of the logarithm of the number of points in the intersection of S and a Hamming ball around x.

3.2 Related work

In [66], a round elimination lemma was given, which applies to a class of problems with certain self-reducibility properties. The lemma is then is used to get lower bounds for various problems including the greater-than and the predecessor problems. This result was later tightened in [82] to get better bounds for the aforementioned problems. Different round elimination arguments were also used in [56, 40, 71, 65, 33, 9] for various communication complexity lower bounds and most recently in [20] and [22] for obtaining lower bounds for the gapped Hamming distance problem.

In parallel and independent of the present form of this paper Brody et al. [21] have also established an $\Omega(n \log^{(r)} n)$ lower bound for the r-round communication complexity of the exists-equal problem with parameter n. Their result applies for protocols with a polynomially small error probability like $1/n$. This stronger assumption on the protocol allows for simpler proof techniques, namely the information complexity based direct sum technique developed in several papers including [1, 25], but it is not enough to create an example where solving the OR of n communication problems requires more than n times the communication of solving a single instance. Indeed, even in the shared random source model one needs $\log n$ bits of communication (independent of the number of rounds) to achieve $1/n$ error in a single equality problem.

3.3 Structure of the present chapter

The general round elimination presented in Section 3.7 is technically involved, but the lower bound on the one-round protocols can also be obtained in a more elementary way. As the one round case exhibits the most dramatic super-linear increase in the communication cost and also generalizes the lower bound in [67], we include this combinatorial argument separately in Section 3.5.

We start in Section 3.4 with our protocols for the sparse set disjointness. Note that the exists-equal problem is a special case of sparse set disjointness, so our protocols work also for the exists-equal problem. In the rest of the paper we establish matching lower bounds showing that the complexity of our protocols are within a constant factor to optimal for both the exists-equal and the sparse set disjointness problems, and for any number of rounds. In Section 3.5 we give an elementary proof for the case of single round protocols. In Section 3.6 we develop our isoperimetric inequality and in Section 3.7 we use it in our round elimination proof to get the lower bound for multiple round protocols. Finally in Section 3.8 we point toward possible extensions of the result of this chapter.

3.4 The upper bound for Disj_k^n

Recall that in the communication problem Disj_k^m, each of the two players is given a subset of $[m]$ of size at most k and they communicate in order to determine whether their sets are disjoint or not. In early 1990s, Håstad and Wigderson [73, 48, 47] discovered a randomized protocol that solves the Disj_k^m problem with $O(k)$ bits of communication and has constant one-sided error probability. The protocol takes $O(\log k)$ rounds. Let us briefly review this protocol as this is the starting point of our protocol.

3.4.1 The Håstad Wigderson protocol

Let $S, T \subseteq [m]$ be the inputs of Alice and Bob. Observe that if they find a set Z satisfying $S \subseteq Z \subseteq [m]$, then Bob can replace his input T with $T' = T \cap Z$ as $T' \cap S = T \cap S$. The main observation is that if S and T are disjoint, then a random set $Z \supseteq S$ will intersect T in a uniform random subset, so one can expect $|T'| \approx |T|/2$. In the Håstad-Wigderson protocol the players alternate in finding a random set that contains the current input of one of them, effectively halving the other player's input. If in this process the input of one of the players becomes empty, they know the original inputs were disjoint. If, however, the sizes of their inputs do not show the expected exponential decrease in time, then they declare that their inputs intersect. This introduces a small one sided error. Note that one of the two outcomes happens in $O(\log k)$ rounds. An important observation is that Alice can describe a random set $Z \supseteq S$ to Bob using an expected $O(|S|)$ bits by making use of the joint random source. This makes the total communication $O(k)$.

3.4.2 Our protocol

In our protocol proving the next theorem, we do almost the same, but we choose the random sets $Z \supseteq S$ not uniformly, but from a biased distribution favoring ever smaller sets. This makes the size of the input sets of the players decrease much more rapidly, but describing the random set Z to the other player becomes more costly. By carefully balancing the parameters we optimize for the total communication given any number of rounds. When the number of rounds reaches $\log^* k - O(1)$ the communication reaches its minimum of $O(k)$ and the error becomes exponentially small.

Theorem 3.4.1. *For any $r \leq \log^* k$, there is an r-round probabilistic protocol for* Disj_k^m *with $O(k \log^{(r)} k)$ bits total communication. There is no error for intersecting input sets, and the probability of error for disjoint sets can be made $O(1/\exp^{(r)}(c \log^{(r)} k) + \exp(-\sqrt{k})) \ll 1/k$ for any constant $c > 1$.*

For $r = \log^* k - O(1)$ rounds this means an $O(k)$-bit protocol with error probability $O(\exp(-\sqrt{k}))$.

Proof. We start with the description of the protocol. Let S_0 and S_1 be the input sets of Alice and Bob, respectively. For $1 \leq i \leq r$, i even Alice sends a message describing a set $Z_i \supset S_i$ based on her "current input" S_i and Bob updates his "current input" S_{i-1} to $S_{i+1} := S_{i-1} \cap Z_i$. In odd numbered rounds the same happens with the role of Alice and Bob reversed. We depart from the Håstad-Wigderson protocol in the way we choose the sets Z_i: Using the shared random source the players generate l_i random subsets of $[m]$ containing each element of $[m]$ independently and with probability p_i. We will set these parameters later. The set Z_i is chosen to be the first such set containing S_i. Alice or Bob (depending on the parity of i) sends the index of this set or ends the protocol by sending a special error signal if none of the generated sets contain S_i. The protocol ends with declaring the inputs disjoint if the error signal is never sent and we have $S_{r+1} = \emptyset$. In all other cases the protocol ends with declaring "not disjoint".

This finishes the description of the protocol except for the setting of the parameters. Note that the error of the protocol is one-sided: $S_0 \cap S_1 = S_i \cap S_{i+1}$ for $i \leq r$, so intersecting inputs cannot yield $S_{r+1} = \emptyset$.

We set the parameters (including k_i used in the analysis) as follows:

$$\begin{aligned}
u &= (c+1)\log^{(r)} k, \\
p_i &= \frac{1}{\exp^{(i)} u} && \text{for } 1 \le i \le r, \\
l_1 &= k\exp(ku), \\
l_i &= k2^{k/2^{i-4}} && \text{for } 2 \le i \le r, \\
k_0 &= k_1 = k, \\
k_i &= \frac{k}{2^{i-4}\exp^{(i-1)} u} && \text{for } 2 \le i \le r, \\
k_{r+1} &= 0.
\end{aligned}$$

The message sent in round $i > 1$ has length $\lceil \log(l_i+1) \rceil < k/2^{i-4} + \log k + 1$, thus the total communication in all rounds but the first is $O(k)$. The length of the first message is $\lceil \log(l_1+1) \rceil \le ku + \log k + 1$. The total communication is $O(ku) = O(ck\log^{(r)} k)$ as claimed (recall that c is a constant).

Let us assume the input pair is disjoint. To estimate the error probability we call round i *bad* if an error message is sent or a set S_{i+1} is created with $|S_{i+1}| > k_{i+1}$. If no bad round exists we have $S_{r+1} = \emptyset$ and the protocol makes no error. In what follows we bound the probability that round i is bad assuming the previous rounds are not bad and therefore having $|S_j| \le k_j$ for $0 \le j \le i$.

The probability that a random set constructed in round i contains S_i is $p_i^{-|S_i|} \ge p_i^{-k_i}$. The probability that none of the l_i sets contains S_i and thus an error message is sent is therefore at most $(1 - p_i^{k_i})^{l_i} < e^{-k}$.

If no error occurs in the first bad round i, then $|S_{i+1}| > k_{i+1}$. Note that in this case $S_{i+1} = S_{i-1} \cap Z_i$ contains each element of S_{i-1} independently and with probability p_i. This is because the choice of Z_i was based on it containing S_i, so it was independent of its intersection with S_{i-1} (recall that $S_i \cap S_{i-1} = S_1 \cap S_0 = \emptyset$). For $i < r$ we use the Chernoff bound. The expected size of S_{i+1} is $|S_{i-1}|p_i \le k_{i-1}p_i \le k_{i+1}/2$, thus the probability of $|S_{i+1}| > k_{i+1}$ is at most $2^{-k_{i+1}/4}$. Finally for the last round $i = r$ we use the simpler estimate $p_r k_{r-1} \le k/\exp^{(r)} u$

for $|S_{r+1}| > k_{r+1} = 0$.

Summing over all these estimates we obtain the following error bound for our protocol:

$$\Pr[\text{error}] \leq re^{-k} + \frac{k}{\exp^{(r)} u} + \sum_{i=2}^{r} 2^{-k_i/4}.$$

In case $k_r \geq 4\sqrt{k}$ this error estimate proves the theorem. In case $k_r < 4\sqrt{k}$ we need to make a minor adjustments in the setting of our parameters. We take j to be the smallest value with $k_j < 4\sqrt{k}$, modify the parameters for round j and stop the protocol after this round declaring "disjoint" if $S_{j+1} = \emptyset$ and "intersecting" otherwise. The new parameters for round j are $k'_j = 4\sqrt{k}$, $p'_j = 2^{-2\sqrt{k}}$, $l'_j = k2^{8k}$. This new setting of the parameters makes the message in the last round linear in k, while both the probability that round $j-1$ is bad because it makes $|S_j| > k'_j$, or the probability that round j is bad for any reason (error message or $S_{j+1} \neq \emptyset$) is $O(2^{-\sqrt{k}})$. This finishes the analysis of our protocol. □

3.5 Lower bound for single round protocols

In this section we give an combinatorial proof that any single round randomized protocol for the exists-equal problem with parameters n and $t = 4n$ has complexity $\Omega(n \log n)$ if its error probability is at most 1/3. As pointed out in the Introduction, to our knowledge this is the fist established case when solving the OR of n instances of a communication problem requires strictly more than n times the complexity needed to solve a single such instance.

We start with with a simple and standard reduction from the randomized protocol to the deterministic one and further to a large set of inputs that makes the first (and in this case only) message fixed. These steps are also used in the general round elimination argument therefore we state them in general form.

Let $\epsilon > 0$ be a small constant and let P be an 1/3-error randomized protocol for the exists-equal problem with parameters n and $t = 4n$. We repeat the protocol P in parallel taking the majority output, so that the number of rounds does not change, the length of the messages is multiplied by a constant and the error probability decreases below ϵ. Now we fix the coins of of this ϵ-error protocol in a way to make the resulting deterministic protocol err on at most ϵ fraction of the possible inputs. Denote the deterministic protocol we obtain by Q.

Lemma 3.5.1. *Let Q be a deterministic protocol for the* EE_n *problem that makes at most ϵ error on the uniform distribution. Assume Alice sends the first message of length c. There exists an $S \subset [t]^n$ of size $\mu(S) = 2^{-c-1}$ such that the first message of Alice is fixed when $x \in S$ and we have* $\Pr_{y\sim\mu}[Q(x,y) \neq \mathrm{EE}(x,y)] \leq 2\epsilon$ *for all $x \in S$.*

Proof. Note that the quantity $e(x) = \Pr_{y\sim\mu}[Q(x,y) \neq \mathrm{EE}(x,y)]$, averaged over all x, is the error probability of Q on the uniform input, hence is at most ϵ. Therefore for at least half of x, we have $e(x) \leq 2\epsilon$. The first message of Alice partitions this half into at most 2^c subsets. We pick S to consist of $t^n/2^{c+1}$ vectors of the same part: at least one part must have this many elements. □

We fix a set S as guaranteed by the lemma. We assume we started with a single round protocol, so $Q(x, y) = Q(x', y)$ whenever $x, x' \in S$. Indeed, Alice sends the same message by the choice of S and then the output is determined by Bob, who has the same input in the two cases.

We call a pair (x, y) *bad* if $x \in S$, $y \in [t]^n$ and Q errs on this input, i.e., $Q(x, y) \neq \mathrm{EE}(x, y)$. Let b be the number of bad pairs. By Lemma 3.5.1 each $x \in |S|$ is involved in at most $2\epsilon t^n$ bad pairs, so we have

$$b \leq 2\epsilon |S| t^n.$$

We call a triple (x, x', y) *bad* if $x, x' \in S$, $y \in [t]^n$, $\mathrm{EE}(x, y) = 1$ and $\mathrm{EE}(x', y) = 0$. The proof is based on double counting the number z of bad triples. Note that for a bad triple (x, x', y) we have $Q(x, y) = Q(x', y)$ but $\mathrm{EE}(x, y) \neq \mathrm{EE}(x', y)$, so Q must err on either (x, y) or (x', y) making one of these pairs bad. Any pair (bad or not) is involved in at most $|S|$ bad triples, so we have

$$z \leq b|S| \leq 2\epsilon |S|^2 t^n.$$

Let us fix arbitrary $x, x' \in S$ with $\mathrm{Match}(x, x') \leq n/2$. We estimate the number of $y \in [t]^n$ that makes (x, x', y) a bad triple. Such a y must have $\mathrm{Match}(x, y) > \mathrm{Match}(x', y) = 0$. To simplify the calculation we only count the vectors y with $\mathrm{Match}(x, y) = 1$. The match between y and x can occur at any position i with $x_i \neq x'_i$. After fixing the coordinate $y_i = x_i$ we can pick the remaining coordinates y_j of y freely as long as we avoid x_j and x'_j. Thus we have

$$|\{y \mid (x, x'y) \text{ is bad}\}| \geq (n - \mathrm{Match}(x, y))(t-2)^{n-1} \geq (n/2)(t-2)^{n-1} > t^n/14,$$

where in the last inequality we used $t = 4n$. Let s be the size of the Hamming ball $B_{n/2}(x) = \{y \in [t]^n \mid \mathrm{Match}(x, y) > n/2\}$. By the Chernoff bound we have $s < t^n/n^{n/2}$ (using $t = 4n$ again). For a fixed x we have at least $|S| - s$ choices for $x' \in S$ with $\mathrm{Match}(x, x') \leq n/2$ when the above bound for triples apply. Thus we have

$$z \geq |S|(|S| - s)t^n/14.$$

Combining this with the lower bound on the number of bad triples we get

$$28\epsilon|S| \geq |S| - s.$$

Therefore we conclude that we either have large error $\epsilon > 1/56$ or else we have $|S| \leq 2s < 2t^n/n^{n/2}$. As we have $|S| = t^n/2^{c+1}$ the latter possibility implies

$$c \geq n \log n/2 - 2.$$

Summarizing we have the following.

Theorem 3.5.2. *A single round randomized protocol for* EE_n *with error probability* $1/3$ *has complexity* $\Omega(n \log n)$*. A single round deterministic protocol for* EE_n *that errs on at most* $1/56$ *fraction of the inputs has complexity at least* $n \log n/2 - 2$*.*

3.6 An isoperimetric inequality on the discrete grid

The isoperimetric problem on the Boolean cube $\{0,1\}^n$ proved extremely useful in theoretical computer science. The problem is to determine the set $S \subseteq \{0,1\}^n$ of a fixed cardinality with the smallest "perimeter", or more generally, to establish connection between the size of a set and the size of its boundary. Here the boundary can be defined in several ways. Considering the Boolean cube as a graph where vertices of Hamming distance 1 are connected, the *edge boundary* of a set S is defined as the set of edges connecting S and its complement, while the *vertex boundary* consists of the vertices outside S having a neighbor in S.

Harper [41] showed that the vertex boundary of a Hamming ball is smallest among all sets of equal size, and the same holds for the edge boundary of a subcube. These results can be generalized to other cardinalities [43]; see the survey by Bezrukov [11].

Consider the metric space over the set $[t]^n$ equipped with the Hamming distance. Let f be a concave function on the nonnegative integers and $1 \leq M < n$ be an integer. We consider the following value as a generalized perimeter of a set $S \subseteq [t]^n$:

$$\mathop{\mathbb{E}}_{x \sim \mu} \left[f\left(|B_M(x) \cap S| \right) \right],$$

where $B_M(x) = \{y \in [t]^n \mid \text{Match}(x,y) \geq M\}$ is the radius $n - M$ Hamming ball around x. Note that when $M = n-1$ and f is the counting function given as $f(0) = 0$ and $f(l) = 1$ for $l > 0$ (which is concave), the above quantity is the size of the vertex boundary of S up to some normalization. For other concave functions f and parameters M this quantity can still be considered a measure of how "spread out" the set S is. We conjecture that n-dimensional boxes minimize this measure in every case.

Conjecture 3.6.1. *Let $1 \leq k \leq t$ and $1 \leq M < n$ be integers. Let S be an arbitrary subset of $[t]^n$ of size k^n and $P = [k]^n$. We have*

$$\mathop{\mathbb{E}}_{x \sim \mu} [f\left(|B_M(x) \cap P| \right)] \leq \mathop{\mathbb{E}}_{x \sim \mu} [f\left(|B_M(x) \cap S| \right)].$$

Even though a proof of Conjecture 3.6.1 remained elusive, in Theorem 3.6.5, we prove an approximate version of this result, where, for technical reasons, we have to restrict our attention to a small fraction of the coordinates. Having this weaker result allows us to prove our communication complexity lower bound in the next section but proving the conjecture here would simplify this proof.

3.6.1 A shifting argument

We start the technical part of this section by introducing the notation we will use. For $x, y \in [t]^n$ and $i \in [n]$ we write $x \sim_i y$ if $x_j = y_j$ for $j \in [n] \setminus \{i\}$. Observe that $\sim_i$ is an equivalence relation. A set $K \subseteq [t]^n$ is called an *i-ideal* if $x \sim_i y$, $x_i < y_i$ and $y \in K$ implies $x \in K$. We call a set $K \subseteq [t]^n$ an *ideal* if it is an i-ideal for all $i \in [n]$.

For $i \in [n]$ and $x \in [t]^n$ we define $\mathrm{down}_i(x) = (x_1, \ldots, x_{i-1}, x_i - 1, x_{i+1}, \ldots, x_n)$. We have $\mathrm{down}_i(x) \in [t]^n$ whenever $x_i > 1$. Let $K \subseteq [t]^n$ be a set, $i \in [n]$ and $2 \leq a \in [t]$. For $x \in K$, we define $\mathrm{down}_{i,a}(x, K) = \mathrm{down}_i(x)$ if $x_i = a$ and $\mathrm{down}_i(x) \notin K$ and we set $\mathrm{down}_{i,a}(x, K) = x$ otherwise. We further define $\mathrm{down}_{i,a}(K) = \{\mathrm{down}_{i,a}(x, K) \mid x \in K\}$. For $K \subseteq [t]^n$ and $i \in [n]$ we define

$$\mathrm{down}_i(K) := \{y \in [t]^n \mid y_i \leq |\{z \in K \mid y \sim_i z\}|\} .$$

Finally for $K \subseteq [t]^n$ we define

$$\mathrm{down}(K) := \mathrm{down}_1\left(\mathrm{down}_2\left(\ldots \mathrm{down}_n(K) \ldots\right)\right) .$$

The following lemma states few simple observations about these down operations.

Lemma 3.6.2. *Let $K \subseteq [t]^n$ be a set and let $i, j \in [n]$ be integers. The following hold.*

(i) $\mathrm{down}_i(K)$ *can be obtained from K by applying several operations* $\mathrm{down}_{i,a}$.

(ii) $|\mathrm{down}_{i,a}(K)| = |K|$ *for each* $2 \leq a \leq t$, $|\mathrm{down}_i(K)| = |K|$ *and* $|\mathrm{down}(K)| = |K|$.

(iii) $\mathrm{down}_i(K)$ *is an i-ideal and if K is a j-ideal, then* $\mathrm{down}_i(K)$ *is also a j-ideal.*

(iv) $\mathrm{down}(K)$ *is an ideal. For any* $x \in \mathrm{down}(K)$ *we have* $P := [x_1] \times [x_2] \times \cdots \times [x_n] \subseteq \mathrm{down}(K)$ *and there exists a set $T \subseteq K$ with* $P = \mathrm{down}(T)$.

Proof. For statement (i) notice that as long as K is not an i-ideal one of the operations $\mathrm{down}_{i,a}$ will not fix K and hence will decrease $\sum_{x\in K} x_i$. Thus a finite sequence of these operations will transform K into an i-ideal. It is easy to see that the operations $\mathrm{down}_{i,a}$ preserve the number of elements in each equivalence class of $\sim_i$, thus the i-ideal we arrive at must indeed be $\mathrm{down}_i(K)$.

Statement (ii) follows directly from the definitions of each of these down operations.

The first claim of statement (iii), namely that $\mathrm{down}_i(K)$ is an i-ideal, is trivial from the definition. Now assume $j \neq i$ and K is a j-ideal, $y \in \mathrm{down}_i(K)$ and $y_j > 1$. To see that $\mathrm{down}_i(K)$ is a j-ideal it is enough to prove that $\mathrm{down}_j(y) \in \mathrm{down}_i(K)$. Since $y \in \mathrm{down}_i(K)$, there are y_i distinct vectors $z \in K$ that satisfy $z \sim_i y$. Considering the vectors $\mathrm{down}_j(z) \sim_i \mathrm{down}_j(y)$ and using that these distinct vectors are in the j-ideal K proves that $\mathrm{down}_j(y)$ is indeed contained in $\mathrm{down}_i(K)$.

By statement (iii), $\mathrm{down}(K)$ is an i-ideal for each $i \in [n]$. Therefore $\mathrm{down}(K)$ is an ideal and the first part of statement (iv), that is, $P \subseteq K'$ follows. We prove the existence of suitable T by induction on the dimension n. The base case $n = 0$ (or even $n = 1$) is trivial. For the inductive step consider $K' = \mathrm{down}_2(\mathrm{down}_3(\ldots \mathrm{down}_n(K)\ldots))$. As $x \in \mathrm{down}(K) = \mathrm{down}_1(K')$, we have distinct vectors $x^{(k)} \in K'$ for $k = 1, \ldots, x_1$, satisfying $x^{(k)} \sim_1 x$. Notice that the construction of K' from K is performed independently on each of the $(n-1)$-dimensional "hyperplanes" $S^l = \{y \in [t]^n \mid y_1 = l\}$ as none of the operations $\mathrm{down}_2, \ldots, \mathrm{down}_n$ change the first coordinate of the vectors. We apply the inductive hypothesis to obtain the sets $T^{(k)} \subseteq S^{x_1^{(k)}} \cap K$ such that $\mathrm{down}_2(\ldots \mathrm{down}_n(T^{(k)})\ldots) = \{x_1^{(k)}\} \times [x_2] \times \cdots \times [x_n]$. Using again that these sets are in distinct hyperplanes and the operations $\mathrm{down}_2, \ldots, \mathrm{down}_n$ act separately on the hyperplanes S^l, we get for $T := \cup_{k=1}^{x_1} T^{(k)}$ that

$$\mathrm{down}_2(\ldots \mathrm{down}_n(T)\ldots) = \{x_1^{(k)} \mid k \in [x_1]\} \times [x_2] \times \cdots \times [x_n].$$

Applying down_1 on both sides finishes the proof of this last part of the lemma. □

For sets $x \in [t]^n$, $I \subseteq [n]$, and integer $M \in [n]$ we define $B_{I,M}(x) = \{y \in [t]^n \mid \mathrm{Match}(x_I, y_I) \geq M\}$. The projection of $B_{I,M}$ to the coordinates in I is the Hamming ball of radius $|I|-M$ around the projection of x.

