Minimum Distance of Error Correcting Codes versus Encoding Complexity, Symmetry, and Pseudorandomness

by

Louay M.J. Bazzi

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2003

Author .....
Department of Electrical Engineering and Computer Science
August 22, 2003

Certified by .....
Sanjoy K. Mitter
Professor of Electrical Engineering
Thesis Supervisor

Certified by .....
Daniel A Spielman
Associate Professor of Mathematics
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Minimum Distance of Error Correcting Codes versus Encoding Complexity, Symmetry, and Pseudorandomness

by

Louay M.J. Bazzi

Submitted to the Department of Electrical Engineering and Computer Science
on August 22, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

We study the minimum distance of binary error correcting codes from the following perspectives:

- The problem of deriving bounds on the minimum distance of a code given constraints on the computational complexity of its encoder.
- The minimum distance of linear codes that are symmetric in the sense of being invariant under the action of a group on the bits of the codewords.
- The derandomization capabilities of probability measures on the Hamming cube based on binary linear codes with good distance properties, and their variations.

Highlights of our results include:

- A general theorem that asserts that if the encoder uses linear time and sub-linear memory in the general binary branching program model, then the minimum distance of the code cannot grow linearly with the block length when the rate is nonvanishing.
- New upper bounds on the minimum distance of various types of Turbo-like codes.
- The first ensemble of asymptotically good Turbo like codes. We prove that depth-three serially concatenated Turbo codes can be asymptotically good.
- The first ensemble of asymptotically good codes that are ideals in the group algebra of a group. We argue that, for infinitely many block lengths, a random ideal in the group algebra of the dihedral group is an asymptotically good rate half code with a high probability.
- An explicit rate-half code whose codewords are in one-to-one correspondence with special hyperelliptic curves over a finite field of prime order where the number of zeros of a codeword corresponds to the number of rational points.

- A sharp $O(k^{-1/2})$ upper bound on the probability that a random binary string generated according to a k-wise independent probability measure has any given weight.
- An assertion saying that any sufficiently log-wise independent probability measure looks random to all polynomially small read-once DNF formulas.
- An elaborate study of the problem of derandomizability of AC_0 by any sufficiently polylog-wise independent probability measure.
- An elaborate study of the problem of approximability of high-degree parity functions on binary linear codes by low-degree polynomials with coefficients in fields of odd characteristics.

Thesis Supervisor: Sanjoy K. Mitter
Title: Professor of Electrical Engineering

Thesis Supervisor: Daniel A Spielman
Title: Associate Professor of Mathematics

Acknowledgments

I am very lucky to have Prof. Sanjoy Mitter and Prof. Daniel Spielman as my co-advisors, and Prof. Madhu Sudan on my thesis committee.

The good advice, support, and encouragement of Prof. Sanjoy Mitter go far beyond this thesis to cover all my experience at MIT. I am greatly indebted to Sanjoy for the unique freedom he gave me in choosing research problems, the ideal and caring environment he provided for working on those problem, and his caring advice and continuous support that spanned all my graduate research and learning experience. I will be forever grateful to Sanjoy.

I am greatly thankful to Prof. Daniel Spielman for his good advice, valuable supervision, support, and encouragement. I would like to thank Dan for all the things I learned from him ranging from the research process to the communication skills. My experience with Dan was profoundly influential in clarifying and shaping my research objectives, and directing my research path.

I wish to thank Prof. Madhu Sudan for his encouragement, and many simulating discussions and useful pointers over the last three years. I am greatly thankful to Madhu also for his kindness, and for being always available for meetings and discussions.

I am thankful also to other people with home I had helpful discussions on problems related to this thesis. Thanks to Enrico Bombieri, Coustantine Caramanis, Noam Elkies, Johan de Jong, Amin Shokrollahi, and Rudiger Urbanke.

I would like to thank also Tom Richardson and Rudiger Urbanke for an exciting summer internship at Bell labs in 1999.

I wish to thank also the lecturers of many influential courses I took at MIT including Professors Sigurdur Helgason, Lars Hesselholt, Michael Hopkins, Victor Guillemin, David Karger, Amos Lapidoth, George Lusztig, Michael Sipser, Daniel Spielman, Gang Tian, Ravi Vakil, and George Verghese.

My experience at MIT would not have been nearly as enjoyable without my friends Mahdi, Saad, Ibrahim, el-Lakkis, Ghassan, Fadi, and all my friends in the Lebanese

gang at MIT, over a period ranging from the Toscis-stairs-boom episode until the post-coffee-house days.

I dedicate this thesis to my close and loving family to whom I owe it all. To my parents Hanan and Muhammad-Jamil, and to my brothers and sisters Abbass, Abd-al-Karim, Hiba, Mona, and Zeina.

This research was supported by NSF grant CCR-0112487, ARO-MURI Grant DAAD19-00-1-0466 (Data Fusion in Large Arrays of Microsensors), the NSF-KDI grant ECS-9873451, the MURI Grant: Vision Strategies and ATR Performance subcontract no. 654-21256 (with Brown University) and the Department of Defense MURI Grant: Complex Adaptive Networks for Cooperative Control, Subaward #03-132 (with University of Illinois).

Contents

Chapter 1

Introduction

Error correcting codes are essential for the design of reliable communication systems and are playing an increasingly important role in areas of complexity theory such as the study of pseudorandomness. The minimum distance of an error correcting code is the minimum Hamming distance between two distinct codewords. It is a fundamental parameter of code design that determines the maximum number of errors that can be corrected under any decoding algorithm.

In this thesis we study the minimum distance of binary error correcting codes from the following points of view:

- The problem of deriving bounds on the minimum distance of a code given constraints on the computational complexity of its encoder with applications to Turbo-like codes.

- The minimum distance of linear codes that are symmetric in the sense of being invariant under the action of a group on the bits of the codewords.

- The derandomization capabilities of probability measures on the Hamming cube having the small bias property, the limited independence property, or the almost limited independence property. Classical constructions of such probability measures are based purely on binary linear codes with good distance properties.

1.1 Overview of Results

1.1.1 On the minimum distance of Turbo-like codes

The low-complexity and near-capacity performance of Turbo codes has led to a revolution in coding theory. However, the most useful Turbo codes have been observed to have low minimum distance.

We derive in Chapter 2 worst-case upper bounds on the minimum distance of parallel concatenated Turbo codes, serially concatenated Turbo codes, repeat-accumulate codes, repeat-convolute codes, and generalizations of these codes obtained by allowing non-linear and large-memory constituent codes.

We show that parallel-concatenated Turbo codes and repeat-convolute codes with sublinear memory are asymptotically bad.

We also show that depth-two serially concatenated codes with constant-memory outer codes and sublinear-memory inner codes are asymptotically bad.

In contrast, we prove that depth-three serially concatenated codes obtained by concatenating a repetition code with two accumulator codes through random permutations can be asymptotically good.

We generalize in Chapter 3 the bound corresponding to parallel-concatenated Turbo codes and repeat-convolute codes to the much more general setting of an arbitrary encoder that uses linear-time and sublinear memory.

The results reported in Chapter 2 will appear in a joint work with M. Mahdiam and D. Spielman [BMS03] that contains also others results that hold in the special setting of linear parallel concatenated Turbo codes.

1.1.2 Encoding complexity versus minimum distance

A natural extension of the problem in Chapter 2 is the the following question: What can we say about the growth of the minimum distance of a binary error correcting code given constraints on the computational complexity of its encoder?

We focus in Chapter 3 mainly on the time-space complexity of the encoder. In

this setting, the question is a natural tradeoff question between the parameters: code minimum distance, code rate, encoding time, and encoding space.

We establish a bound on the minimum distance of a binary error correcting code given constraints on the computational time-space complexity of its encoder in the general binary branching program model.

The bound we obtain implies a general theorem that asserts that if the encoder uses linear time and sublinear space in the most general sense, then the minimum distance of the code cannot grow linearly with the block length when the rate is nonvanishing, i.e., the the code cannot be asymptotically good.

Our argument is based on branching program techniques introduced by Ajtai [Ajt99]. We consider also the case when the encoder is a constant-depth AND-OR circuit.

The results reported in Chapter 3 will appear in a joint work with S. Mitter [BM03a].

1.1.3 Some symmetric codes with good distance

Linear codes that are symmetric in the sense of being invariant under the action of some group on the bits of the codewords have been studied extensively before, yet we still know very little about how the group structure can be exploited in order to establish bounds on the minimum distance or to come up with efficient decoding algorithms. One example of such codes are codes that are invariant under the action of some group on itself. When the group is cyclic these are cyclic codes. Another example is when we have a group acting on more than one copy of itself. When the group is cyclic these are quasi-cyclic codes. The main reason behind looking at such codes is the presence of an underlying symmetry structure. An ideal goal one hopes to achieve is to come up with an explicit construction of codes which achieves the binary GV (Gilbert-Varshamov) bound. This is a very open question since no such codes are known. Even explicitly constructing new asymptotically good codes is very desirable since there are only two known classes of constructions: concatenated algebraic geometric codes and their variations, and expander codes.

We study in Chapter 4 randomized and explicit constructions of binary linear codes that are invariant under the action of some group on the bits of the codewords. We study a nonabelian randomized construction corresponding to the action of the dihedral group on a single copy of itself, a randomized abelian construction based on the action of an abelian group on a number of disjoint copies of itself, and a related explicit construction.

Cyclic codes have been extensively studied over the last 40 years, yet it is still an open question whether there exist asymptotically good binary cyclic codes. We argue that by using a group slightly stronger than a cyclic group, namely the dihedral group, the existence of asymptotically good codes that are invariant under the action of the group on itself can be guaranteed. In particular, we show that, for infinitely many block lengths, a random ideal in the binary group algebra of the dihedral group is an asymptotically good rate-half code with a high probability.

We argue also that a random code that is invariant under the action of an abelian group G of odd order on k disjoint copies of itself satisfies the rate-$1/k$ binary Gilbert-Varshamov bound with a high probability under a condition on the family of groups. The underlying condition is in terms of the growth of the smallest dimension of a nontrivial $\mathbb{F}_2$-representation of the group and is satisfied by roughly most abelian groups of odd order, and specifically by almost all cyclic groups of prime order.

The explicit code we study is a specific nondegenerate element of above codes ensemble in the setting when G is cyclic of prime order p and $k = 2$. It is based on quadratic residues. For nondegeneracy reasons, we conjecture that this explicit code is asymptotically good and probably achieves the binary GV bound. We show that the codewords in this specific code are in one to one correspondence with special hyperelliptic curves over the finite field of order p, where the number of zeros of a codeword corresponds to the number of rational points. This suggests a conjecture about a bound tighter than the general estimates obtainable from Weil's theorem for the underlying class of curves.

The results reported in Chapter 4 will appear in a joint work with S. Mitter [BM03b].

1.1.4 On the pseudorandomness based on minimum distance

The notion of indistinguishability was introduced in the eighties by [BM82, Yao82]. A probability measure μ on $\{0,1\}^n$ is said to ϵ-fool a boolean function $f : \{0,1\}^n \to \{0,1\}$ if the probability that f takes the value 1, when x is selected according μ, is ϵ-close to the true probability that f takes the value 1 when x is selected uniformly at random.

In the late eighties and early nineties, the following basic pseudorandomness notions was introduced by [Vaz86, NN93] as special purpose generators to derandomize some randomized algorithms whose analyses can be made to work when only limited independence is assumed. A probability measure μ on $\{0,1\}^n$ is said to have the *δ-almost k-wise independence* property if μ can $\delta/2$-fool all parity functions on k or fewer of the n bits. The δ-almost n-wise independence is called the *δ-bias* property. The 0-almost k-wise independence property is called the *k-wise independence* property. Saying that μ has the k-wise independence property is equivalent to saying that any k or fewer of the n binary random variables are statistically independent, and each of those random variables is equally likely to be 0 or 1.

Classical constructions of such probability measures are based on linear codes with good distance properties [Vaz86, NN93, AGHP92]. For instance, if C is a block-length-n binary linear code whose dual has minimum distance above k, then the uniform distribution on the codewords of C is k-wise independent as a probability measure on $\{0,1\}^n$.

We study in Chapter 5 the derandomization capabilities of probability measures on the Hamming cube having the k-wise independence property, the δ-bias property, or the δ-almost k-wise independence property. Mostly, the questions we consider are about statements that hold for any probability measure having one of those properties. The exceptions are when we focus on linear-codes-based k-wise independent probability measures.

The δ-almost k-wise independence property is the weakest of these properties, and it is necessarily satisfied by any pseudorandom generator for suitable values of

k and δ. The k-wise independence property is stronger, but when k is relatively small, the two notions are loosely speaking equivalent in the sense that statements about foolablity by the k-wise independence property can be translated to statements about foolablity by the δ-almost k-wise independence property. The δ-bias property is stronger than the δ-almost k-wise independence property, and it is necessarily satisfied by any pseudorandom generator for log-depth circuits or randomized bounded-space computations for suitable values of δ. Thus, in general, understanding the power and the limitations of such pseudorandomness properties is of fundamental importance due to their basic nature.

We note first that linear-programming duality can be used to get a purely analytical characterization of the class of boolean function that can be fooled by the δ-almost k-wise independence property. The characterization is necessary and sufficient and is in terms of tight average sandwichability between real valued functions with low degree and small L_1-norm in the Fourier domain.

Then we characterize the location of classical linear-codes-based constructions of k-wise independent probability measures in the convex polytope of all such measures, and its subpolytope consisting of those measures whose Fourier transform is nonnegative.

In terms of limitations, we prove that the exponentially-small-bias property is not sufficient to fool small log-depth circuits nor the weakest branching programs.

From a concrete viewpoint, we prove first that any sufficiently log-wise independent probability measure looks random to all polynomially small read-once DNF formulas. The setting is naturally extendable to almost k-wise independent probability measures. We give an application related to the distribution of quadratic-residues.

Then we establish a very sharp upper bound on the probability that a random binary string generated according to a k-wise independent probability measure has any given weight. The setting is naturally extendable to almost k-wise independent probability measures. We give applications related to the distribution of quadratic-residues and the weight distribution of linear codes.

We consider also the problem of derandomizability of AC_0 by arbitrary k-wise in-

dependent probability measures, when k is made polylogaritmically large enough. We reduce this problem to a conjecture about the the symmetry of the optimum of some symmetric optimization problem with linear constraints and a nonlinear objective function.

Finally, we study of the problem of approximability of high-degree parity functions on high-dual-distance binary linear codes by low-degree polynomials with coefficients in fields of odd characteristics. This problem has applications to the ability of binary linear codes with sufficiently large dual distance to derandomize AC_0, or low-degree polynomial equations on binary input variables with coefficients in small finite fields of odd order. Among other results, we relax this problem into essentially a single low-dimensional low-complexity linear program in terms of Krawtchouk polynomials. The problem of bounding the optimum of the linear program remains open.

1.2 Error correcting codes basic language

In this section, we assemble various basic notions, definitions, and classical conventions about error correcting codes that we will be using. For an introduction to the theory of error correcting codes see [Lin99, MS92, Spi96n, Sud01]. See also [Gal63, Sti93, VY00] for specialized treatments, and [PHB98] for a partially exhaustive treatment.

All codes we will consider in this thesis are binary codes. So, unless otherwise specified, a code means a *binary code*.

One way to specify a binary code is by an injective map $\{0,1\}^m \to \{0,1\}^n$ called the *encoder*. The encoder maps binary strings of length m to binary strings of length n. We call m the *message length*, n the *block length*, and the ratio m/n the *rate*. The strings in the image of the encoder are the *codewords*. The *code* is the image of encoder in $\{0,1\}^n$, i.e., the set of codewords. The encoder is called *systematic* if the n codeword-coordinates can be permuted in such a way that the projection of the encoder on the first m codeword-coordinates is the identity map.

If the encoder is not specified then a block length n code simply means a subset of

$\{0,1\}^n$, i.e., a set of binary strings of length n that are called *codewords*. In this case, the *rate* of the code is defined as $r = \frac{1}{n}\log M$, where M is the number of codewords and log means here and elsewhere in this thesis the binary logarithm. Usually, the encoder is not specified only when the code is linear because in such a case there is a relatively acceptable way of encoding in quadratic time. A code is called *linear* if is $\mathbb{F}_2$-linear, i.e., linear as an $\mathbb{F}_2$-vector space, or equivalently if it is a subgroup of $(\mathbb{Z}/2\mathbb{Z})^n$, which we simply denote by $\mathbb{Z}_2^n$. Associated with a linear code are the dual code, and matrices called a generator matrix and a parity check matrix. The *dual code* of block-length-n linear code Q is another block-length-n linear code denoted by $Q^\perp$, and defined as the set of all binary strings y of length n s.t. $\sum_{i=1}^n x_i y_i = 0 \pmod 2$, for each codeword x of Q. A *generator matrix* of a linear code is any matrix realization of an $\mathbb{F}_2$-linear encoder of the code. A *parity check* matrix of a linear code is any matrix whose Null space is the code as an $\mathbb{F}_2$-vector space.

In Chapters 2 and 3, a code will be specified by an encoder, which is not necessarily linear. By abuse of notation we will call the encoder a code when there is no possibility of confusion. In Chapters 4 and 5, a code will be a linear code that is specified as a set of strings.

The *minimum distance* of a code is the minimum Hamming distance between two distinct codewords, where the *Hamming distance* between two binary strings of the same length is the number of positions where they disagree. When the code is linear, the minimum distance of the code is also equal to the minimum Hamming weight of a nonzero codeword, where the *Hamming weight* (which we simply denote by *weight*) of a binary string means the Hamming distance between this string and the all zeros string, i.e., the number of nonzero coordinates of the string.

The *minimum relative distance* of a code is its minimum distance normalized by the block length, i.e., the ratio of the minimum distance and the block length.

When speaking about a code, we always mean implicitly that we have an *infinite family of codes* indexed by the block length. We do not require that each positive integer be a block length, we simply require that there are infinitely many block lengths. The *rate* (*minimum relative distance*, respectively) of an infinite family of

codes means the lim-inf of the rate (minimum relative distance, respectively) of a code in the family as the block length tends to infinity.

The rate of a code reveals the amount of redundancy added, and the minimum distance reveals the maximum number of errors that can be corrected under any decoding algorithm in a worst case sense.

Rate and minimum distance are conflicting parameters. An infinite family of codes is called *asymptotically good* if both its rate and its minimum distance are strictly positive. This is equivalent to saying that the fraction of redundancy added is bounded by a constant, and the minimum distance of the code grows linearly with the block length. If an infinite family of codes is not asymptotically good, it is called *asymptotically bad*.

By abuse of notation, when asymptotic statements are made, a code means implicitly an infinite family of codes. For instance, "an asymptotically good code" means "an asymptotically good infinite family of codes", and "a code satisfying the GV bound" means "an infinite family of codes satisfying the GV bound" (See the GV bound definition below).

The main objective of combinatorial coding theory is to construct asymptotically good codes with efficient encoding and decoding algorithms, where complexity is measured in terms of the growth of the block length.

We say that a family of codes of rate r and minimum relative distance δ *satisfies or achieves the binary GV (Gilbert-Varshamov) bound* if $r \geq 1 - h(\delta)$, where h is the *binary entropy function*, i.e., $h(x) = -x \log x - (1-x) \log (1-x)$.