Lemma 3.6.3. *Let $I \subseteq [n]$, $M \in [n]$ and let f be a concave function on the nonnegative integers. For arbitrary $K \subseteq [t]^n$ we have*

$$\mathop{\mathbb{E}}_{x\sim\mu}[f(|B_{I,M}(x) \cap \mathrm{down}(K)|)] \le \mathop{\mathbb{E}}_{x\sim\mu}[f(|B_{I,M}(x) \cap K|)].$$

Proof. By Lemma 3.6.2(i), the set $\mathrm{down}(K)$ can be obtained from K by a series of operations $\mathrm{down}_{i,a}$ with various $i \in [n]$ and $2 \le a \le t$. Therefore, it is enough to prove that the expectation in the lemma does not increase in any one step. Let us fix $i \in [n]$ and $2 \le a \le t$. We write $N_x = B_{I,M}(x) \cap K$ and $N'_x = B_{I,M}(x) \cap \mathrm{down}_{i,a}(K)$ for $x \in [t]^n$. We need to prove that

$$\mathop{\mathbb{E}}_{x\sim\mu}[f(|N_x|)] \ge \mathop{\mathbb{E}}_{x\sim\mu}[f(|N'_x|)].$$

Note that $|N_x| = |N'_x|$ whenever $i \notin I$ or $x_i \notin \{a, a-1\}$. Thus, we can assume $i \in I$ and concentrate on $x \in [t]^n$ with $x_i \in \{a, a-1\}$. It is enough to prove $f(|N_x|) + f(|N_y|) \ge f(|N'_x|) + f(|N'_y|)$ for any pair of vectors $x, y \in [t]^n$, satisfying $x_i = a$, and $y = \mathrm{down}_i(x)$.

Let us fix such a pair x, y and set $C = \{z \in K \setminus \mathrm{down}_{i,a}(K) \mid \mathrm{Match}(x_I, z_I) = M\}$. Observe that $N_x = N'_x \cup C$ and $N'_x \cap C = \emptyset$. Similarly, observe that $N'_y = N_y \cup \mathrm{down}_{i,a}(C)$ and $N_y \cap \mathrm{down}_{i,a}(C) = \emptyset$. Thus we have $|N'_x| = |N_x| - |C|$ and $|N'_y| = |N_y| + |\mathrm{down}_{i,a}(C)| = |N_y| + |C|$.

The inequality $f(|N_x|) + f(|N_y|) \ge f(|N'_x|) + f(|N'_y|)$ follows now from the concavity of f, the inequalities $|N'_x| \le |N_y| \le |N'_y|$ and the equality $|N_x| + |N_y| = |N'_x| + |N'_y|$. Here the first inequality follows from $\mathrm{down}_{i,a}(N'_x) \subseteq \mathrm{down}_{i,a}(N_y)$, the second inequality and the equality comes from the observations of the previous paragraph. □

3.6.2 Projecting to a subset of the coordinates

Lemma 3.6.4. *Let $K \subseteq [t]^n$ be arbitrary. There exists a vector $x \in K$ having at least $n/5$ coordinates that are greater than $k := \frac{t}{2}\mu(K)^{5/(4n)}$.*

Proof. The number of vectors that have at most $n/5$ coordinates greater than k can be upper bounded as

$$\binom{n}{n/5} t^{n/5} k^{4n/5} = t^n \binom{n}{n/5} (k/t)^{4n/5} = |K| \frac{\binom{n}{n/5}}{2^{4n/5}},$$

where in the last step we have substituted $\frac{k}{t} = \frac{1}{2}\mu(K)^{5/(4n)}$ and $\mu(K) = |K|/t^n$. Estimating $\binom{n}{n/5} \leq 2^{n\mathrm{H}_2(1/5)}$, we obtain that the above quantity is less than $|K|$. Therefore, there must exists an $x \in K$ that has at least $n/5$ coordinates greater than k. □

Theorem 3.6.5. *Let S be an arbitrary subset of $[t]^n$. Let $k = \frac{t}{2}\mu(S)^{5/(4n)}$ and $M = nk/(20t)$. There exists a subset $T \subset S$ of size $k^{n/5}$ and $I \subset [n]$ of size $n/5$ such that, defining $N_x = \{x' \in T \mid \mathrm{Match}(x_I, x'_I) \geq M\}$, we have*

- *(i)* $\Pr_{x\sim\mu}[N_x = \emptyset] \leq 5^{-M}$ *and*
- *(ii)* $\mathbb{E}_{x\sim\mu}[\log|N_x|] \geq (n/5 - M)\log k - n\log k/5^M$, *where we take* $\log 0 = -1$ *to make the above expectation exist.*

Proof. By Lemma 3.6.2(ii), we have $|\operatorname{down}(S)| = |S|$. By Lemma 3.6.4, there exists an $x \in \operatorname{down}(S)$ having at least $n/5$ coordinates that are greater than k. Let $I \subset [n]$ be a set of $n/5$ coordinates such that $x_i \geq k$ for a fixed $x \in \operatorname{down}(S)$. By Lemma 3.6.2(iv), $\operatorname{down}(S)$ is an ideal and thus it contains the set $P = \prod_i P_i$, where $P_i = [k]$ for $i \in I$ and $P_i = \{1\}$ for $i \notin I$. Also by Lemma 3.6.2(iv), there exists a $T \subseteq S$ such that $P = \operatorname{down}(T)$. We fix such a set T. Clearly, $|T| = k^{n/5}$.

For a vector $x \in [t]^n$, let $h(x)$ be the number of coordinates $i \in I$ such that $x_i \in [k]$. Note that $\mathbb{E}_{x\sim\mu}[h(x)] = 4M$ and $h(x)$ has a binomial distribution. By the Chernoff bound we have $\Pr_{x\sim\mu}[h(x) < M] < 5^{-M}$. For x with $h(x) \geq M$ we have $|B_{I,M}(x) \cap P| \geq k^{n/5-M}$, but for $h(x) < M$ we have $B_{I,M}(x) \cap P = \emptyset$. With the unusual convention $\log 0 = -1$ we have

$$\begin{aligned}\mathop{\mathbb{E}}_{x\sim\mu}[\log|B_{I,M}(x) \cap P|] &\geq \Pr[h(x) \geq M](n/5 - M)\log k - \Pr[h(x) < M] \\ &> (n/5 - M)\log k - n\log k/5^M\end{aligned}$$

We have $\mathrm{down}(T) = P$ and our unusual log is concave on the nonnegative integers, so Lemma 3.6.3 applies and proves statement (ii):

$$\begin{aligned}\mathop{\mathbb{E}}_{x\sim\mu}[\log|N_x|] &\geq \mathop{\mathbb{E}}_{x\sim\mu}[\log|B_{I,M}(x)\cap P|]\\ &\geq (n/5 - M)\log k - n\log k/5^M.\end{aligned}$$

To show statement (i), we apply Lemma 3.6.3 with the concave function f defined as $f(0) = -1$ and $f(l) = 0$ for all $l > 0$. We obtain that

$$\begin{aligned}\Pr_{x\sim\mu}[N_x = \emptyset] &= -\mathop{\mathbb{E}}_{x\sim\mu}[f(|N_x|)]\\ &\leq -\mathop{\mathbb{E}}_{x\sim\mu}[f(|B_{I,M}(x)\cap P|)]\\ &= \Pr_{x\sim\mu}[B_{I,M}(x)\cap P = \emptyset]\\ &< 5^{-M}.\end{aligned}$$

This completes the proof. □

3.7 *Lower bound for* Disj_k^n

In this section we prove our main lower bound result for the exists-equal problem, which implies a corresponding lower bound for the k-disjointness problem.

Theorem 3.7.1. *For any $r \leq \log^* n$, an r-round probabilistic protocol for EE_n with error probability at most $1/3$ sends at least one message of size $\Omega(n \log^{(r)} n)$.*

Note that the $r = 1$ round case of this theorem was proved as Theorem 3.5.2 in Section 3.5. The other extreme, which immediately follows from Theorem 4.0.4, is the following.

Corollary 3.7.2. *Any randomized protocol for EE_n with maximum message size $O(n)$ and error $1/3$ has at least $\log^* n - O(1)$ rounds.*

Theorem 4.0.4 is a direct consequence of the corresponding statement on deterministic protocols with small distributional error on uniform distribution; see Theorem 3.7.9 at the end of this section. Indeed, we can decrease the error of a randomized protocol below any constant $\epsilon > 0$ for the price of increasing the message length by a constant factor, then we can fix the coins of this low error protocol in a way that makes the resulting deterministic protocol Q err in at most ϵ fraction of the possible inputs. Applying Theorem 3.7.9 to the protocol Q proves Theorem 4.0.4.

In the rest of this section we use a round-elimination argument to prove Theorem 3.7.9, that is, we will use Q to solve smaller instances of the exists-equal problem in a way that the first message is always the same, and hence can be eliminated.

Suppose Alice sends the first message of c bits in protocol Q. By Lemma 3.5.1, there exists a $S \subset [t]^n$ of size $\mu(S) = 2^{-c-1}$ such that the first message of Alice is fixed when $x \in S$ and we have $\Pr_{y\sim\mu}[Q(x, y) \neq \mathrm{EE}(x, y)] \leq 2\epsilon$ for all $x \in S$. Fix such a set S and let $k := t/2^{\frac{5(c+1)}{4n}+1}$ and $M := nk/(20t)$. By Theorem 3.6.5, there exists a $T \subset S$ of size $k^{n/5}$ and $I \subset [n]$ of size $n/5$ such that defining

$$N_x = \{y \in T \mid \mathrm{Match}(x_I, y_I) \geq M\}$$

we have $\Pr_{x\sim\mu}[N_x = \emptyset] \le 5^{-M}$ and $\mathbb{E}_{x\sim\mu}[\log|N_x|] \ge (n/5 - M)\log k - n\log k/5^M$. Let us fix such sets T and I. Note also that Theorem 3.6.5 guarantees that T is a strict subset of S. Designate an arbitrary element of $S \setminus T$ as x'_e.

3.7.1 Embedding the smaller problem

Let $n' := M/10$ and $t' := 4n'$. Suppose Alice and Bob are given an instance (u, v) of the $\mathrm{EE}_{n'}$ problem, where $u, v \in [t']^{n'}$. To compute $\mathrm{EE}(u, v)$, through a random process, Alice and Bob will map (u, v) to random vectors (X', Y), where X' and Y are supported on $\in [t]^n$, and then run the protocol on $\mathrm{EE}(X', Y)$. The players embed a smaller instance $u, v \in [t']^{n'}$ of the exists-equal problem in EE_n concentrating on the coordinates I determined above. We set $n' := M/10$ and $t' := 4n'$. Optimally, the same embedding should guarantee low error probability for all pairs of inputs, but for technical reasons we need to know the number of coordinate agreements $\mathrm{Match}(u, v)$ for the input pairs (u, v) in the smaller problem having $\mathrm{EE}_{n'}(u, v) = 1$. Let $R \ge 1$ be this number, so we are interested in inputs $u, v \in [t']^{n'}$ with $\mathrm{Match}(u, v) = 0$ or R. We need this extra parameter so that we can eliminate a non-constant number of rounds and still keep the error bound a constant. For results on constant round protocols one can concentrate on the $R = 1$ case.

In order to solve the exist-equal problem with parameters t' and n' Alice and Bob use the shared random source to turn their input $u, v \in [t']^{n'}$ into longer random vectors $X', Y \in [t]^n$, respectively, and apply the protocol Q above to solve this exists-equal problem for these larger inputs. Here we informally list the main requirements on the process generating X' and Y. We require these properties for the random vectors $X', Y \in [t]^n$ generated from a fixed pair $u, v \in [t']^{n'}$ satisfying $\mathrm{Match}(u, v) = 0$ or R.

(P1) $\mathrm{EE}(X', Y) = \mathrm{EE}(u, v)$ with large probability,

(P2) $\mathrm{supp}(X') = T \cup \{x'_e\}$ and

(P3) $Y \mid X'$ is distributed almost uniformly

Combining these properties with the fact that $\Pr_{y\sim\mu}[Q(x,y) \neq \mathrm{EE}(x,y)] \leq 2\epsilon$ for each $x \in S$, we will argue that for the considered pairs of inputs $Q(X',Y)$ equals $\mathrm{EE}(u,v)$ with large probability, thus the combined protocol solves the small exists-equal instance with small error, at least for input pairs with $\mathrm{Match}(u,v) = 0$ or R. Furthermore, by Property (P2) the first message of Alice will be fixed and hence does not need to be sent, making the combined protocol one round shorter.

The random variables X' and Y are constructed as follows. Let $m := 2n/(MR)$ be an integer. Each player repeats his or her input (u and v, respectively) m times, obtaining a vector of size $n/(5R)$. Then using the shared randomness, the players pick $n/(5R)$ uniform random maps $m_i \colon [t'] \to [t]$ independently and apply m_i to ith coordinate. Furthermore, the players pick a uniform random 1-1 mapping $\pi \colon [n/(5R)] \to I$ and use it to embed the coordinates of the vectors they constructed among the coordinates of the vectors X and Y of length n. The remaining $n - n/(5R)$ coordinates of X is picked uniformly at random by Alice and similarly, the remaining $n - n/(5R)$ coordinates of Y is picked uniformly at random by Bob. Note that the marginal distribution of both X and Y are uniform on $[t]^n$. If $\mathrm{Match}(u,v) = 0$ the vectors X and Y are independent, while if $\mathrm{Match}(u,v) = R$, then Y can be obtained by selecting a random subset of I of cardinality mR, copying the corresponding coordinates of X and filling the rest of Y uniformly at random.

This completes the description of the random process for Bob. However Alice generates one more random variable X' as follows. Recall that $N_x = \{z \in T \mid \mathrm{Match}(z_I, x_I) \geq M\}$. The random variable X' is obtained by drawing $x \sim X$ first and then choosing a uniform random element of N_x. In the (unlikely) case that $N_x = \emptyset$, Alice chooses $X' = x'_e$.

Note that X' either equals x'_e or takes values from T, hence Property (P2) holds. In the next lemma we quantify and prove Property (P1) as well.

Lemma 3.7.3. *Assume $n \geq 3$, $M \geq 2$ and $u, v \in [t']^{n'}$. We have*

(i) if $\mathrm{Match}(u,v) = 0$ *then* $\Pr[\mathrm{EE}(X',Y) = 0] > 0.77$*;*

(ii) if $\mathrm{Match}(u,v) = R$*, then* $\Pr[\mathrm{EE}(X',Y) = 1] \geq 0.80$.

Proof. For the first claim, note that when $\mathrm{Match}(u, v) = 0$, the random variables X and Y are independent and uniformly distributed. We construct X' based on X, so its value is also independent of Y. Hence $\Pr[\mathrm{EE}(X', Y) = 0] = (1 - 1/t)^n$. This quantity goes to $e^{-1/4}$ since $t = 4n$ and is larger than 0.77 when $n \geq 3$. This establishes the first claim.

For the second claim let $J = \{i \in I \mid X_i = Y_i\}$ and $K = \{i \in I \mid X'_i = X_i\}$. By construction, $|J| = \mathrm{Match}(X_I, Y_I) \geq mR$ and $|K| = \mathrm{Match}(X'_I, X_I) \geq M$ unless $N_X = \emptyset$. By our construction, each $J \subset I$ of the same size is equally likely by symmetry, even when we condition on a fix value of X and X'. Thus we have $\mathbb{E}[|J \cap K| \mid N_X \neq \emptyset] \geq mRM/|I| = 10$ and $\Pr[J \cap K = \emptyset \mid N_X \neq \emptyset] < e^{-10}$. Note that X is distributed uniformly over $[t]^n$, therefore by Theorem 3.6.5(i) the probability that $N_X = \emptyset$ is at most 5^{-M}. Note that $\mathrm{Match}(X', Y) \geq |J \cap K|$ and thus $\Pr[\mathrm{EE}(X', Y) = 0] \leq \Pr[J \cap K = \emptyset] \leq \Pr[J \cap K = \emptyset \mid N_X \neq \emptyset] + \Pr[N_X = \emptyset] \leq e^{-10} + 5^{-M}$. This completes the proof. □

We quantify the correlation of X' and Y stated in Property (P3) by their mutual information. This mutual information argument is postponed to the next subsection; here we show how such a bound to the mutual information implies that the error introduced by Q is small.

Lemma 3.7.4. *Let $\gamma = \Pr[Q(X', Y) \neq \mathrm{EE}(X', Y)]$. If $\gamma \geq 2\epsilon$, then we have $\mathbf{D}_2(\gamma \,\|\, 2\epsilon) \leq \mathbf{I}(X' : Y)$.*

Proof. For all fixings $x \in \mathrm{supp}(X')$ and a distribution ν on $[t]^n$, define

$$e_x(\nu) := \Pr_{y \sim \nu}[Q(x, y) \neq \mathrm{EE}(x, y)].$$

By the definition of mutual information and the conditional divergence,

$$\begin{aligned}\mathbf{I}(X' : Y) &= \mathbf{D}(Y \mid X' \,\|\, Y)\\ &= \mathop{\mathbb{E}}_{x \sim X'} \mathbf{D}(Y \mid X' = x \,\|\, Y)\\ &\geq \mathop{\mathbb{E}}_{x \sim X'} \mathbf{D}_2(e_x(\mathrm{dist}(Y \mid X' = x)) \,\|\, e_x(\mu))\\ &\geq \mathbf{D}_2\left(\mathop{\mathbb{E}}_{x \sim X'} e_x(\mathrm{dist}(Y \mid X' = x)) \,\middle\|\, \mathop{\mathbb{E}}_{x \sim X'} e_x(\mu)\right)\\ &= \mathbf{D}_2\left(\gamma \,\middle\|\, \mathop{\mathbb{E}}_{x \sim X'} e_x(\mu)\right)\\ &\geq \mathbf{D}_2(\gamma \,\|\, 2\epsilon)\end{aligned}$$

where the first inequality is the data processing inequality, the second inequality follows from the convexity of $\mathbf{D}(\cdot \,\|\, \cdot)$ and the last inequality follows from the guarantee $e_x(\mu) \leq 2\epsilon$ for all x that is provided by Lemma 3.5.1 and the assumption of the present lemma that $\gamma \geq 2\epsilon$. □

3.7.2 Establishing the low correlation property

We quantify Property (P3) using the mutual information. If $\mathrm{Match}(u, v) = R$ our process generates X and Y with the expected number $\mathbb{E}[\mathrm{Match}(X_I, Y_I)]$ of matches only slightly more than the minimum mR. We lose most of these matches with Y when we replace X by X' and only an expected constant number remains. A constant number of forced matches with X' within I restricts the number of possible vectors Y but it only decreases the entropy by $O(1)$. The calculations in this subsection make this intuitive argument precise.

Recall that X and Y are correlated due to the random process with which Alice and Bob generate them and X' is obtained from X. To understand $\mathbf{I}(X' : Y)$

Lemma 3.7.5. *For any $u, v \in [t']^{n'}$ it holds that $\mathbf{I}(X : X') \leq M \log k + n \log k / 5^M$.*

Proof. We have

$$\mathbf{I}(X : X') = \mathrm{H}(X') - \mathrm{H}(X' \mid X)$$

and $\mathrm{H}(X') \leq \log|\mathrm{supp}(X')| = \log(|T| + 1) \leq \frac{n}{5} \log k + 1$.

Observe that $\mathrm{H}(X' \mid X) = \mathbb{E}_{x\sim\mu}[\log|N_x|]$, where $\log 0$ is now taken to be 0. From Theorem 3.6.5(ii) we get $\mathrm{H}(X' \mid X) \geq \frac{n}{5}\log k - M\log k - n\log k/5^M$. Plugging in, the $\frac{n}{5}\log k$ terms cancel and we get the statement of the lemma. □

Lemma 3.7.6. *Let X', Y be as constructed above. The following hold.*

(i) If $\mathrm{Match}(u,v) = 0$ *we have* $\mathbf{I}(X' : Y) = 0$

(ii) If $M > 100\log n$ *and* $\mathrm{Match}(u,v) = R$ *we have* $\mathbf{I}(X' : Y) = O(1)$.

Proof. Part (i) holds as Y is independent of X' whenever $\mathrm{EE}(u,v) = 0$ by construction.

For part (ii) recall that if $\mathrm{Match}(u,v) = R$ one can construct X and Y by uniformly selecting a size mR set $L \subseteq I$ and selecting X and Y uniformly among all pairs satisfying $X_L = Y_L$. Recall that L is the set of coordinates the mR matches between u^m and v^m were mapped. These are the "intentional matches" between X_I and Y_I. Note that there may be also "unintended matches" between X_I and Y_I, but not too many: their expected number is $(n/5 - mR)/t < 1/20$. As given any fixed L, the marginal distribution of both X and Y are still uniform, so in particular X is independent of L and so is X' constructed from X. Let us expand $\mathbf{I}(X'L : Y)$ using the chain rule in two different ways obtaining

$$\begin{aligned}\mathbf{I}(X'L : Y) &= \mathbf{I}(X' : Y) + \mathbf{I}(L : Y \mid X') \\ &= \mathbf{I}(L : Y) + \mathbf{I}(X' : Y \mid L)\,. \end{aligned} \tag{3.7.1}$$

Since the first term of (3.7.1) is zero by independence of Y and L, we conclude

$$\begin{aligned}\mathbf{I}(X' : Y) &= \mathbf{I}(X' : Y \mid L) - \mathbf{I}(L : Y \mid X') \\ &= \mathbf{I}(X' : Y \mid L) - \mathbf{I}(L : X'Y)\,, \end{aligned} \tag{3.7.2}$$

where the second inequality follows again by the chain rule and the fact that X' and L are independent. Let us understand the terms of (3.7.2) one by one. First we expand the first term by the chain rule, obtaining

$$\mathbf{I}(X' : Y \mid L) = \mathbf{I}(X' : Y_L \mid L) + \mathbf{I}\Big(X' : Y_{[n]\setminus L} \mid LY_L\Big)$$

however since $Y_{[n]\setminus L}$ is uniformly distributed for any fixed L, X' and Y_L, the second term on the right hand side is zero. By construction we have $X_L = Y_L$, thus

$$\begin{aligned}\mathbf{I}(Y_L : X' \,|\, L) &= \mathbf{I}(X_L : X' \,|\, L)\\ &\leq \frac{mR}{n/5}\mathbf{I}(X_I : X')\\ &\leq 10\log k + \frac{MR}{5^{M-1}}\log k,\end{aligned}$$

where the first inequality follows by Lemma 2.3.5 as L is a uniform and independent of X and X' and the second inequality follows from Lemma 3.7.5 that we will prove shortly and the formula defining m.

Here, when condition on L, the correlation of X' and Y is roughly $10\log k$ bits, which is significantly more than the constant bound we seek. Next we will see that all but a constant bits of this correlation comes from having observed what L is. The next term, $\mathrm{H}(L)$ is easy to compute as L is a uniform subset of I of size mR:

$$\mathrm{H}(L) = \log\binom{n/5}{mR}$$

It remains to bound the term $\mathrm{H}(L \,|\, Y, X')$. Let $Z = \{i \mid i \in I \text{ and } X'_i = Y_i\}$. Note that Z can be derived from X', Y (as I is fixed) hence $\mathrm{H}(L \,|\, Y, X') \leq \mathrm{H}(L \,|\, Z)$. Further, let $C = |Z \setminus L|$. We obtain

$$\begin{aligned}\mathrm{H}(L \,|\, Y, X') &\leq \mathrm{H}(L \,|\, Z) \leq \mathrm{H}(L \,|\, Z, C) + \mathrm{H}(C)\\ &< \mathop{\mathbb{E}}_{Z,C}\left[\log\binom{n/5 - |Z| + C}{mR - |Z| + C}\right] + \mathop{\mathbb{E}}_{Z,C}\left[\log\binom{|Z|}{C}\right] + 2\end{aligned}$$

where we used $\mathrm{H}(C) < 2$. Note that for any fixed $x' \in T$ and $x \in \mathrm{supp}(X \,|\, X' = x')$, we have

$$\mathbb{E}[|Z| - C \,|\, X = x, X' = x'] = \mathrm{Match}(x_I, x'_I)mR/(n/5) \geq 10$$

as $\mathrm{Match}(x_I, x'_I) \geq M$ by definition. Hence we have

$$\log\binom{n/5}{mR} - \log\binom{n/5 - |Z| + |C|}{mR - |Z| + |C|} \geq 10\log\frac{n}{5m} - O(1),$$

$$\mathop{\mathbb{E}}_{Z,C}\left[\log\binom{|Z|}{C}\right] \le \mathbb{E}[|Z|] < 20.$$

Summing the estimates above for the various parts of $\mathrm{H}(Y \mid X')$ the statement of the lemma follows. □

It remains to prove the following simple lemma that "reverses" the conditional entropy bound in Theorem 3.6.5(ii):

3.7.3 The round elimination lemma

Let ν_n be the uniform distribution on $[t]^n \times [t]^n$, where we set $t = 4n$. The following lemma gives the base case of the round elimination argument.

Lemma 3.7.7. *Any 0-round deterministic protocol for* EE_n *has at least 0.22 distributional error on* ν_n*, when* $n \ge 1$*.*

Proof. The output of the protocol is decided by a single player, say Bob. For any given input $y \in [t]^n$ we have $3/4 \le \Pr_{x\sim\mu}[\mathrm{EE}(x,y) = 0] < e^{-1/4} < 0.78$. Therefore the distributional error is at least 0.22 for any given y regardless of the output Bob chooses, thus the overall error is also at least 0.22. □

Now we give our full round elimination lemma.

Lemma 3.7.8. *Let* $r > 0, c, n$ *be an integers such that* $c < \frac{4}{5}n\log n$*. There is a constant* $0 < \epsilon_0 < 1/200$ *such that if there is an* r*-round deterministic protocol with* c*-bit messages for* EE_n *that has* ϵ_0 *error on* ν_n*, then there is an* $(r-1)$*-round deterministic protocol with* $O(c)$*-bit messages for* $\mathrm{EE}_{n'}$ *that has* ϵ_0 *error on* $\nu_{n'}$*, where* $n' = \Omega(n/2^{\frac{5c}{4n}})$*.*

Proof. We start with an intuitive description of our reduction. Let us be given the deterministic protocol Q for EE_n that errs on an ϵ_0 fraction of the inputs. To solve an instance (u, v) of the smaller $\mathrm{EE}_{n'}$ problem the players perform the embedding procedure described in previous subsection k_0 times independently for each parameter $R \in [R_0]$. Here k_0 and R_0 are constants we set later. They perform the protocol Q in parallel for each of the k_0R_0 pairs of inputs they generated. Then they take the majority of the k_0 outputs for a fixed parameter R. We show that this result gives the correct value of $\mathrm{EE}(u, v)$ with large probability provided that $\mathrm{Match}(u, v) = 0$ or R. Finally they take the OR of these results for the R_0 possible values of R. By the union bound this gives the correct value $\mathrm{EE}(u, v)$ with large probability provided $\mathrm{Match}(u, v) \leq R_0$. Fixing the random choices of the reduction we obtain a deterministic protocol. The probability of error for the uniform random input can only grow by the small probability that $\mathrm{Match}(u, v) > R_0$ and we make sure it remains below ϵ_0. The rest of the proof makes this argument precise.

For random variables X' and Y constructed in Section 3.7.1, Lemma 3.7.6 guarantees that $\mathrm{H}(Y \mid X') \geq n \log t - \alpha_0$ for some constant α_0, as long as $M > 100 \log n$ and $\mathrm{Match}(u, v) = R$. Let ϵ_0 be a constant such that $\mathbf{D}_2(1/10 \,\|\, 2\epsilon_0) > 200(\alpha_0 + 1)$. Note that such ϵ_0 can be found as $\mathbf{D}_2(1/10 \,\|\, \epsilon)$ tends to infinity as ϵ goes to 0. We can bound $\Pr_{(x,y)\sim\nu_m}[\mathrm{Match}(x, y) \geq l] \leq 1/(4^l l!)$ for all $m \geq 1$. We set R_0 such that $\Pr_{(x,y)\sim\nu_m}[\mathrm{Match}(x, y) \geq R_0] \leq \epsilon_0/2$ for all $m \geq 1$.

Let Q be a deterministic protocol for EE_n that sends $c < (n \log n)/2$ in each round and that has ϵ_0 error on ν_n. Let S be as constructed in Lemma 3.5.1 and let M be as defined in Theorem 3.6.5. We have $M = \frac{n}{40} 2^{\frac{-5(c+1)}{4n}}$ as $t = 4n$ and $\mu(S) = 2^{-(c+1)}$ by Lemma 3.5.1. Note that by our choice of c, we have $M > 100 \log n$, hence the hypotheses of Lemma 3.7.6 are satisfied.

Let $n' = M/10 = \frac{n}{400} 2^{\frac{-5(c+1)}{4n}}$. Now we give a randomized protocol Q' for $\mathrm{EE}_{n'}$. Suppose the players are given an instance of $\mathrm{EE}_{n'}$, namely the vectors $(u, v) \in [4n']^{n'} \times [4n']^{n'}$. Let $k_0 = 10 \log(R_0 + 1/\epsilon_0)$. For $R \in [R_0]$ and $k \in [k_0]$, the players construct the vectors $X'_{R,k}$ and $Y_{R,k}$ as described in Section 3.7.1 with parameter R and with fresh randomness for each of the $R_0 k_0$ procedures. The players run $R_0 k_0$ instances of protocol Q in parallel, on inputs $X'_{R,k}, Y_{R,k}$ for $R \in [R_0]$ and $k \in [k_0]$. Note that the first message of the first player, Alice, is fixed for all instances of Q by Property (P2) and Lemma 3.5.1. Therefore, the second player, Bob, can start the protocol assuming Alice has sent the fixed first message. After the protocols finish, for each $R \in [R_0]$, the last player who received a message computes b_R as the majority of $Q(X'_{R,k}, Y_{R,k})$ for $k \in [k_0]$. Finally, this player outputs 0 if $b_R = 0$ for all

$R \in [R_0]$ and outputs 1 otherwise.

Suppose now that $\mathrm{EE}(u,v) = 0$. By Lemma 3.7.3(i), we have $\Pr[\mathrm{EE}(X'_{R,k}, Y_{R,k}) = 0] \geq 0.77$ for each R and k. Recall that that $Y_{R,k}$ is distributed uniformly for each R and k and since $\mathrm{EE}(u,v) = 0$, it is independent of $X'_{R,k}$. Therefore, by $X'_{R,k} \in S$ (Property (P2)) and the fact that $\Pr_{y \sim \mu}[Q(x,y) \neq \mathrm{EE}(x,y)] \leq 2\epsilon_0$ for all $x \in S$ as per Lemma 3.5.1, we obtain $\Pr[Q(X'_{R,k}, Y_{R,k}) = 0] \geq 0.77 - 2\epsilon_0 > 0.76$. By the Chernoff bound we have $\Pr[b_R = 1] < \epsilon_0/(2R_0)$, and by the union bound $\Pr[Q' \text{ outputs } 0] \geq 1 - \epsilon_0/2$.

Let us now consider the case $\mathrm{Match}(u,v) = R$ for some $R \in [R_0]$. Fix any $k \in [k_o]$ and set $X' = X'_{R,k}$, $Y = Y_{R,k}$. By Lemma 3.7.3(ii), $\Pr[\mathrm{EE}(X', Y) = 1] \geq 0.80$. By Lemma 3.7.6, $\mathbf{I}(X' : Y) \leq \alpha_0$ and that Y is distributed uniformly at random. By Lemma 3.7.4 and our choice of ϵ_0, we have $\Pr[\mathrm{EE}(X', Y) \neq Q(X', Y)] < 1/10$. Furthermore, by Lemma 3.7.3(ii), $\Pr[\mathrm{EE}(u,v) \neq \mathrm{EE}(X', Y)] < 0.20$ hence with probability at least 0.70 we have $\mathrm{EE}(u,v) = Q(X', Y)$. This happens independently for all the values of $k \in [k_0]$, so by the Chernoff bound and our choice of k_0, we have $\Pr[Q' \text{ outputs } 0] \leq \Pr[b_R = 0] < \epsilon_0/2$.

Finally, $\Pr_{(u,v) \sim \nu_{n'}}[\mathrm{Match}(u,v) \geq R_0] \leq \epsilon_0/2$ by our choice of R_0. Note that the protocol Q' uses a shared random bit string, say W, in the construction of the vectors $X'_{R,k}$ and $Y_{R,k}$. Hence, overall, we have

$$\Pr_{W,(u,v) \sim \nu_{n'}}[\mathrm{EE}(u,v) = Q'(u,v)] \geq 1 - \epsilon_0$$

Since we measure the error of the protocol under a distribution, we can fix W to a value without increasing the error under the aforementioned distribution by the so called easy direction of Yao's lemma. Namely, there exists a $w \in \mathrm{supp}(W)$ such that

$$\Pr_{(u,v) \sim \nu_{n'}}[\mathrm{EE}(u,v) = Q'(u,v) \,|\, W = w] \geq 1 - \epsilon_0$$

Fix such w. Observe that Q' is a $(r-1)$-round protocol for $\mathrm{EE}_{n'}$ where $n' = \frac{n}{400} 2^{\frac{-5(c+1)}{4n}} = \Omega(n/2^{\frac{5c}{4n}})$ and it sends at most $R_0 k_0 c = O(c)$ bits in each message. Furthermore, Q' is deterministic and has at most ϵ_0 error on $\nu_{n'}$ as desired. □

Theorem 3.7.9. *There exists a constant ϵ_0 such that for any $r \leq \log^* n$, an r-round deterministic protocol for EE_n which has ϵ_0 error on ν_n sends at least one message of size $\Omega(n \log^{(r)} n)$.*

Proof. Suppose we have an r-round protocol with c-bit messages for EE_n that has ϵ_0 error on ν_n, where $c = \gamma n \log^{(r)} n$ for some $\gamma < 4/5 - o(1)$. By Lemma 3.7.8, this protocol can be converted to an $r-1$ round protocol with αc-bit messages for $\mathrm{EE}_{n'}$ that has ϵ_0-error on $\nu_{n'}$, where $n' = \beta n/2^{5c/4n}$ for some $\alpha, \beta > 0$. We only need to verify that $\alpha c \leq \gamma n' \log^{(r-1)} n'$. We have

$$\begin{aligned}
\gamma n' \log^{(r-1)} n' &= \gamma\beta n/2^{5c/4n} \log^{(r-1)}(\beta n/2^{5c/4n}) \\
&= \gamma\beta n/2^{\frac{5\gamma}{4}\log^{(r)} n} \log^{(r-1)}(\beta n/2^{5c/4n}) \\
&\geq \gamma\beta n \left(\log^{(r-1)} n\right)^{1-\frac{5\gamma}{4}-o(1)} \\
&\geq \gamma\alpha n \log^{(r)} n
\end{aligned}$$

for $\gamma < 4/5 - o(1)$ and large enough n. Therefore, by iteratively applying Lemma 3.7.8 we obtain a 0-round protocol for $\mathrm{EE}_{\bar{n}}$ that makes ϵ_0 error on $\nu_{\bar{n}}$ for some $\bar{n}$ satisfying $\gamma\bar{n}^2 = \gamma\bar{n}\log^{(0)}\bar{n} \geq c\alpha^r$. Therefore $\bar{n} \geq 1$ and since $\epsilon_0 < 0.22$, the protocol we obtain contradicts Lemma 3.7.7, showing that the protocol we started with cannot exists. □

Remark 3.7.10. *We note that in the proof of Theorem 4.0.4, to show that a protocol with small communication does not exist, we start with the given protocol and apply the round elimination lemma (i.e., Lemma 3.7.8) r times to obtain a 0-round protocol with small error probability, which is shown to be impossible by Lemma 3.7.7. Alternatively, one can apply the round elimination $r-1$ times to obtain a 1-round protocol with $o(n \log n)$ communication for EE_n, which is ruled out by Theorem 3.5.2.*

3.8 Discussion

The r-round protocol we gave in Section 3.4 solves the sparse set disjointness problem in $O(k \log^{(r)} k)$ total communication. As we proved in Section 3.7 this is optimal. The same, however, cannot be said of the error probability. With the same protocol, but with more careful setting of the parameters the exponentially small error $O(2^{-\sqrt{k}})$ of the $\log^* k$-round protocol can be further decreased to $2^{-k^{1-o(1)}}$.

For small (say, constant) values of r this protocol cannot achieve exponentially small error error without the increase in the complexity if the universe size m is unbounded. But if m is polynomial in k (or even slightly larger, $m = \exp^{(r)}(O(\log^{(r)} k))$), we can replace the last round of the protocol by one player deterministically sending his or her entire "current set" S_r. With careful setting of the parameters in other rounds, this modified protocol has the same $O(k \log^{(r)} k)$ complexity but the error is now exponentially small: $O(2^{-k/\log k})$. Note that in our lower bound on the r-round complexity of the sparse set disjointness we we use the exists-equal problem with parameters $n = k$ and $t = 4k$. This corresponds to the universe size $m = tn = 4k^2$. In this case any protocol solving the exists-equal problem with $1/3$ error can be strengthened to exponentially small error using the same number of rounds and only a constant factor more communication.

Our lower and upper bounds match for the exists-equal problem with parameters n and $t = \Omega(n)$, since the upper bounds were established without any regard of the universe size, while the lower bounds worked for $t = 4n$. Extensions of the techniques presented in this paper give matching bounds also in the case $3 \leq t < n$, where the r-round complexity is $\Theta(n \log^{(r)} t)$ for $r \leq \log^* t$. Note, however, that in this case one needs to consider significantly more complicated input distributions and a more refined isoperimetric inequality, that does not permit arbitrary mismatches. The $\Omega(n)$ lower bound applies for the exists-equal problem of parameters n and $t \geq 3$ regardless of the number of rounds, as the disjointness problem on a universe of size n is a sub-problem. For $t = 2$ the situation is drastically different, the exists-equal problem with $t = 2$ is equivalent to a single equality problem.

Finally a remark on using the joint random source model of randomized protocols throughout the paper. By a result of Newman [69] our protocols of Section 3.4 can be made to work in private coin model (or even if one of the players is forced to behave deterministically) by increasing the first message length by $O(\log\log(N) + \log(1/\epsilon))$ bits, where $N = \binom{m}{k}$ is the number of possible inputs. In our case this means adding the term $O(\log\log m) + o(k)$ to our bound of $O(k \log^{(r)} k)$, since our protocols make at least $\exp(-k/\log k)$ error. This additional cost is insignificant for reasonably small values of m, but it is necessary for large values as the equality problem, which is an instance of disjointness, requires $\Omega(\log\log m)$-bits in the private coin model.

Note also that we achieve a super-linear increase in the communication for OR of n instances of equality even in the private coin model for $r = 1$. For $r \geq 2$, no such increase happens in the private coin model as communication complexity of EE_n^t is at most $O(n \log\log t)$ however a single equality problem requires $\Omega(\log\log t)$ bits.

3.9 Chapter notes

The results presented in this chapter are obtained with Gábor Tardos and published in our joint paper [80] in FOCS 2013.

Chapter 4

THE BLAKLEY-DIXON-ERDOS-SIMONOVITS CONJECTURE

Suppose that some initial heat configuration $u\colon \Omega \to \mathbb{R}_+$ is given over a finite space Ω and the configuration evolves according to the map $w \mapsto Sw$ in each time step $t = 0, 1, \ldots$, for some symmetric stochastic matrix $S\colon \Omega \times \Omega \to \mathbb{R}_+$. Assume that we are interested in the amount of heat contained in a certain region $R \subseteq \Omega$ and how this quantity changes over time. In notation, assuming $\|u\|_2 = 1$ for normalization purposes and $v(x) := \mathbb{1}_{x\in R}/|R|^{1/2}$ for $x \in \Omega$, we would like to understand how

$$m_t := \left\langle v, S^t u \right\rangle$$

changes as a function of t. In this paper we derive local bounds that $\{m_t\}_{t=0}^{\infty}$ must obey for any S, u and v satisfying the symmetry, magnitude and positivity constraints above (in fact our bounds work for any countable Ω, arbitrary non-negative unit vector v and symmetric non-negative S). Our first bound $m_{t+2} \geq m_t^{1+2/t}$ answers a question of Blakley and Dixon from 1966 (see Conjecture 4.0.2 below) which was later conjectured independently by Erdős and Simonovits in 1982 also (see Conjecture 4.0.6 in Section 4.0.2).

Moreover we establish a tight connection between such bounds and the well-studied k-Hamming distance problem [72, 88, 30, 6, 37, 46, 15, 23, 14, 3] and the k-Hamming weight problem [2, 16, 23] and obtain the first tight bounds for respectively the communication complexity and parity decision tree complexity of them. Our tight $\Omega(k \log(k/\delta))$ lower bound for the δ-error communication complexity of the k-Hamming distance problem (that applies whenever $k^2 < \delta n$) answers affirmatively a conjecture stated in [14] (Conjecture 1.4). Prior to our work, the best impossibility results for this problem were an $\Omega(k \log^{(r)} k)$ bits lower bound ($\log^{(r)} z$ being the r nested applications of logarithm) that applies to any randomized

r-round communication protocol [80], and an $\Omega(k \log(1/\delta))$ lower bound that applies to any δ-error randomized protocol for $k < \delta n$ [14].

Our parity decision tree lower bound shows that any δ-error parity decision tree solving the k-Hamming weight problem has size $\exp \Omega\left(k \log(k/\delta)\right)$, which directly implies an $\Omega(k \log(k/\delta))$ bound on the depth of any such decision tree. Previously no nontrivial lower bound was known for the parity decision tree size of this problem and an $\Omega(k \log(1/\delta))$ bound on the parity decision tree depth followed from the communication complexity bound of [14]. Prior to [14], the best bound on the parity decision tree depth was $\Omega(k)$, derived in [15] and [16].

Either by combining our communication complexity lower bound with the reduction technique developed in [15] or by combining our parity decision tree lower bound with a reduction given in [12], one obtains an $\Omega(k \log(k/\delta))$ bound for any (potentially adaptive) property tester for the δ-error probability k-linearity testing problem. This establishes the correct bound for this problem which was studied extensively [35, 39, 16, 12, 23, 15] since [35] or earlier.

4.0.1 Motivating our bounds on m_t

We would like to provide some intuition as to why one should expect

$$m_{t+2} \geq m_t^{1+2/t}, \text{ and} \tag{4.0.1}$$

$$m_{t+2} \geq m_t^{1+2/t} \cdot \min\left\{t^{1-\epsilon}, \delta \frac{m_t^{1-2/t}}{m_{t-2}}\right\} \tag{4.0.2}$$

to hold for appropriate ϵ, δ. Recalling that S is a symmetric matrix with maximum eigenvalue 1, we may write $S = QDQ^{\intercal}$ for an orthonormal matrix Q having columns q_x, $x \in \Omega$ and a diagonal matrix D with entries $\lambda_x \leq 1$, $x \in \Omega$. Plugging this into $m_t = \langle v, S^t u\rangle$, we get

$$m_t = \sum_{x \in \Omega} \lambda_x^t \langle u, q_x\rangle \langle v, q_x\rangle . \tag{4.0.3}$$

For sake of analogy let us drop our assumption that S, u, v are coordinate-wise nonnegative for a moment but instead assume that each summand in the right hand side of Eq. (4.0.3) is nonnegative by some coincidence. In this case we can consider $\{m_t\}_{t=0}^{\infty}$ as the moment sequence of a random variable supported on $[0,1]$ that takes the value $|\lambda_x|$ with probability $|\langle u, q_x\rangle \langle v, q_x\rangle|$ and the value 0 with probability $1-\sum_x |\langle u, q_x\rangle \langle v, q_x\rangle|$ (which is nonnegative by Cauchy-Schwarz inequality). This would imply that $\{m_t\}_{t=0}^{\infty}$ is *completely monotone* by Hausdorff's characterization [44] and therefore log-convex (e.g., [70], Section 2.1, Example 6).

One particular implication of the log-convexity of $\{m_t\}_{t=0}^{\infty}$, that $\frac{1}{t}\log m_t + \frac{t-1}{t}\log m_0 \geq \log m_1$, when combined with the fact $0 \leq m_0 \leq 1$ (that follows from our assumption on the terms of Eq. (4.0.3)), leads to $m_t \geq m_1^t$. In 1958, Mandel and Hughes showed that if $u = v$, rather surprisingly, one can trade the assumption that the summands of Eq. (4.0.3) are nonnegative with the assumption that S and $u = v$ are coordinate-wise nonnegative and still obtain the conclusion $m_t \geq m_1^t$:

Theorem 4.0.1 (Mandel and Hughes [64]). *Let u be a nonnegative unit vector and S be a symmetric matrix with nonnegative entries. For an integer*[1] $t \geq 1$ *we have* $\langle u, S^t u\rangle \geq \langle u, Su\rangle^t$.

A more general implication of the log-convexity of $\{m_t\}_{t=0}^{\infty}$ and that $m_0 \leq 1$ is that for $k \geq t$, $\frac{t}{k}\log m_k + \frac{k-t}{k}\log m_0 \geq \log m_t$, therefore $m_k^t \geq m_t^k$. In 1966, Blakley and Dixon [18] investigated whether $m_k^t \geq m_t^k$ holds in the case $u = v$ when the nonnegativity assumption on the summands of Eq. (4.0.3) is replaced by the coordinate-wise nonnegativity of S, $u = v$. They note that the inequality $m_k^t \geq m_t^k$ fails when k and t have different parity and otherwise holds true under the restriction $m_t \geq e^{-4t}$. While the following is not explicitly stated as a conjecture in [18], they write

> if $t > 1$, [...] we cannot show that the inequality Eq. (4.0.1) holds for each nonnegative $|\Omega|$-vector u if S is nonnegative.

so with the earlier caveat we attribute the following to Blakley and Dixon [18]:

[1] Since $u = v$ here, the summands inside Eq. (4.0.3) are nonnegative when t is even so this theorem is most interesting for t odd.

Conjecture 4.0.2 (Blakley and Dixon [18])**.** *Let $S\colon \Omega \times \Omega \to \mathbb{R}_+$ be a symmetric matrix with nonnegative entries and let $u\colon \Omega \to \mathbb{R}_+$ be a nonnegative unit vector. For positive integers $k \geq t$ of the same parity, we have*

$$\left\langle u, S^k u \right\rangle^t \geq \left\langle u, S^t u \right\rangle^k .$$

In Section 4.1 we prove the following theorem which shows that a generalization of Conjecture 4.0.2 holds true.