The existence of families codes satisfying the binary GV bound follows easily by counting. In fact, a random linear code satisfies the GV bound with a high probability. A random linear code means the linear code whose generator matrix is selected uniformly at random. Since we will be using the expression "high probability", it is appropriate here to explain its meaning. When saying that an outcome of a probabilistic experiment happens with a *high probability*, we mean that we have an infinite family of probabilistic experiments indexed by the integers n in some infinite set together with an infinite family of outcomes such that the probability that an outcome

happens approaches 1 as n tends to infinity.

It is an old open question whether the binary GV bound is tight, i.e., it is not known if there are families of codes with better rate versus minimum distance tradeoffs[1]. Some people believe it is tight. The best known upper bound is the MRRW (McEliece, Rodemich, Rumsey, and Welch) bound [MRRW77] on Delsarte LP (Linear Programming) bound [Del73]. The MRRW bound says that any infinite family of codes of rate r and minimum relative distance δ must satisfy $r \leq 1 - h(1/2 - \sqrt{\delta(1-\delta)})$.

Another old open question is how to explicitly construct codes that satisfy the GV bound. Here it is appropriate to make it clear what *explicit* means. For simplicity, assume that we are talking about linear codes. The weakest notion of an explicit linear code (i.e., an infinite family of linear codes) requires that there is a polynomial time algorithm that, when given any integer n, outputs a generator matrix for the code at block length n if n is a feasible block length. Note that complexity is measured here in terms of n, i.e., not in terms of the length of the representation of n. This notion is satisfactory, but it can be strengthened by requiring that the algorithm uses logarithmic space only. It is trendy however to use the word explicit in the sense of algebraically explicit which, in addition to being algorithmically explicit, means that that the construction is number theoretic in nature.

[1]Recall that we are not talking about nonbinary codes. The analog of the GV bound over nonbinary alphabets is known to be not tight. The construction of algebraic geometric codes from modular curves, by Tsfasman, Valdut, and Zink [TVZ82], beats the GV bound on alphabets of size 49. One interpretation of this phenomenon is that the Hamming distance is not the sharpest metric in the nonbinary alphabets case.

Chapter 2

On the Minimum Distance of Turbo-Like Codes

We derive in this chapter worst-case upper bounds on the minimum distance of parallel concatenated Turbo codes, serially concatenated Turbo codes, repeat-accumulate codes, repeat-convolute codes, and generalizations of these codes obtained by allowing non-linear and large-memory constituent codes.

We show that parallel-concatenated Turbo codes and repeat-convolute codes with sublinear memory are asymptotically bad.

We also show that depth-two serially concatenated codes with constant-memory outer codes and sublinear-memory inner codes are asymptotically bad.

In contrast, we prove that depth-three serially concatenated codes obtained by concatenating a repetition code with two accumulator codes through random permutations can be asymptotically good.

We will generalize in Chapter 3 the bound corresponding to parallel-concatenated Turbo codes and repeat-convolute codes to the much more general setting of an arbitrary encoder that uses linear-time and sublinear memory.

2.1 Introduction

The low-complexity and near-capacity performance of Turbo codes [BGT93, VY00] has led to a revolution in coding theory. However, the most useful Turbo codes have been observed to have low minimum distance. In this chapter, we provide general conditions under which many constructions of turbo-like codes, including families of serially-concatenated Turbo-like codes [BDMP98] and Repeat-Accumulate (RA) codes [DJM98, JM99, KU97], must be asymptotically bad. We also present a simple family of depth-3 serially concatenated Turbo-like codes that are asymptotically good.

Our work is motivated by the analyses of randomly constructed parallel and serially concatenated Turbo codes by Kahale and Urbanke [KU97] and of Turbo codes with two branches by Breiling [Bre01]. Kahale and Urbanke provided estimates on the probable minimum distance of randomly generated parallel concatenated Turbo codes with a constant number of branches. They also provided similar estimates for the minimum distance of the random concatenation of two convolutional codes with bounded memory. Breiling proved that the parallel concatenation of two convolutional codes with bounded memory always has logarithmic minimum distance. We note that both of these bounds are for linear codes with low memory.

These analyses naturally lead to the following four questions:

- **Better than random?** Do there exist asymptotically good parallel concatenated Turbo codes with more than two branches or do there exist asymptotically good repeat-convolute or repeat-accumulate codes?

  Note that the result of Breiling only applies to Turbo codes with two branches and the results of Kahale and Urbanke do not preclude the existence of codes that are better than the randomly generated codes.

- **Larger memory?** What happens if we allow the memories of the constituent convolutional codes to grow with the block length?

  All the previous bounds become vacuous if the memory even grows logarithmically with the block length.

- **Nonlinearity?** Can the minimum distance of Turbo-like codes be improved by the use of nonlinear constituent encoders, such as automata encoders?

- **Concatenation depth?** Can one obtain asymptotically good codes by serially concatenating a repetition code with two levels of convolutional codes?

We will give essentially negative answers to the first three questions and a positive answer to the last one. For parallel concatenations and depth-2 serial concatenations of convolutional and automata codes, we prove upper bounds on the minimum distance of the resulting codes in terms of the memories of the constituent codes. We show that parallel concatenated codes and repeat-convolute codes are asymptotically bad if their constituent codes have sublinear memory. This bound even holds if the constituent codes are nonlinear. We also show that depth-two serially concatenated convolutional codes are asymptotically bad if their inner code has sublinear memory and their outer code has constant memory. In contrast, we show that depth-three concatenations of constant-memory codes can be asymptotically good. In particular, we prove this for the random concatenation of a repetition code with two accumulator codes.

2.1.1 Turbo-like codes

The fundamental components of the codes we consider in this chapter are convolutional codes (as block codes) and their nonlinear generalizations, which we call *automata codes*. The fundamental parameter of a convolutional code that we will measure is its *memory*—the number of registers in its encoder. The memory can also be defined to be the binary logarithm of the number of states in the encoder's state diagram. A general automata encoder is obtained by considering an encoder with any deterministic state diagram. We will consider automata encoders that read one bit at each time step, and output a constant number of bits at each time step. These are also described as deterministic automata or transducers with one input bit and a constant number of output bits on each transition. We will again define the memory of an automata encoder to be the binary logarithm of its number of states.

Given k convolutional codes $Q_1, \ldots, Q_k$, a message length n, and k permutations $\pi_1, \ldots, \pi_k$ of length n, we can define the *parallel concatenated Turbo code with k branches* [BGT93, VY00] $P_{Q_1,\ldots,Q_k,\pi_1,\ldots,\pi_k}$, to be the code that encodes a binary message x to $(x, Q_1(\pi_1(x)), \ldots, Q_k(\pi_k(x)))$, where $\pi_i(x)$ denotes the permutation of the bits in x according to π_i and $Q_i(y)$ denotes the output of the convolutional code Q_i on input y.

Given an integer k, we define the repeat-k-times code, r_k, to be the code that just repeats each of its input bits k times. Given a convolutional code Q, a message length n, and a permutation π of length kn, we define the *repeat-convolute code* [DJM98], $C_{k,\pi,Q}$ to be the code that maps an input $x \in \{0,1\}^n$ to $(x, Q(\pi(r_k(x))))$. That is, each bit of the input is repeated k times, the resulting kn bits are permuted, and then fed through the convolutional encoder. We also assume that the input x is outputed as well. While some implementations do not include x in the output, its exclusion cannot improve the minimum distance so we assume it appears. The number k is called the *repetition factor* of the code.

When the convolutional code Q is the accumulator (*i.e.,* the map $Q(x)_j = \sum_{i=1}^{j} x_i$), this code is called a *repeat-accumulate (RA) code* [DJM98]. We remark that a parallel concatenated Turbo code with k branches can be simulated by a repeat-convolute code with repetition factor k whose encoder is a product of the encoders in the parallel code.

Given two convolutional encoders Q_o and Q_i that output h_o and h_i bits per time step respectively, an integer n, and a permutation π of length $h_o n$, we define the depth-two *serially concatenated Turbo code* [BDMP98, VY00] $C_{Q_o,Q_i,\pi}$ to be the rate $1/h_o h_i$ code that maps an input $x \in \{0,1\}^n$ to the codeword $Q_i(\pi(Q_o(x)))$. The codes Q_o and Q_i are called *outer* and *inner* codes, respectively. A classical example of serially concatenated Turbo codes, and that considered in [KU97], is a rate 1/4 code given by the map $(\pi(x, L_o(x)), L_i(\pi((x, L_o(x)))))$, where L_o and L_i are rate-1 convolutional codes. This fits into our framework with $Q_o(x) = (x, L_o(x))$ and $Q_i(x) = (x, L_i(x))$.

One can allow greater depth in serial concatenation. The only codes of greater depth that we consider will be repeat-accumulate-accumulate codes (RAA). These

are specified by a repetition factor k, an integer n, and two permutations π_1 and π_2 of length kn. Setting Q_1 and Q_2 to be accumulators, the resulting code maps an input x to $Q_2(\pi_2(Q_1(\pi_1(r_k(x)))))$.

We can generalize each of these constructions by allowing the component codes to be automata codes. In this case, we will refer to the resulting codes as *parallel concatenated Turbo-like codes*, *repeat convolute-like codes*, and *serially concatenated Turbo-like codes*. We refer to all the codes in this family as *Turbo-like* codes.

In practice, some extra bits are often appended to the input x of a Turbo-like code so as to guarantee that some of the encoders return to the zero state. As this addition does not substantially increase the minimum distance of the resulting code, we will not consider this technicality in this chapter. Note that this technicality easily fits in the encoding model that we will study in the next chapter.

2.1.2 Previous work

Kahale and Urbanke [KU97] proved that if one builds a parallel concatenated Turbo code from a random interleaver and convolutional encoders of memory at most M, then the resulting code has minimum distance at most $\tilde{O}(2^M n^{1-2/k})$ [1] and at least $\Omega(n^{1-2/k})$ with high probability. For rate 1/4 serially concatenated Turbo codes of the form mentioned in the previous section with a random interleaver, they proved that the resulting code has minimum distance at most $\tilde{O}(2^{M_i} n^{1-2/d_o})$ and at least $\Omega(n^{1-2/d_o})$ with high probability, where d_o is the free distance of the outer code and M_i is the inner code memory.

For parallel concatenated Turbo codes with two branches, Breiling [Bre01] proved that no construction could be much better than a random code: if the constituent codes have memory M, then the minimum distance of the resulting code is $O(2^M \log n)$.

Serially concatenated codes of depth greater than 2 were studied by Pfister and Siegel [PS99], who performed experimental analyses of the serial concatenation of repetition codes with l levels of accumulators connected by random interleavers, and theoretical analyses of of concatenations of a repetition code with certain rate-1 codes

[1] $\tilde{O}(f(n))$ means $O(f(n)\log^{O(1)} n)$.

for large l. Their experimental results indicate that the average minimum distance of the ensemble starts becoming good for $l \geq 2$, which is consistent with our theorem. For certain rate-1 codes and l going to infinity, they proved their codes could become asymptotically good.

In many arguments, we use techniques introduced Ajtai [Ajt99] to prove time-space trade-offs for branching programs.

2.1.3 Summary of results

In Section 2.2, we upper bound the minimum distance of repeat-convolute-like codes. We prove that repeat-convolute-like codes of message length n, memory M, and repetition factor k have minimum distance at most $O(n^{1-1/k}M^{1/k})$, and therefore such codes are asymptotically bad when k is constant and M is sublinear in n. Note that M sublinear in n corresponds to the case when the size of the corresponding trellis is subexponential, and so it includes the cases in which the codes have natural subexponential time iterative decoding algorithm. As parallel concatenated Turbo-like codes in which the component codes have memory M can be encoded by repeat-convolute codes with memory kM, we find that these are also asymptotically bad for k constant and M sublinear in n. This proof uses techniques introduced by Ajtai [Ajt99] for obtaining time-space trade-offs for branching programs. Comparing our upper bound with the $\tilde{O}(2^M n^{1-2/k})$ high-probability upper bound of Kahale and Urbanke for parallel concatenated codes, we see that our bound has a much better dependence on M and a slightly worse dependence on k. A similar relation holds between our bound and the $O(2^M \log n)$ upper bound of Breiling [Bre01].

In Section 2.3.1, we study serially concatenated Turbo-like codes with two levels, and prove that if the outer code has memory M_o and the inner code has memory M_i, then the resulting code has minimum distance at most $O(n^{1-1/h_o(M_o+2)}M_i^{1/h_o(M_o+2)})$. Accordingly, we see that such codes are asymptotically bad when M_o, h_o and h_i are constants and M_i is sublinear in n. The proof uses similar techniques to those used in Section 2.2. When specialized to the classical rate 1/4 construction of serially concatenated Turbo codes considered by Kahale and Urbanke [KU97], our bound on

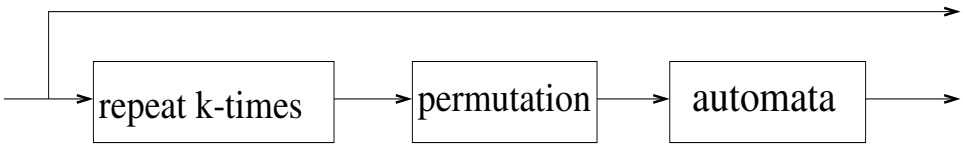

Figure 2-1: Repeat-convolute-like code.

the minimum distance becomes $O(n^{1-1/(2M_o+4)}M_i^{1/(2M_o+4)})$. Comparing this with the high-probability $\tilde{O}(n^{1-2/d_o}2^{M_i})$ upper bound of Kahale and Urbanke, we see that our bound is better in terms of M_i, and comparable in terms of d_o.

Finally, in Section 2.3.2, we show that serially concatenated codes of depth greater than two can be asymptotically good, even if the constituent codes are repetition codes and accumulators. In particular, we prove that randomly constructed RAA codes are asymptotically good with constant probability.

We conclude with some open questions.

2.2 An upper bound on the minimum distance of repeat-convolute-like codes

In this section we consider codes that are obtained by serially concatenating a repeat-k-times code r_k with any code Q that can be encoded by an automata (transducer) with at most 2^M states and one output bit per transition. More precisely, if Q is such an encoder, π is a permutation, and r_k is the repeat-k-times map, we define $C_{k,\pi,Q}$ to be the code that maps a string x to $C_{k,\pi,Q}(x) := (x, Q(\pi(r_k(x))))$. See Figure 2-1

This class of codes contains repeat-convolute codes and repeat-accumulate code when Q is a convolutional code. It also contains parallel concatenated Turbo codes: a parallel concatenated Turbo code with k branches and memory M can be encoded by a repeat-convolute-like code with repetition factor k and memory kM by interleaving the permutations on the k branches.

Theorem 2.2.1 *Let $k \geq 2$ be a constant integer, Q an automata encoder with at most 2^M states, n an integer, and π a permutation of length kn.*

If $n \geq 2^k kM$, then the minimum distance of the code $C_{k,\pi,Q}$ is at most

$$3k^2 n^{1-1/k} M^{1/k} + 2^k kM + k + 1.$$

Proof. To prove this theorem, we make use of the techniques introduced by Ajtai [Ajt99] for proving time-space trade-offs for branching programs. In particular, for an input x of length n, the encoding action of Q is naturally divided into kn time steps in which the automata reads a bit of $\pi(x)$, outputs a bit, and changes state. For convenience, we will let $I = \{1, \ldots, kn\}$ denote the set of time steps, and we will let $s_i(x)$ denote the state of Q on input $\pi(r_k(x))$ at the end of the i'th time step.

Let C denote the code $C_{k,\pi,Q}$. To prove the claimed bound on the minimum distance of C, we will prove the existence of two input strings, x and y, a set $U \subset \{1, \ldots, n\}$ of size at most $2^k k(M+1)$, and $J \subset I$ of size at most $3k^2 n^{1-1/k}(M+1)^{1/k}+k$ such that x and y may only differ on bits with indices in U and $s_i(x)$ and $s_j(x)$ may only differ on time steps with indices in J.

To construct the set J, we first divide the set of time steps I into b consecutive intervals, where b is a parameter we will specify later. We choose these intervals so that each has size $\lfloor kn/b \rfloor$ or $\lceil kn/b \rceil$. For example, if $k = 2$, $n = 4$, and $b = 3$ we can divide $I = \{1, \ldots, 8\}$ into the intervals $[1, 3]$, $[4, 6]$, and $[7, 8]$.

For each index of an input bit $i \in \{1, \ldots, n\}$, we let S_i denote the set of time intervals in which Q reads input bit i. As each bit appears k times, the sets S_i each have size at most k. As there are b intervals, there are at most b^k possible k-sets of intervals, where by a k-set we mean a set of cardinality k. So, there exists a set $U \subset \{1, \ldots, n\}$ of size at least n/b^k and a set of intervals, S, such that for all $i \in U$, $S_i = S$. Let U be such a set with $|U| = \left\lceil n/b^k \right\rceil$ and let T be the corresponding set of intervals. Let $l = |T|$. The set J will be the union of the intervals in T.

Let $t_1, \ldots, t_l$ be the last times in the time intervals in T (e.g., in the above example the last time of the interval $[4, 6]$ is 6). For each $x \in \{0, 1\}^n$, that is zero outside U, we consider the vector of states of Q at times $t_1, \ldots, t_l$ on input $\pi(r_k(x))$: $\{s_{t_i}(x)\}_{i=1}^{l}$. As the number of such possible sequences is at most 2^{Ml} and the number of x that

are zero outside U is $2^{|U|}$, if

$$2^{|U|} > 2^{Ml}, \tag{2.1}$$

then there should exist two different strings x and y that are both zero outside of U and such that $s_{t_i}(x) = s_{t_i}(y)$ for $i = 1, \ldots, l$. To make sure that (2.1) is satisfied, we set

$$b = \left\lceil \left(\frac{n}{kM}\right)^{1/k} \right\rceil - 1.$$

Our assumption that $n \geq 2^k kM$ ensures that $b \geq 1$. Now, since

- x and y agree outside U,
- the bits in U only appear in time intervals in T, and
- Q traverses the same states at the ends of time intervals in T on inputs $\pi(r_k(x))$ and $\pi(r_k(y))$,

Q must traverse the same states at all times in intervals outside T on inputs $\pi(r_k(x))$ and $\pi(r_k(y))$. Thus, the bits output by Q in time steps outside intervals in T must be the same on inputs $\pi(r_k(x))$ and $\pi(r_k(y))$. So $Q(\pi(r_k(x)))$ and $Q(\pi(r_k(y)))$ can only disagree on bits output during times in the intervals in T, and hence on at most $l\lceil kn/b \rceil$ bits. This means that the distance between $C(x)$ and $C(y)$ is at most

$$\begin{aligned}
|U| + l\lceil kn/b \rceil &\leq \left\lceil \frac{n}{b^k} \right\rceil + k\lceil kn/b \rceil, \text{ as } |U| = \lceil n/b^k \rceil \text{ and } l \leq k, \\
&\leq \frac{n}{b^k} + 1 + \frac{k^2 n}{b} + k \\
&\leq \frac{n}{\left\lceil \left(\frac{n}{kM}\right)^{1/k} - 1 \right\rceil^k} + \frac{k^2 n}{\left(\frac{n}{kM}\right)^{1/k} - 1} + k + 1 \\
&\leq \frac{n}{\left(\left(\frac{n}{kM}\right)^{1/k} - 1\right)^k} + \frac{k^2 n}{\left(\frac{n}{kM}\right)^{1/k}} \frac{\left(\frac{n}{kM}\right)^{1/k}}{\left(\frac{n}{kM}\right)^{1/k} - 1} + k + 1 \\
&\leq \frac{n}{\left(\frac{n}{kM}\right)} \left(\frac{\left(\frac{n}{kM}\right)^{1/k}}{\left(\frac{n}{kM}\right)^{1/k} - 1} \right)^k + \frac{k^2 n}{\left(\frac{n}{kM}\right)^{1/k}} \frac{\left(\frac{n}{kM}\right)^{1/k}}{\left(\frac{n}{kM}\right)^{1/k} - 1} + k + 1 \\
&\leq \frac{n}{\left(\frac{n}{kM}\right)} 2^k + 2\frac{k^2 n}{\left(\frac{n}{kM}\right)^{1/k}} + k + 1, \text{ as } n \geq 2^k kM
\end{aligned}$$

$$\begin{aligned} &\leq 2^k kM + 2k^2 n^{1-1/k} M^{1/k} k^{1/k} + k + 1, \\ &\leq 3k^2 n^{1-1/k} M^{1/k} + 2^k kM + k + 1, \end{aligned}$$

as $k^{1/k} \leq 3/2$. ■

Corollary 2.2.2 *Let k be a constant. Then, every repeat-convolute code with input length n and memory M and repetition factor k and every parallel concatenated Turbo code with input length n, convolutional encoder memory M and k branches has minimum distance $O(n^{1-1/k}M^{1/k})$. Thus, such codes cannot be asymptotically good for M sublinear in n.*

This means that if we allow M to grow like $\log n$, or even like $n^{1-\epsilon}$ for some $\epsilon > 0$, the minimum relative distance of the code will still go to zero. Moreover, M sublinear in n corresponds to the case in which the size of the corresponding trellis is subexponential, and therefore it includes all the cases in which such codes have subexponential-time iterative decoding algorithms.

It is interesting to compare our bound with that obtained by Kahale and Urbanke [KU97], who proved that a randomly chosen Turbo code with k branches has minimum distance $\tilde{O}(2^M n^{1-2/k})$ with high probability. Theorem 2.2.1 has a much better dependence on M and a slightly worse dependence on n. A similar comparison can be made with the bound of Breiling [Bre01], who proved that every parallel concatenated Turbo code with $k = 2$ branches has minimum distance $O(2^M \log n)$.

2.3 The minimum distance of serially concatenated Turbo-like codes

In this section, we consider codes that are obtained by serially concatenating convolutional codes and, more generally, automata codes. In Section 2.3.1, we prove an upper bound on the minimum distance of the concatenation of a low-memory outer automata encoder with an arbitrary inner automata encoder. In particular, we prove

that if the memory of the outer code is constant and the memory of the inner code is sublinear, then the code is asymptotically bad. In contrast, in Section 2.3.2, we prove that if the input is first passed through a repetition code and a random permutation, then the code is asymptotically good with constant probability, even if both convolutional encoders are accumulators.

2.3.1 An upper bound on the minimum distance when the outer code is weak

In this section, we consider the serial concatenation of automata codes. We assume that each automata outputs a constant number of bits per transition. This class of codes includes the standard serially concatenated Turbo codes, and includes those introduced by Benedetto, Divsalar, Montorsi and Pollara [BDMP98] and studied by Kahale and Urbanke [KU97]. If the outer code has constant memory and the inner code has sublinear memory, then our bound implies that the code cannot be asymptotically good.

Formally, we assume that Q_o (Q_i, respectively) is an automata encoder with at most 2^{M_o} (2^{M_i}, respectively) states and h_o (h_i, respectively) output bits per transition. For an integer n and a permutation π of length $h_o n$, we define $C_{Q_o,Q_i,\pi}$ to be the code that encodes a string $x \in \{0,1\}^n$ to the string $C_{Q_o,Q_i,\pi}(x) := Q_i(\pi(Q_o(x))) \in \{0,1\}^{h_o h_i n}$. We will assume without loss of generality that Q_o, Q_i, and π are such that this mapping is an injective mapping. The encoders Q_o and Q_i are called the outer and inner encoders, respectively.

Theorem 2.3.1 *Let Q_o be an automata encoder with at most 2^{M_o} states that outputs h_o bits at each time step, and let Q_i be an automata encoder with at most 2^{M_i} states that outputs h_i bits at each time step. For any positive integer n and any permutation π of length nh_o, the minimum distance of the code $C_{Q_o,Q_i,\pi}$ is at most*

$$3h_o^2 h_i (M_o + 2) n^{1 - \frac{1}{h_o(M_o+2)}} M_i^{\frac{1}{h_o(M_o+2)}}.$$

In particular, if M_o is constant (and h_i and h_0 are constants), the minimum distance of the code $C_{Q_o,Q_i,\pi}$ is

$$O(n^{1-\frac{1}{h_o(M_o+2)}} M_i^{\frac{1}{h_o(M_o+2)}}),$$

and consequently any such family of codes $C_{k,\pi,Q}$ is asymptotically bad as long as M_i is sublinear in n.

Proof. The proof follows the same outline as the proof of Theorem 2.2.1. We begin by setting $I_o = \{1, \ldots, n\}$ to be the set of times steps in the computation of Q_o on input $x \in \{0,1\}^n$, and setting $I_i = \{1, \ldots, h_o n\}$ to be the set of times steps in the computation of Q_i on input $\pi(Q_o(x)) \in \{0,1\}^{h_o n}$. We similarly, let $\left\{s_o^{(t)}(x)\right\}_{t \in I_o}$ denote the sequence of states traversed by Q_o on input x and $\left\{s_i^{(t)}(x)\right\}_{t \in I_i}$ denote the sequence of states traversed by Q_i on input $\pi(Q_o(x))$.

To prove the claimed bound on the minimum distance of $C_{Q_o,Q_i,\pi}$, we will prove the existence of two distinct input strings x and y, a set $V \subset \{1, \ldots, n\}$, a set $J_o \subset I_o$, and a set $J_i \subset I_i$ such that x and y are both 0 on bits not in V, $s_o^{(t)}(x)$ and $s_o^{(t)}(y)$ only differ for $t \in J_o$, and $s_i^{(t)}(x)$ and $s_i^{(t)}(y)$ only differ for $t \in J_i$. The minimum distance bound will then follow from an upper bound on the size of J_i.

To construct these sets, we make use of parameters m_o and m_i to be determined later. We first partition the set I_o into $b_o \stackrel{\text{def}}{=} \lfloor n/m_o \rfloor$ intervals each of size m_o or $m_o + 1$, and we partition the set I_i into $b_i \stackrel{\text{def}}{=} \lfloor n h_o/m_i \rfloor$ intervals each of size m_i or $m_i + 1$.

As Q_o outputs at most $(m_o + 1)h_o$ bits during the time steps in an interval in I_o, the bits output by Q_o during an interval in I_o are read by Q_i during at most $(m_o+1)h_o$ intervals in I_i. As there are fewer than $(b_i)^{(m_o+1)h_o}$ sets of at most $(m_o+1)h_o$ intervals in I_i, there exists a set of at least $b_o/(b_i)^{(m_o+1)h_o}$ intervals in I_o such that all the bits output by Q_o during these intervals are read by Q_i during a single set of at most $(m_o + 1)h_o$ intervals in I_i. Let U denote the set of at least $b_o/(b_i)^{(m_o+1)h_o}$ intervals in I_o and let T denote the corresponding set of at most $(m_o + 1)h_o$ intervals in I_i. We then let V denote the set of input bits read by Q_o during the intervals in U. As all the intervals in I_o have size at least m_o, we have $|V| \geq m_o |U|$. The set J^o will be the

union of the intervals in T and J^i will be the union of the intervals in U.

Let $\{u_j\}_{j=1}^{|U|}$ and $\{t_j\}_{j=1}^{|T|}$ denote the last time steps in the intervals in U and T respectively. For each $x \in \{0,1\}^n$ that is zero outside V, we consider $\left(s_o^{(u_j)}(x)\right)_{j=1}^{|U|}$, the sequence of states traversed by Q_o on x at times $u_1, \ldots, u_{|U|}$, and, $\left(s_i^{(t_j)}(x)\right)_{j=1}^{|T|}$, the sequence of states traversed by Q_i on input $\pi(Q_o(x))$ at times $t_1, \ldots, t_{|T|}$. There are at most $2^{M_o|U|}2^{M_i|T|}$ such pairs of sequences. So, if

$$2^{M_o|U|}2^{M_i|T|} < 2^{|V|}, \tag{2.2}$$

then there are two distinct x and y in $\{0,1\}^n$ that are both 0 outside V and a pair of sequences $\left(s_i^{(t_j)}\right)_{j=1}^{|T|}$ and $\left(s_o^{(u_j)}\right)_{j=1}^{|U|}$ such that $s_i^{(t_j)}(x) = s_o^{(t_j)}(y) = s_i^{(t_j)}$ for all $1 \leq j \leq |T|$ and $s_o^{(u_j)}(x) = s_o^{(u_j)}(y) = s_o^{(u_j)}$ for all $1 \leq j \leq |U|$. This means that the bits output and states traversed by Q_o on inputs x and y are the same at time steps outside the time intervals in U, and therefore the bits output and states traversed by Q_i on inputs $\pi(Q_o(x))$ and $\pi(Q_o(y))$ are the same outside time steps in intervals in T. Thus

$$0 < d(C_{Q_i,Q_o,\pi}(x), C_{Q_i,Q_o,\pi}(y)) \leq m_i h_i \, |T| \leq (m_o + 1) m_i h_o h_i. \tag{2.3}$$

As this bound assumes (2.2), we will now show that for

$$\begin{aligned} m_o &= M_o + 1, \text{ and} \\ m_i &= 3h_o n^{1-\frac{1}{(M_o+2)h_o}} (M_i)^{\frac{1}{(M_o+2)h_o}}, \end{aligned}$$

this assumption is true.

Our setting of m_o reduces (2.2) to

$$|U| \geq |T| \, M_i,$$

which would be implied by

$$\frac{b_o}{b_i^{(m_o+1)h_o}} > (m_o+1)h_o M_i. \tag{2.4}$$

To derive this inequality, we first note that since $x^{2/x} < 3$ for $x \geq 1$,

$$m_i > h_o n^{1-\frac{1}{(M_o+2)h_o}} (((m_o+1)h_o)^2 M_i)^{\frac{1}{(M_o+2)h_o}}.$$

Rearranging terms, we find this implies

$$\left(\frac{n}{(m_o+1)^2 h_o^2 M_i}\right)^{\frac{1}{(m_o+1)h_o}} > \frac{nh_o}{m_i} \geq b_i.$$

Again rearranging terms, we obtain

$$n > b_i^{(m_o+1)h_o}(m_o+1)^2 h_o^2 M_i \geq b_i^{(m_o+1)h_o}(m_o+1)m_o h_o M_i + m_o,$$

which implies

$$\left\lfloor \frac{n}{m_o} \right\rfloor > b_i^{(m_o+1)h_o}(m_o+1)h_o M_i.$$

By now dividing both sides by $b_i^{(m_o+1)h_o}$ and recalling $b_o = \left\lfloor \frac{n}{m_o} \right\rfloor$, we derive (2.4).

Finally, the bound on the minimum distance of the code now follows by substituting the chosen values for m_o and m_i into (2.3).

■

We now compare this with the high-probability upper bound of $\tilde{O}(n^{1-2/d_o}2^{M_i})$ on the minimum distance of rate 1/4 random serially concatenated codes obtained by Kahale and Urbanke [KU97]. In their case, we have $h_o = h_i = 2$, and our upper bound becomes $O(n^{1-1/(2M_o+4)}\ M_i^{1/(2M_o+4)})$. We note that the dependence of our bound on d_o is comparable, and the dependence of our bound on M_i is much better.

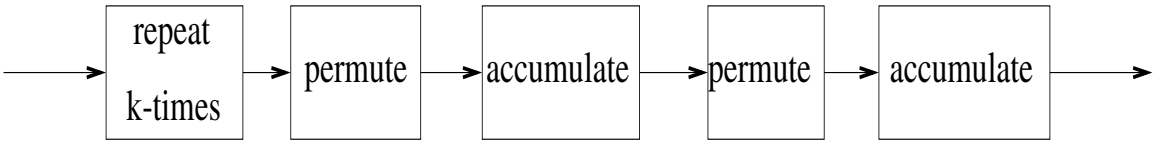

Figure 2-2: RAA code.

2.3.2 A strong outer code: when serially concatenated Turbo-like codes become asymptotically good

The proof technique used in Theorem 2.3.1 fails if the outer code is not a convolution code or encodable by a small finite automata. This suggests that by strengthening the outer code one might be able to construct asymptotically good codes. In fact, we will prove that the serial concatenation of an outer repeat-accumulate code with an inner accumulator yields an asymptotically good code with some positive probability.

Let $k \geq 2$ be an integer, r_k be the repeat-k-times map, Q_1 and Q_2 be accumulators[2], n be an integer, and π_1 and π_2 be permutations of length kn. We define C_{k,π_1,π_2} to be the code that maps input strings $x \in \{0,1\}^n$ to $C_{k,\pi_1,\pi_2}(x) := Q_2(\pi_2(Q_1(\pi_1(r_k(x)))))$. We call C_{k,π_1,π_2} an RAA (Repeat, Accumulate, and Accumulate) code We note that this code has rate $1/k$. See Figure 2-2.

In contrast with the codes analyzed in Theorem 2.3.1, these RAA codes have a repeat-accumulate code, $C_{k,\pi_1}(y) = Q_1(\pi_1(r_k(x))$ where those analyzed in Theorem 2.3.1 merely have an automata encoder.

Theorem 2.3.2 *Let $k \geq 2$ and n be integers, and let π_1 and π_2 be permutations of length kn chosen uniformly at random. Then for each constant $\delta > 0$, there exists a constant $\epsilon > 0$ and an integer n_0, such that the RAA code C_{k,π_1,π_2} has minimum distance at least ϵn with probability at least $1 - \delta$ for all $n \geq n_0$.*

So specifically, there exists an infinite family of asymptotically good RRA codes.

[2]While Q_1 and Q_2 are identical as codes, we give them different names to indicate their different roles in the construction.

Proof. Conditions bounding the size of ϵ will be appear throughout the proof.

Let $E_{\epsilon n}$ denote the expected number of nonzero codewords in C_{k,π_1,π_2} of weight less than or equal to ϵn. Taking a union bound over inputs and applying linearity of expectation, we see that the probability the minimum distance of C_{k,π_1,π_2} is less than ϵn is at most $E_{\epsilon n}$. Thus, we will bound this probability by bounding $E_{\epsilon n}$.

To bound $E_{\epsilon n}$, we use techniques introduced by Divsalar, Jin and McEliece [DJM98] for computing the expected input-output weight enumerator of random Turbo-like codes. For an accumulator code of message length N, let $A_{w,h}^{(N)}$ denote the number of inputs of weight w on which the output of the accumulator has weight h. Divsalar, Jin and McEliece [DJM98] prove that

$$A_{w,h}^{(N)} = \binom{N-h}{\lfloor w/2 \rfloor} \binom{h-1}{\lceil w/2 \rceil - 1}, \tag{2.5}$$

where $\binom{a}{b}$ is defined to be zero if $a < b$. Therefore, if the input to Q is a random string of length N and weight w, the probability that the output has weight h is

$$\frac{A_{w,h}^{(N)}}{\binom{N}{w}} = \frac{\binom{N-h}{\lfloor w/2 \rfloor}\binom{h-1}{\lceil w/2 \rceil - 1}}{\binom{N}{w}}. \tag{2.6}$$

Now consider a fixed input x to the encoder for C_{k,π_1,π_2}. If x has length n and weight w and π_1 is a random permutation of length kn, then $\pi_1(r_k(x))$ is a random string of length kn and weight kw. This random string is the input to the accumulator Q_1. Therefore, by (2.6), for any h_1 the probability that the output of Q_1 has weight h_1 is $A_{kw,h_1}^{(kn)} / \binom{kn}{kw}$. If this happens, the input to Q_2 will be a random string of weight h_1, and therefore, again by (2.6), the probability that the output of Q_2 has weight h will be equal to $A_{h_1,h}^{(kn)} / \binom{kn}{h_1}$. Thus, for any fixed input string x of weight w, and any fixed h_1 and h, the probability over the choice of π_1 and π_2 that the output of Q_1 has weight h_1 and the output of Q_2 (which is also the output of C_{k,π_1,π_2}) has weight h is equal to

$$\frac{A_{kw,h_1}^{(kn)} A_{h_1,h}^{(kn)}}{\binom{kn}{kw}\binom{kn}{h_1}}.$$

Thus, by the linearity of expectation, the expected number of nonzero codewords

of C_{k,π_1,π_2} of weight at most ϵn equals

$$E_{\epsilon n} = \sum_{w=1}^{n} \sum_{h_1=0}^{kn} \sum_{h=1}^{\epsilon n} \frac{\binom{n}{w} A^{(kn)}_{kw,h_1} A^{(kn)}_{h_1,h}}{\binom{kn}{kw}\binom{kn}{h_1}} = \sum_{h_1=1}^{2\epsilon n} \sum_{w=1}^{2h_1/k} \sum_{h=1}^{\epsilon n} \frac{\binom{n}{w} A^{(kn)}_{kw,h_1} A^{(kn)}_{h_1,h}}{\binom{kn}{kw}\binom{kn}{h_1}},$$

as the terms with $\lceil h_1/2 \rceil > h$ or $\lceil kw/2 \rceil > h_1$ are zero. Using the inequalities $\binom{x}{y} \leq (ex/y)^y$, $\binom{x}{\lfloor y/2 \rfloor} \leq (4ex/y)^y$ and $\binom{x}{\lceil y/2 \rceil - 1} \leq (4ex/y)^y$, for positive integers x and y, we bound this sum by

$$\begin{aligned}
E_{\epsilon n} &= \sum_{h_1=1}^{2\epsilon n} \sum_{w=1}^{2h_1/k} \sum_{h=1}^{\epsilon n} \frac{\binom{n}{w}\binom{kn-h_1}{\lfloor kw/2 \rfloor}\binom{h_1-1}{\lceil kw/2 \rceil - 1}\binom{kn-h}{\lfloor h_1/2 \rfloor}\binom{h-1}{\lceil h_1/2 \rceil - 1}}{\binom{kn}{kw}\binom{kn}{h_1}} \\
&\leq \sum_{h_1=1}^{2\epsilon n} \sum_{w=1}^{2h_1/k} \sum_{h=1}^{\epsilon n} \frac{\binom{n}{w}\left(\frac{4ekn}{kw}\right)^{kw/2}\left(\frac{4eh_1}{kw}\right)^{kw/2}\left(\frac{4ekn}{h_1}\right)^{\lfloor h_1/2 \rfloor}\left(\frac{4eh}{h_1}\right)^{\lceil h_1/2 \rceil - 1}}{\left(\frac{n}{w}\right)^{kw}\left(\frac{kn}{h_1}\right)^{h_1}} \\
&= \sum_{h_1=1}^{2\epsilon n} \sum_{w=1}^{2h_1/k} \sum_{h=1}^{\epsilon n} \binom{n}{w}\left(\frac{4e\sqrt{h_1}}{\sqrt{kn}}\right)^{kw}\left(\frac{h}{kn}\right)^{\lceil h_1/2 \rceil}\frac{(4e)^{h_1-1}h_1}{h}
\end{aligned}$$