Theorem 4.0.3. *Let $S\colon \Omega \times \Omega \to \mathbb{R}_+$ be a symmetric matrix with nonnegative entries and $u, v\colon \Omega \to \mathbb{R}_+$ be nonnegative unit vectors. For positive integers $k \geq t$ of the same parity, we have*

$$\left\langle v, S^k u \right\rangle^t \geq \left\langle v, S^t u \right\rangle^k .$$

It goes without saying that Eq. (4.0.1) is equivalent to Theorem 4.0.3 as we can rearrange Eq. (4.0.1) to $m_{t+2}^{1/(t+2)} \geq m_t^{1/t}$ and apply it iteratively to obtain $m_k^{1/k} \geq \cdots \geq m_{t+2}^{1/(t+2)} \geq m_t^{1/t}$ whenever $k \geq t$ and k, t have the same parity. Moreover, while defining Eq. (4.0.1) we assumed S to be substochastic only to illustrate our interpretation of the inequality: indeed any nonnegative S can be scaled to be substochastic as both sides of Eq. (4.0.1) are $(t+2)$-homogeneous in S.

In Theorem 4.0.1 and Theorem 4.0.3 we observed that increasingly more general implications of the log-convexity of $\{m_t\}_{t=0}^{\infty}$ can be derived by only assuming the coordinate-wise nonnegativity of S, u and v. One may naturally wonder if the coordinate-wise nonnegativity of S, u and v implies the log-convexity of $\{m_t\}_{t=0}^{\infty}$ in its entirety. Unfortunately the following example shows that this is far from the truth.

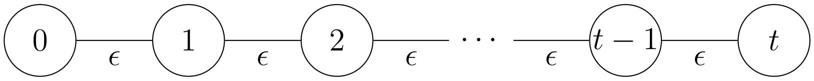

Figure 4.1: $\Omega = \{0, 1, \ldots, t\}$, $S(i, i+1) = S(i+1, i) = \epsilon$ for $i = 0, \ldots, t-1$.

Consider the transition matrix S on $\Omega = \{0, 1, \dots, t\}$ such that $S(i, i+1) = S(i+1, i) = \epsilon$ for $i = 0, 1, \dots, t-1$ and $S(i, j) = 0$ elsewhere. Let u and v be the point masses respectively on states 0 and t; namely $u = [1, 0, \dots, 0]^\intercal$ and $v = [0, 0, \dots, 1]^\intercal$. We have $m_{t-2} = 0$, $m_t = \epsilon^t$ and $m_{t+2} = t\epsilon^{t+2}$. Therefore $m_{t-2}m_{t+2} = 0 \not\geq \epsilon^{2t} = m_t^2$. In this example the log-convexity breaks (in the strongest possible way) because the states 0 and t are separated by t hops according to S and the point mass at state 0 cannot reach state t before the tth time step.

Our next theorem shows that such reachability issues are essentially the only way the log-convexity property can fail to hold:

Theorem 4.0.4. *For every $\epsilon > 0$ there is a $\delta > 0$ such that for any symmetric matrix $S \colon \Omega \times \Omega \to \mathbb{R}_+$ and unit vectors $u, v \colon \Omega \to \mathbb{R}_+$ with nonnegative entires, defining m_t as before, we have*

$$\frac{m_{t+2}}{m_t^{1+2/t}} \geq \min\left\{t^{1-\epsilon}, \delta \frac{m_t^{1-2/t}}{m_{t-2}}\right\}, \qquad \forall t \geq 2. \tag{4.0.4}$$

In other words, Theorem 4.0.4 shows that one can recover a truncated version of the log-convexity of $\{m_t\}_{t=0}^{\infty}$ from just the coordinate-wise nonnegativity assumption of S, u and v. We stress that Theorem 4.0.4 is tight up to the appearance of ϵ and the choice of $\delta = \delta(\epsilon)$. A direct calculation on Figure 4.1 for time steps $t, t+2, t+4$ shows that Eq. (4.0.4) cannot be improved to

$$\frac{m_{t+2}}{m_t^{1+2/t}} \geq \min\left\{t^{1-2/t}, \left(\frac{1+\eta}{2}\right) \frac{m_t^{1-2/t}}{m_{t-2}}\right\}$$

for $\eta > 0$.

4.0.2 Related work on m_t

Almost simultaneously with the work of Mandel and Hughes [64], Mulholland and Smith also prove Theorem 4.0.1 in [68] and moreover they characterize the equality conditions of the inequality. Independently, in 1965, Blakley and Roy [17] prove the same inequality and characterize the equality conditions and [62] provides an alternative proof to that of [68] in 1966. We remark that Theorem 4.0.1 is most commonly referred to as the Blakley-Roy bound or "Sidorenko's conjecture for walks". Note these results show that Conjecture 4.0.2 is true whenever t divides k. Finally in 2012, Pate shows that $m_t \geq m_1^t$ without the restriction $u = v$:

Theorem 4.0.5 (Pate [74])**.** *Let $S\colon \Omega \times \Omega \to \mathbb{R}_+$ be a symmetric matrix with nonnegative entries and let $u, v\colon \Omega \to \mathbb{R}_+$ be nonnegative unit vectors. It holds that*

$$\left\langle v, S^{2t+1}u \right\rangle \geq \langle v, Su \rangle^{2t+1},$$

with equality if and only if $\langle v, S^{2t+1}u \rangle = 0$ or $Su = \lambda v$ and $Sv = \lambda u$ for some $\lambda \in \mathbb{R}_+$.

This result already shows that Theorem 4.0.3 is true when t divides k but such a bound does not have any implications for our applications in complexity theory. In [34], Erdős and Simonovits conjecture the following.

Conjecture 4.0.6 (Erdős and Simonovits [34], Conjecture 6)**.** *For a graph $G = (V, E)$, let $w_k(G)$ be the number length k walks in G divided by $|V|$. For an undirected graph G, we have $w_k(G)^t \geq w_t(G)^k$ for $k > t$ of the same parity.*

This conjecture was recalled in a recet book [84] by Täubig as Conjecture 4.1. Note that Conjecture 4.0.6 is a specialization of Conjecture 4.0.2 to S having 0-1 entries and $u = \mathbf{1}/\sqrt{|V|}$ therefore our Theorem 4.0.3 verifies Conjecture 4.0.6 as well.

4.0.3 Our results in complexity theory

Here we list our results in complexity theory; see **??** for the definition of the models and the problems. The following theorem (which was already known [14]) is a consequence of Theorem 4.0.3 and uses the standard corruption technique in communication complexity.

Theorem 4.0.7. *Any two party δ-error randomized protocol solving the k-Hamming distance problem over length-n strings communicates at least $\Omega(k \log(1/\delta))$ bits for $k^2 \leq \delta n$.*

The next is our main result for the communication complexity of the k-Hamming distance problem and is a consequence of Theorem 4.0.4. This result cannot be obtained by the standard corruption technique and requires a suitable modification similar to [83].

Theorem 4.0.8. *Any two party δ-error randomized protocol solving the k-Hamming distance problem over length-n strings communicates at least $\Omega(k \log(k/\delta))$ bits for $k^2 \leq \delta n$.*

Theorem 4.0.9. *Any δ-error parity decision tree deciding the k-Hamming weight predicate over length-n strings has size $\exp \Omega\left(k \log(k/\delta)\right)$ for $k^2 < \delta n$.*

Corollary 4.0.10. *Any δ-error probability property tester for k-linearity requires $\Omega(k \log(k/\delta))$ queries.*

Note the bound $m_k^t \geq m_t^k$ obtained in [18] under the condition $m_t \geq e^{-4t}$ does not have any implications for the communication complexity of the k-Hamming distance problem as our reduction crucially uses the fact that u and v are arbitrary, however it does lead to an $\exp \Omega(k)$ lower bound for the parity decision tree size of the k-Hamming weight problem when combined with our reduction.

Remark 4.0.11. *Note that in Theorem 4.0.3, when $u = v$ and either both k, t are even or S is positive semidefinite, the summands of Eq. (4.0.3) become nonnegative and the inequality holds trivially. For our application in communication complexity we crucially use the fact that u and v are arbitrary and for our application in parity decision trees, one can do away with $u = v$ but only at the expense of having to choose k, t odd. In both results the S we choose has eigenvalues 1 and -1 with equal multiplicities and therefore far away from being positive semidefinite. In either case, the implications of Theorem 4.0.3 in complexity theory follow from the interesting cases of this theorem.*

4.1 Monotonicity of $t \mapsto m_{2t}^{1/(2t)}$ and $t \mapsto m_{2t+1}^{1/(2t+1)}$

In this section we prove Theorem 4.0.3 which we restate here (with additional equality conditions) for the convenience of the reader. Recall this theorem confirms Conjecture 4.0.2 and Conjecture 4.0.6.

Theorem 4.0.3 (restated)**.** *Let $S\colon \Omega \times \Omega \to \mathbb{R}_+$ be a symmetric matrix with nonnegative entries and $u, v\colon \Omega \to \mathbb{R}_+$ be nonnegative unit vectors. For positive integers $k \geq t$ of the same parity, we have*

$$\left\langle v, S^k u \right\rangle^t \geq \left\langle v, S^t u \right\rangle^k, \tag{4.1.1}$$

with equality if and only if $\left\langle v, S^k u \right\rangle = 0$ or $Su = \lambda v$ and $Sv = \lambda u$ for some $\lambda \in \mathbb{R}_+$ when t is odd and $u = v$ is an eigenvector of S^2 when t is even.

We prove Theorem 4.0.3 by an information theoretic argument. Define the distributions $\mu := u/\left\|u\right\|_1$ and $\nu := v/\left\|v\right\|_1$. Since either side of Eq. (4.1.1) is kt-homogeneous in S, we may assume that S is substochastic by scaling as needed. Having fixed this normalization, we view Eq. (4.1.1) as a statement about random walks on Ω that start from a state sampled according to μ or ν and evolve according to the transition matrix S.

4.1.1 Reference random walks

Let $\Omega_\circ = \Omega \cup \{r\}$ for some state $r \notin \Omega$ and t be a positive integer. Recall that $\mu = u/\left\|u\right\|_1$ and $\nu = v/\left\|v\right\|_1$. We start by defining random walks F^t, B^t on $\Omega_\circ$ that evolve in discrete time steps $-1, 0, 1, \ldots, t, t+1$.

The random walk F^t starts at r and transitions to a state $x \in \Omega$ with probability $\mu(x)$ at time step -1. In steps $0, 1, \ldots, t-1$, the random walk proceeds according to the transition matrix S. At the time step t, each state $x \in \Omega$ transitions to r with probability $\nu(x)$ and transitions to an arbitrary state in Ω with probability $1 - \nu(x)$ (say, all of them to the same arbitrary state). We view F^t as a joint random variable $F^t = (F^t_{-1}, F^t_0, \ldots, F^t_{t+1})$, where F^t_i is the location of the walk in time step i.

The random walk B^t proceeds backwards in time. At time step $t+1$ the walk B^t starts at r and transitions to a state $x \in \Omega$ with probability $\nu(x)$. In time steps $t, t-1, \ldots, 1$, the random walk proceeds as prescribed by S. At time step 0, each state $x \in \Omega$ transitions to r with probability $\mu(x)$ and to an arbitrary state in Ω with probability $1-\mu(x)$. Similarly, B^t denotes the joint random variable $B^t = (B^t_{-1}, B^t_0, \ldots, B^t_{t+1})$, where B^t_i is the location of the walk at time step i.

The following facts about F^t and B^t are immediate. The random variables F^t_{-1} and B^t_{t+1} are fixed to a single value r. The random variables F^t, B^t are Markovian, namely, $\operatorname{dist}(F^t_i \mid F^t_{i-1}, \ldots, F^t_{-1}) = \operatorname{dist}(F^t_i \mid F^t_{i-1})$ and $\operatorname{dist}(B^t_{i-1} \mid B^t_i, \ldots B^t_{t+1}) = \operatorname{dist}(B^t_{i-1} \mid B^t_i)$ for $i \in \{0, \ldots, t+1\}$.

4.1.2 Random walks returning to the origin

Assume that $\Pr[F^t_{t+1} = r] > 0$. Let X be the walk F^t conditioned on $F^t_{t+1} = r$. Note that X is a random variable on the sample space $\Omega_\circ^{t+3}$. The next two lemmas explicitly calculate the distribution of X.

For a matrix $M\colon \Omega\times\Omega \to \mathbb{R}_+$, functions $f, g\colon \Omega \to \mathbb{R}_+$, and $x, y \in \Omega$ we use the shorthands

$$\begin{aligned} M(f,y) &:= \sum_{x\in\Omega} f(x)M(x,y) = (M^\intercal f)(y) \\ M(x,g) &:= \sum_{y\in\Omega} M(x,y)g(y) = (Mg)(x) \\ M(f,g) &:= \sum_{x,y\in\Omega} f(x)M(x,y)g(y) = f^\intercal M g, \end{aligned}$$

where the last expression in each line is understood as a matrix vector multiplication.

Lemma 4.1.1. *Under our assumption $S^t(\mu,\nu) > 0$,*

(i) we have $\Pr[X_i = x] = \frac{S^i(\mu,x)S^{t-i}(x,\nu)}{S^t(\mu,\nu)}$, *and*

(ii) if $S^{t-i}(x,\nu) > 0$, we have $\Pr[X_{i+1} = y \mid X_{\le i} = x_{\le i}] = \frac{S(x_i,y)S^{t-i-1}(y,\nu)}{S^{t-i}(x_i,\nu)}$.

Proof. From the definition of F^t (cf. Section 4.1.1), we have

$$\Pr[F_i^t = x] = S^i(\mu, x) \tag{4.1.2}$$

$$\Pr[F_{t+1}^t = r \mid F_i^t = x] = S^{t-i}(x, \nu) \tag{4.1.3}$$

$$\Pr[F_{t+1}^t = r] = S^t(\mu, \nu) \tag{4.1.4}$$

$$\Pr[F_{i+1}^t = y \text{ and } F_{t+1}^t = r \mid F_i^t = x] = S(x, y)S^{t-i-1}(y, \nu). \tag{4.1.5}$$

Using Bayes' rule with Eq. (4.1.2), (4.1.3) and (4.1.4) gives (i). Combining Eq. (4.1.3), (4.1.5) and the observation that F^t is Markovian gives (ii). □

With Lemma 4.1.1 we confirm that the random variable $X = (X_{-1}, X_0, \ldots, X_{t+1})$ is Markovian; in particular a time inhomogeneous random walk on $\Omega_\circ$. Next we observe that the random variable B^t conditioned on $B_{-1}^t = r$ is precisely X also.

Lemma 4.1.2. *Under our assumption* $S^t(\mu, \nu) > 0$*,*

(i) we have $\operatorname{dist}(X) = \operatorname{dist}(B^t \mid B_{-1}^t = r)$*, and*

(ii) if $S^i(\mu, x) > 0$*, we have* $\Pr[X_{i-1} = y \mid X_{\geq i} = x_{\geq i}] = \frac{S(x_i, y)S^{i-1}(y, \mu)}{S^i(x_i, \mu)}$.

Proof. For any $x \in \Omega_\circ^{t+3}$ with $x_{t+1} = r$,

$$\begin{aligned}
\Pr[X = x] &= \frac{\mu(x_0)\prod_{i=1}^t S(x_{i-1}, x_i)\nu(x_t)}{S^t(\mu, \nu)} \\
&= \frac{\nu(x_t)\prod_{i=1}^t S(x_i, x_{i-1})\mu(x_0)}{S^t(\mu, \nu)} && \text{(as } S \text{ is symmetric)} \\
&= \frac{\Pr[B^t = x]}{\Pr[B_{t+1}^t = r]} = \Pr[B^t = x \mid B_{t+1}^t = r] && \text{(by Bayes' rule).}
\end{aligned}$$

This proves (i). Given (i), the proof of (ii) is the same as Lemma 4.1.1(ii). □

Lemma 4.1.3. *We have* $\mathbf{D}(X \,\|\, F^t) = \mathbf{D}(X \,\|\, B^t) = -\log S^t(\mu, \nu)$.

Proof. Recall that $\Pr[F_{t+1}^t = r] = S^t(\mu, \nu)$. Since X is obtained from F^t by conditioning on $F_{t+1}^t = r$, the equality criteria of Lemma 2.3.2 are fulfilled and thus $\mathbf{D}(X \,\|\, F^t) = -\log S^t(\mu, \nu)$. The derivation of $\mathbf{D}(X \,\|\, B^t)$ is identical as per Lemma 4.1.2(i). □

4.1.3 Longer random walks

Let J be an integer valued random variable taking the values $\{1, 2, \ldots, t\}$, each with equal probability. For each fixing j of J we perform a random walk $Z \,|\, J = j$ on $\Omega_\circ$ that evolves in time steps $-1, 0, 1, \ldots, t, t+1, t+2, t+3$ as follows.

The random walk starts at r and for each time step $-1 \leq i < j$, proceeds according to the transition kernel $\operatorname{dist}(X_{i+1} \,|\, X_i)$. At time step j, the random walk proceeds according to $\operatorname{dist}(X_{j-1} \,|\, X_j)$ and in time steps $j < i \leq t+3$ proceeds according to the transition kernel $\operatorname{dist}(X_{i-1} \,|\, X_{i-2})$. We view Z as a joint random variable $Z = (Z_{-1}, Z_0, \ldots, Z_{t+3})$, where Z_i denotes the location of the random walk at time step i.

Lemma 4.1.4. *For $-1 \leq i \leq j$, we have $\operatorname{dist}(Z_i \,|\, J = j) = \operatorname{dist}(X_i)$ and for $j < i \leq t+3$, $\operatorname{dist}(Z_i \,|\, J = j) = \operatorname{dist}(X_{i-2})$.*

Proof. This follows from the fact that

$$\operatorname{dist}(F^t \,|\, F^t_{t+1} = r) = \operatorname{dist}(X) = \operatorname{dist}(B^t \,|\, B^t_{-1} = r)$$

and that X is an actual random walk (i.e., Markovian) on $\Omega_\circ$.

To be more explicit, we have $\operatorname{dist}(X_i) = \operatorname{dist}(Z_i)$ for $i \leq j$ since both X and Z start at r in time step -1 and evolve according to the transition kernel $\operatorname{dist}(X_{i+1} \,|\, X_i)$ for $i = -1, \ldots, j-1$. Since $\operatorname{dist}(X_j) = \operatorname{dist}(Z_j)$ and Z proceeds according to $\operatorname{dist}(X_{i-1} \,|\, X_j)$ at time step j, by Lemma 4.1.2, $\operatorname{dist}(X_{j-1}) = \operatorname{dist}(Z_{j+1})$. Finally in time steps $i > j$, we have $\operatorname{dist}(Z_i) = \operatorname{dist}(X_{i-2})$ since $\operatorname{dist}(X_{j-1}) = \operatorname{dist}(Z_{j+1})$ and Z proceeds according to $\operatorname{dist}(X_{i-1} \,|\, X_{i-2})$. □

From this we can deduce that Z always ends up in r at time step $t+3$. We next argue that if X does not diverge too much from the reference random walk F^t, then Z does not diverge too much from F^{t+2}.

Lemma 4.1.5. *We have*

$$\begin{aligned}\mathbf{D}\Big(Z \,|\, J \,\Big\|\, F^{t+2}\Big) = \frac{t+2}{t}\mathbf{D}\Big(X \,\Big\|\, F^t\Big) &- \frac{1}{t}\Big(\mathbf{D}\Big(X_0 \,\Big\|\, F^t_0\Big) + \mathbf{D}\Big(X_{t+1} \,|\, X_t \,\Big\|\, F^t_{t+1} \,|\, F^t_t\Big)\Big) \\ &- \frac{1}{t}\Big(\mathbf{D}\Big(X_t \,\Big\|\, B^t_t\Big) + \mathbf{D}\Big(X_{-1} \,|\, X_0 \,\Big\|\, B^t_{-1} \,|\, B^t_0\Big)\Big).\end{aligned}$$

Proof. For a fixing j of J, we have

$$\begin{aligned}\mathbf{D}\Big(Z \,|\, J = j \,\Big\|\, F^{t+2}\Big) = \sum_{i=-1}^{j-1} \mathbf{D}\Big(X_{i+1} \,|\, X_i \,\Big\|\, F^{t+2}_{i+1} \,|\, F^{t+2}_{i}\Big) + \mathbf{D}\Big(X_{j-1} \,|\, X_j \,\Big\|\, F^{t+2}_{j+1} \,|\, F^{t+2}_{j}\Big) \\ + \sum_{i=j+1}^{t+2} \mathbf{D}\Big(X_{i-1} \,|\, X_{i-2} \,\Big\|\, F^{t+2}_{i+1} \,|\, F^{t+2}_{i}\Big)\,,\end{aligned}$$

where we have used the chain rule for divergence (cf. Lemma 2.3.3), the fact that $Z \,|\, J = j$ and F^{t+2} are Markovian and Lemma 4.1.4. Recalling that F^t and B^t evolve according to S in time steps $0, 1, \ldots, t-1$, and $\mathrm{dist}(F^t_{t+1} \,|\, F^t_t) = \mathrm{dist}(F^{t+2}_{t+3} \,|\, F^{t+2}_{t+2})$, we write

$$\begin{aligned}\mathbf{D}\Big(Z \,|\, J = j \,\Big\|\, F^{t+2}\Big) &= \sum_{i=-1}^{t} \mathbf{D}\Big(X_{i+1} \,|\, X_i \,\Big\|\, F^{t}_{i+1} \,|\, F^{t}_{i}\Big) \\ &\qquad + \mathbf{D}\Big(X_{j-1} \,|\, X_j \,\Big\|\, B^{t}_{j-1} \,|\, B^{t}_{j}\Big) + \mathbf{D}\Big(X_{j} \,|\, X_{j-1} \,\Big\|\, F^{t}_{j} \,|\, F^{t}_{j-1}\Big) \\ &= \mathbf{D}\Big(X \,\Big\|\, F^t\Big) + \mathbf{D}\Big(X_{j-1} \,|\, X_j \,\Big\|\, B^{t}_{j-1} \,|\, B^{t}_{j}\Big) + \mathbf{D}\Big(X_{j} \,|\, X_{j-1} \,\Big\|\, F^{t}_{j} \,|\, F^{t}_{j-1}\Big)\end{aligned}$$

again by the chain rule for divergence (Lemma 2.3.3) and the fact that X and F^t are Markovian. Now taking an expectation over all $j \in \mathrm{supp}(J)$, we have

$$\begin{aligned}\mathbf{D}\Big(Z \,|\, J \,\Big\|\, F^{t+2}\Big) &= \frac{1}{t}\sum_j \mathbf{D}\Big(Z \,|\, J = j \,\Big\|\, F^{t+2}\Big) \\ &= \mathbf{D}\Big(X \,\Big\|\, F^t\Big) + \frac{1}{t}\sum \Big(\mathbf{D}\Big(X_{j-1} \,|\, X_j \,\Big\|\, B^{t}_{j-1} \,|\, B^{t}_{j}\Big) + \mathbf{D}\Big(X_{j} \,|\, X_{j-1} \,\Big\|\, F^{t}_{j} \,|\, F^{t}_{j-1}\Big)\Big) \\ &= \mathbf{D}\Big(X \,\Big\|\, F^t\Big) + \frac{\mathbf{D}(X \,\|\, B^t) - \mathbf{D}(X_t \,\|\, B^t_t) - \mathbf{D}\Big(X_{-1} \,|\, X_0 \,\Big\|\, B^t_{-1} \,|\, B^t_0\Big)}{t} \\ &\qquad + \frac{\mathbf{D}(X \,\|\, F^t) - \mathbf{D}(X_0 \,\|\, F^t_0) - \mathbf{D}\Big(X_{t+1} \,|\, X_t \,\Big\|\, F^t_{t+1} \,|\, F^t_t\Big)}{t}.\end{aligned}$$

Since $\mathbf{D}(X \,\|\, B^t) = \mathbf{D}(X \,\|\, F^t)$ by Lemma 4.1.3, collecting the $\mathbf{D}(X \,\|\, F^t)$ terms we finish the proof. □

Finally, we lower bound the negative terms in the statement of Lemma 4.1.5.