The summand in the above expression is at maximum when $h = \epsilon n$. Therefore,

$$\begin{aligned}
E_{\epsilon n} &\leq \epsilon n \sum_{h_1=1}^{2\epsilon n}\left(\frac{\epsilon n}{kn}\right)^{\lceil h_1/2 \rceil}\frac{h_1(4e)^{h_1-1}}{\epsilon n}\sum_{w=1}^{2h_1/k}\binom{n}{w}\left(\frac{4e\sqrt{h_1}}{k\sqrt{n}}\right)^{kw} \\
&\leq \sum_{h_1=1}^{2\epsilon n} h_1\left(4e\sqrt{\epsilon/k}\right)^{h_1}\sum_{w=1}^{2h_1/k}\binom{n}{w}\left(\frac{4e\sqrt{h_1}}{k\sqrt{n}}\right)^{kw} \\
&\leq \sum_{h_1=1}^{2\epsilon n} h_1\left(4e\sqrt{\epsilon/k}\right)^{h_1}\sum_{w=1}^{2h_1/k}\left(\frac{ne}{w}\right)^{w}\left(\frac{4e\sqrt{h_1}}{k\sqrt{n}}\right)^{kw} \\
&= \sum_{h_1=1}^{2\epsilon n} h_1\left(4e\sqrt{\epsilon/k}\right)^{h_1}\sum_{w=1}^{2h_1/k}\left(\frac{e\left(\frac{4e}{k}\right)^{k} n^{1-k/2}h_1^{k/2}}{w}\right)^{w} \\
&\leq \sum_{h_1=1}^{2\epsilon n} h_1\left(4e\sqrt{\epsilon/k}\right)^{h_1}\frac{2h_1}{k}e^{\left(\frac{4e}{k}\right)^{k}n^{1-k/2}h_1^{k/2}}, \text{ as } \left(\frac{y}{x}\right)^{x} \leq e^{y/e} \\
&\leq \frac{2}{k}\sum_{h_1=1}^{2\epsilon n}\left(4e^2\sqrt{\epsilon/k}\right)^{h_1} e^{\left(\left(\frac{4e^2}{k}\right)^{k}n^{1-k/2}\right)h_1^{k/2}}, \qquad (2.7)
\end{aligned}$$

since $h_1^2 \leq e^{h_1}$ for all $h_1 \geq 1$. To bound (2.7), note that the sum has the form

$$S = \sum_{x=1}^{m} \alpha^x e^{\beta x^l},$$

where $\alpha = 4e^2\sqrt{\epsilon/k}$, $\beta = \left(\frac{4e^2}{k}\right)^k n^{1-k/2}$, $l = \frac{k}{2}$, and $m = 2\epsilon n$. If we can guarantee that

$$\alpha^{x+1} e^{\beta(x+1)^l} \leq \frac{1}{2}\alpha^x e^{\beta x^l}, \tag{2.8}$$

for all $x = 1, \ldots, m-1$, we can use the bound

$$S \leq 2\alpha e^{\beta}. \tag{2.9}$$

We can express (2.8) as $\beta((x+1)^l - x^l) \leq \ln\frac{1}{2\alpha}$. Thus (2.8) holds for all the desired values of x if $\beta((m+1)^l - m^l) \leq \ln\frac{1}{2\alpha}$, or equivalently

$$\beta m^l \left(\left(1 + \frac{1}{m}\right)^l - 1\right) \leq \ln\frac{1}{2\alpha},$$

which can be guaranteed when

$$2l\beta m^{l-1} \leq \ln\frac{1}{2\alpha} \quad \text{and} \quad l \leq m, \tag{2.10}$$

via the bounds

$$\left(1 + \frac{1}{m}\right)^l \leq e^{l/m} \leq 1 + (e-1)\frac{l}{m} \leq 1 + 2\frac{l}{m},$$

where we need $l \leq m$ in the second inequality. Going back to (2.7), we get via (2.9) and (2.10) that

$$E_{\epsilon n} \leq \frac{2}{k} 2(4e^2\sqrt{\epsilon/k}) e^{\left(\frac{4e^2}{k}\right)^k n^{1-k/2}} = \frac{16e^2\sqrt{\epsilon}}{k\sqrt{k}} e^{\left(\frac{4e^2}{k}\right)^k n^{1-k/2}}, \tag{2.11}$$

when

$$2\frac{k}{2}\left(\frac{4e^2}{k}\right)^k n^{1-k/2}(2\epsilon n)^{k/2-1} \leq \ln\left(\frac{1}{8e^2}\sqrt{\frac{k}{\epsilon}}\right) \quad \text{and} \quad \frac{k}{2} \leq 2\epsilon n,$$

or, equivalently, when

$$\ln \frac{1}{\epsilon} \geq \left(2k \left(\frac{4e}{k} \right)^k 2^{k/2-1} \right) \epsilon^{k/2-1} - 2 \ln \frac{\sqrt{k}}{8e^2} \quad \text{and} \quad \frac{k}{2} \leq 2\epsilon n. \tag{2.12}$$

It follows from (2.11) and (2.12), that for each $k \geq 2$, and for each constant $\delta > 0$, there is constant $\epsilon > 0$ such $E_{\epsilon n} < \delta$ when n is sufficiently large.

■

While the constants we obtain are not particularly sharp, they are sufficient to prove the existence of asymptotically good families of serially concatenated turbo-like codes of depth 3. This result should be compared with the work of Pfister and Siegel [PS99], who performed experimental analyses of the serial concatenation of repetition codes with l levels of accumulators connected by random interleavers, and theoretical analyses of of concatenations of a repetition code with certain rate-1 codes for large l. Their experimental results indicate that the average minimum distance of the ensemble starts becoming good for $l \geq 2$, which is consistent with our theorem. For certain rate-1 codes and l going to infinity, they proved their codes could become asymptotically good. In contrast, we prove this for $l = 2$ and accumulator codes.

2.4 Open questions

- Can the RAA codes described in Section 2.3.2 be efficiently decoded by iterative decoding, or any other algorithm?

- Can one obtain depth-3 serially concatenated codes with better minimum distance by replacing the accumulators in the RAA codes with small convolutional codes? Also, can one improve the minimum distance bounds on the RAA codes?

- If one allows the memory of the outer code in a depth-2 serially concatenated code to grow logarithmically with the block length, can one obtain an asymptotically good code?

Chapter 3

Encoding Complexity Versus Minimum Distance

We establish in this chapter a bound on the minimum distance of a binary error correcting code given constraints on the computational time-space complexity of its encoder in the general binary branching program model.

The bound we obtain asserts that if the encoder uses linear time and sublinear memory in the most general sense, then the minimum distance of the code cannot grow linearly with the block length when the rate is nonvanishing, i.e., the code cannot be asymptotically good.

The bound extends the bound we obtained in Chapter 2 on the minimum distance of parallel-concatenated Turbo codes and repeat-convolute codes to the much the more general setting of an arbitrary encoder that uses linear-time and sublinear memory. Our argument is based on branching program techniques introduced by Ajtai [Ajt99]. We also consider the case of constant-depth AND-OR circuits encoders with unbounded fanin.

3.1 Introduction

In this chapter, we consider the following question:

> What can we say about the growth of the minimum distance of a binary

error correcting code given constraints on the computational complexity of its encoder?

We concentrate mainly on the time-space complexity of the encoder. In this setting, the question is a natural tradeoffs question between the parameters: code minimum distance, code rate, encoding time, and encoding space.

From a practical perspective, this question is important since there are popular error correcting codes that have low time-space encoding complexity. We are referring here to Turbo codes, or more precisely to parallel concatenated Turbo codes introduced by Berrou, Glavieux, and Thitimajshima in [BGT93], and repeat-convolute codes introduced by Divsalar, Jin, and McEliece in [DJM98]. This low time-space encoding complexity is crucial for the corresponding iterative decoding algorithms because these algorithms process the state space representation of the encoder. Sharp bounds on the minimum distance of Turbo codes were first obtained by Kahale and Urbanke [KU97] when the underlying interleavers are chosen uniformly at random and the memory of the constituent convolutional codes is is bounded by a constant. In Chapter 2, we derived strong bounds on the minimum distance of Turbo like codes in variety of cases. One of these cases is the well structured setting of generalized repeat-convolute codes where the convolutional code is replaced by an arbitrary automaton. We argued that such codes are asymptotically bad when the memory of the automaton is sublinear and the number of repetitions is constant.

In this chapter, we extend this particular result to the much more general setting where the encoder is a binary branching program, or equivalently a nonuniform random-access machine with binary input registers.

We establish a general theorem that asserts that if the encoder is a binary branching program that uses linear time and sublinear space, then the minimum distance of the code cannot grow linearly with the block length when the rate is nonvanishing. In other words, the code cannot be asymptotically good in such a case, which is a rather surprising result. In general we derive a bound relating the involved parameters.

Our proof is based on the branching programs techniques introduced in the recent breakthrough of Ajtai [Ajt99].

We also consider the case of constant-depth AND-OR circuits encoders with unbounded fanin. We conclude with a conjecture about the strongest possible time-space tradeoffs for encoding asymptotically good codes.

3.1.1 Branching program encoders

By a code we mean in this chapter an infinite family of binary codes where each code is specified by an encoder. See Section 1.2 for the basic notions and conventions.

Consider a code specified by an encoding map $C : \{0,1\}^n \to \{0,1\}^m$. By a *branching program encoder* (binary by default, i.e., 2-way) B computing C we mean a connected directed acyclic graph with a single source and multiple sinks, together with a set of n binary *input variables* and a set of m binary *output variables* satisfying the following. There are exactly two arrows leaving each non-sink node, the first labeled with a one and the second with a zero. Every non-sink node is associated with an input variable. Some of the nodes are associated with output variables (possibly more than one variable per node), in which case the nodes are labeled by zeros or ones. The nodes of the graph are called states, the source is called the *start state*, and the sinks are called the *end states*. The branching program B computes C as follows. The computation starts by reading the value of the variable associated with the start state and moving according to its value to the next state and so on by reading more bits, while outputting when an output node is reached, until an end state is reached. We may want to assume that on any input each output variable will be set at least once, or we can assume that the output variables are arbitrarily preset. We stress here that we are allowing the branching program to set an output variable more than once.

The *computation of the branching program on an input* is the corresponding sequence of states starting with the start state and ending with an end state. The length of a computation is its number of states, and its time is its number of states plus the total number of times each output variable is set along the way.

The *length* of the branching program is the maximum length of a computation. The *time* t of the branching program is the maximum time of a computation. The

size S of the branching program is the total number of states. The *memory* or the *space* M of a branching program is $M = \log S$.

Remark 3.1.1 The codes encodable by such general branching programs correspond to those encodable by *random-access machines* with bounded read-write space and binary input registers, where the time of the branching program is the worst case running time, and its size is $\Theta(2^M)$, M being the number read-write bits. Note that the machine has two types of read-only bits, those corresponding to the input message, and those corresponding to the code description. Complexity is measured in terms of the number of bits of the first type. Note that we are not restricting the size of the read-only bits corresponding to the code description. This is not a problem since we are deriving lower bounds. See Example 3.1.5 for some consequences of this unlimited read-only space.

3.1.1.1 Some special types of branching programs

The branching program is called *leveled* if the states are divided into an ordered collection of sets each called a *level* where edges are between consecutive levels only. In such a case, the *width* of the branching program is the maximum number of states per level.

The branching program is called *oblivious* if the input variables (and the output variables) are read (respectively, set) in the same order regardless of the input under consideration. Thus an oblivious branching program is naturally leveled in such a way that all the nodes in the same level read the same input variables, and set the same output variables.

The branching program is called a *read-k-times* branching program if each input variable is read at most k times on any input.

The branching program is called a *write-w-times* if at most w output variables are set per state.

3.1.1.2 Examples

Example 3.1.2 The trellis of a convolutional code is an oblivious, read-once, and write-once branching program. The trellis of a systematic convolutional code is an oblivious, read-once, and write-2-times branching program.

Example 3.1.3 Parallel concatenated Turbo codes are encodable by low complexity oblivious branching programs as follows. A *parallel concatenated Turbo code* [BGT93] C with a constant number k of branches, message length n, and memory M is specified by k permutations $\pi_1, \ldots, \pi_k$ each on n bits and a rate 1 convolutional code Q (the *component code*) of memory M_0. For x in $\{0,1\}^n$, C encodes x as $C(x) = (x, Q(\pi_1(x)), \ldots, Q(\pi_k(x)))$, where $\pi_i(x)$ is the string obtained by permuting the bits of x according to the permutation π_i, and $Q(y)$ is the output of the convolutional encoder Q on the input string y. Thus C is naturally encodable by an oblivious read-k-times write-2-times branching program B. The states of B are copies of those of the automaton, and B is naturally leveled. B has length kn and time $\Theta(n)$. The width of B is 2^{M_0}, and its size is at most $kn2^{M_0}$. Note that the same holds if we follow the technicality of appending a terminating sequence to the input.

Example 3.1.4 Repeat-convolute codes fit in the same picture. A *repeat-convolute code* [DJM98] consists of a repeat-k-times code, a convolutional code, and a permutation. More precisely, a repeat-convolute code C of message length n and memory M is specified by a constant integer k, a permutation π on kn bits, and a convolutional encoder Q of memory M. For x in $\{0,1\}^n$, C encodes x as $C(x) = (x, Q(\pi(r(x))))$, where r is the repeat-k-times map, i.e., $r(x)$ is the concatenation of k copies of x. As in Example 3.1.3, C is naturally encodable by a leveled, oblivious, read-k-times, write-2-times, length-kn, time-$\Theta(n)$, and width-2^{M_0} branching program whose size is at most $kn2^{M_0}$.

Example 3.1.5 Any code can be trivially encoded in the binary branching program model in linear time and linear space by a tree branching program that on any input, outputs the whole codeword at the corresponding leaf. This makes sense in the

random-access machine with binary input registers encoding model because we are not counting the read-only space needed to store the code description (See Remark 3.1.1). It is worth noting here that when this read-only space is taken into consideration, we know from the work of Spielman [Spi96] that there exists an asymptotically good code encodable in linear time and linear space.

Example 3.1.6 Any linear code is naturally encodable by a leveled, oblivious, width-2, quadratic-time, and write-once branching program.

3.1.2 Main result

Theorem 3.1.7 *Let $C : \{0,1\}^n \to \{0,1\}^m$ be a code (i.e., an injective function) encodable (i.e., computable) by a branching program B of size $S = S(n)$, time $t = t(n)$, and length $l = l(n)$ (so $l \leq t$).*
If $t(n) = \Theta(n)$, then the minimum distance of C is

$$O\left(\left(\frac{\log S}{n}\right)^{\frac{1}{\lceil l/n \rceil}} n\right).$$

Therefore, C is asymptotically bad when $S(n) = 2^{o(n)}$ and $t(n) = O(n)$.
More generally, if $t(n) = \Omega(n)$, then the minimum distance of C is

$$O\left(\left(\frac{t}{n}\right)^3 \left(\frac{\log S}{n}\right)^{\frac{n}{2t}} n\right).$$

Thus, C is asymptotically bad also when $S(n) = 2^{O(n^{1-\epsilon_1})}$ and $t(n) = O(n \log^{1-\epsilon_2} n)$, for all $\epsilon_1, \epsilon_2 > 0$.

Note that l does not appear in the bound in the more general case since t is an upper bound on l.

In other words linear time and sublinear space for encoding imply that the code is asymptotically bad, i.e., the minimum distance cannot grow linearly with the block length when the rate is nonvanishing.

Note that when the time is linear, the sublinear memory requirement is asymptotically tight for encoding asymptotically good codes (See Example 3.1.5).

3.1.2.1 Application to Turbo-like codes

By applying Theorem 3.1.7 to parallel concatenated Turbo codes and repeat-convolute codes (See Examples 3.1.3 and 3.1.4), we can recover the corresponding bound in Chapter 2 as follows.

The minimum distance of a parallel concatenated Turbo code with a constant number k of branches, message length n, and memory M_0 is $O(n^{1-1/k}M_0^{1/k})$ because the size of the corresponding branching program is at most $kn2^{M_0}$. Similarly, the minimum distance of a repeat-convolute code with k repetitions, message length n, and memory M_0 is $O(n^{1-1/k}M_0^{1/k})$.

So both types of codes will be asymptotically bad in any reasonable setting, i.e., as long as M_0 is sublinear in n. Note that the situation when M_0 is sublinear in n corresponds to the case when the underlying trellis has subexponential size, i.e., when the corresponding iterative Turbo decoding algorithm has subexponential running time.

3.2 Proof of Theorem 3.1.7

3.2.1 Ajtai proof techniques for the Hamming distance problem

We use branching program techniques introduced by Ajtai in [Ajt99]. More specifically, we are referring to the branching program techniques that Ajtai introduced to show that there is no $O(n)$-time and $o(n \log n)$-space R-way branching program, $R = n^c$ (c some absolute constant), that decides on the Hamming distance problem: given n strings in $\{0,1\}^{\log R}$, decide whether any (distinct) two of them are at $\lambda \log R$ Hamming distance apart (λ another absolute constant related to c).

Even though this is a decision problem in the setting of R-way branching programs,

while ours is not a decision problem and is in the setting of 2-way branching programs, the techniques introduced by Ajtai are behind the proof we describe below.

We refer the reader to Ajtai's paper [Ajt99].

3.2.2 Objects under consideration and terminologies

We will start by making the branching program leveled [1]. Recall from Section 3.1.1 that this means that the states are partitioned into l consecutive sets of states each called a level in such a way that edges (i.e., transitions) occur only between consecutive levels.

We will divide the branching program into blocks. By *divide*, we mean partition, and by a *block* we mean a set of consecutive levels. For a given block, we will be looking at states in the lower boundary level of the block. By the *lower boundary level of a block* we mean mean the last (with respect to the levels ordering) level (which is a set of states) in the block (which is a set of levels).

Given an input, we will be looking at the computation of the branching program on this input, which as we explained in Section 3.1.1 is defined to be the corresponding sequence of states starting with the start state and ending with an end state. So, in the leveled case, each computation takes exactly l steps, i.e., it contains exactly l states.

Fix an input x, and consider the corresponding computation of the branching program B on x. Fix also a set L of levels or a set T of blocks. By an input bit or variable (respectively, output bit or variable) accessed or read (respectively, set) in L or T *during the computation* of B on x (or equivalently by L setting the input bit during the computation and so on ...), all that we mean is that there is a state in the

[1] This can be done by a classical procedure. Construct a leveled directed graph of l levels where each level consists of a copy of all the nodes of the original branching program together with the related output labels. Connect the nodes in each two consecutive levels according to the the graph of the original branching program. Associate the end states not in the last level with arbitrary input variables. Connect the end states in any tow consecutive levels by two arrows labeled respectively by one and zero. Finally, remove all the nodes (together with the related edges) that are not accessible from the start sate in the first level or can not reach an end state in the last level. The start state of the new branching program is the remaining state in the first level, and its end states are those remaining in the last level.

computation that belongs to a level in L or a level in a block in T where the value of the input variable is read in order to move to another state (respectively, the value of the output variable is set).

Finally, by a *computation which contains a sequence of states*, we mean that each state in this sequence appears in the computation. Note that here the order does not matter since the states in a computation are distinct because the branching program is acyclic.

3.2.3 The oblivious case argument

Recall that an oblivious branching program is naturally leveled in such a way that all the nodes in the same level read the same input variables, and set the same output variables.

Since the proof of Theorem 3.1.7 is relatively long, it is instructive to look first at the very special case when B is oblivious. This case is very restrictive compared to a general branching program. To restrict the setting further, assume that B is read-k-times and write-w-times, where $k = O(1)$ and $w = O(1)$.

The argument we used in Chapter 2 to bound the minimum distance of repeat-convolute codes was in the setting of automata. More specifically, we studied the case of a repeat-convolute code where the convolutional code is replaced by an arbitrary automaton. Even though the automata setting is less general than the case we are considering in this section, the argument naturally extends as follows.

Assume that B is a read-k-times, write-w-times, and oblivious branching program, where $k = O(1)$ and $w = O(1)$. Thus $n \leq l \leq kn$ and $m \leq wn$. We want to argue that the minimum distance of C is $O(n(\frac{\log S}{n})^{1/k})$.

Let W be the width of B, thus $W \leq S$. We will exhibit two distinct input strings that map to two codewords at distance $O(n(\frac{\log W}{n})^{1/k})$ apart. We will do this by finding a nonempty set of input variables U, a subset J of levels, and two distinct strings x_1 and x_2 in $\{0,1\}^n$ such that x_1 and x_2 agree outside U, and the computations of B on x_1 and x_2 agree outside J. This will give us the desired bound on the minimum distance. J will be constructed as a union of intervals from a partition of B that we

define next.

Partition B into b consecutive blocks, each consisting of s_1 or s_2 levels, where $s_1 = \lfloor kn/b \rfloor$ and $s_2 = \lceil kn/b \rceil$. Assume for now that b is arbitrary as long as $s_1 \geq 1$. We will optimize on the integer b later.

Each of the n input variables is read by B in at most k blocks. Recall that B is oblivious. Thus, for any specific variable, these blocks will be the same irrespective of the setting of the input variables. There are at most b^k k-set of blocks. Here by a k-set of blocks, we mean a set of blocks of cardinality at most k. So there are at least n/b^k input variables that are read by B in the same k-set of blocks. Let U be such a set of input variables with $|U| = \lceil n/b^k \rceil$, T be such a k-set of blocks, thus $1 \leq |T| \leq k$. The set J we mentioned above is the union of the blocks in T.

Consider the lower boundary levels $L_1, \ldots, L_{|T|}$ of the blocks in T ordered by the level index, and let Q be the set of strings in $\{0,1\}^n$ that are zero outside U, thus $|Q| = 2^{|U|}$. There are at most W^k state sequences in $L_1 \times \ldots \times L_{|T|}$, and for each x in Q the computation of B on x contains such a sequence. So if we can guarantee that $2^{|U|} > W^k$, we get that there should be a sequence of states $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ in $L_1 \times \ldots \times L_{|T|}$ and two different strings x_1 and x_2 in Q such that the computation of B on both x_1 and x_2 contains $\{\mathfrak{s}_i\}_{i=1}^{|T|}$. Since x_1 and x_2 agree outside U, the computation of Q on x_1 and x_2 are exactly the same outside the blocks in T. The reason is that both computations are in similar states each time the branching program leaves an intervals in T. Thus $C(x_1)$ and $C(x_2)$ can only differ in the blocks in T. This means that the distance between $C(x_1)$ and $C(x_2)$ is at most $|T|s_2w \leq k\lceil kn/b \rceil w$, since $|T| \leq k$, and $s_2 = \lceil kn/b \rceil$.

This bound holds under the assumption that $2^{|U|} > W^k$, which can be guaranteed if $2^{n/b^k} > W^k$. So, choose $b = \lceil (\frac{n}{k \log W})^{1/k} \rceil - 1$. Note that the only other constraints we have on b are $1 \leq b \leq kn$, and the selected value satisfies these constraints when W is not exponential in n. Note also that if W is exponential in n, the statement of the theorem is trivial. By replacing this value of b in the upper bound $k\lceil kn/b \rceil w$ on the distance between $C(x_1)$ and $C(x_2)$, and using S as a upper bound on W, we get that the minimum distance of C is $O(n(\frac{\log S}{n})^{1/k})$. Note that we used here also that

$k = O(1)$, $w = O(1)$, and $C(x_1) \neq C(x_2)$ because $x_1 \neq x_2$ and C is injective.

This proof is short and simple. But when B is not oblivious, the proof does not go through. The main reason is that we cannot construct U and T regardless of the setting of the input variables as we did above. When the branching program is oblivious, the read-k-times and the write-w-times restrictions are not fundamental. When it is not oblivious, they become restrictive. For example, in the general branching program model, depending on the input, a very large number of the output variables may be set in a particular state, or a particular input variable may be read a very large number of times. The point to keep in mind is that the oblivious assumption is very restrictive.

We will sketch in the next section how to handle the general situation. The proof is longer and more sophisticated. This is not strange since the statement we are proving is much more general. The reader is encouraged to go carefully over the above argument before proceeding to the general case.

3.2.4 Proof technique

We follow the techniques introduced by Ajtai [Ajt99] in the setting of the Hamming distance problem.

We want to find two input strings x_1 and x_2 such that $C(x_1)$ and $C(x_2)$ are close to each other.

The first step is to make the branching program leveled without affecting its input-output behavior. Next, we divide the branching program into blocks each consisting of consecutive levels whose number will be suitably selected later and whose sizes are as uniform as possible.

To exhibit x_1 and x_2, we will find a set T of blocks such that:

- the size of T is small,
- the computations of B on x_1 and x_2 are exactly the same in the blocks outside T, and

- not too many output bits of $C(x_1)$ (respectively $C(x_2)$) are set in any of the blocks in T during the computation of B on x_1 (respectively x_2).

Thus $C(x_1)$ and $C(x_2)$ can only disagree on the few output bits that are set in T.

To find such x_1, x_2, and T, we find first T together with a set Q' of input strings in $\{0,1\}^n$ and a sequence $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ of states in the lower boundary levels of the blocks in T in such a way that for each x in Q':

- the computation of B on x contains $\{\mathfrak{s}_i\}_{i=1}^{|T|}$,
- not too many output bits of $C(x)$ are set in any of the blocks in T during the computation of B on x, and
- the number of variables in x that are accessed only in the blocks in T during the computation of B on x is large.

We will eventually find the desired x_1 and x_2 inside Q' as follows.

We modify the branching program B again so that B is forced to to pass through a state in the sequence $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ each time it attempts to leave a lower boundary level of a block in T, but without affecting its input-output behavior on Q'.

Using T, define an equivalence relation on $\{0,1\}^n$ by relating two strings if:

- they share the same set of input variables that are not read during the computation of B in blocks outside T, and
- they agree on the values of their bits outside this set.

Thus each equivalence class $[x]$ is determined by a set $I_{[x]}$ of input variables and a setting of the variables outside $I_{[x]}$.

We forced the computation of B to contain the states $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ on all inputs so that we get $|[x]| = 2^{|I_{[x]}|}$, and hence the size of each equivalence class $[x]$ can be guaranteed to be large when $I_{[x]}$ is large.

Since for each input string in Q', the number of variables that are accessed only in the blocks in T during the computation of B is large, we get that the equivalence class of each input in Q' is large.

By considering the set Ω of sufficiently large equivalence classes so that the equivalence classes of all the elements of Q' are guaranteed to be elements of Ω, our problem reduces to selecting the number of blocks so that $|Q'|$ is strictly larger than $|\Omega|$, and hence there are distinct x_1 and x_2 in Q' that have the same equivalence class. The fact that $[x_1] = [x_2]$ means that the computations of B on x_1 and x_2 are exactly the same outside the blocks in T, and hence $C(x_1)$ and $C(x_2)$ can only disagree on the output bits that are set inside the blocks in T.

By construction the number of those output bits will be small. Moreover, and since C is injective, $C(x_1)$ and $C(x_2)$ are distinct. The distance between $C(x_1)$ and $C(x_2)$ will be the desired bound on the minimum distance of C.

3.2.5 Proof outline

Assume for moment that $t = \Theta(n)$. We will deal with the more general case when we are done by working more carefully with the constants. So say that

$$t = an \text{ and } l = cn, \tag{3.1}$$

where $a, c \geq 1$ are constants ($a, c \geq 1$ because C is injective).

A) We modify the branching program so that it is leveled. See the footnote in Section 3.2.2. The modified branching program computes the same function, i.e., B computes C. The length of the resulting branching program B is l, its time is t, its size is at most Sl, and its width is at most S. The difference is that now edges occur only between consecutive levels, and each computation takes exactly l steps.

B) Partition B into b consecutive blocks, each consisting of s_1 or s_2 levels, where we define

$$s_1 \stackrel{\text{def}}{=} \left\lfloor \frac{l}{b} \right\rfloor = \left\lfloor \frac{cn}{b} \right\rfloor \text{ and } s_2 \stackrel{\text{def}}{=} \left\lceil \frac{l}{b} \right\rceil = \left\lceil \frac{cn}{b} \right\rceil. \tag{3.2}$$

Assume for now that in general $1 \leq b \leq l$ so that $1 \leq s_1, s_2 \leq l$. We will optimize on the integer b later.

C) **Lemma 3.2.1** *There exist:*

a) absolute constants $h, \alpha > 0$,

b) $Q' \subset \{0,1\}^n$ *such that*

$$|Q'| \geq \frac{2^n}{(Sb)^k}, \tag{3.3}$$

where

$$k = \lfloor c \rfloor, \tag{3.4}$$

c) a set of blocks T,

$$1 \leq |T| \leq k, \tag{3.5}$$

d) and a sequence $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ *of states in the lower boundary levels of the blocks in* T,

such that for each x *in* Q'*:*

1) the computation of B *on* x *contains* $\{\mathfrak{s}_i\}_{i=1}^{|T|}$,

2) at most

$$w = \frac{hs_1t}{l} \tag{3.6}$$

output bits of $C(x)$ *are set in each block in* T *during the computation of* B *on* x*, and*

3) the number of variables in x *that are accessed only in the blocks in* T *during the computation of* B *on* x *is at least*

$$\frac{\alpha n}{b^k}.$$

Proof. See Section 3.2.6. ■

D) Now we modify the branching program B again so that B is forced to to pass through a state in $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ each time it attempts to leave a lower boundary level of a block in T, while guaranteeing that B behaves exactly like the old B on the inputs in Q', i.e., it computes $C(x)$ for each x in Q'.

We can do this by simply connecting (on both inputs) all the states in the level above that of $\mathfrak{s}_i$ to $\mathfrak{s}_i$, for each i. Note that B need not compute an injective function anymore, so it may not read all the input variables on some inputs. It may also leave some of the output variables unset, but this is not a problem since we can assume that the output variables were arbitrarily preset.

Note that this step is essential for what follows. See Section 3.2.4 for the big picture.

E) Finally, we bound in Section 3.2.7 the minimum distance of C by exhibiting distinct x_1 and x_2 in Q' such that the distance between $C(x_1)$ and $C(x_2)$ is $O(n(\frac{\log S}{n})^{\frac{1}{k}})$.

F) In Section 3.2.8, we explain how to drop the assumption $t = \Theta(n)$.

3.2.6 Proof of Lemma 3.2.1

Consider any input x in $\{0,1\}^n$.

- Let $k \geq 1$ be an integer, and let $h > 0$. We will set k then h as we continue.
- Let R_x be the set consisting of all the blocks that sets at most

$$w \stackrel{\text{def}}{=} \frac{hs_1t}{l}$$

bits of $C(x)$ during the computation of B on x.

- Let D_x be the set of input variables that are read in at most k states during the computation of B on x.
- And let D'_x be the set of input variables in D_x that are read only in blocks in R_x during the computation of B on x.

First recall from (3.1) that $a, c \geq 1$ are the constants satisfying

$$t \stackrel{\text{def}}{=} an \text{ and } l \stackrel{\text{def}}{=} cn.$$

Recall also from (3.2) that

$$s_1 \stackrel{\text{def}}{=} \left\lfloor \frac{l}{b} \right\rfloor = \left\lfloor \frac{cn}{b} \right\rfloor \geq 1 \text{ and } s_2 \stackrel{\text{def}}{=} \left\lceil \frac{l}{b} \right\rceil = \left\lceil \frac{cn}{b} \right\rceil .$$

Some bounds:

- From the definition of R_x, we must have $w(b - |R_x|) \leq t$, thus

$$|R_x| \geq b - \frac{l}{hs_1} \geq b\left(1 - \frac{2}{h}\right), \tag{3.7}$$

  where the first inequality follows from $w = hs_1 t/l$, and the second from the bound $s_1 = \lfloor l/b \rfloor \geq l/(2b)$.

- Since C is injective, each input variable must be read at least once, so from the definition of D_x, we must have have

$$|D_x| + (k+1)(n - |D_x|) \leq l,$$

  i.e., $|D_x| \geq n(1 - \frac{c-1}{k})$ because $l = cn$. Thus if we set

$$k \stackrel{\text{def}}{=} \lfloor c \rfloor,$$

  we get

$$|D_x| \geq n(1 - \epsilon), \text{ where } \epsilon \stackrel{\text{def}}{=} \frac{c-1}{k} < 1. \tag{3.8}$$

- The number of input variables read in blocks outside R_x is at most

$$(b - |R_x|)s_2 \leq \frac{2bs_2}{h} \leq n\frac{4c}{h},$$

  where the first inequality follows from (3.7), and the second from the bound $s_2 = \lceil cn/b \rceil \leq 2cn/b$. Thus, by the definition of D'_x, we must have

$$|D'_x| \geq |D_x| - \frac{4c}{h} \geq n\left(1 - \epsilon - \frac{4c}{h}\right),$$

where the second inequality follows from (3.8).

Let h be sufficiently large such that

$$\alpha \stackrel{\text{def}}{=} 1 - \epsilon - \frac{4c}{h} > 0. \tag{3.9}$$

Note that this implies also that $1 - \frac{2}{h} > 0$ since $c \geq 1$, i.e

$$|R_x| > 0,$$

by (3.7).

To sum up, we have fixed some constants $h, \alpha > 0$, and specified $k \stackrel{\text{def}}{=} \lfloor c \rfloor$, so that

$$1 \leq |R_x| \leq b \text{ and } |D'_x| \geq \alpha n. \tag{3.10}$$

Now, keep the definition of R_x in mind, ignore D_x, and recall that D'_x is a set of input variables such that:

- each input variable in D'_x is read in at most k levels during the computation of B on x, and
- each of those levels belongs to a block in R_x.

Recall also that so far we are fixing an input x in $\{0,1\}^n$.

Consider all the k-sets in R_x, i.e., the subsets of R_x of size at most k. Each input variable in D'_x is read in such a k-set during the computation on x, and there are at most $|R_x|^k$ such k-set, so there are at least

$$\frac{|D'_x|}{|R_x|^k} \geq \frac{\alpha n}{b^k}$$

variables in D'_x read in the same k-set of blocks, where we have used (3.10) to obtain the estimate. Let U_x be such a set of variables in D'_x, and let T_x be such a k-set of blocks in R_x. So

$$1 \leq |T_x| \leq k \text{ and } |U_x| \geq \frac{\alpha n}{b^k}.$$

Note that T_x is nonempty since C is injective.

For each x in $\{0,1\}^n$, there is such a U_x and T_x. Fix any such correspondence. There are at most b^k such T_x, so there is a subset $Q \subset \{0,1\}^n$ and a k-set of blocks T such that

$$|Q| \geq \frac{2^n}{b^k},$$

and $T = T_x$ for each x in Q. Now consider the lower boundary levels $L_1, \ldots, L_{|T|}$ of the blocks in T ordered by the level index. There are at most S^k state sequences in $L_1 \times \ldots \times L_{|T|}$, and for each x in Q, the computation of B on x contains such a sequence, so there is a sequence $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ of states in $L_1 \times \ldots \times L_{|T|}$ and a subset $Q' \subset Q$ such that

$$|Q'| \geq \frac{|Q|}{S^k} \geq \frac{2^n}{(Sb)^k},$$

and the computation of B on x contains $\{\mathfrak{s}_i\}_{i=1}^{|T|}$, for each x in Q'.

3.2.7 Bounding the minimum distance

Now we are ready to find the two distinct messages x_1 and x_2 that C map to close codewords.

Using T, for each x in $\{0,1\}^n$, let I_x be the set of input variables that are not read during the computation of B on x in blocks outside T. Note that we need a double negation ("not read" and "outside") since some of the input variables may not be read at all because we modified the branching program in (D).

So, by (C.3), for each x in Q',

$$|I_x| \geq \frac{\alpha n}{b^k}. \tag{3.11}$$

Using T, define the equivalence relation $\sim$ on $\{0,1\}^n$ by $x \sim y$ if:

- $I_x = I_y$, and
- x agree with y on the bits outside I_x.

In other words, $x|_{I_{[x]}} = y|_{I_{[y]}}$, where $[x]$ means the equivalence class of x.

Given any x in $\{0,1\}^n$, each $y \sim x$ can only disagree with x on I_x. Conversely, if y disagrees with x only inside I_x, it must be the case that $y \sim x$. To see why this is true, note that we forced in (D) all the computations of B to leave the blocks in T in the same states: the the sequence of states $\{\mathfrak{s}_i\}_{i=1}^{|T|}$ that we exhibited in (C.1). So the computations of B on x and y are exactly the same outside the blocks in T, and hence any bit accessed on x outside T will be accessed on y outside T and none of the bits in I_x will be accessed on y outside T. It follows that

$$|[x]| = 2^{|I_x|}.$$

Thus, by (3.11), for each x in Q',

$$|[x]| \geq 2^{\alpha n/b^k}.$$

Let Ω be the set of equivalence classes the size of each being at least $2^{\alpha n/b^k}$. So, $[x]$ is in Ω for each x in Q'. Besides, since the equivalence classes are disjoint, we must have $|\Omega| 2^{\alpha n/b^k} \leq 2^n$, i.e.,

$$|\Omega| \leq \frac{2^n}{2^{\alpha n/b^k}}. \tag{3.12}$$

If we can guarantee that

$$|Q'| > |\Omega|, \tag{3.13}$$

we get that there should be $x_1 \neq x_2$ in Q' such that $[x_1] = [x_2]$. The fact that $[x_1] = [x_2]$ means that the computations of B on x_1 and x_2 are exactly the same outside the blocks in T, and hence $C(x_1)$ and $C(x_2)$ can only disagree on the output bits that are set inside the blocks in T. But, by (C.2), we constructed Q' in such way that the computation of B on any x in Q' can set at most w bits of $C(x)$ in each block in T. Thus $C(x_1)$ and $C(x_2)$ can disagree on at most

$$2|T|w \leq 2k\frac{hs_1a}{c} \leq \frac{2khcna}{bc} = \frac{2khan}{b} \tag{3.14}$$

bits, where the first inequality follows from $|T| \leq k$ (by (3.5)) and $w = hs_1t/l =$

$hs_1 a/c$ (by (3.6) and (3.1)), and the second follows from $s_1 = \lfloor l/b \rfloor = \lfloor cn/b \rfloor \leq cn/b$ (by (3.2)).