Lemma 4.1.6. *We have*

$$\mathbf{D}\Big(X_0 \,\Big\|\, F^t_0\Big) + \mathbf{D}\Big(X_{-1} \,|\, X_0 \,\Big\|\, B^t_{-1} \,|\, B^t_0\Big) \geq \mathrm{H}_2(\mu) := -\log \|\mu\|_2^2\,, \text{ and} \tag{4.1.6}$$

$$\mathbf{D}\Big(X_t \,\Big\|\, B^t_t\Big) + \mathbf{D}\Big(X_{t+1} \,|\, X_t \,\Big\|\, F^t_{t+1} \,|\, F^t_t\Big) \geq \mathrm{H}_2(\nu) := -\log \|\nu\|_2^2\,, \tag{4.1.7}$$

where $\mathrm{H}_2(\cdot)$ *denotes the second order Rényi entropy.*

Proof. We only prove the first inequality as the second one is symmetric. By Lemma 4.1.1, we have

$$\mathbf{D}\left(X_0 \,\middle\|\, F_0^t\right) = \sum_{x\in\Omega} \frac{\mu(x)S^t(x,\nu)}{S^t(\mu,\nu)} \log \frac{S^t(x,\nu)}{S^t(\mu,\nu)} \quad \text{and}$$
$$\mathbf{D}\left(X_{-1} \mid X_0 \,\middle\|\, B_{-1}^t \mid B_0^t\right) = \sum_{x\in\Omega} \frac{\mu(x)S^t(x,\nu)}{S^t(\mu,\nu)} \log \frac{1}{\mu(x)}.$$

Let $\Psi = \mathrm{supp}(X_0)$. By adding the two terms we get

$$\begin{aligned}
\mathbf{D}\left(X_0 \,\middle\|\, F_0^t\right) + \mathbf{D}\left(X_{-1} \mid X_0 \,\middle\|\, B_{-1}^t \mid B_0^t\right) &= -\sum_{x\in\Psi} \frac{\mu(x)S^t(x,\nu)}{S^t(\mu,\nu)} \log \frac{\mu(x)S^t(\mu,\nu)}{S^t(x,\nu)} \\
&\geq -\log \sum_{x\in\Psi} \frac{\mu(x)^2 S^t(x,\nu) S^t(\mu,\nu)}{S^t(\mu,\nu) S^t(x,\nu)} && (4.1.8) \\
&= -\log \sum_{x\in\Psi} \mu(x)^2 \\
&\geq -\log \sum_{x\in\Omega} \mu(x)^2 \ , && (4.1.9)
\end{aligned}$$

where the first inequality is by concavity of $z \mapsto \log z$ and the second inequality is true as the summands are nonnegative. □

4.1.4 Combining the inequalities

Proof of Theorem 4.0.3. Note that Z_{-1} is fixed to r by definition (cf. Section 4.1.3) and Z_{t+3} is fixed to r by Lemma 4.1.4. Therefore by Lemma 2.3.2 we have

$$\begin{aligned}
-\log S^{t+2}(\mu,\nu) &\leq \mathbf{D}\left(Z \,\middle\|\, F^{t+2}\right) && (4.1.10) \\
&= \mathbf{D}\left(Z \mid J \,\middle\|\, F^{t+2}\right) - \mathbf{I}(J : Z) && (4.1.11) \\
&\leq \mathbf{D}\left(Z \mid J \,\middle\|\, F^{t+2}\right) && (4.1.12) \\
&\leq \frac{t+2}{t}\mathbf{D}\left(X \,\middle\|\, F^t\right) + \frac{\log \|\mu\|_2^2 + \log \|\nu\|_2^2}{t}.
\end{aligned}$$

Here Eq. (4.1.11) follows from the chain rule for the divergence (Lemma 2.3.3) and the definition of mutual information (cf. Eq. (2.3.4)), Eq. (4.1.12) follows from the nonnegativity of mutual information and the last line follows from Lemma 4.1.5 and Lemma 4.1.6. Plugging in $\mathbf{D}(X \parallel F^t) = -\log S^t(\mu, \nu)$, provided by Lemma 4.1.3, we obtain

$$-\log S^{t+2}(\mu,\nu) \leq -\frac{t+2}{t}\log S^t(\mu,\nu) + \frac{\log\|\mu\|_2^2 + \log\|\nu\|_2^2}{t}.$$

Arranging, we get

$$\|\mu\|_2^2 \, \|\nu\|_2^2 \left\langle \nu, S^{t+2}\mu \right\rangle^t \geq \left\langle \nu, S^t \mu \right\rangle^{t+2}$$

and substituting $\mu = u/\left\|u\right\|_1$, $\nu = v/\left\|v\right\|_1$, and recalling that u, v are unit vectors, we obtain

$$\begin{aligned} \left\langle v, S^{t+2}u \right\rangle^t &\geq \left\langle v, S^t u \right\rangle^{t+2}, \text{ i.e.,} \\ m_{t+2} &\geq m_t^{1+2/t}. \end{aligned} \tag{4.1.13}$$

By applying this inequality iteratively, we get $\left\langle v, S^k u \right\rangle^t \geq \langle v, S^t u\rangle^k$ or written differently $m_k^{1/k} \geq m_t^{1/t}$ as long as $k > t$ and k, t have the same parity.

Next we characterize the equality conditions of Eq. (4.1.13). Let us verify the 'if' direction of the statement. Clearly if $\left\langle v, S^k u \right\rangle = 0$ then we have $\left\langle v, S^k u \right\rangle = \langle v, S^t u\rangle$ by the first part of the theorem and the fact that $\langle v, S^t u\rangle \geq 0$. If $S^2 u = \lambda^2 u$, then $m_{2t} = \lambda^{2t}$ and if $Su = \lambda v$ and $Sv = \lambda u$, then $m_{2t+1} = \lambda^{2t+1}$, therefore in both t even and t odd cases the inequality holds with equality.

Conversely, if $0 \neq \left\langle v, S^{k+2}u \right\rangle = \left\langle v, S^k u \right\rangle$, then the inequalities (4.1.8), (4.1.9), (4.1.10), and (4.1.11) must hold with equality. Combining the assumption that Eq. (4.1.9) and (4.1.8) hold with equality with the strict concavity of $z \mapsto \log z$ and Jensen's lemma, we get that $S^t \nu = \lambda_1 \mu + \sigma_1$ for some $\lambda_1 \leq 1$ and $\sigma_1 \in \mathbb{R}_+^\Omega$ satisfying $\operatorname{supp}(\sigma_1) \cap \operatorname{supp}(\mu) = \emptyset$. This also means that $\Pr[X_0 = x] = \mu(x)^2/\left\|\mu\right\|_2^2$ for $x \in \Omega$ and a similar and symmetrical argument shows that $\Pr[X_t = x] = \nu(x)^2/\left\|\nu\right\|_2^2$ for $x \in \Omega$. Assuming Eq. (4.1.11) holds with equality leads to $\mathbf{I}(Z : J) = 0$, which in turn shows that $\operatorname{dist}(X_i) = \operatorname{dist}(X_{i+2})$ for $i = 0, \ldots, t-2$. Let $X^{k+2} := \left(F^{t+2} \,|\, F_{t+3}^{t+2} = r\right)$. From our assumption $0 \neq \left\langle v, S^{k+2}u \right\rangle = \left\langle v, S^k u \right\rangle$ we conclude that $\text{dist}(Z) = \text{dist}(X^{t+2})$ as X^{t+2} is the minimally divergent distribution from F^{t+2} among distributions on walks ending at state r. Since $\text{dist}(X_2^{t+2}) = \text{dist}(X_0^{t+2}) = \text{dist}(X_0)$, and $S^t\nu = \lambda_1\mu + \sigma_1$ we get that $S^2\mu = \lambda_2\mu + \sigma_2$ for some $\lambda_2 \leq 1$ and $\sigma_2 \in \mathbb{R}_+^\Omega$ satisfying $\text{supp}(\sigma_2) \cap \text{supp}(\mu) = \emptyset$. Now we will show that it must be that $\sigma_2 = 0$. Suppose for sake of contradiction that $\sigma_2(z) > 0$ for some $z \notin \text{supp}(\mu)$. There exists $x, y \in \Omega$ so that $\mu(x)S(x,y)S(y,z) > 0$. If $y \in \text{supp}(X_1)$ then adding to a walk $w \in \text{supp}(X)$ with $w_1 = y$ the loop $(y,z)(z,y)$ we obtain a length $t+2$ walk which is not in the support of X^{t+2} as $\text{dist}(X_2^{t+2}) = \text{dist}(X_2) = \text{dist}(X_0)$, which contradicts the fact that X^{t+2} is defined as $F^{t+2} \mid F_{t+3}^{t+2} = r$. If on the other hand $y \notin \text{supp}(X_1)$, adding the loop $(x,y)(y,x)$ to a walk with $w_0 = x$ leads to a walk which is not in the support of X^{t+2}, which is a contradiction. Having established $S^2\mu = \lambda\mu$, we complete the proof for even t by recalling that $\text{dist}(X_0) = \text{dist}(X_t)$ therefore $\mu = \nu$. For t odd, the last argument shows that $S\mu = \lambda_3\nu + \sigma_3$ for some $\lambda_3 \leq 1$ and $\sigma_3 \in \mathbb{R}_+^\Omega$ satisfying $\text{supp}(\sigma_3) \cap \text{supp}(\mu) = \emptyset$. It remains to show that $\sigma_3 = 0$ by using the assumption that $\text{dist}(Z) = \text{dist}(X^{t+2})$. Suppose $\sigma_3(y) > 0$ for some $y \notin \text{supp}(\nu)$. There exists $x \in \Omega$ such that $\mu(x)S(x,y) > 0$. Adding the loop $(x,y)(y,x)$ to a walk w with $w_{t-1} = x$ leads to a length $t+2$ walk which is not in the support of X^{t+2}. A symmetrical argument shows that $\lambda'\mu = S\nu$. This completes the proof. □

4.2 Near log-convexity of $t \mapsto m_{2t}$ and $t \mapsto m_{2t+1}$

In this section we would like to prove the following improvement to Eq. (4.1.13): for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$m_{t+2} \geq m_t^{1+2/t} \cdot \min\left\{t^{1-\epsilon}, \left\lceil \delta \frac{m_t^{1-2/t}}{m_{t-2}} \right\rceil\right\}, \quad \forall t \geq 2. \tag{4.2.1}$$

Recall that in proving Eq. (4.1.13), in line (4.1.12), we used the relaxation $\mathbf{I}(J : Z) \geq 0$. Note that J is uniformly distributed on $[t]$ therefore has $\log t$ bits of entropy and provided that it is possible to infer J from Z (i.e., it is possibly to locate the time reversal we have inserted in Z) the $\mathbf{I}(J : K)$ term appears to be large enough to recover the factor

$$\min\left\{t^{1-\epsilon}, \left\lceil \delta \frac{m_t^{1-2/t}}{m_{t-2}} \right\rceil\right\}.$$

Note moreover that intuitively we are able to infer J from Z better when $\frac{m_t^{1-2/t}}{m_{t-2}}$ is high, as in such cases on average for a time step $i \in [t]$ and a typical $x \sim X_i$, the distributions $\mathrm{dist}(X_{i-1} \,|\, X_i = x)$ and $\mathrm{dist}(X_{i+1} \,|\, X_i = x)$ should be far from each other, as otherwise we can argue that there should be many $t-2$ walks as follows. If $\mathrm{dist}(X_{i-1} \,|\, X_i = x)$ and $\mathrm{dist}(X_{i+1} \,|\, X_i = x)$ are close to each other, there should be many $p \in \Omega$ which has high probability in both these distributions. Sample such a p, and attach to it a walk sampled from $X_{-1}X_1 \ldots X_{i-1} \,|\, X_{i-1} = p$ and another walk sampled from $X_{i+1}X_{i+2} \ldots X_{t+1} \,|\, X_{i+1} = p$, which leads to a length $t-2$ walk returning to the origin. However if m_{t-2} is low, this should not happen and therefore $\mathrm{dist}(X_{i-1} \,|\, X_i = x)$ and $\mathrm{dist}(X_{i+1} \,|\, X_i = x)$ on average should be far apart, which means that we can notice when we take a step backwards in time and therefore infer J. In particular, Figure 4.1 gives such an example where $m_{t-2} = 0$ and we can always recover J with certainty from a sample from Z: whenever we take a step to the left, it must be that we are at time step J.

Given this discussion, a direct approach to proving Eq. (4.2.1) appears to bound

$$\mathbf{I}(Z : J) \geq \log \min\left\{t^{1-\epsilon}, \left\lceil \delta \frac{m_t^{1-2/t}}{m_{t-2}} \right\rceil\right\}. \tag{4.2.2}$$

Unfortunately, this approach does not seem to work as we demonstrate with an example in the full version of this paper. The problem here appears to be that we fix a single distribution Z to explore the two cases of Eq. (4.2.1). In our final approach, we pick different distributions depending on the case we would like to prove. Namely, if $\mathbf{I}(J : Z) \geq (1-\epsilon)\log t$, then carrying out the calculations in Eq. (4.1.10) through (4.1.13) with the assumption $\mathbf{I}(J : Z) \geq (1-\epsilon)\log t$, we prove the first case, namely $m_{t+2} \geq t^{1-\epsilon} m_t^{1+2/t}$ using the distribution given by Z. If $\mathbf{I}(J : Z) < (1-\epsilon)\log t$ on the other hand, we demonstrate two new random variables W, Y which are distributed respectively on length $t+2$ and length $t-2$ walks so that $\mathbf{D}(W \,\|\, F^{t+2}) + \mathbf{D}(Y \,\|\, F^{t-2}) \leq -2\log S^t(\mu,\nu) - \log\delta$, which implies that $m_{t-2}m_{t+2} \geq \delta m_t^2$. While W and Y are constructed by modifying X in suitable ways, which is how Z was constructed also, we do so with the hindsight of having inspected what causes $\mathbf{I}(J : Z)$ to be smaller than $(1-\epsilon)\log t$. It is precisely this adaptivity which enables this approach to overcome the difficulties encountered by the one suggested in Eq. (4.2.2).

If $\mathbf{I}(J : Z) \geq (1-\epsilon)\log t$, by plugging this into Eq. (4.1.12) and carrying out the following calculations, we get $m_{t+2} \geq t^{1-\epsilon} \cdot m_t^{1+2/t}$. Therefore it remains to show there exists a $\delta > 0$ such that assuming $\mathbf{I}(J : Z) < (1-\epsilon)\log t$, we have $m_{t+2}m_{t-2} \geq \delta m_t^2$. To do so, we will demonstrate random variables W and Y supported on walks that start from $r \in \Omega_\circ$ and return to r after spending respectively $t+2$ and $t-2$ time steps in Ω such that $\mathbf{D}(W \,\|\, F^{t+2}) + \mathbf{D}(Y \,\|\, F^{t-2}) \leq -2\log S^t(\mu,\nu) - \log\delta$. Notice that by Lemma 2.2.1 this indeed implies that $m_{t+2}m_{t-2} \geq \delta m_t^2$. The random variables W and Y will be mixtures of $\Theta(t)$ random walks, in particular, they are not Markovian in general.

For brevity let us set $\mu_i^x := \operatorname{dist}(X_{i-1} \,|\, X_i = x)$ and $\nu_i^x := \operatorname{dist}(X_{i+1} \,|\, X_i = x)$. Let U be the unary encoding of J: a length t bit vector of which only the Jth coordinate is set. First we would like to understand the contribution of each bit of U to $\mathbf{I}(Z : J) = \mathbf{I}(Z : U)$. Using the chain rule, we write

$$
\begin{aligned}
(1-\epsilon)\log t &> \mathbf{I}(U:Z) \\
&= \sum_{i=1}^{t} \mathbf{I}(U_i : Z \,|\, U_{<i}) && (4.2.3) \\
&= \sum_{i=1}^{t} \frac{t-i+1}{t} \mathbf{I}(U_i : Z \,|\, U_{<i} = 0) && (4.2.4) \\
&\geq \sum_{i=1}^{t} \frac{t-i+1}{t} \mathbf{I}(U_i : Z_i Z_{i+1} \,|\, U_{<i} = 0) && (4.2.5) \\
&= \sum_{i=1}^{t} \frac{1}{t} \mathop{\mathbb{E}}_{x\sim X_i} \mathbf{D}(\mu_i^x \,\|\, \lambda_i \mu_i^x + (1-\lambda_i)\nu_i^x) \\
&\qquad + \sum_{i=1}^{t} \frac{t-i}{t} \mathop{\mathbb{E}}_{x\sim X_i} \mathbf{D}(\nu_i^x \,\|\, \lambda_i \mu_i^x + (1-\lambda_i)\nu_i^x) && (4.2.6)
\end{aligned}
$$

where we set $\lambda_i := 1/(t-i+1)$, which is the probability that $U_i = 1 \,|\, U_{<i} = 0$. Here, Eq. (4.2.3) follows from the chain rule, Eq. (4.2.4) is true because if $U_{<i} \neq 0$ then $U_i = 0$ (as U has a single coordinate that is one) and consequently the mutual information is zero, and Eq. (4.2.5) is the data processing inequality. Next we lower bound Eq. (4.2.6) by its first term (which is valid since μ_i^x, ν_i^x are distributions hence the second term of Eq. (4.2.6) is nonnegative), obtaining

$$
(1-\epsilon)\log t > \mathop{\mathbb{E}}_{i\sim J} \mathop{\mathbb{E}}_{x\sim X_i} \mathbf{D}(\mu_i^x \,\|\, \lambda_i \mu_i^x + (1-\lambda_i)\nu_i^x) \ . \tag{4.2.7}
$$

To simplify the presentation, here we only provide the proof of Theorem 4.0.4 for $\epsilon > 7/8$ which demonstrates the ideas in their simplest form. This bound already implies all our results in complexity theory, with a constant factor loss of no more than 8. The proof for any $\epsilon > 0$ can be found in the full version of this paper.

4.2.1 The bound for $\epsilon > 7/8$

If we condition on the event $i \in \{1, \ldots, \lceil t/2 \rceil\}$, this expectation increases by a factor of at most 2; namely

$$
\mathop{\mathbb{E}}_{i\sim[t/2]} \mathop{\mathbb{E}}_{x\sim X_i} \mathbf{D}(\mu_i^x \,\|\, \lambda_i \mu_i^x + (1-\lambda_i)\nu_i^x) < 2(1-\epsilon)\log t.
$$

By Markov's inequality

$$\Pr_{i\sim[t/2],x\sim X_i}[\mathbf{D}(\mu_i^x \,\|\, \lambda_i\mu_i^x + (1-\lambda_i)\nu_i^x) \geq 8(1-\epsilon)\log t] < 1/4,$$

so it follows that there is a set $T \subseteq [\lceil t/2\rceil]$ of size at least $\lfloor t/4\rfloor$ such that if $i \in T$ we have

$$\Pr_{x\sim X_i}[\mathbf{D}(\mu_i^x \,\|\, \lambda_i\mu_i^x + (1-\lambda_i)\nu_i^x) \geq 8(1-\epsilon)\log t] < 1/2. \tag{4.2.8}$$

For each $i \in T$ let X_i' be the random variable obtained from X_i by conditioning on those $x \in \operatorname{supp}(X_i)$ satisfying $\mathbf{D}(\mu_i^x \,\|\, \lambda_i\mu_i^x + (1-\lambda_i)\nu_i^x) < 8(1-\epsilon)\log t$. Furthermore, for each $i \in T$ and $x \in \operatorname{supp}(X_i')$, we construct distributions $\pi_i^x \colon \Omega \to \mathbb{R}_+$ to be specified later. Let P_i be sampled by $x \sim X_i'$ first and then picking $p \sim \pi_i^x$.

4.2.2 The distributions W and Y

Let K be an integer sampled uniformly at random from the set T (constructed in the previous section). For each fixing k of K, the random variables $W \mid K = k$ and $Y \mid K = k$ are random walks (i.e., they are Markovian) constructed as follows. We first pick $x, p \sim X_k'P_k$. The walk $Y \mid K = k$ is generated by concatenating a sample from $X_{-1}X_0 \ldots X_{k-1} \mid X_{k-1} = p$ and an independent sample from $X_{k+1} \ldots X_{t+1} \mid X_{k+1} = p$. The walk W is generated by concatenating a sample from $X_{-1}X_0 \ldots X_k \mid X_k = x$, the path (x,p) and (p,x') for an independent sample $x' \sim (X_k' \mid P_k = p)$ and an independent sample from $X_k \ldots X_{t+1} \mid X_k = x'$.

For $k \in T$ we define another random walk $\check{X}^k = (\check{X}^k_{-1}, \ldots, \check{X}^k_{t+1})$, only to be used in the analysis of W and Y. We sample $x \sim X_k'$ and set $\check{X}^k_k = x$. We pick the rest of the coordinates of $\check{X}^k$ according to the distribution $X \mid X_k = x$. Note that for any $k \in T$, we have

$$\mathbf{D}(\check{X}^k \,\|\, X) = \mathbf{D}(X_k' \,\|\, X_k) \leq 1$$

by Eq. (4.2.8) and Lemma 2.3.2 and the fact that both X and $\check{X}^k$ are Markovian.

Lemma 4.2.1. *We have*

$$\mathbf{D}\Big(W \mid K = k \,\Big\|\, F^{t+2}\Big) + \mathbf{D}\Big(Y \mid K = k \,\Big\|\, F^{t-2}\Big) \leq -2\log S^t(\mu,\nu) + 2 + \mathop{\mathbb{E}}_{x\sim X_k'} \mathbf{D}(\pi_k^x \,\|\, \mu_k^x) + \mathop{\mathbb{E}}_{x\sim X_k'} \mathbf{D}(\pi_k^x \,\|\, \nu_k^x)\,.$$

Proof. We have

$$\mathbf{D}\Big(W \mid K = k \,\Big\|\, F^{t+2}\Big) = \mathbf{D}(\check{X}^k \,\|\, F^t) + \mathbf{D}(P_k \mid X'_k \,\|\, F_{k+1} \mid F_k) + \mathbf{D}(X'_k \mid P_k \,\|\, F_{k+1} \mid F_k)$$

and further

$$\begin{aligned}\mathbf{D}\Big(Y \mid K = k \,\Big\|\, F^{t-2}\Big) &+ \mathbf{D}(P_k \mid X'_k \,\|\, F_{k+1} \mid F_k) + \mathbf{D}(X'_k \mid P_k \,\|\, F_{k+1} \mid F_k) \\ &= \mathbf{D}(\check{X}^k \,\|\, F^t) + \mathop{\mathbb{E}}_{x \sim X'_k} \mathbf{D}(\pi_k^x \,\|\, \mu_k^x) + \mathop{\mathbb{E}}_{x \sim X'_k} \mathbf{D}(\pi_k^x \,\|\, \nu_k^x)\,.\end{aligned}$$

Summing up the two inequalities and substituting $\mathbf{D}(\check{X}^k \,\|\, X) \leq 1$ we get the result. □

At this point, in light of Lemma 4.2.1, we could pick each π_k^x so that it minimizes $\mathbf{D}(\pi_k^x \,\|\, \mu_k^x) + \mathbf{D}(\pi_k^x \,\|\, \nu_k^x)$: the unique minimizer is given by $\pi_k^x = \sqrt{\mu_k^x \nu_k^x} / \langle \sqrt{\mu_k^x}, \sqrt{\nu_k^x} \rangle$. However doing so leads to W, Y which diverge from the F walk by more than a constant, and therefore is not good enough for our needs. To obtain better random variables W and Y, we crucially use the fact that W is a mixture of $\Theta(t)$ random walks. Namely, if we consider the entropy coming from the $\mathbf{I}(W : K)$ term also, a better strategy for picking the distributions π_k^x becomes available. By contrast, we do not use the fact that Y is a mixture and, in fact, it can be replaced by $Y \mid K = k_0$ where $k_0 = \arg\min_k \mathbf{D}(Y \mid K = k \,\|\, F^{t-2})$, however the averaged quantity $\mathbf{D}(Y \mid K \,\|\, F^{t-2})$ is far more convenient to work with.

4.2.3 The contribution of $\mathbf{I}(K : W)$

Similar to Eq. (4.2.7), we would like to understand the contribution of each time step $t \in T$ to $\mathbf{I}(K : W)$. Let V be the unary encoding of K: a length t bit vector of which only the Vth coordinate is set. Using the chain rule for mutual information

$$\begin{aligned}\mathbf{I}(W : V) &= \sum_{i \in T} \mathbf{I}(V_i : W \mid V_{<i}) \\ &\geq \mathop{\mathbb{E}}_{k \sim K} \mathop{\mathbb{E}}_{x \sim X'_k} \mathbf{D}(\pi_k^x \,\|\, \eta_i \pi_k^x + (1 - \eta_i) \widetilde{\upsilon}_k^x)\,,\end{aligned}$$

where $\eta_k = 1/\operatorname{rank}_T(k)$ and $\widetilde{\upsilon}_k^x := \mathbb{E}_{j>k : j \in T} \operatorname{dist}(\check{X}_{k+1}^j \mid \check{X}_k^j = x)$. Here $\operatorname{rank}_T(i)$ denotes the position of $i \in T$ when the elements of T are sorted in decreasing order. By Eq. (4.2.8), and the definition of X'_k, we have $\widetilde{v}_k^x(y) \le 2\nu_k^x(y)$ for all $y \in \Omega$. Therefore we conclude that

$$\mathbf{I}(W : K) \ge \mathop{\mathbb{E}}_{k\sim K} \mathop{\mathbb{E}}_{x\sim X'_k} \mathbf{D}(\pi_k^x \,\|\, \eta_k \pi_k^x + 2(1-\eta_k)\nu_k^x)\,. \tag{4.2.9}$$

Note in the above divergence expression the reference measure is not a probability distribution, which our definition permits (cf. Eq. (2.3.1)).

Recall our goal in this section is to upper bound $\mathbf{D}(W \,\|\, F^{t+2}) + \mathbf{D}(Y \,\|\, F^{t-2}) + 2\log S^t(\mu,\nu)$ by $\log 1/\delta$. Let us write

$$\begin{aligned} \mathbf{D}\Big(W \,\Big\|\, F^{t+2}\Big) + & \mathbf{D}\Big(Y \,\Big\|\, F^{t-2}\Big) + 2\log S^t(\mu,\nu) \\ &\le \mathbf{D}\Big(W \mid K \,\Big\|\, F^{t+2}\Big) + \mathbf{D}\Big(Y \mid K \,\Big\|\, F^{t-2}\Big) - \mathbf{I}(K : W) + 2\log S^t(\mu,\nu) \\ &\le 2 + \mathop{\mathbb{E}}_{k\sim K, x\sim X'_k} \mathbf{D}\Big(\pi_x^k \,\Big\|\, \mu_x^k\Big) + \mathbf{D}\Big(\pi_x^k \,\Big\|\, \nu_x^k\Big) - \mathbf{I}(K : W) \\ &\le 2 + \mathop{\mathbb{E}}_{k\sim K, x\sim X'_k} \mathop{\mathbb{E}}_{y\sim\pi_k^x} \log \frac{\eta_k \pi_k^x(y)^2 + 2(1-\eta_k)\nu_k^x(y)\pi_k^x(y)}{\mu_k^x(y)\nu_k^x(y)}, \end{aligned} \tag{4.2.10}$$

where the second inequality follows from Lemma 4.2.1 and the last inequality is obtained by plugging in Eq. (4.2.9). Note that the function $z \mapsto z\log(az^2 + bz)$ is strictly convex in $\mathbb{R}_+$ whenever $ab > 0$, therefore for each k, x there is a unique minimizer $(\pi_k^x)^*$ of Eq. (4.2.10), which can be calculated, say, using Lagrange multipliers. However, instead of the minimizer, we work with a simple approximation of it. For each $k \in T$ and $x \in \operatorname{supp}(X'_k)$, we let

$$\Psi_k^x := \left\{ y \in \Omega \;\middle|\; \nu_k^x(y) \ge \frac{\lambda_k}{1-\lambda_k}\mu_k^x(y) \right\}.$$

By definition of X'_k, we have $\mathbf{D}(\mu_i^x \,\|\, \lambda_i \mu_i^x + (1-\lambda_i)\nu_i^x) < 8(1-\epsilon)\log t$. Let $\gamma = 1 - 8(1-\epsilon)$, which is positive by our assumption $\epsilon > 7/8$. By Markov's inequality, and the fact that $\lambda_k \le 2/t$, we get

$$\mu_k^x(\Psi_k^x) \ge \gamma$$

for large enough t. Let π_k^x be $\mu_k^x \mid \Psi_k^x$, namely we have $\pi_k^x(y) = \mu_k^x(y)/\mu_k^x(\Psi_k^x)$ if $y \in \Psi_k^x$, and $\pi_k^x(y) = 0$ otherwise. Continuing from Eq. (4.2.10), we have

$$\begin{aligned} &\le 2 + \mathop{\mathbb{E}}_{k\sim K, x\sim X'_k} \mathop{\mathbb{E}}_{y\sim\pi_k^x} \log \frac{\eta_k \pi_k^x(y)^2 + 2(1-\eta_k)\nu_k^x(y)\pi_k^x(y)}{\mu_k^x(y)\nu_k^x(y)} \\ &\le 2 + \mathop{\mathbb{E}}_{k\sim K} \log\left(\frac{\eta_k(1-\lambda_k)}{\lambda_k\gamma^2} + \frac{2}{\gamma}\right), \end{aligned} \tag{4.2.11}$$

where the second inequality is true by definition of Ψ_k^x and π_k^x. Now we argue that the expectation term in Eq. (4.2.11) is maximized when T is the set containing the smallest $|T|$ elements of $[\lceil t/2 \rceil]$. To see this suppose there is an $i \notin T$ which is smaller than the maximum element of T. Let j be the smallest item in T which is greater than i. We see that the expectation term increases if we replace T by $T \setminus \{j\} \cup \{i\}$ as $\log\left(\frac{C(1-\lambda_k)}{\lambda_k \gamma^2} + \frac{2}{\gamma}\right)$ is decreasing in k and the ranks do not change after swapping j with i. Therefore,

$$\begin{aligned}
\mathbf{D}\left(W \,\middle\|\, F^{t+2}\right) + &\mathbf{D}\left(Y \,\middle\|\, F^{t-2}\right) + 2\log S^t(\mu,\nu) \\
&\leq 2 + \log\left(\prod_{i=1}^{|T|} \frac{t/2+3i}{i\gamma^2}\right)^{1/|T|} \\
&= \log\frac{12}{\gamma^2} + \log\left(\prod_{i=1}^{|T|} \frac{t/6+i}{i}\right)^{1/|T|} \\
&\leq \log\frac{12}{\gamma^2} + \log\binom{2|T|}{|T|}^{1/|T|} && (4.2.12) \\
&\leq \log\frac{48}{\gamma^2}, && (4.2.13)
\end{aligned}$$

where the $\binom{2|T|}{|T|}$ is the middle binomial coefficient, in the second inequality we use the fact $|T| > t/6$, and the last inequality is true as $\binom{2n}{n} < 2^{2n}$. Therefore it is enough to choose $\epsilon > 7/8$ and $\delta \leq \frac{(1-8(1-\epsilon))^2}{48} = \frac{4}{3}(\epsilon - {}^7\!/_8)^2$. We have established the following.

Theorem 4.0.4 (restated)**.** *For any $\epsilon > 7/8$ there is a $\delta > 0$ such that $m_{t+2} \geq t^{1-\epsilon} m_t^{1+2/t}$ unless $m_{t+2} m_{t-2} \geq \delta m_t^2$.*

4.3 Chapter notes

The results of this chapter were obtained in our FOCS 2018 paper [79].

Chapter 5

THE K-HAMMING DISTANCE PROBLEM

In this section we study the k-Hamming distance problem in communication complexity and it's incarnations in related computation models.

5.1 Communication complexity

In a two player communication problem the players, named Alice and Bob, receive separate inputs, respectively $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, and they communicate in order to compute the value $f(x, y)$ of a function $f \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ (known to both players). In an r-round protocol, the players can take at most r turns alternately sending each other a message (that is, a bit string) and the last player to receive a message declares the output of the protocol. A protocol can be *deterministic* or *randomized*; in the latter case the players can base their actions on a common random source and we measure the *error probability*: the maximum over inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$, of the probability that the output of the protocol differs from $f(x, y)$. The *communication cost* of a protocol is the maximum, over the inputs and the random string, of the total number of bits sent between the players. For a function $f \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, an integer r and $\delta \in [0, 1]$, we denote by $R^r_\delta(f)$ the minimum over all protocols for f having r-rounds and error probability at most δ, of the communication cost incurred. We define $R_\delta(f)$ similarly, but we take the maximum over δ-error protocols with no restriction on the number of rounds it uses.

In the k-Hamming distance problem, denoted Ham^n_k, the players receive length-n bit strings, respectively $x, y \in \{0, 1\}^n$, and are required determine if $\|x - y\|_1 \leq k$ or not. There is a well known one-round communication protocol which accomplishes this with error probability δ by communicating $O(k \log{(k/\delta)})$ bits.

Theorem 5.1.1 (e.g., Huang, Shi, Zhang and Zhu [46])**.** *It holds that*

$$R^1_\delta(\mathrm{Ham}^n_k) = O(\min\{k\log(k/\delta), k\log(n/k)\}).$$

Highly related to the Ham^n_k is the k-disjointness problem Disj^n_k, wherein the players each receive a k-subset of $[n]$ and their goal is to determine if their sets intersect. Notice that Disj^n_k can be seen as a promise version of Ham^n_{2k-2} where each player is guaranteed to have a string with Hamming weight k: the sets are disjoint if and only if the Hamming distance between the characteristic vectors of the sets is more than $2k-2$. Therefore any upper bound for the Ham^n_k carries over to Disj^n_k and any lower bound for Disj^n_k carries over to Ham^n_k. Around 1993, Håstad and Wigderson [48] showed that there is a more efficient protocol for Disj^n_k than that implied by Theorem 5.1.1, which communicates only $O(k)$ bits, but over $O(\log k)$ rounds.

On the lower bounds side, the result of [55] implies that $\Omega(k)$ bits is needed for these problems even if one uses arbitrarily large number of round protocols. In [23] it was shown that any 1-round protocol for Disj^n_k needs to communicate at least $\Omega(k\log k)$ bits when $k^2 < n$ (this result was proven later in [?] also). In Theorem 3.2 of [78], an $\Omega(k\log(1/\delta))$ bound for 1-round complexity of Disj^n_k was shown even when Bob receives just one element (i.e., the indexing problem) for $k < \delta n$ and a slightly more general result was shown in [53]. Finally, in [80] the communication complexity of Disj^n_k was settled as presented in Chapter 5 of this thesis.

$$R^r_{1/3}(\mathrm{Disj}^n_k) = \Theta(k\log^{(r)} k)$$

for $1 \le r \le \log^* k$ and $k^2 < n$. Their upper bound solves the disjointness problem with error probability at most $1/\exp k + 1/\exp^{(r)}(c\log^{(r)} k)$ for any $c > 1$ by communicating $O(k\log^{(r)} k)$ bits over r rounds. In fact bulk of the bits is sent in the first round and the rest of the rounds amount to an $O(k)$ bits of communication. Taking $r = \log^* k$, this leads to an $O(k)$ bits protocol with error probability that is exponentially small in k. Their lower bound shows that at least one message of size $\Omega(k\log^{(r)} k)$ bits needs to be sent by any r-round protocol, even if it has error probability $1/3$. Prior to this work, this lower bound provided the strongest

Problem	**Upper bound**	**Rounds**	**Error**	**Lower bound**	**Reference**
Ham_k^n	$O(k\log(k/\delta))$	1	δ		Folklore, [46]
	appl. ↓	any	δ	$\Omega(k\log(1/\delta))$	[14]
		any	δ	$\Omega(k\log(k/\delta))$	This work
Disj_k^n	$O(k\log(k/\delta))$	1	δ		Folklore
	$O(k)$	$O(\log k)$	$1/3$	applies ↑	[48]
	$O(k\log^{(r)} k)$	r	$1/\exp^{(r)}\left(c\log^{(r)} k\right)$		[80]
	$O(k)$	$\log^* k$	$1/\exp k$		[80]
		r	$1/3$	$\Omega(k\log^{(r)} k)$	[80]
		1	$1/3$	$\Omega(k\log k)$	[23, **?**]
		1	δ	$\Omega(k\log(1/\delta))$	[78, 53]
		any	$1/3$	$\Omega(k)$	[55]

Table 5.1: Known bounds for Ham_k^n and Disj_k^n. Each upper bound for Ham_k^n applies to Disj_k^n and each lower bound for Disj_k^n applies to Ham_k^n.

lower bound for Ham_k^n also, along with the incomparable bound of $\Omega(k\log(1/\delta))$ due to [14] which holds for any number of rounds, which we discuss shortly.

To summarize the above results, the 1-round communication complexity of both Disj_k^n and Ham_k^n is $\Theta(k\log(k/\delta))$ by [23, 78, 53] and [46]. We know that Disj_k^n can be solved much more efficiently if one is allowed multiple rounds: firstly the $\log k$ factor can be removed [48] and secondly the error probability can be brought down to $\exp(-k)$, by using no more than $\log^* k$ rounds [80]. It is an interesting question whether similar efficiency improvements can be obtained for Ham_k^n also, by using multiple rounds. The first separation of Disj_k^n and Ham_k^n was proven in [14], which shows that $\Omega(k\log(1/\delta))$ lower bound holds for any protocol solving Ham^n_k. Therefore in Ham^n_k, we get no improvements in error probability by interactive communication. It remained an open question whether *any* improvement can be made at all to the 1-round protocol by communicating interactively. In this work we answer this question negatively:

Theorem 4.0.8 (restated)**.** *For $k^2 < \delta n$ we have $R_\delta(\mathrm{Ham}^n_k) = \Omega(k \log(k/\delta))$. The bound applies even to protocols that may output an arbitrary answer when $\|x - y\|_1 \notin \{k-2, k, k+2\}$.*

Before we proceed with proving Theorem 4.0.8, let us first warm up by showing that Theorem 4.0.3 implies an $\Omega(k \log(1/\delta))$ lower bound on $R_\delta(\mathrm{Ham}^n_k)$. To do so, let us review the so called *corruption bound* method. Let $f \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be the function the players would like to compute with Alice having received $x \in \mathcal{X}$ and Bob $y \in \mathcal{Y}$. For a protocol P for f, define the matrix $A_P \colon \mathcal{X} \times \mathcal{Y} \to [0,1]$ such that $A_P(x,y)$ is the probability that the protocol outputs 1 on input (x,y). It is well known and not difficult to see that if P has communication cost c, then A_P is the average of matrices each of which is the sum of at most 2^c rank 1 matrices $uv^\intercal$ with $u \in \{0,1\}^{\mathcal{X}}$ and $v \in \{0,1\}^{\mathcal{Y}}$. Therefore to show the communication cost of a protocol P is more than c, it suffices to argue A_P lies outside 2^c times the polytope

$$\mathcal{T} := \mathrm{conv}\left\{uv^\intercal \mid u \in \{0,1\}^{\mathcal{X}}, v \in \{0,1\}^{\mathcal{Y}}\right\},$$

where conv denotes the convex hull. By convexity, A_p lies outside of $2^c\mathcal{T}$ if and only if there is a hyperplane (with normal H) separating the two; namely that $\langle A_P, H\rangle > 2^c \langle R, H\rangle$ for all vertices R of the polytope $\mathcal{T}$.

Let $\mu_k \colon \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}_+$ be the distribution on pairs (x,y) obtained as follows. Sample x uniformly at random and obtain y by flipping k coordinates of x chosen uniformly at random and with replacement (here if a coordinate gets flipped twice it reverts back to its initial value).

Theorem 4.0.7 (restated)**.** *For $k^2 < \delta n$ we have $R_\delta(\mathrm{Ham}^n_k) = \Omega(k \log(1/\delta))$. The bound applies even to protocols that may output an arbitrary answer when $\|x - y\|_1 \notin \{k, k+2\}$.*

Proof. Suppose we have a randomized protocol for Ham_k^n with error probability δ. Form the matrix A, where $A(x,y)$ is the probability that the protocol reports $\|x-y\|_1 \le k$ on input (x,y).

Set $H = \mu_k - \mu_{k+2}/(3\delta)$. Let us first argue that $\langle A, H\rangle \ge 1/3$. Note that when we sample k elements from $[n]$ uniformly at random, by a union bound, the probability that there is a collision is at most $\binom{k}{2}/n$, therefore μ_k chooses a pair (x,y) at distance k with probability at least $1-\binom{k}{2}/n$. Hence, $\langle A, \mu_k\rangle \ge (1-\delta)(1-\binom{k}{2}/n) > 1-3\delta/2$, where in the last step we used $k^2 < \delta n$. Similarly, we have $\langle A, \mu_{k+2}\rangle \le \delta + \binom{k+2}{2}/n \le 3\delta/2$, thus it follows that $\langle A, H\rangle \ge 1/3$ for $\delta \le 1/9$.

Next we argue that $\langle R, H\rangle < (3\delta)^{k/2}$ for any $R = uv^\intercal$ with $u, v \in \{0,1\}^n$. If $\langle R, \mu_k\rangle < (3\delta)^{k/2}$, we are done as $\langle R, \mu_{k+2}\rangle \ge 0$ and is a negative term in $\langle R, H\rangle$. If $\langle R, \mu_k\rangle \ge (3\delta)^{k/2}$ on the other hand, observing $\langle R, \mu_k\rangle = \left\langle v, W^k u\right\rangle/2^n$, where W is the normalized adjacency matrix of the Hamming cube, we have by Theorem 4.0.3

$$\left(\frac{\|u\|_2\,\|v\|_2}{2^n}\right)^{2/k}\left\langle v, W^{k+2}u\right\rangle \ge \left\langle v, W^k u\right\rangle^{1+2/k}.$$

Note $\|u\|_2\,\|v\|_2 \le 2^n$ since u, v are 0-1 vectors, therefore $\langle R, \mu_{k+2}\rangle \ge 3\delta\,\langle R, \mu_k\rangle$ and hence $\langle R, H\rangle \le 0$. In either case we have shown that $\langle R, H\rangle < (3\delta)^{k/2}$. This implies an $\log((3\delta)^{-k/2}/3) = \Omega(k\log(1/\delta))$ bits lower bound on $R_\delta(\mathrm{Ham}_k^n)$. □

Interestingly, Theorem 4.0.8 cannot be proved by a direct application of the corruption method described above. If we assume that the protocol is supposed to output 1 on inputs $\|x-y\|_1 \le k$, then there are vertices of the polytope $\mathcal{T}$ for which the $\Omega(k\log(1/\delta))$ bound of Theorem 4.0.7 is tight. If we assume that the protocol is supposed to output 1 on inputs $\|x-y\|_1 > k$ on the other hand, no bound above $\Omega(k)$ can be obtained, as there are vertices for which this is tight. If we insist however that the protocol outputs 1 for $\|x-y\|_1 = k$ and 0 for $\|x-y\|_1 \in \{k-2, k+2\}$ then a protocol with cost smaller than $O(k\log(k/\delta))$ would be in violation of the near log-convexity principle we established in Theorem 4.0.4 as we argue next. Of course, if we had a δ-error randomized protocol P for Ham_k^{n+2} outputting 1 when $\|x-y\|_1 \in \{k-2, k\}$ and 0 if $\|x-y\|_1 = k+2$ (but without any guarantees for other types of inputs), then given inputs $a, b \in \{0,1\}^n$ Alice and Bob can run P (say, in parallel) on instances $(00a, 00b)$ and $(00a, 11b)$ and declare $\|a-b\|_1 = k$ if P returns 1 on $(00a, 00b)$ and 0 on $(00a, 11b)$. This would lead to a protocol with twice the error probability and communication cost of P, deciding between $\|a-b\|_1 = k$, $\|a-b\|_1 = k-2$ and $\|a-b\|_1 = k+2$. The table below shows that P outputting 1 on $(00a, 00b)$ and 0 on $(00a, 11b)$ implies $\|a-b\|_1 = k$ or at least one invocation of P erred.

Input	$k-2$	k	$k+2$
$(00a, 00b)$	1	1	0
$(00a, 11b)$	1	0	?

Proof of Theorem 4.0.8. Suppose we have a δ-error randomized protocol that outputs 1 when $\|x-y\|_1 = k$ and 0 when $\|x-y\|_1 = k-2$ or $\|x-y\|_1 = k+2$.

Form the matrix A, where $A(x, y)$ is the probability that the protocol reports that $\|x-y\|_1 = k$ on input (x, y). Let $\alpha_1, \alpha_2 > 0$ be some reals so that Theorem 4.0.4 implies $m_{t+2} \geq t^{\alpha_1} m^{1+2/t}$ or $m_{t+2} m_{t-2} \geq \alpha_2 {m_t}^2$ for m_t defined in statement of this theorem.

Set $H = \mu_k - (\mu_{k-2} + \mu_{k+2})/(6\delta)$. Let us argue first that $\langle A, H\rangle \geq 1/3$. One can verify that $\langle A, \mu_k\rangle \geq (1-\delta)(1 - \binom{k}{2}/n) > 1 - 3\delta/2$ and $\langle A, \mu_{k-2}\rangle + \langle A, \mu_{k-2}\rangle < 3\delta$. Hence $\langle A, H\rangle \geq 1/3$ for $\delta \leq 1/9$.