Moreover, $C(x_1)$ and $C(x_2)$ must disagree on at least one bit since $x_1 \neq x_2$, and C is injective.

Using (3.3) and (3.12), condition (3.13) can be guaranteed to hold if

$$2^{\alpha n/b^k} > (Sb)^k,$$

which is fulfilled when

$$\frac{1}{b} > \left(\frac{k \log (Scn)}{\alpha n}\right)^{1/k},$$

since $b \leq l = cn$. If we can select b so that this holds, we get that the minimum distance of C is at most $2khan/b$. The only restriction we have on b is $1 \leq b \leq l$, so we set

$$b \stackrel{\text{def}}{=} \left\lceil \left(\frac{\alpha n}{k \log (Scn)}\right)^{1/k} \right\rceil - 1. \tag{3.15}$$

This is always below l, and it cannot go below 1 unless $S \geq 2^{\alpha n/k 2^k - \log (nc)}$ in which case the statement of the theorem is trivial. Thus, via (3.14), the minimum distance of C is at most

$$\frac{2khan}{\left\lceil \left(\frac{\alpha n}{k \log (Scn)}\right)^{1/k} \right\rceil - 1} = O\left(n \left(\frac{\log S}{n}\right)^{\frac{1}{k}}\right).$$

3.2.8 Dropping the linear time assumption

Now we drop the assumption that $t = \Theta(n)$, thus a and c need not be constants. Since $l \leq t$, we use a as an upper bound on c. Assume that a grows with n and assume also that it is $O(\log n)$ since otherwise the statement of the theorem is trivial. We will not set $k = \lfloor c \rfloor$. Going back to (3.8) and (3.9), we have

$$\alpha = 1 - \frac{c-1}{k} - \frac{4c}{h} > 1 - \frac{a-1}{k} - \frac{4a}{h},$$

a value that we need to keep bounded away from zero by a positive constant. Set $k = \lceil 2(a-1) \rceil$ and $h = \lceil 16a \rceil$, thus $\alpha > 1/4$.

By using the same choice of b in (3.15) and the same bound on the minimum distance of C in (3.14), but with the new values of k and h and the bound $c \leq a$, we get that the minimum distance of C is at most

$$\frac{2\lceil 2(a-1)\rceil\lceil 16a\rceil an}{\left\lceil \left(\frac{n/4}{\lceil 2(a-1)\rceil \log(San)}\right)^{1/\lceil 2(a-1)\rceil}\right\rceil - 1} = O\left(a^3 n \left(\frac{\log S}{n}\right)^{\frac{1}{2a}}\right).$$

3.3 When the encoder is a constant-depth AND-OR circuit

To outline the boundaries of the picture, we consider the same problem but from the perspective of the circuit complexity of the encoder. Here we note that not much can be said other than what is essentially expected. Since we know from [Spi96] that there are asymptotically good codes that are encodable by linear-size and logarithmic-depth circuits, we are left with constant-depth circuits encoders with unbounded fanin.

Note first that if $C : \{0,1\}^n \to \{0,1\}^m$ is a code (i.e., an injective map in general), we say that C is encodable by a depth d AND-OR circuit with unbounded fanin if each of the m output bits is computable by a depth d circuit where: 1) the only allowed gated are AND/OR gates with possibly negated inputs, and 2) the number of inputs per gate is unbounded. The size of the circuit is the total number of gates.

We argue by a direct application of Hastad switching Lemma that a polynomial size constant-depth circuit cannot encode an asymptotically good code (actually as long as the circuit size is subexponential in a power of the block length inversely proportional to the circuit depth). This is not surprising since in the special case of linear codes, a small depth circuit encoder corresponds to a code with a low density generator matrix.

Lemma 3.3.1 Hastad switching Lemma [Has86]*: Let $f : \{0,1\}^n \to \{0,1\}$ be computable by an unbounded-fanin depth-d AND-OR circuit of size M. Consider a random restriction ρ that independently keeps each input bit unset with a probability $p = 1/(20k)^d$, sets it to 1 with a probability $1 - p/2$, and to 0 with a probability*

$1 - p/2$. *Then the probability, over the choice of* ρ*, that* f*, when restricted to the values set by* ρ*, cannot be evaluated by a decision tree of depth* k *is at most* $M2^{-2k}$.

Note that a decision tree computing a binary function $b : \{0,1\}^n \to \{0,1\}$ on n variables is a binary tree where each node is associated with one of the input variables, and each leaf is associated with a 0 or 1 setting of the single output variable. This implies that if we fix any setting of the input variables, there are at most k variables that, when negated, will affect the value of b, where k is the depth of the tree. In other words, when k is small, b has low sensitivity. Thus if a code $\{0,1\}^n \to \{0,1\}^m$ (an injective map) is encodable by m decision trees each of depth k, a direct counting argument shows that its minimum distance can be at most km/n. Hastad switching Lemma essentially reduces the circuit case to this situation.

Theorem 3.3.2 *Let* $C : \{0,1\}^n \to \{0,1\}^m$*,* $m = \Theta(n)$*, be a code (i.e., in general an injective map) encodable by an unbounded fanin AND-OR circuit of size* S *and depth* l*, then the minimum distance of* C *is*

$$O((20)^l \log^{l+1} mS).$$

Thus, C *is asymptotically bad when* $l = O(1)$ *and* $S = 2^{o(m^{1/l})}$.

Proof. Let $x_1, \ldots, x_n$ be the input variables, $A_1, \ldots, A_m$ the circuits that compute the output variables $y_1, \ldots, y_m$, and $a = m/n$. Thus $a \geq 1$ is constant, the size of each A_i is at most S, and the depth of each A_i is at most l.

Hit the x_i's with a random restriction ρ that keeps each x_i unset with a probability $p = 1/(20k)^l$, sets x_i to 1 with a probability $1-p/2$, and to 0 with a probability $1-p/2$.

Then, for each A_i, from Hastad switching Lemma, the probability that A_i does not collapse to a decision tree of depth k is at most $S2^{-2k}$. Thus the probability that one of the A_i's does not collapse, or the number of remaining (unset) variables is below $np/2$ is at most

$$P = mS2^{-2k} + \frac{4(1-p)}{np},$$

where the later term comes from the Chebychev inequality.

Fix $k = \log Sm$ so that $P \leq 1/(Sm) + 4(20 \log Sm)^l/n < 1$ when n is large enough and S is subexponential in $n^{1/l}$. Note that when S is exponential in $n^{1/l}$, the statement of the theorem is trivial.

So, fix any restriction ρ with the property that:

- the set I of input variables left unset by ρ has size at least $np/2$, and
- each of the A_i's collapses under ρ to a decision tree T_i of depth k, where $k = \log Sm$ and $p = 1/(20k)^l$.

Consider any setting of the variables in I, and let I_i be the set of variables in I read by T_i on this setting. Each I_i contains at most k variables, and the output of T_i can only be affected when we change some of the variables in I_i. So there should be a variable in I that appears in at most

$$\frac{\sum_j |I_j|}{|I|} \leq \frac{km}{np/2} = \frac{2ka}{p}$$

of the I_i's . By flipping this variable, we can affect at most $2ka/p$ output bits, and at least one output bit since C is injective. Hence the minimum distance of C is at most

$$\frac{2ka}{p} = O((20)^l \log^{l+1} Sm).$$

■

3.4 Open questions

Using branching program techniques introduced by Ajtai [Ajt99], we argued in Theorem 3.1.7 that there are no asymptotically good codes that are encodable in linear time and sublinear space in the most general sense. When the time is linear, the sublinear memory requirement is asymptotically tight for encoding asymptotically good codes (See Example 3.1.5).

On the other extreme, quadratic encoding time is achievable by random linear codes while requiring minimal encoding memory in the branching program model

(See Example 3.1.6).

We conjecture that in general

Conjecture 3.4.1 *If* $C : \{0,1\}^n \to \{0,1\}^m$, $m = O(n)$, *is a code (an injective map), that is computable by a branching program of memory* M *and time* T, *where* $MT = o(n^2)$, *then the minimum distance of* C *must be* $o(n)$.

Proving the conjecture or finding the correct time-space tradeoffs for encoding asymptotically good codes when the encoding time is superlinear and subquadratic is very desirable.

Chapter 4

Some symmetric codes with good distance

We study in this chapter randomized and explicit constructions of binary linear codes that are invariant under the action of some group on the bits of the codewords. We study a nonabelian randomized construction corresponding to the action of the dihedral group on a single copy of itself, a randomized abelian construction based on the action of an abelian group on a number of disjoint copies of itself, and a related explicit construction.

Cyclic codes have been extensively studied over the last 40 years, yet it is still an open question whether there exist asymptotically good binary cyclic codes. We argue that by using a group slightly stronger than a cyclic group, namely the dihedral group, the existence of asymptotically good binary codes that are invariant under the action of the group on itself can be guaranteed. In particular, we show that, for infinitely many block lengths, a random ideal in the binary group algebra of the dihedral group is an asymptotically good rate-half code with a high probability.

We argue also that a random code that is invariant under the action of an abelian group G of odd order on k disjoint copies of itself satisfies the rate-$1/k$ binary (GV) Gilbert-Varshamov bound with a high probability under a condition on the family of groups. The underlying condition is in terms of the growth of the smallest dimension of a nontrivial $\mathbb{F}_2$-representation of the group and is satisfied roughly by most abelian

groups of odd order, and specifically by almost all cyclic groups of prime order.

The explicit code we study is a specific nondegenerate element of above codes ensemble in the setting when G is cyclic of prime order p, and $k = 2$. It is based on quadratic residues. For nondegeneracy reasons, we conjecture that this explicit code is asymptotically good and probably achieves the binary GV bound. We show that the codewords in this specific code are in one to one correspondence with special hyperelliptic curves over the finite field of order p, where the number of zeros of a codeword corresponds to the number of rational points. This suggests a conjecture about a bound tighter than the general estimates obtainable from Weil's theorem for the underlying class of curves.

4.1 Introduction

Linear codes that are symmetric in the sense of being invariant under the action of some group on the bits of the codewords have been studied extensively before, yet we still know very little about how the group structure can be exploited in order to establish bounds on the minimum distance or to come up with decoding algorithms.

One example of such codes are codes that are invariant under the action of some group on itself. When the group is cyclic these are cyclic codes. Another example is when we have a group acting on more than one copy of itself. When the group is cyclic these are quasi-cyclic codes.

structure to come up with fast decoding algorithms.

The main reason behind looking at such codes is the presence of an underlying symmetry structure. An ideal goal one hopes to achieve is to come up with an explicit construction of codes up the binary GV (Gilbert-Varshamov) bound. This is a very open question since no such codes are known. Even explicitly constructing new asymptotically good codes is very desirable since there are only two known classes of constructions:

- Concatenated algebraic geometric codes: Justesen [Jus72], Goppa [Gop70], Tsfasman, Valdut, and Zink [TVZ82], Elkies [Elk01], and their variations such as

the variation of Justesen's codes based on chinese-remainder codes (see [GRS00] and the references therein).

- Expander codes: Alon, Bruck, Naor, Naor, and Roth [ABN+92], Sipser and Spielman [SS96], and Spielman [Spi96].

Another fundamental goal is to be able to use the group structure to come up with fast decoding algorithms.

4.1.1 Preliminaries

4.1.1.1 Binary linear codes

Unless otherwise specified, by a code, we mean in this chapter an infinite family of binary linear codes. See Section 1.2 for the basic definitions and conventions.

4.1.1.2 Group algebras

Let R be a finite ring with identity. The ring R is called *simple* if it has no proper two sided ideal or equivalently if it is isomorphic to a matrix algebra over some division ring that must be a finite field since R is finite. The ring R is called *semisimple* if its radical is zero, or equivalently if R is the direct sum of 2-sided ideals that are simple as rings in which case the decomposition is unique.

Let G be a finite group, F a finite field. The *group algebra* $F[G]$ of G over F is the F-algebra consisting of formal sums $\sum_{g \in G} f(g)g$ over F, $f : G \to F$. The group algebra $F[G]$ is semisimple if and only if the characteristic of F does not divide the order of G.

See [Bur65, CR62, McD74] for general background.

4.1.1.3 Group action codes

What we mean by a binary linear code invariant under the action of some group is as follows. Consider an action ρ of a finite group G on a finite set S, and say that a

(binary $\mathbb{F}_2$-linear) code C is ρ-invariant if it satisfies the following. Let M be the $|S|$-dimensional $\mathbb{F}_2$-vector space written as the set of formal sums $\sum_{s\in S} f(s)s$, $f : S \to \mathbb{F}_2$. Consider the induced action of G on M by (say left) translation $g : f(x) \mapsto f(gx)$. Then we say that C is ρ-invariant if C is a subset of M closed under addition and closed under translation by the elements of G. In other words, C is ρ-invariant if C is an $\mathbb{F}_2[G]$-submodule of M (again with the left multiplication convention). Note that if $\sum_{s\in S} f(s)s$ is an element of C, then the vector representation of the corresponding codeword is $(f(s))_{s\in S}$. Note also that when talking about the asymptotic properties of a group action code, we implicitly mean that we have an infinite family of group actions $\{\rho_n\}_{n\in I}$, with the group G_n acting on the set S_n via ρ_n. The family is indexed by the block length $n = |S_n|$ of the ρ_n-invariant code C_n.

4.1.1.4 Hyperelliptic curves

For general background on hyperelliptic curves over finite fields, see for instance Section 6.2 in [Sti93].

4.1.2 Group action codes literature

4.1.2.1 Cyclic and abelian codes

Binary abelian codes are invariant under the action of an abelian group G on a single copy of itself, i.e., they are ideals in the binary group algebra $\mathbb{F}_2[G]$. Cyclic codes correspond to the special case when G is cyclic. These codes, and specifically cyclic codes, have been extensively studied over the last 40 years. See for instance [PHB98]. Yet, it is still an open question whether there exist asymptotically good binary cyclic or abelian codes in general.

4.1.2.2 Codes in the binary group algebra of the dihedral group

These codes are invariant under the action of the dihedral group D_m on itself, i.e., they are ideals in the binary group algebra $\mathbb{F}_2[D_m]$. The Dihedral group D_m contains

$2m$ elements. It is generated by α and β subject to the relations $\alpha^2 = 1$, $\beta^m = 1$, and $\alpha\beta = \beta^{-1}\alpha$.

Codes in the binary group algebra of the dihedral group were introduced by MacWilliams [Mac69] in the setting of self dual codes. As far as we know, nothing was known before our work about their asymptotic distance properties.

4.1.2.3 Quasi-cyclic codes

Quasi-cyclic codes are invariant under the action of a cyclic group on k disjoint copies of itself, i.e., they are $\mathbb{F}_2[\mathbb{Z}/m\mathbb{Z}]$-submodules of $\mathbb{F}_2[\mathbb{Z}/m\mathbb{Z}]^k$.

Quasi-cyclic codes were first studied by Chen, Peterson, and Weldon [CPW69] in the setting when $m = p$ is prime. The result in [CPW69] says that if 2 is a primitive root of p (i.e., 2 generates $\mathbb{F}_p^\times$), a random quasi-cyclic code (i.e., an $\mathbb{F}_2[\mathbb{Z}/p\mathbb{Z}]$-submodule of $\mathbb{F}_2[\mathbb{Z}/p\mathbb{Z}]^k$ generated by a random element of $\mathbb{F}_2[\mathbb{Z}/p\mathbb{Z}]^k$) achieves the GV bound with a high probability. Without assuming the ERH (Extended Riemann Hypothesis), it is not known whether there are infinitely many primes with the above property. A later result by Kasami [Kas74] shows that if instead of working in $\mathbb{Z}/p\mathbb{Z}$, if we work in $\mathbb{Z}/p_0^l\mathbb{Z}$, where l can be varied and p_0 is fixed to be the largest known prime such that 2 is a primitive root of p_0, a random quasi-cyclic code achieves a slightly weaker bound than the GV bound.

A subsequent work by Chepyzhov [Che92], which was mentioned to us recently by Barg [Bar01], shows that in the cyclic prime case the condition in [CPW69] that requires 2 to be a primitive root of p can be relaxed to requiring that the size of the multiplicative group generated by 2 in $\mathbb{F}_p^\times$ grows faster than $\log p$ and hence the ERH can be avoided as it is not hard to show that there are infinitely many such primes.

4.1.2.4 Quadratic residues codes

Let p be a prime such that 2 is a quadratic residue, i.e., $p = \pm 1 \pmod 8$. Consider the decomposition $x^p - 1 = (x-1)q(x)\bar{q}(x)$ over $\mathbb{F}_2$ where $q(x) = \prod_{i \in Q}(x - \beta^i)$, $\bar{q}(x) = \prod_{i \in \bar{Q}}(x - \beta^i)$, Q is the set of quadratic residues modulo p, $\bar{Q} = \mathbb{F}_p^\times \backslash Q$, and β is a primitive p'th root of 1 in an extension field of $\mathbb{F}_2$. Binary quadratic residues codes

are the ideals of $\mathbb{F}_2[\mathbb{Z}/p\mathbb{Z}] = \mathbb{F}_2[x]/(x^p - 1)$ generated by one the polynomial $q(x)$, $\bar{q}(x)$ or one of their product with with the polynomial $x - 1$.

Other than being cyclic codes, these codes are invariant under the action of the subgroup $\left\{\begin{pmatrix} a & * \\ 0 & a^{-1} \end{pmatrix}\right\}_{a \neq 0}$ of $PSL_2(\mathbb{F}_p)$ on $\mathbb{F}_p$ by affine transformations. They are also extendible from $\mathbb{F}_p$ to $\mathbb{F}_p \cup \{\infty\}$ in such a way they are invariant under the action of $PSL_2(\mathbb{F}_p)$ by fractional linear transformations on $\mathbb{F}_p \cup \{\infty\}$. See [Lin99, War74, PHB98].

It is not known if binary quadratic residue codes can be asymptotically good.

One of the codes we will be studying in this chapter can be related to special nonbinary quadratic residue codes over $\mathbb{F}_4$.

4.1.2.5 Cayley graphs codes

Sipser and Spielman [SS96] constructed explicit binary asymptotically good Low density parity check codes based on the explicit constructions of Cayley graph expanders of Lubotzky, Phillips, and Sarnak [LPS88], and Margulis [Mar88]. The underlying Cayley graph group is $PSL_2(\mathbb{F}_p)$, p prime. These codes are realized as unbalanced bipartite graphs in such a way that the codewords are defined on the edges of the Cayley graph. They are invariant under the action of $PSL_2(\mathbb{F}_p)$ on more than one copy of itself.

4.1.3 Summary of Results

4.1.3.1 Asymptotically good codes in the group algebra of the dihedral group

The most natural class of group action codes are those that are invariant under the action of a group G on itself, i.e., those that are ideals in the binary group algebra $\mathbb{F}_2[G]$ of a group G. The case when G is cyclic (respectively, abelian) corresponds to the case of cyclic (respectively, abelian) codes. Such codes are very well studied, yet it is still an open question whether there exist asymptotically good cyclic or abelian codes. The case when G is nonabelian was studied and introduced by MacWilliams

[Mac69] in the setting of the dihedral group D_m. However, it was not noted that this group algebra contains asymptotically good codes.

Our result in Section 4.4 says that if we use a group slightly stronger than a cyclic group, and namely the dihedral group, the existence of asymptotically good codes can be guaranteed in the group algebra. In particular, we show that for infinitely many m's, a random ideal in $\mathbb{F}_2[D_m]$ is an asymptotically good rate $1/2$ binary code. The first condition we need on m is that the smallest size of the multiplicative group generated by 2 in $\mathbb{F}_p^\times$, as p runs over the prime divisors of m (or equivalently the smallest dimension of a nontrivial $\mathbb{F}_2$-representation of $\mathbb{Z}/m\mathbb{Z}$), grows asymptotically faster than $\log m$. We require also for simplicity another condition and we argue that it is satisfied by all the primes $p = \pm 5 \pmod 8$. By random here we mean according to some specific distribution based on the $\mathbb{F}_2$-representations of D_m. The implicit bound on the relative minimum distance is $h^{-1}(1/4)$, where h is the binary entropy function.

As far as we know, this is the first provably good randomized construction of codes that are ideals in the group algebra of a group. We are not also aware if it was previously known that there exists asymptotically good codes that are ideals in the group algebra of a group.

We leave the corresponding analysis till the end of this chapter since it is based on the analysis of the quasi-abelian case that we summarize next.

4.1.3.2 Quasi-abelian codes up to the GV bound

Rather than considering the action of a group G on itself, one can consider the action of G on k disjoint copies of itself. This means looking at codes that are $\mathbb{F}_2[G]$-submodules of $\mathbb{F}_2[G]^k$. When G is cyclic, these are quasi-cyclic codes.

We consider the case when G is an abelian group of odd order. Our result in Section 4.2 says that if the dimension $L(G)$ of the smallest irreducible $\mathbb{F}_2$-representation of G grows faster than logarithmically in the order of the G, then an $\mathbb{F}_2[G]$-submodules of $\mathbb{F}_2[G]^k$ generated by a random element of $\mathbb{F}_2[G]^k$ achieves the GV bound at rate $1/k$ with a high probability. Here random means almost uniformly in a suitable sense.

Roughly, almost all abelian group of odd order satisfy the above condition. This includes almost all cyclic groups of prime order. Since G is abelian, $L(G)$ depends only on the order of G, and it is the smallest size of the multiplicative group generated by 2 in $\mathbb{F}_p^\times$, where p runs over the prime divisors of m.

Comparing our result with the literature on quasi-cyclic codes surveyed in Section 4.1.2.3, we see that the innovation in our result is in the fact that it holds for abelian groups that are not necessarily cyclic of prime order which has the advantage of supplying more block lengths. Our condition on the order of the group is a generalization of the condition of Chepyzhov [Che92] from cyclic groups of prime order to arbitrary abelian groups of odd order. Another related work, which was announced for the first time in the same workshop [DCC01] in which our results were presented, is that of Meshulam and Wigderson [MW02, MW03] who arrived to a similar condition on the distribution of the dimensions of the irreducible representations of abelian groups in the setting of expander graphs constructions.

In Section 4.2.1, we tune the construction in the rate 1/2 case, i.e., when $k = 2$. We show that under the same condition on the order of G, and over the choice of a uniformly random element b of $\mathbb{F}_2[G]$, the $\mathbb{F}_2[G]$-submodule generated by $(b, 1)$ in $\mathbb{F}_2[G]^2$ achieves the GV bound at rate 1/2 with a high probability.

4.1.3.3 An explicit construction based on quadratic residues

Of special interest to us is the above tuned construction in the case when the group G is cyclic of prime order.

Let $G = \mathbb{Z}/p\mathbb{Z}$, p prime. Define the PQC (Prime Quasi-Cyclic) codes ensemble as follows. The codes in the PQC ensemble are in one to one correspondence with the subsets A of $\mathbb{F}_p$. For each such A, there is a code in the PQC ensemble given by the set of all pairs $(r_S r_A, r_S)$ as S spans the subsets of $\mathbb{F}_p$. Here $r_A \stackrel{\text{def}}{=} \sum_{g \in A} g \in R \stackrel{\text{def}}{=} \mathbb{F}_2[G]$. We know that, for almost all the primes p, a random subset A of $\mathbb{F}_p$ gives a code in the PQC ensemble that achieves the GV bound at rate 1/2 with a high probability.

We consider in Section 4.3 the problem of explicitly constructing a good code from the above ensemble, i.e., the problem of explicitly constructing a good set A.

We study specifically the natural explicit code that corresponds to setting A to be the set of quadratic residues in $\mathbb{F}_p$. Call this code the QQR (Quasi-cyclic Quadratic Residues) Code. For nondegeneracy reasons, we conjecture that this explicit code is asymptotically good and probably achieves the binary GV bound at rate 1/2. The intuition is the following. The explicit construction problem consists of finding a good A. We know that almost all the A's are good and we want to construct a good one. What are the bad A's? Characterizing the bad ones is hard. But, intuitively, a bad A seems to be a one that is too small, too large, or is degenerate in some sense under addition modulo p. So a promising candidate for A seems to be a moderately sized set that can be defined without using addition at all. Since p is a prime, there is a natural choice for such an A: a multiplicative subgroup of $\mathbb{F}_p^\times$ whose size is around half the size of $\mathbb{F}_p^\times$. But, there is only one such subgroup: the set of quadratic residues mod p.

We present a preliminary analysis of the minimum distance of the QQR code which (when the prime p is special enough) reduces the problem of bounding its minimum distance to obtaining a bound on the the maximum number of rational points on a curve from a family of hyperelliptic curves over $\mathbb{F}_p$. We show that the codewords in this codes are in one to one correspondence with special hyperelliptic curves over $\mathbb{F}_p$, where the number of zeros in a codewords is up to an additive $O(1)$ term equal to the number of $\mathbb{F}_p$-rational points on the corresponding curve. This is when the prime $p = \pm 5 \pmod 8$. The curves are of the form $y^2 = f(x)$, where $f(x)$ is a nonconstant square free polynomial of even degree in $\mathbb{F}_p[x]$ that has all its zeros in $\mathbb{F}_p$. We argue that maximizing over the odd degree polynomials can only affect the bound by an additive $O(1)$ term. This extends the family of curves to those of the form $y^2 = f(x)$ where $f(x)$ a is a nonconstant square free polynomial in $\mathbb{F}_p[x]$ that has all its zeros in $\mathbb{F}_p$.

The general bound of Weil for bounding the number of rational point is inadequate in our situation since it becomes trivial when the genus of the curve becomes large. Other than being in the special setting of hyperelliptic curves, the more important additional special features in our case are:

- $f(x)$ splits completely over $\mathbb{F}_p$, and

- $\mathbb{F}_p$ is a field of prime order, which is a situation where it is not known whether Weil's bound can be tight for any curve.

This leads to the following conjecture, whose correctness is equivalent to the asymptotic goodness of the QQR code with β as its minimum relative distance:

Conjecture 4.3.6: There exists $\beta > 0$ such that for any prime p (or for infinitely many primes p) and for any non-constant square free polynomial $f(x)$ that splits completely over $\mathbb{F}_p$, the number of $\mathbb{F}_p$-rational points on the hyperelliptic curve $y^2 = f(x)$ is smaller than $2(1-\beta)p$.

To support the conjecture we note in Section 4.3.4 that the classical special class of curves where f has only two non-zero coefficients and where Weil's bound can become tight over square fields cannot have too many rational points in our prime field setting.

The splitting condition may be only needed to handle the high genus cases. A stronger statement might be true:

Conjecture 4.3.7: There exists $\beta > 0$ such that for any prime p (or for infinitely many primes p) and for any nonconstant square free polynomial $f(x)$ whose degree is sublinear in p, the number of $\mathbb{F}_p$-rational points on the hyperelliptic curve $y^2 = f(x)$ is smaller than $2(1-\beta)p$?

This basically means improving Weil's bound in the setting of hyperelliptic curves over prime fields.

The explicit binary codes we are looking at are not essentially new codes in the sense that they can be related after minor modifications to a special class of non-binary classical cyclic quadratic residue codes over $\mathbb{F}_4$ when the prime is special enough as we explain in Section 4.3.5. What is new is that this special case comes from a code ensemble where the random is good, and where the code we are studying seems highly non-degenerate. Our technical contribution is the reduction of the minimum distance problem to a problem about points on curves. The importance of the conjecture is that its correctness leads to at least a new family of explicit asymptotically good binary

codes. From the perspective of the construction origin, it is tempting to speculate also that the construction achieves the binary GV bound. Proving the conjecture or finding counter examples for all but finitely many primes is very desirable.

Comparing further with the literature, we note that the relation between codes and curves we are talking about is very far from algebraic geometric codes since in our setting each codeword corresponds to a curve whose number of rational points is related to the weight of the codeword. We note also that the relation between codes and character sums has been explored before but in the setting of fields of order a power of 2 (See [PHB98]). The relation we are indicating is different as it is over fields of prime order.

4.2 A randomized construction from abelian groups actions

We establish in this section the Claims of Section 4.1.3.2. We consider the case when G is an abelian group of odd order. We argue in Theorems 4.2.1 and 4.2.4 that if the dimension $L(G)$ of the smallest irreducible $\mathbb{F}_2$-representation of G grows faster than logarithmically in the order of the G, then an $\mathbb{F}_2[G]$-submodule of $\mathbb{F}_2[G]^k$ generated by a random element of $\mathbb{F}_2[G]^k$ achieves the GV bound with a high probability. Since G is abelian, $L(G)$ depends only on the order of G, and it is the smallest size of the multiplicative group generated by 2 in $\mathbb{F}_p^\times$, where p runs over the prime divisors of m. See Lemma 4.2.5. We note that roughly, almost all abelian groups of odd order satisfy the above condition. Finally, we tune the construction in Section 4.2.1 in a way that will be of special interest to us in Section 4.3.

Theorem 4.2.1 *Let G be a finite abelian group of odd order m, and consider its binary group algebra*

$$\mathbb{F}_2[G] \stackrel{\text{def}}{=} \{\sum_{g\in G} f(g)g | f : G \to \mathbb{F}_2\}.$$

Consider the randomized construction of codes

$$C_{a,b} = \{(fa, fb) | f \in \mathbb{F}_2[G]\},$$

where a, b are selected uniformly at random from $\mathbb{F}_2[G]$.

Let $L(G)$ be the smallest dimension of a non-trivial $\mathbb{F}_2$-representation of G.

Let $\delta > 0$ nbe such that $h(\delta) \leq \frac{1}{2} - \frac{\log m}{2L(G)}$. Then the probability that the minimum relative distance of the code $C_{a,b}$ is below δ or the rate of $C_{a,b}$ is below $\frac{1}{2} - \frac{1}{2m}$ is at most $2^{-2L(G)(1/2-h(\delta))+5\log m}$, where h is the binary entropy function.

Therefore, if $L(G)$ grows asymptotically faster than $\log m$, then the code $C_{a,b}$ achieves the GV bound for rate $1/2$ with a high probability.

Note that $L(G)$ is the lowest dimension of a nontrivial $\mathbb{F}_2[G]$-module, and more specifically the lowest dimension of a nontrivial irreducible ideal in $\mathbb{F}_2[G]$. By a trivial $\mathbb{F}_2[G]$-module we mean a R-module M such that $rm = m$, $\forall m \in M$ and $r \in \mathbb{F}_2[G]$.

Proof. Let $R \stackrel{\text{def}}{=} \mathbb{F}_2[G]$. Let P be the probability that $C_{a,b}$ has dimension below $m-1$ and minimum distance below $2m\delta$, where δ is say below $1/2$ for the moment. P is at most the probability that there is an $f \in R$, $f \neq 0$ and $f \neq e_0 \stackrel{\text{def}}{=} \sum_{g \in G} g$, such that the event

$$E(f, a, b) : 0 \leq w(fa) + w(fb) < 2m\delta$$

happens. This is because $(e_0 a, e_0 b)$ is either $(e_0, 0)$, $(0, e_0)$, (e_0, e_0), or $(0, 0)$, and thus $w(e_0 a) + w(e_0 b) = m$, $2m$, or 0. The first two values are above $2\delta m$ and the last can only decrease the rank by 1. Thus, by the union bound on f,

$$P \leq \sum_{f \in R, f \neq 0, e_0} Pr_{a,b}[E(f, a, b)] \leq \sum_{l=2}^{m} |D_l| \max_{f \in D_l} Pr_{a,b}[E(f, a, b)]. \tag{4.1}$$

where

$$D_l = \{f \in R | dim_{\mathbb{F}_2} fR = l\},$$

and fR is the ideal generated by f in R. Note that we excluded the case $l = 0$ and $l = 1$ since they can only happen when $f = 0$ and $f = e_0$ respectively.

For all $f \neq e_0$, the ideal fR is nontrivial, so $dim_{\mathbb{F}_2} fR \geq L(G)$. Thus

$$D_l = \emptyset \text{ for all } 2 \leq l < L(G). \tag{4.2}$$

Let

$$\Omega_l = \{I \text{ an ideal of } R \mid dim_{\mathbb{F}_2} I = l\},$$

so we have

$$|D_l| \leq 2^l |\Omega_l| \tag{4.3}$$

Consider any l, and any $f \in D_l$. We have

$$\begin{aligned} Pr_{a,b}[E(f,a,b)] &\leq \sum_{r_1,r_2 \in fR \text{ s.t. } 0 \leq w(r_1)+w(r_2)<2m\delta} Pr_{a,b}[fa = r_1 \text{ and } fb = r_2] \\ &\leq 2^{-2l} \sum_{w_1,w_2 \geq 0; w_1+w_2<2\delta m} |I^{(w_1)}||I^{(w_2)}|, \end{aligned} \tag{4.4}$$

where $I = fR$, and if I is an ideal, by $I^{(w)}$ we mean

$$I^{(w)} \stackrel{\text{def}}{=} \{r \in I | w(r) = w\}.$$

The 2^{-2l} term is the value of $Pr_{a,b}[fa = r_1 \text{ and } fb = r_2]$; indeed, for any $r \in fR$,

$$Pr_a[fa = r] = \frac{|Ker\ \Phi_f|}{|R|} = \frac{1}{|fR|} = 2^{-l},$$

where $\Phi_f : R \twoheadrightarrow fR$ is given by $a \mapsto fa$.

Replacing (4.2), (4.3), and (4.4) in (4.1), we get

$$P \leq \sum_{l=L(G)}^{m} 2^{-l} |\Omega_l| \max_{I \in \Omega_l} \sum_{w_1,w_2 \geq 0; w_1+w_2<2\delta m} |I^{(w_1)}||I^{(w_1)}|. \tag{4.5}$$

Note that so far we have not used any property that depends on G being abelian. Note also that the maximum above can be replaced by an expected value, but we will

not need that.

Lemma 4.2.2 *If I is an ideal in R of dimension l, then $|I^{(w)}| \leq 2^{lh(w/m)}$, where h is the binary entropy function.*

Proof. This follows from the work of Piert [Pir85] and Shparlinsky [Shp86]. In fact this holds when $R = \mathbb{F}_2[G]$, and G is an arbitrary group of size m. The result in [Pir85, Shp86] says the following. Let C be a subset of $\{0,1\}^J$ of size 2^l, J an index set of size m. Say that a subset A of J is an information set of C if the projection map from C to $\{0,1\}^A$ is a bijection, thus $|A| = l$. Say that C is balanced, if there exists $r \geq 1$ and information sets $A_1, \ldots, A_u$ of C such that for all i in J, the number of j's such that $i \in A_j$ is exactly r. Note that the A_i's need not be distinct. The result says that if C is balanced, then the number of vectors in C of weight w is at most $2^{lh(w/m)}$. The proof is a double counting argument. This is directly applicable to the case when C is an ideal in $\mathbb{F}_2[G]$. The reason is that since C is linear it must contain an information set $S \subset G$ of size l, and since C is invariant under the action of G, the $\{Sg\}_{g \in G}$ are informations sets also. These information sets make C balanced because for each a in G, the number of g's such that $a \in Sg$ is exactly $|S|$. ▼

Lemma 4.2.3 $|\Omega_l| \leq m^{l/l_0+1}$, *where* $l_0 = L(G)$.

Proof. Here we use the fact that G is a abelian. In general, since $|G|$ is odd, R is semisimple. Let $R = R_0 \oplus R_1 \oplus \ldots \oplus R_s$ be the unique decomposition of R into indecomposable 2-sided ideals. The R_i's are simple rings. Since G is abelian the R_i's are irreducible and they are the only irreducible ideals in R (Each R_i is actually a field with its idempotent as a unit element). Thus each ideal in R is of the form $\oplus_{i \in A} R_i$ for some subset A of $\{0, 1, \ldots, s\}$. This fact is the reason behind the bound on $|\Omega_l|$; if G was non abelian then $|\Omega_l|$ can be much larger than this because each R_i may contain many irreducible ideals. Without loss of generality, say that R_0 is the trivial one dimensional ideal, i.e., $R_0 = (\sum_{g \in G} g)R$. Thus, for each $i \neq 0$, the dimension of R_i is at least $l_0 = L(G)$. So, $1 + l_0(s-1) \leq m$. If I is an ideal of dimension l, then it

is a direct sum of at most $l/l_0 + 1$ of the R_i's. There are at most s^{l/l_0+1} such direct sum, so $|\Omega_l| \leq s^{l/l_0+1} \leq m^{l/l_0+1}$. ▼

Note that we can get a sharper bound, but this is sufficient for our purpose.

Replacing the estimates in Lemmas 4.2.2 and 4.2.3 in (4.5), we get

$$\begin{aligned} P &\leq \sum_{l=l_0}^{m} 2^{-l} m^{l/l_0+1} \sum_{w_1,w_2\geq 0; w_1+w_2<2\delta m} 2^{l(h(w_1/m)+h(w_2/m))} \\ &\leq \sum_{l=l_0}^{m} 2^{-l} m^{l/l_0+1} (2\delta m)^2 2^{2lh(\delta)} \qquad \text{(since } h \text{ is convex)} \\ &\leq \sum_{l=l_0}^{m} 2^{-2l\left(\frac{1}{2}-h(\delta)-\frac{\log m}{2l_0}\right)+3\log m}. \end{aligned}$$

If $\frac{1}{2} - h(\delta) - \frac{\log m}{2l_0} \geq 0$, we get

$$\begin{aligned} P &\leq 2^{-2l_0(\frac{1}{2}-h(\delta)-\frac{\log m}{2l_0})+4\log m} \\ &= 2^{-2l_0(\frac{1}{2}-h(\delta))+5\log m}. \end{aligned}$$

This completes the proof of Theorem 4.2.1. ■

Note that the fact that the estimate of Lemma 4.2.3 fails for nonabelian groups does not mean that they do not lead to good codes in the setting of this randomized construction. All that it means is that the argument may need some modifications. But in all cases, it will become clear in Section 4.4 that the reason why Lemma 4.2.3 fails for nonabelian groups makes them subject to a more natural randomized construction.