We upper bound $\langle R, H\rangle$ for some rank-1 matrix $R = uv^{\intercal}$ with 0-1 values. Let W be the normalized adjacency matrix of the Hamming cube graph. Observe that $\langle R, \mu_k\rangle = \left\langle v, W^k u\right\rangle / 2^n$. By Theorem 4.0.4, either $\langle R, \mu_{k+2}\rangle \langle R, \mu_{k-2}\rangle \geq \alpha_2 \langle R, \mu_k\rangle^2$ or

$$\left(\frac{\|u\|_2 \|v\|_2}{2^n}\right)^{2/k} \langle R, \mu_{k+2}\rangle \geq k^{\alpha_1} \langle R, \mu_k\rangle^{1+2/k}.$$

In the former case,

$$\frac{\langle R, \mu_{k+2}\rangle + \langle R, \mu_{k-2}\rangle}{2} \geq \sqrt{\langle R, \mu_{k+2}\rangle \langle R, \mu_{k-2}\rangle} \geq \sqrt{\alpha_2} \langle R, \mu_k\rangle,$$

which implies that $\langle R, H\rangle < 0$ whenever $\delta < 2\sqrt{\alpha_2}/6$ (recall α_2 is a constant). In the latter case we get $\langle R, H\rangle < 0$ unless $6\delta \langle R, \mu_{k+2}\rangle \leq \langle R, \mu_k\rangle$, which implies, recalling $\|v\|_2 \|u\|_2 \leq 2^n$, that $k^{\alpha_1} \langle R, \mu_k \rangle^{2/k} < 6\delta$. From this we get

$$\langle R, H \rangle \leq \langle R, \mu_k \rangle \leq \left(\frac{6\delta}{k^{\alpha_1}} \right)^{k/2} ,$$

and hence $\langle R, H \rangle < \left(\frac{6\delta}{k^{\alpha_1}} \right)^{k/2}$ in every case and $R_\delta(\mathrm{Ham}_k^n) = \Omega(k \log(k/\delta))$ whenever $k^2 < \delta n$. □

For a protocol P, denote by $\Pi = \Pi(x, y)$ the random variable entailing all the messages communicated between the players on input (x, y). So far we have considered the communication cost of a protocol which is the maximum length of Π over all inputs and the configurations of the random source (these together determine the value of Π). When a distribution μ on the inputs is available, we may speak of a more refined notion of cost, *the internal information cost*, for a protocol P which is defined as

$$\mathsf{IC}_\mu(P) := \mathbf{I}(\Pi : Y \mid X) + \mathbf{I}(\Pi : X \mid Y) ,$$

where $(X, Y) \sim \mu$. Combining our Theorem 4.0.8 with a result of [57] which relates information and communication costs of a protocol under suitable circumstances, one can conclude that any randomized protocol for Ham_k^n has information cost $\Omega(k \log k)$ as well, under the distribution $\mu = (\mu_k + \mu_{k-2} + \mu_{k+2})/3$. However we note that instead of using Theorem 4.0.4 black-box, taking a closer look at the proof of Theorem 4.0.3 and not performing the relaxation provided in Lemma 4.1.6, we get the following more directly.

Theorem 5.1.2. *Let P be a protocol outputting 1 on pairs (x, y) having $\|x - y\|_1 = k$ with probability $1 - \delta$ and outputting 0 on pairs (x, y) having $\|x - y\|_1 \in \{k - 2, k + 2\}$ with probability $1 - \delta$. We have $\mathsf{IC}_{\mu_k}(P) = \Omega(k \log(k/\delta))$.*

Let us finally mention another highly related problem, the so called the gap Hamming distance problem. In GHD_k^n, each of the players receive a bit string, respectively $x, y \in \{0, 1\}^n$, with the promise that either $\|x - y\|_1 \leq k$ or $\|x - y\|_1 \geq k + \sqrt{k}$. Their goal is to determine which is the case for any given input. In [24], an $\Omega(k)$ lower bound for this problem was shown, which applies to protocols with any number of rounds. Here we conjecture an improvement to this bound and argue that it would follow from a natural analogue of Theorem 4.0.4 for continuous time Markov chains, which we discuss in Section 5.7.

Conjecture 5.1.3. *For $k < \delta n$, we have $R_\delta(\mathrm{GHD}_k^n) = \Omega(k \log(1/\delta))$.*

5.2 Parity decision trees

In the parity decision tree model, we are given a string $x \in \mathbb{F}_2^n$ and our goal is to determine whether x satisfies a fixed predicate $P\colon \mathbb{F}_2^n \to \{0,1\}$ by only making linear measurements of the form $\langle x, y\rangle$ for some $y \in \mathbb{F}_2^n$ we get to choose. Here, the inner product is over $\mathbb{F}_2^n$, and therefore we get a single bit answer for every measurement we make.

Such measurements can be identified by binary decision trees wherein each internal node is labeled by a $y \in \mathbb{F}_2^n$ denoting the linear measurement $\langle x, y\rangle$ we would make at that node and each leaf is labeled by a YES or a NO denoting the final decision we arrive. Given such a tree and an x, the output of the decision tree is obtained by a root to leaf walk, where at each internal node v with label y_v, we perform the measurement $\langle x, y_v\rangle$ and walk to the left child of v if $\langle x, y_v\rangle = 0$ and to its right child if $\langle x, y_v\rangle = 1$. If a leaf node is reached, the label of the node is taken as the answer of the decision tree. Two quantities we are concerned with are the depth and the size (i.e., the total number of nodes) of the tree.

A δ-error randomized decision tree is a distribution ν over decision trees such that for any fixed x, the sampled decision tree outputs the correct answer with probability at least $1-\delta$, where the randomness is over the choice of the decision tree from ν. The depth and the size of a randomized decision tree can be taken as the maximum over the decision trees in the support of ν (here, one can also take the average depth or size also; our result on decision tree size actually lower bounds this potentially smaller quantity).

For a predicate $P\colon \mathbb{F}_2^n \to \{0,1\}$, let $\mathrm{PD}_\delta(P)$ be the minimum, over all randomized decision trees T computing P with probability $1-\delta$, of the depth of T. Let $\mathrm{PS}_\delta(P)$ be the minimum, over all randomized decision trees T computing P with probability $1-\delta$, of the size of T.

The following inequalities are immediate

$$\begin{aligned} R_\delta(P \circ \oplus) &\leq 2\mathrm{PD}_\delta(P), \qquad (5.2.1)\\ \log \mathrm{PS}_\delta(P) &\leq \mathrm{PD}_\delta(P), \end{aligned}$$

where $P \circ \oplus$ is the two player communication game in which the two players are given strings $x, y \in \mathbb{F}_2^n$ and are required to calculate $P(x+y)$. We remark that $\log \mathrm{PS}_\delta$ is incomparable to R_δ in general.

Here we study the predicate H_k^n which equals 1 if and only if the Hamming weight of its input is precisely k. By Eq. (5.5.1) and a padding argument similar to the one we gave before the proof of Theorem 4.0.8, each lower bound for Ham_k^n listed in Table 5.1 applies to $\mathrm{PD}_\delta(H_k^n)$ as well. In [16] another direct $\Omega(k)$ bound for $\mathrm{PD}_\delta(H_k^n)$ was shown. In [12], showing an $\Omega(k \log k)$ lower bound to a variant of $\mathrm{PD}_\delta(H_k^n)$ to obtain tight bounds for k-linearity problem (see Section 5.6) was suggested. Finally, our Theorem 4.0.8 shows that $\mathrm{PD}_\delta(H_k^n) = \Omega(k \log(k/\delta))$, which is tight. Next we show the same bound holds even for $\log \mathrm{PS}_\delta(H_k^n)$.

Theorem 4.0.9 (restated)**.** *For $k^2 < \delta n$, $\log \mathrm{PS}_\delta(H_k^n) = \Omega(k \log(k/\delta))$.*

Proof. The proof is very similar to that of Theorem 4.0.8, so we only describe the differences.

Let T be a δ-error randomized parity decision tree computing H_k^n. Form $A\colon \mathbb{F}_2^n \to [0,1]$ so that $A(x)$ is the probability T outputs 1 on input $x \in \mathbb{F}_2^n$. Define the polytope

$$\mathcal{P} := \mathrm{conv}\left\{x \mapsto \mathbb{1}[Bx = c] \mid B \in \mathbb{F}_2^{n\times n}, c \in \mathbb{F}_2^n\right\}$$

whose vertices are indicator functions for affine subspaces of $\mathbb{F}_2^n$. Given a randomized parity decision tree, for each fixing of the randomness, the set of inputs that end up in a particular leaf of it is an affine subspace in $\mathbb{F}_2^n$. Therefore if T has at most s leaves, then A is inside $s\mathcal{P}$. It remains to demonstrate a hyperplane with normal H so that $\langle A, H\rangle > s\langle V, H\rangle$ for any vertex V of the polytope $\mathcal{P}$ for $s = \exp \Omega(k \log(k/\delta))$.

Let μ_k be a distribution on $\mathbb{F}_2^n$ obtained as follows. Start with the 0 vector, and flip a coordinate chosen uniformly at random with replacement k times. Here, flipping a coordinate an even number of times leaves it as 0. Set $H = \mu_k - (\mu_{k-2} + \mu_{k+2})/(6\delta)$.

First observe that $\langle A, \mu_k \rangle > (1-\delta)(1 - \binom{k}{2}/n) > 1 - 3\delta/2$ and $\langle A, \mu_{k+2} \rangle + \langle A, \mu_{k-2} \rangle < 3\delta$ so $\langle A, H \rangle \geq 1/3$ for $\delta \leq 1/9$. Next we would like to upper bound $\langle V, H \rangle$ for an indicator function V of an affine subspace $\{x \in \mathbb{F}_2^n \mid Bx = c\}$. The key observation is

$$\langle V, \mu_k \rangle = \left\langle \mathbb{1}_c, S^k \mathbb{1}_0 \right\rangle \tag{5.2.2}$$

where S is a stochastic matrix describing the following transition: For any $x \in \mathbb{F}_2^n$, sample a column y of $B \in \mathbb{F}_2^{n \times n}$ uniformly at random and transition to $x + y$. Namely, the right hand side of Eq. (5.5.2) describes the following probability. We start with the 0 vector in $\mathbb{F}_2^n$ and in each time step sample a uniform random column y of B and add y to the current state. We measure the probability of reaching $c \in \mathbb{F}_2^n$ at time step k. Having observed Eq. (5.5.2), and that $\|\mathbb{1}_0\|_2 = \|\mathbb{1}_c\|_2 = 1$, the rest of the proof is identical to that of Theorem 4.0.8: by Theorem 4.0.4, we either have

$$\left\langle \mathbb{1}_c, S^{k+2} \mathbb{1}_0 \right\rangle \left\langle \mathbb{1}_c, S^{k-2} \mathbb{1}_0 \right\rangle \geq \alpha_2 \left\langle \mathbb{1}_c, S^k \mathbb{1}_0 \right\rangle^2$$

or

$$\left\langle \mathbb{1}_c, S^{k+2} \mathbb{1}_0 \right\rangle \geq k^{\alpha_1} \left\langle \mathbb{1}_c, S^k \mathbb{1}_0 \right\rangle^{1+2/k}.$$

In either event, we conclude that $\langle V, H \rangle \leq \left(\frac{6\delta}{k^{\alpha_1}}\right)^{k/2}$. This completes the proof. □

Note in Theorem 4.0.8, we use Theorem 4.0.4 with a simple and fixed S (i.e., the standard random walk on the Hamming cube), but with complicated vectors u, v that come from the particular communication protocol whose communication cost we would like to lower bound. By contrast, in Theorem 4.0.9 the vectors u, v are simple point masses on states 0 and c but the matrix S is a convolution random walk on the Hamming cube that comes from the particular decision tree whose size we lower bound.

5.3 Property testing

In the property testing model, given black box access to an otherwise unknown function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$, our goal is to tell apart whether $x \in P$ for some fixed set of functions P or $\|f - g\|_1 \geq \epsilon 2^n$ for any $g \in P$. Here, the black box queries are done by providing an input $x \in \mathbb{F}_2^n$ to the function and observing $f(x)$.

A function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ is called k-linear if f is given by

$$f(x) = \sum_{i \in S} x_i$$

for some $S \subseteq [n]$ of size at most k. By combining our communication complexity lower bound Theorem 4.0.8 with the reduction technique developed in [15] or by combining our parity decision tree lower bound Theorem 4.0.9 with a reduction given in [12], one obtains the following.

Corollary 4.0.10 (restated)**.** *Any δ-error property testing algorithm for k-linearity with $\epsilon = 1/2$ requires $\Omega(k \log(k/\delta))$ queries.*

In fact through this, one obtains similar lower bounds to property testing for k-juntas, k-term DNFs, size-k formulas, size-k decision trees, k-sparse $\mathbb{F}_2$-polynomials; see [13, 26].

5.4 *The lower bound for* Ham_k^n

Theorem 4.0.8 (restated)**.** *For $k^2 < \delta n$ we have $R_\delta(\mathrm{Ham}_k^n) = \Omega(k \log(k/\delta))$. The bound applies even to protocols that may output an arbitrary answer when $\|x-y\|_1 \notin \{k-2, k, k+2\}$.*

Before we proceed with proving Theorem 4.0.8, let us first warm up by showing that Theorem 4.0.3 implies an $\Omega(k \log(1/\delta))$ lower bound on $R_\delta(\mathrm{Ham}_k^n)$. To do so, let us review the so called *corruption bound* method. Let $f \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be the function the players would like to compute with Alice having received $x \in \mathcal{X}$ and Bob $y \in \mathcal{Y}$. For a protocol P for f, define the matrix $A_P \colon \mathcal{X} \times \mathcal{Y} \to [0,1]$ such that $A_P(x,y)$ is the probability that the protocol outputs 1 on input (x,y). It is well known and not difficult to see that if P has communication cost c, then A_P is the average of matrices each of which is the sum of at most 2^c rank 1 matrices $uv^\intercal$ with $u \in \{0,1\}^{\mathcal{X}}$ and $v \in \{0,1\}^{\mathcal{Y}}$. Therefore to show the communication cost of a protocol P is more than c, it suffices to argue A_P lies outside 2^c times the polytope

$$\mathcal{T} := \mathrm{conv}\left\{uv^\intercal \mid u \in \{0,1\}^{\mathcal{X}}, v \in \{0,1\}^{\mathcal{Y}}\right\},$$

where conv denotes the convex hull. By convexity, A_p lies outside of $2^c\mathcal{T}$ if and only if there is a hyperplane (with normal H) separating the two; namely that $\langle A_P, H\rangle > 2^c \langle R, H\rangle$ for all vertices R of the polytope $\mathcal{T}$.

Let $\mu_k \colon \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}_+$ be the distribution on pairs (x,y) obtained as follows. Sample x uniformly at random and obtain y by flipping k coordinates of x chosen uniformly at random and with replacement (here if a coordinate gets flipped twice it reverts back to its initial value).

Theorem 4.0.7 (restated)**.** *For $k^2 < \delta n$ we have $R_\delta(\mathrm{Ham}_k^n) = \Omega(k \log(1/\delta))$. The bound applies even to protocols that may output an arbitrary answer when $\|x-y\|_1 \notin \{k, k+2\}$.*

Proof. Suppose we have a randomized protocol for Ham_k^n with error probability δ. Form the matrix A, where $A(x,y)$ is the probability that the protocol reports $\|x-y\|_1 \le k$ on input (x,y).

Set $H = \mu_k - \mu_{k+2}/(3\delta)$. Let us first argue that $\langle A, H\rangle \geq 1/3$. Note that when we sample k elements from $[n]$ uniformly at random, by a union bound, the probability that there is a collision is at most $\binom{k}{2}/n$, therefore μ_k chooses a pair (x, y) at distance k with probability at least $1 - \binom{k}{2}/n$. Hence, $\langle A, \mu_k\rangle \geq (1-\delta)(1-\binom{k}{2}/n) > 1 - 3\delta/2$, where in the last step we used $k^2 < \delta n$. Similarly, we have $\langle A, \mu_{k+2}\rangle \leq \delta + \binom{k+2}{2}/n \leq 3\delta/2$, thus it follows that $\langle A, H\rangle \geq 1/3$ for $\delta \leq 1/9$.

Next we argue that $\langle R, H\rangle < (3\delta)^{k/2}$ for any $R = uv^\intercal$ with $u, v \in \{0,1\}^n$. If $\langle R, \mu_k\rangle < (3\delta)^{k/2}$, we are done as $\langle R, \mu_{k+2}\rangle \geq 0$ and is a negative term in $\langle R, H\rangle$. If $\langle R, \mu_k\rangle \geq (3\delta)^{k/2}$ on the other hand, observing $\langle R, \mu_k\rangle = \left\langle v, W^k u\right\rangle/2^n$, where W is the normalized adjacency matrix of the Hamming cube, we have by Theorem 4.0.3

$$\left(\frac{\|u\|_2\|v\|_2}{2^n}\right)^{2/k}\left\langle v, W^{k+2}u\right\rangle \geq \left\langle v, W^k u\right\rangle^{1+2/k}.$$

Note $\|u\|_2\|v\|_2 \leq 2^n$ since u, v are 0-1 vectors, therefore $\langle R, \mu_{k+2}\rangle \geq 3\delta\langle R, \mu_k\rangle$ and hence $\langle R, H\rangle \leq 0$. In either case we have shown that $\langle R, H\rangle < (3\delta)^{k/2}$. This implies an $\log((3\delta)^{-k/2}/3) = \Omega(k\log(1/\delta))$ bits lower bound on $R_\delta(\mathrm{Ham}_k^n)$. □

Interestingly, Theorem 4.0.8 cannot be proved by a direct application of the corruption method described above. If we assume that the protocol is supposed to output 1 on inputs $\|x-y\|_1 \leq k$, then there are vertices of the polytope $\mathcal{T}$ for which the $\Omega(k\log(1/\delta))$ bound of Theorem 4.0.7 is tight. If we assume that the protocol is supposed to output 1 on inputs $\|x-y\|_1 > k$ on the other hand, no bound above $\Omega(k)$ can be obtained, as there are vertices for which this is tight. If we insist however that the protocol outputs 1 for $\|x-y\|_1 = k$ and 0 for $\|x-y\|_1 \in \{k-2, k+2\}$ then a protocol with cost smaller than $O(k\log(k/\delta))$ would be in violation of the near log-convexity principle we established in Theorem 4.0.4 as we argue next. Of course, if we had a δ-error randomized protocol P for Ham_k^{n+2} outputting 1 when $\|x-y\|_1 \in \{k-2, k\}$ and 0 if $\|x-y\|_1 = k+2$ (but without any guarantees for other types of inputs), then given inputs $a, b \in \{0,1\}^n$ Alice and Bob can run P (say, in parallel) on instances $(00a, 00b)$ and $(00a, 11b)$ and declare $\|a-b\|_1 = k$ if P returns 1 on $(00a, 00b)$ and 0 on $(00a, 11b)$. This would lead to a protocol with twice the error probability and communication cost of P, deciding between $\|a-b\|_1 = k$, $\|a-b\|_1 = k-2$ and $\|a-b\|_1 = k+2$. The table below shows that P outputting 1 on $(00a, 00b)$ and 0 on $(00a, 11b)$ implies $\|a-b\|_1 = k$ or at least one invocation of P erred.

Input	$k-2$	k	$k+2$
$(00a, 00b)$	1	1	0
$(00a, 11b)$	1	0	?

Proof of Theorem 4.0.8. Suppose we have a δ-error randomized protocol that outputs 1 when $\|x-y\|_1 = k$ and 0 when $\|x-y\|_1 = k-2$ or $\|x-y\|_1 = k+2$.

Form the matrix A, where $A(x,y)$ is the probability that the protocol reports that $\|x-y\|_1 = k$ on input (x,y). Let $\alpha_1, \alpha_2 > 0$ be some reals so that Theorem 4.0.4 implies $m_{t+2} \geq t^{\alpha_1} m^{1+2/t}$ or $m_{t+2} m_{t-2} \geq \alpha_2 {m_t}^2$ for m_t defined in statement of this theorem.

Set $H = \mu_k - (\mu_{k-2} + \mu_{k+2})/(6\delta)$. Let us argue first that $\langle A, H\rangle \geq 1/3$. One can verify that $\langle A, \mu_k\rangle \geq (1-\delta)(1-\binom{k}{2}/n) > 1 - 3\delta/2$ and $\langle A, \mu_{k-2}\rangle + \langle A, \mu_{k-2}\rangle < 3\delta$. Hence $\langle A, H\rangle \geq 1/3$ for $\delta \leq 1/9$.

We upper bound $\langle R, H\rangle$ for some rank-1 matrix $R = uv^\intercal$ with 0-1 values. Let W be the normalized adjacency matrix of the Hamming cube graph. Observe that $\langle R, \mu_k\rangle = \left\langle v, W^k u\right\rangle / 2^n$. By Theorem 4.0.4, either $\langle R, \mu_{k+2}\rangle \langle R, \mu_{k-2}\rangle \geq \alpha_2 \langle R, \mu_k\rangle^2$ or

$$\left(\frac{\|u\|_2 \|v\|_2}{2^n}\right)^{2/k} \langle R, \mu_{k+2}\rangle \geq k^{\alpha_1} \langle R, \mu_k\rangle^{1+2/k}.$$

In the former case,

$$\frac{\langle R, \mu_{k+2}\rangle + \langle R, \mu_{k-2}\rangle}{2} \geq \sqrt{\langle R, \mu_{k+2}\rangle \langle R, \mu_{k-2}\rangle} \geq \sqrt{\alpha_2} \langle R, \mu_k\rangle,$$

which implies that $\langle R, H\rangle < 0$ whenever $\delta < 2\sqrt{\alpha_2}/6$ (recall α_2 is a constant). In the latter case we get $\langle R, H\rangle < 0$ unless $6\delta \langle R, \mu_{k+2}\rangle \leq \langle R, \mu_k\rangle$, which implies, recalling $\|v\|_2 \|u\|_2 \leq 2^n$, that $k^{\alpha_1} \langle R, \mu_k\rangle^{2/k} < 6\delta$. From this we get

$$\langle R, H\rangle \leq \langle R, \mu_k\rangle \leq \left(\frac{6\delta}{k^{\alpha_1}}\right)^{k/2},$$

and hence $\langle R, H\rangle < \left(\frac{6\delta}{k^{\alpha_1}}\right)^{k/2}$ in every case and $R_\delta(\mathrm{Ham}_k^n) = \Omega(k\log(k/\delta))$ whenever $k^2 < \delta n$. □

For a protocol P, denote by $\Pi = \Pi(x, y)$ the random variable entailing all the messages communicated between the players on input (x, y). So far we have considered the communication cost of a protocol which is the maximum length of Π over all inputs and the configurations of the random source (these together determine the value of Π). When a distribution μ on the inputs is available, we may speak of a more refined notion of cost, *the internal information cost*, for a protocol P which is defined as

$$\mathsf{IC}_\mu(P) := \mathbf{I}(\Pi : Y \mid X) + \mathbf{I}(\Pi : X \mid Y)\,,$$

where $(X, Y) \sim \mu$. Combining our Theorem 4.0.8 with a result of [57] which relates information and communication costs of a protocol under suitable circumstances, one can conclude that any randomized protocol for Ham_k^n has information cost $\Omega(k\log k)$ as well, under the distribution $\mu = (\mu_k + \mu_{k-2} + \mu_{k+2})/3$. However we note that instead of using Theorem 4.0.4 black-box, taking a closer look at the proof of Theorem 4.0.3 and not performing the relaxation provided in Lemma 4.1.6, we get the following more directly.

Theorem 5.4.1. *Let P be a protocol outputting 1 on pairs (x, y) having $\|x - y\|_1 = k$ with probability $1 - \delta$ and outputting 0 on pairs (x, y) having $\|x - y\|_1 \in \{k - 2, k + 2\}$ with probability $1 - \delta$. We have $\mathsf{IC}_{\mu_k}(P) = \Omega(k\log(k/\delta))$.*

Let us finally mention another highly related problem, the so called the gap Hamming distance problem. In GHD_k^n, each of the players receive a bit string, respectively $x, y \in \{0, 1\}^n$, with the promise that either $\|x - y\|_1 \leq k$ or $\|x - y\|_1 \geq k + \sqrt{k}$. Their goal is to determine which is the case for any given input. In [24], an $\Omega(k)$ lower bound for this problem was shown, which applies to protocols with any number of rounds. Here we conjecture an improvement to this bound and argue that it would follow from a natural analogue of Theorem 4.0.4 for continuous time Markov chains, which we discuss in Section 5.7.

Conjecture 5.4.2. *For $k < \delta n$, we have $R_\delta(\mathrm{GHD}_k^n) = \Omega(k\log(1/\delta))$.*

5.5 Parity decision trees

In the parity decision tree model, we are given a string $x \in \mathbb{F}_2^n$ and our goal is to determine whether x satisfies a fixed predicate $P\colon \mathbb{F}_2^n \to \{0,1\}$ by only making linear measurements of the form $\langle x, y\rangle$ for some $y \in \mathbb{F}_2^n$ we get to choose. Here, the inner product is over $\mathbb{F}_2^n$, and therefore we get a single bit answer for every measurement we make.

Such measurements can be identified by binary decision trees wherein each internal node is labeled by a $y \in \mathbb{F}_2^n$ denoting the linear measurement $\langle x, y\rangle$ we would make at that node and each leaf is labeled by a YES or a NO denoting the final decision we arrive. Given such a tree and an x, the output of the decision tree is obtained by a root to leaf walk, where at each internal node v with label y_v, we perform the measurement $\langle x, y_v\rangle$ and walk to the left child of v if $\langle x, y_v\rangle = 0$ and to its right child if $\langle x, y_v\rangle = 1$. If a leaf node is reached, the label of the node is taken as the answer of the decision tree. Two quantities we are concerned with are the depth and the size (i.e., the total number of nodes) of the tree.

A δ-error randomized decision tree is a distribution ν over decision trees such that for any fixed x, the sampled decision tree outputs the correct answer with probability at least $1-\delta$, where the randomness is over the choice of the decision tree from ν. The depth and the size of a randomized decision tree can be taken as the maximum over the decision trees in the support of ν (here, one can also take the average depth or size also; our result on decision tree size actually lower bounds this potentially smaller quantity).

For a predicate $P\colon \mathbb{F}_2^n \to \{0,1\}$, let $\mathrm{PD}_\delta(P)$ be the minimum, over all randomized decision trees T computing P with probability $1-\delta$, of the depth of T. Let $\mathrm{PS}_\delta(P)$ be the minimum, over all randomized decision trees T computing P with probability $1-\delta$, of the size of T. The following inequalities are immediate

$$R_\delta(P \circ \oplus) \le 2\mathrm{PD}_\delta(P), \tag{5.5.1}$$
$$\log \mathrm{PS}_\delta(P) \le \mathrm{PD}_\delta(P),$$

where $P \circ \oplus$ is the two player communication game in which the two players are given strings

$x, y \in \mathbb{F}_2^n$ and are required to calculate $P(x+y)$. We remark that $\log \mathrm{PS}_\delta$ is incomparable to R_δ in general.

Here we study the predicate H_k^n which equals 1 if and only if the Hamming weight of its input is precisely k. By Eq. (5.5.1) and a padding argument similar to the one we gave before the proof of Theorem 4.0.8, each lower bound for Ham_k^n listed in Table 5.1 applies to $\mathrm{PD}_\delta(H_k^n)$ as well. In [16] another direct $\Omega(k)$ bound for $\mathrm{PD}_\delta(H_k^n)$ was shown. In [12], showing an $\Omega(k \log k)$ lower bound to a variant of $\mathrm{PD}_\delta(H_k^n)$ to obtain tight bounds for k-linearity problem (see **??**) was suggested. Finally, our Theorem 4.0.8 shows that $\mathrm{PD}_\delta(H_k^n) = \Omega(k \log(k/\delta))$, which is tight. Next we show the same bound holds even for $\log \mathrm{PS}_\delta(H_k^n)$.

Theorem 4.0.9 (restated)**.** *For $k^2 < \delta n$, $\log \mathrm{PS}_\delta(H_k^n) = \Omega(k \log(k/\delta))$.*

Proof. The proof is very similar to that of Theorem 4.0.8, so we only describe the differences.

Let T be a δ-error randomized parity decision tree computing H_k^n. Form $A\colon \mathbb{F}_2^n \to [0,1]$ so that $A(x)$ is the probability T outputs 1 on input $x \in \mathbb{F}_2^n$. Define the polytope

$$\mathcal{P} := \mathrm{conv}\left\{x \mapsto \mathbb{1}[Bx = c] \mid B \in \mathbb{F}_2^{n\times n}, c \in \mathbb{F}_2^n\right\}$$

whose vertices are indicator functions for affine subspaces of $\mathbb{F}_2^n$. Given a randomized parity decision tree, for each fixing of the randomness, the set of inputs that end up in a particular leaf of it is an affine subspace in $\mathbb{F}_2^n$. Therefore if T has at most s leaves, then A is inside $s\mathcal{P}$. It remains to demonstrate a hyperplane with normal H so that $\langle A, H\rangle > s\langle V, H\rangle$ for any vertex V of the polytope $\mathcal{P}$ for $s = \exp \Omega(k \log(k/\delta))$.

Let μ_k be a distribution on $\mathbb{F}_2^n$ obtained as follows. Start with the 0 vector, and flip a coordinate chosen uniformly at random with replacement k times. Here, flipping a coordinate an even number of times leaves it as 0. Set $H = \mu_k - (\mu_{k-2} + \mu_{k+2})/(6\delta)$.

First observe that $\langle A, \mu_k\rangle > (1-\delta)(1-\binom{k}{2}/n) > 1 - 3\delta/2$ and $\langle A, \mu_{k+2}\rangle + \langle A, \mu_{k-2}\rangle < 3\delta$ so $\langle A, H\rangle \geq 1/3$ for $\delta \leq 1/9$. Next we would like to upper bound $\langle V, H\rangle$ for an indicator function V of an affine subspace $\{x \in \mathbb{F}_2^n \mid Bx = c\}$. The key observation is

$$\langle V, \mu_k\rangle = \left\langle \mathbb{1}_c, S^k \mathbb{1}_0\right\rangle \tag{5.5.2}$$

where S is a stochastic matrix describing the following transition: For any $x \in \mathbb{F}_2^n$, sample a column y of $B \in \mathbb{F}_2^{n\times n}$ uniformly at random and transition to $x+y$. Namely, the right hand side of Eq. (5.5.2) describes the following probability. We start with the 0 vector in $\mathbb{F}_2^n$ and in each time step sample a uniform random column y of B and add y to the current state. We measure the probability of reaching $c \in \mathbb{F}_2^n$ at time step k. Having observed Eq. (5.5.2), and that $\|\mathbb{1}_0\|_2 = \|\mathbb{1}_c\|_2 = 1$, the rest of the proof is identical to that of Theorem 4.0.8: by Theorem 4.0.4, we either have

$$\left\langle \mathbb{1}_c, S^{k+2}\mathbb{1}_0 \right\rangle \left\langle \mathbb{1}_c, S^{k-2}\mathbb{1}_0 \right\rangle \geq \alpha_2 \left\langle \mathbb{1}_c, S^{k}\mathbb{1}_0 \right\rangle^2$$

or

$$\left\langle \mathbb{1}_c, S^{k+2}\mathbb{1}_0 \right\rangle \geq k^{\alpha_1} \left\langle \mathbb{1}_c, S^{k}\mathbb{1}_0 \right\rangle^{1+2/k}.$$

In either event, we conclude that $\langle V, H\rangle \leq \left(\frac{6\delta}{k^{\alpha_1}}\right)^{k/2}$. This completes the proof. □

Note in Theorem 4.0.8, we use Theorem 4.0.4 with a simple and fixed S (i.e., the standard random walk on the Hamming cube), but with complicated vectors u, v that come from the particular communication protocol whose communication cost we would like to lower bound. By contrast, in Theorem 4.0.9 the vectors u, v are simple point masses on states 0 and c but the matrix S is a convolution random walk on the Hamming cube that comes from the particular decision tree whose size we lower bound.

5.6 *Property testing*

In the property testing model, given black box access to an otherwise unknown function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$, our goal is to tell apart whether $x \in P$ for some fixed set of functions P or $\|f-g\|_1 \geq \epsilon 2^n$ for any $g \in P$. Here, the black box queries are done by providing an input $x \in \mathbb{F}_2^n$ to the function and observing $f(x)$.

A function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ is called k-linear if f is given by

$$f(x) = \sum_{i\in S} x_i$$

for some $S \subseteq [n]$ of size at most k. By combining our communication complexity lower bound Theorem 4.0.8 with the reduction technique developed in [15] or by combining our parity decision tree lower bound Theorem 4.0.9 with a reduction given in [12], one obtains the following.

Corollary 4.0.10 (restated)**.** *Any δ-error property testing algorithm for k-linearity with $\epsilon = 1/2$ requires $\Omega(k \log(k/\delta))$ queries.*

In fact through this, one obtains similar lower bounds to property testing for k-juntas, k-term DNFs, size-k formulas, size-k decision trees, k-sparse $\mathbb{F}_2$-polynomials; see [13, 26].

5.7 Discussion

We showed that for a symmetric matrix $S\colon \Omega \times \Omega \to \mathbb{R}_+$ and unit vectors $u, v\colon \Omega \to \mathbb{R}_+$, defining $m_t = \langle v, S^t u\rangle$ for $t = 0, 1, \ldots$, we have

$$m_{t+2} \geq m_t^{1+2/t}, \text{ and} \tag{5.7.1}$$

$$m_{t+2} \geq m_t^{1+2/t} \cdot \min \left\{ t^{1-\epsilon}, \left\lceil \delta \frac{m_t^{1-2/t}}{m_{t-2}} \right\rceil \right\} \tag{5.7.2}$$

and argued that Eq. (5.7.2) and (5.7.1), in this order, are best viewed as gradual weakenings of the log-convexity of $\{m_t\}_{t=0}^{\infty}$. We conjecture that a similar principle holds true for continuous time Markov chains as well.

Call a function $f\colon \mathbb{R}_+ \to [0,1]$, whose logarithm is continuously twice differentiable (i.e., $\log f \in C^2(\mathbb{R}_+)$), *nearly-log-convex* if $x^2(\log f)''(x) \geq 2 \log f(x)$ for $x \in \mathbb{R}_+$. Note that $\log f \leq 0$, therefore this is a weakening of the usual log-convexity definition, which requires $(\log f)'' \geq 0$.

Conjecture 5.7.1. *Let $S\colon \Omega \times \Omega \to \mathbb{R}_+$ be a symmetric substochastic matrix and $u, v\colon \Omega \to \mathbb{R}_+$ be unit vectors. The function*

$$t \mapsto \left\langle v, e^{t(S-I)} u \right\rangle$$

is nearly-log-convex.

By an argument similar to the proof of Theorem 4.0.8, one can show the following.

Theorem 5.7.2. *Conjecture 5.7.1 implies Conjecture 5.4.2.*

5.8 Chapter notes

The results of this chapter have been published in FOCS 2018 in [79].

BIBLIOGRAPHY

[1] Farid Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 157:139–159, 1996.

[2] Anil Ada, Omar Fawzi, and Hamed Hatami. Spectral norm of symmetric functions. In Anupam Gupta, Klaus Jansen, José Rolim, and Rocco Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 338–349, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[3] Andris Ambainis, William Gasarch, Aravind Srinivasan, and Andrey Utis. Lower bounds on the deterministic and quantum communication complexity of hamming-distance problems. *ACM Trans. Comput. Theory*, 7(3):10:1–10:10, June 2015.

[4] M. Azizoğlu and Ö. Öğecioğlu. Extremal sets minimizing dimension-normalized boundary in Hamming graphs. *SIAM Journal on Discrete Mathematics*, 17(2):219–236, 2003.

[5] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *FOCS*, pages 337–347. IEEE Computer Society, 1986.

[6] Z. Bar-Yossef, T. S. Jayram, R. Krauthgamer, and R. Kumar. Approximating edit distance efficiently. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 550–559, 10 2004.

[7] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[8] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In Schulman [81], pages 67–76.

[9] Paul Beame and Faith E. Fich. Optimal bounds for the predecessor problem and related problems. *Journal of Computer and System Sciences*, 65:2002, 2001.

[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 *FOCS*, pages 477–486. IEEE Computer Society, 2008.

[11] Sergei Bezrukov. Isoperimetric problems in discrete spaces. In *Bolyai Soc. Math. Stud*, pages 59–91, 1994.

[12] Abhishek Bhrushundi, Sourav Chakraborty, and Raghav Kulkarni. Property testing bounds for linear and quadratic functions via parity decision trees. In Edward A. Hirsch, Sergei O. Kuznetsov, Jean-Éric Pin, and Nikolay K. Vereshchagin, editors, *Computer Science - Theory and Applications*, pages 97–110, Cham, 2014. Springer International Publishing.

[13] Eric Blais. Testing juntas nearly optimally. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 151–158. ACM, 2009.

[14] Eric Blais, Joshua Brody, and Badih Ghazi. The Information Complexity of Hamming Distance. In Klaus Jansen, José D. P. Rolim, Nikhil R. Devanur, and Cristopher Moore, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*, volume 28 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 465–489, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[15] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *computational complexity*, 21(2):311–358, 06 2012.

[16] Eric Blais and Daniel Kane. Tight bounds for testing k-linearity. In Anupam Gupta, Klaus Jansen, José Rolim, and Rocco Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 435–446, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[17] G. R. Blakley and Prabir Roy. A hölder type inequality for symmetric matrices with nonnegative entries. *Proceedings of the American Mathematical Society*, 16(6):1244–1245, 1965.

[18] GR Blakley and RD Dixon. Hölder type inequalities in cones. *Journal of Mathematical Analysis and Applications*, 14:1–4, 1966.

[19] Béla Bollobás and Imre Leader. Edge-isoperimetric inequalities in the grid. *Combinatorica*, 11(4):299–314, 1991.

[20] Joshua Brody and Amit Chakrabarti. A multi-round communication lower bound for gap Hamming and some consequences. In *IEEE Conference on Computational Complexity*, pages 358–368. IEEE Computer Society, 2009.

[21] Joshua Brody, Amit Chakrabarti, and Ranganath Kondapally. Certifying equality with limited interaction. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:153, 2012.

[22] Joshua Brody, Amit Chakrabarti, Oded Regev, Thomas Vidick, and Ronald de Wolf. Better Gap-Hamming lower bounds via better round elimination. In Maria J. Serna, Ronen Shaltiel, Klaus Jansen, and José D. P. Rolim, editors, *APPROX-RANDOM*, volume 6302 of *Lecture Notes in Computer Science*, pages 476–489. Springer, 2010.

[23] Harry Buhrman, David Garcia-Soriano, Arie Matsliah, and Ronald de Wolf. The non-adaptive query complexity of testing k-parities, 2012.

[24] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM Journal on Computing*, 41(5):1299–1317, 2012.

[25] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS*, pages 270–278. IEEE Computer Society, 2001.

[26] Sourav Chakraborty, David Garcia-Soriano, and Arie Matsliah. Efficient sample extractors for juntas with applications. In *Proceedings of the 38th International Colloquim Conference on Automata, Languages and Programming - Volume Part I*, ICALP'11, pages 545–556, Berlin, Heidelberg, 2011. Springer-Verlag.

[27] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sums of observations. *Annals of Mathematical Statistics*, 23:409–507, 1952.

[28] F. R. Chung, R. L. Graham, P. Frankl, and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Comb. Theory Ser. A*, 43(1):23–37, September 1986.

[29] G. F. Clements. Sets of lattice points which contain a maximal number of edges. *Proc. Amer. Math. Soc.*, 27:13–15, 1971.

[30] Graham Cormode, Mike Paterson, Süleyman Cenk Sahinalp, and Uzi Vishkin. Communication complexity of document exchange. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '00, pages 197–206, Philadelphia, PA, USA, 2000. Society for Industrial and Applied Mathematics.

[31] Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006.

[32] Anirban Dasgupta, Ravi Kumar, and D. Sivakumar. Sparse and lopsided set disjointness via information theory. In *APPROX-RANDOM*, pages 517–528, 2012.

[33] Pavol Duris, Zvi Galil, and Georg Schnitger. Lower bounds on communication complexity. *Inf. Comput.*, 73(1):1–22, April 1987.

[34] P. Erdős and M. Simonovits. Compactness results in extremal graph theory. *Combinatorica*, 2(3):275–288, 9 1982.

[35] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over general poset domains. In *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing*, STOC '02, pages 474–483, New York, NY, USA, 2002. ACM.

[36] Dmitry Gavinsky. On the role of shared entanglement. *Quantum Information & Computation*, 8(1):82–95, 2008.

[37] Dmitry Gavinsky, Julia Kempe, and Ronald de Wolf. Quantum communication cannot simulate a public coin. *CoRR*, quant-ph/0411051, 2004.

[38] Josiah Willard Gibbs. *Elementary Principles in Statistical Mechanics.* Cambridge University Press, 1902. Cambridge Books Online.

[39] Oded Goldreich. On testing computability by small width obdds. In *Proceedings of the 13th International Conference on Approximation, and 14 the International Conference on Randomization, and Combinatorial Optimization: Algorithms and Techniques*, APPROX/RANDOM'10, pages 574–587, Berlin, Heidelberg, 2010. Springer-Verlag.

[40] Bernd Halstenberg and Rüdiger Reischuk. On different modes of communication (extended abstract). In *STOC*, pages 162–172, 1988.

[41] L.H. Harper. Optimal assignment of numbers to vertices. *J. Soc. Ind. Appl. Math.*, 12:131–135, 1964.

[42] Prahladh Harsha, Rahul Jain, David A. McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010.

[43] Sergiu Hart. A note on the edges of the n-cube. *Discrete Mathematics*, 14(2):157 – 163, 1976.

[44] F. Hausdorff. Summationsmethoden und momentfolgen i. *Mathematische Zeitschrift*, 9:74–109, 1921.

[45] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[46] Wei Huang, Yaoyun Shi, Shengyu Zhang, and Yufan Zhu. The communication complexity of the hamming distance problem. *Information Processing Letters*, 99(4):149 – 153, 2006.

[47] Johan Håstad and Avi Wigderson. The probabilistic communication complexity of disjointness of k-sets is $o(k)$. 1990.

[48] Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(11):211–219, 2007.

[49] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds: extended abstract. In *STOC*, pages 599–608, 2008.

[50] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *ICALP*, pages 300–315, 2003.

[51] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *IEEE Conference on Computational Complexity*, pages 285–296, 2005.

[52] T. S. Jayram and David P. Woodruff. The data stream space complexity of cascaded norms. In *FOCS*, pages 765–774, 2009.

[53] T. S. Jayram and David P. Woodruff. Optimal bounds for johnson-lindenstrauss transforms and streaming problems with sub-constant error. In *SODA*, pages 1–10, 2011.

[54] J. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193, 12 1906.

[55] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[56] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. Discrete Math.*, 3(2):255–265, 1990.

[57] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *SIAM Journal on Computing*, 44(5):1550–1572, 2015.

[58] D. L. Kleitman, M. M. Krieger, and B. L. Rotschild. Configurations maximizing the number of pairs of Hamming-adjacent lattice points. *Studies in Appl. Math.*, 50:115–119, 1971.

[59] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[60] Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, 1997.

[61] John H. Lindsey. Assignment of numbers to vertices. *The American Mathematical Monthly*, 71(5):508–516, 1964.

[62] David London. Inequalities in quadratic forms. *Duke Math. J.*, 33(3):511–522, 09 1966.

[63] Frédéric Magniez, Claire Mathieu, and Ashwin Nayak. Recognizing well-parenthesized expressions in the streaming model. In Schulman [81], pages 261–270.

[64] SPH Mandel and IM Hughes. Change in mean viability at a multiallelic locus in a population under random mating. *Nature*, 182(4627):63, 1958.

[65] Peter Bro Miltersen. Lower bounds for union-split-find related problems on random access machines, 1994.

[66] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998.

[67] Marco Molinaro, David Woodruff, and Grigory Yaroslavtsev. Beating the direct sum theorem in communication complexity with implications for sketching. In *Proceedings of 24th ACM-SIAM Symposium on Discrete Algorithms*, pages 1486–1502, 2013.

[68] H. P. Mulholland and C. A. B. Smith. An inequality arising in genetical theory. *The American Mathematical Monthly*, 66(8):673–683, 1959.

[69] Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67 – 71, 1991.

[70] C. Niculescu and L.E. Persson. *Convex Functions and their Applications: A Contemporary Approach*. CMS Books in Mathematics. Springer New York, 2005.

[71] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM J. Comput.*, 22(1):211–219, 1993.

[72] King F Pang and Abbas El Gamal. Communication complexity of computing the hamming distance. *SIAM J. Comput.*, 15(4):932–947, November 1986.

[73] 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.

[74] Thomas Pate. Extending the hölder type inequality of blakley and roy to non-symmetric non-square matrices. *Transactions of the American Mathematical Society*, 364(8):4267–4281, 2012.

[75] Mihai Pǎtraşcu. Cc4: One-way communication and a puzzle. `http://infoweekly.blogspot.com/2009/04/cc4-one-way-communication-and-puzzle.html`. Accessed: 31/03/2013.

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

[77] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[78] Mert Sağlam. Tight bounds for data stream algorithms and communication problems. Master's thesis, Simon Fraser University, School of Computing Science, 8888 University Drive, Burnaby, B.C. Canada V5A 1S6, 9 2011.

[79] Mert Sağlam. Near log-convexity of measured heat in (discrete) time and consequences. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 967–978, 2018.

[80] Mert Sağlam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. In *54th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2013, Berkeley, CA, October 26-29, 2013*, pages 678–687, 2013.

[81] Leonard J. Schulman, editor. *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*. ACM, 2010.

[82] Pranab Sen and Srinivasan Venkatesh. Lower bounds for predecessor searching in the cell probe model. *Journal of Computer and System Sciences*, 74(3):364 – 385, 2003. Computational Complexity 2003.

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

[84] Hanjo Täubig. *Matrix Inequalities for Iterative Systems*. CRC Press, Taylor & Francis Group, 2017.

[85] A. Wald. Sequential tests of statistical hypotheses. *Ann. Math. Statist.*, 16(2):117–186, 06 1945.

[86] David P. Woodruff. personal communication, 2008.

[87] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM.

[88] Andrew Chi-Chih Yao. On the power of quantum fingerprinting. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 77–81, New York, NY, USA, 2003. ACM.