More generally,

Theorem 4.2.4 *Let G be an abelian group of order m, and consider the randomized codes construction*

$$C_{a_1,\ldots,a_k} = \{(fa_1, \ldots, fa_k) | f \in \mathbb{F}_2[G]\},$$

where $a_1, \ldots, a_k$ are selected uniformly at random from R^*, and R^* is the set of even weight strings in $R \stackrel{\text{def}}{=} \mathbb{F}_2[G]$.

If $L(G)$ is asymptotically larger than $\log m$, then the code $C_{a_1,\ldots,a_k}$ achieves the GV bound for rate $1/k$ with a high probability.

Proof. The proof is by the same argument in Theorem 4.2.1. We need this even weight technicality in order to avoid the dominance of some bad events when k is large enough. The fact that the a_i's have even weight will take care of the case when $f = e_0$ since then $e_0 a_i = 0$ always. ■

Lemma 4.2.5 *Since G is abelian, $L(G)$ depends only on the order of G and is given by*

$$l(m) = \min\{\#\langle 2\rangle_p | p \text{ a prime divisor of } m\},$$

where $\langle 2\rangle_p$ is the multiplicative subgroup generated by 2 in $\mathbb{F}_p^\times$.

Proof. Since G is abelian say that $G = G_1 \times \ldots \times G_t$, $G_i \cong \mathbb{Z}/p_i^{k_i}\mathbb{Z}$, p_i prime. Thus $m = \prod_i p_i^{k_i}$. If $\rho : G \to GL_l(\mathbb{F}_2)$ is a nontrivial $\mathbb{F}_2$-representation of G, then the restriction of ρ on one the G_i's must be nontrivial, thus $L(G) \geq \min_i L(G_i)$. Conversely, given a representation $\rho_i : G_i \to GL_l(\mathbb{F}_2)$ of G_i, we can extend ρ_i to G via $\rho_i(g_1 \ldots g_t) = \rho_i(g_i)$. Thus $L(G) = \min_i L(G_i)$. So we can assume without loss of generality that G is cyclic of order a power of a prime, say G is $\mathbb{Z}/p^k\mathbb{Z}$. Then the dimensions of the irreducible $\mathbb{F}_2$-representations of G are precisely the sizes of the equivalence classes in $(\mathbb{Z}/p^k\mathbb{Z})/\sim$, where $a \sim b$ if $a = 2^i b \pmod{p^k}$ for some i. The trivial representation corresponds to the class consisting of 0. Thus

$$L(\mathbb{Z}/p^k\mathbb{Z}) = \min_{0<a<p^k} h(a, p^k) \quad \text{where} \quad h(a, p^k) = \min_{i\geq 1; a2^i = a(\bmod p^k)} i.$$

Now, $h(a, p^k) = h(1, p^i)$ where $p^i = a/\gcd(p^k, a)$ as it easy to check. Thus

$$L(\mathbb{Z}/m\mathbb{Z}) = \min_{i=1,\ldots,k} h(1, p^i) = h(1, p)$$

because $h(1, p^i) \geq h(1, p)$ for all $i \geq 1$, and hence the claim since $h(1, p) = \#\langle 2\rangle_p$. ■

Now, if $r : \mathbb{Z}_+ \to \mathbb{Z}_+$ is a nondecreasing function, let

$$Z(r) = \{m \in \mathbb{Z}_+ | l(m) \geq r(m)\}.$$

So any family of abelian groups whose orders is in $Z(r)$ leads to rate 1/2 codes up the GV bound as long as $r(m) \gg \log m$. Let $P(r)$ be the set of odd primes in $Z(r)$.

Lemma 4.2.6 *When $r(m) \ll \sqrt{m/\log m}$, $P(r)$ is infinite and contains almost all the primes.*

Proof. This statement appears in Chepyzhov [Che92], but we include a proof for completeness. Say that a prime is bad if it is not in $P(r)$, and let B_n be the set of bad primes smaller than n. If p is a bad prime then there exists integers a and k such that $0 < a < r(p)$ and $2^a - 1 = kp$. Since $r(n)$ is nondecreasing, we have

$$\begin{aligned} |B_n| &\leq \#\{(a,k) \mid 0 < a < r(n) \text{ and } (2^a-1)/k \text{ prime}\} \\ &\leq r(n)\log\left(2^{r(n)}-1\right) \leq r^2(n), \end{aligned}$$

and hence the lemma from the prime numbers density theorem. ■

So we have many infinite families of abelian groups that lead to codes up to the GV bound in the sense of Theorem 4.2.1, for instance:

- The cyclic groups of prime order, where the primes are in $P(r)$, and $r(m) = \log m \log\log m$.
- Any version of the Abelian groups of order pq where $p, q \in P(r)$, $r(m) = \log m \log\log m$, and $p^k, q^k > pq$ for some prespecified constant k.
- Any version of the abelian groups of order p^k where $p \in P(r)$, $r(m) = \log m \log\log m$, and k is a prespecified constant.
- Any version of the abelian groups of order $p^{k(p)}$ where $p \in P(r)$, $r(m) = \frac{\sqrt{m}}{logm}$, and $k(p) < \frac{\sqrt{p}}{log^3 p}$.

4.2.1 Tuning the construction

We can tune the rate 1/2 randomized construction in Theorem 4.2.1 in a way that uses less randomness as follows.

Theorem 4.2.7 *Let G be a finite abelian group of odd order m. Consider the randomized construction of codes*

$$C_b = \{(f, fb) | f \in \mathbb{F}_2[G]\},$$

where b is selected uniformly at random from $\mathbb{F}_2[G]$.

If $l(m) \gg \log m$, then the code C_b achieves the GV bound for rate $1/2$ with a high probability.

Proof. We use the terminologies of Theorem 4.2.1. In the setting of Theorem 4.2.1, $C_b = C_{1,b}$, where b is selected uniformly at random from $R \stackrel{\text{def}}{=} \mathbb{F}_2[G]$.

Let $R^\times$ be the set of invertible elements in R. First we note that

$$\begin{aligned} Pr_{b\in R}[d_{min}(C_b) < 2\delta p] &= Pr_{a\in R^\times, b\in R}[d_{min}(C_{a,ab}) < 2\delta p] \\ &= Pr_{a\in R^\times, b\in R}[d_{min}(C_{a,b}) < 2\delta p]. \end{aligned}$$

This is true because if $b \in R$, then $C_b = C_{a,ab}$ and $aR = R$, $\forall a \in R^\times$. Now, we proceed as in Theorem 4.2.1. Let

$$P = Pr_{a\in R^\times, b\in R}[d_{min}(C_{a,b}) < 2\delta p],$$

so P is the probability that there is an $f \in R$, $f \neq 0$, such that the event

$$E(f, a, b) : 0 \leq w(fa) + w(fb) < 2m\delta$$

happens. Thus

$$P \leq \sum_{f\in R, f\neq 0} Pr_{a\in R^\times, b\in R}[E(f, a, b)] \leq \sum_{l=l(m)}^{m} |D_l| \max_{f\in D_l} Pr_{a\in R^\times, b\in R}[E(f, a, b)],$$

where D_l as in Theorem 4.2.1. Consider any l, and any f in D_l. Then

$$Pr_{a\in R^\times, b\in R}[E(f, a, b)]$$

is at most

$$\sum_{r_1 \in fR^\times, r_2 \in fR \text{ s.t. } 0 \leq w(r_1)+w(r_2) < 2m\delta} Pr_{a \in R^\times}[fa = r_1] Pr_{b \in R}[fb = r_2],$$

and this is at most

$$\max_{r \in fR^\times} Pr_{a \in R^\times}[fa = r] \max_{r \in fR} Pr_{b \in R}[fb = r] \sum_{w_1, w_2 \geq 0; w_1 + w_2 < 2\delta m} |I^{(w_1)}||I^{(w_2)}|,$$

where $I = fR$, and $I^{(w)}$ is as defined in Theorem 4.2.1. As before, the second maximum is exactly 2^{-l}. We will argue that the first is 2^{-l} also.

Say that $R = \oplus_{i=0}^{s} R_i$ is the unique decomposition of R into irreducible ideals. Thus each R_i is a field with its idempotent as a unit element, and $R^\times = \oplus_{i=0}^{s} R_i^\times$, where $R_i^\times$ is the multiplicative group of the field R_i. Let $V \subset [s]$ such that $f = \sum_{i \in V} f_i$, f_i nonzero in R_i. Since each element a of $R^\times$ is of the form $\sum_{i=1}^{s} a_i$, where each a_i is a nonzero element of R_i, we have $fa = \sum_{i \in V} f_i a_i$, hence the number of a's in $R^\times$ such that $fa = r$ is exactly $\prod_{i \notin V} |R_i^\times|$, i.e., all the elements of $fR^\times$ are equally likely to appear as fa when a is selected from $R^\times$, which means that $Pr_{a \in R^\times}[fa = r] = 2^{-l}$.

The rest is exactly as in Theorem 4.2.1. Note that regarding the rate, here we always have $dim C_b = m$.

■

4.3 An explicit construction based on quadratic residues

We elaborate on the explicit construction introduced in Section 4.1.3.3. We study a natural explicit code based on quadratic residues, that we call the QQR code, in the setting of quasi-cyclic codes of prime order from Theorem 4.2.7 restricted to the case when the group is cyclic of prime order. For nondegeneracy reasons, we conjecture that this explicit code is asymptotically good and probably achieves the binary GV bound at rate 1/2. We present a preliminary analysis of the minimum distance

of the QQR code which reduces the problem of bounding its minimum distance to obtaining a bound on the maximum number of rational points on a curve from a family of hyperelliptic curves over $\mathbb{F}_p$ (Corollary 4.3.5, Conjectures 4.3.6 and 4.3.7). We show in Theorem 4.3.3 that (when $p = 3 \pmod 8$) the codewords in the QQR code are in one to one correspondence with special hyperelliptic curves over $\mathbb{F}_p$, where the number of zeros in a codewords is up to an $O(1)$ term equal to the number of $\mathbb{F}_p$-rational points on the corresponding curve. The curves are of the form $y^2 = f(x)$, where $f(x)$ is a nonconstant square free polynomial of even degree in $\mathbb{F}_p[x]$ that has all its zeros in $\mathbb{F}_p$. Then we show in Theorem 4.3.4 that maximizing over the odd degrees polynomials can only affect the bound by an additive $O(1)$ term. This extends the family of curves to those of the form $y^2 = f(x)$, where $f(x)$ is a nonconstant square free polynomial in $\mathbb{F}_p[x]$ that has all its zeros in $\mathbb{F}_p$ (Corollary 4.3.5). We discuss the prime field situation in Section 4.3.4, and we note in Section 4.3.5 that the QQR code can be related to a special class of non-binary cyclic quadratic residue codes over $\mathbb{F}_4$.

Consider the codes ensemble from Theorem 4.2.7 restricted to the case when the group is cyclic of prime order.

Definition 4.3.1 "The Prime Quasi-Cyclic codes (PQC) ensemble": *Let p be a prime. The block-length-$2p$ PQC ensemble consists of the rate-$1/2$ codes*

$$C_A = \{(r_S r_A, r_S) : S \subset \mathbb{F}_p\} \tag{4.6}$$

indexed by the subsets A of $\mathbb{F}_p$.

Here if U is a subset of $\mathbb{F}_p$,

$$r_U \stackrel{\text{def}}{=} \sum_{g \in U} g$$

as an element of the binary group algebra $\mathbb{F}_2[\mathbb{Z}/p\mathbb{Z}]$.

We remind the reader with construction intuition introduced in Section 4.1.3.3. Theorem 4.2.7 implies that a random code from the PQC ensemble achieves the GV bound at rate $1/2$ for almost all the primes. The explicit construction problem consists of finding a good A. We know that almost all the A's are good, and we

want to construct a good one. What are the bad A's? Characterizing the bad ones is hard. But, intuitively, a bad A seems to be a one that is too small, too large, or is degenerate in some sense under addition modulo p. So a promising candidate for A seems to be a moderately sized set that can be defined without using addition at all. Since p is a prime, there is a natural choice for such an A: a multiplicative subgroup of $\mathbb{F}_p^\times$ whose size is around half the size of $\mathbb{F}_p^\times$. But, there is only one such subgroup: the set of quadratic residues mod p. In what follows we start studying the explicit construction resulting from setting A to be the set of quadratic residues modulo p. Our objective is to bound the minimum distance of the resulting code. Call the resulting rate half code the Quasi-cyclic Quadratic Residue (QQR) code.

Definition 4.3.2 "The Quasi-cyclic Quadratic Residue (QQR) code": *Let p be a prime, and let Q be the multiplicative subgroup Q of $\mathbb{F}_p^\times$ of index 2, i.e., Q is the set of quadratic residues mod p. The block-length-$2p$ QQR code is the code C_Q in the PQC ensemble.*

The main point to keep in mind is that the reason that we are looking at this code is that it is a highly nondegenerate choice in an ensemble of codes where almost all the codes achieve the GV bound. So it is very tempting to suspect that this codes achieves the GV bound or is at least asymptotically good.

4.3.1 The minimum distance of the QQR code

When 2 and -1 are both non-quadratic residues mod p, we can exhibit a one-to-one correspondence between the codewords of the QQR code and a family of hyperelliptic curves over $\mathbb{F}_p$.

If $S \subset \mathbb{F}_p$, let

$$f_S(x) \stackrel{\text{def}}{=} \prod_{a \in S} (x - a) \in \mathbb{F}_p[x].$$

Let ψ be the quadratic residues character, i.e.,

$$\psi(a) = \begin{cases} 1 & \text{if } a \in Q \\ -1 & \text{if } a \in \mathbb{F}_q^\times \backslash Q \\ 0 & \text{if } a = 0. \end{cases}$$

Theorem 4.3.3 *Let p be a prime such that 2 and -1 are non-quadratic residues mod p (i.e., $p = 3 \pmod 8$), then the block-length-$2p$ QQR code can be expressed as*

$$\{(r_{\bar{Q}} r_S, r_Q r_S) : S \text{ a subset of } \mathbb{F}_p\}, \tag{4.7}$$

where $\bar{Q}$ is the complement of Q in $\mathbb{F}_p^\times$. Moreover, if $(r_{\bar{Q}} r_S, r_Q r_S)$ is a codeword of the QQR code, then the weight of this codeword can be expressed in terms of a character sum as

$$p - \sum_{a \in \mathbb{F}_p} \psi(f_S(a)) \tag{4.8}$$

if $|S|$ is even, and

$$p + \sum_{a \in \mathbb{F}_p} \psi(f_{\mathbb{F}_p \backslash S}(a)) \tag{4.9}$$

if $|S|$ is odd.

Note that if -1 is a quadratic residue, we get the same relation if we restrict our attention to the the even weight codewords in the QQR code.

Proof. To avoid confusion between addition and multiplication in $\mathbb{F}_p$, we will be working with the polynomial version of the group algebra $R = \mathbb{F}_2[\mathbb{Z}/pZ]$. So $R = \mathbb{F}_2[x]/(x^p - 1)$ and $r_S \in R$ means it is a polynomial $r_S(x) = \sum_{i \in S} x^i$.

The reason why the QQR code as defined in (4.6) with $A = Q$ can be expressed as as in (4.7) is that $(r_Q, 1)$ and $(r_{\bar{Q}}, r_Q)$ generate the same R-submodule in R^2. This is because due to the fact that 2 and -1 are in $\bar{Q}$, we have $r_Q^2 = r_{\bar{Q}}$, and $r_Q^3 = 1$. We have $r_Q^2 = r_{\bar{Q}}$ since $r_Q^2 = r_{2Q}$ because we are in characteristic 2 and $2Q = \bar{Q}$ since $2 \in \bar{Q}$. To see why $r_Q^3 = 1$, note that

$$r_Q^3 + 1 = (r_Q + 1)(1 + r_Q + r_Q^2) = (r_Q + 1) r_{\mathbb{Z}/p\mathbb{Z}}.$$

Now $r_{\mathbb{Z}/p\mathbb{Z}}$ annihilates all the even weight elements in R and absorbs all the odd weight ones, thus $(r_Q+1)r_{\mathbb{Z}/p\mathbb{Z}}=0$ since $|Q|=(p-1)/2$ is odd (because -1 is a nonquadratic residue), i.e., $r_Q^3+1=0$.

If $r_A, r_B \in R$, we can express the weight of $r_A r_B$ as

$$w(r_A r_B)=\sum_{l\in\mathbb{F}_p} parity|A\cap(l-B)|.$$

Let $r_S \in R$. We have

$$p-w(r_Q r_S)-w(r_{\bar{Q}} r_S)=\sum_{l\in\mathbb{F}_p} 1-parity|Q\cap(l-S)|-parity|\bar{Q}\cap(l-S)|.$$

Consider a term of this summation for a fixed l and call it $T_l(S)$.

Assume that $|S|$ is even. We have two cases to consider.

- **Case 1.** When $0\in(l-S)$.

  In this case $T_l(S)=0$. The reason is that $|(l-S)\backslash\{0\}|$ is odd, and hence will intersect either Q or $\bar{Q}$ evenly but not both.

- **Case 2.** When $0\notin(l-S)$.

  In this case $|Q\cap(l-S)|$ is odd if and only if $|\bar{Q}\cap(l-S)|$ is odd since $|l-S|$ is even. Thus

$$T_l(S)=\begin{cases} -1 & \text{if } |\bar{Q}\cap(l-S_f)| \text{ is odd} \\ 1 & \text{if } |\bar{Q}\cap(l-S_f)| \text{ is even.}\end{cases}$$

It follows that

$$T_l(S)=\prod_{a\in l-S}\psi(a)=\psi(\prod_{a\in S}(l-a))=\psi(f_S(l)),$$

and hence (4.8).

If $|S|$ is odd, we get (4.9) since $T_l(S)=-T_l(\mathbb{F}_p\backslash S)$. This is true because $|Q\cap(l-S)|$ and $|Q\cap(l-\mathbb{F}_p\backslash S)|$ have opposite parities since $|Q|=(p-1)/2$ is odd, and similarly for $|\bar{Q}\cap(l-S)|$ and $|\bar{Q}\cap(l-\mathbb{F}_p\backslash S)|$.

This completes the proof of Theorem 4.3.3. ■

Thus the minimum distance of the QQR code is given by

$$p - \max\{|\sum_{a\in\mathbb{F}_p} \psi(f_S(a))| : S \text{ a nonempty subset of } \mathbb{F}_p \text{ of even cardinality }\}.$$

Rather than talking about such character sums, we can talk about the corresponding hyperelliptic curves.

In general if $f(x) \in \mathbb{F}_p[x]$ is a square free polynomial of degree d and leading coefficient f_d, consider the projective nonsingular hyperelliptic curve whose function field is $\mathbb{F}_p(x, y), y^2 = f(x)$. Let g be its genus, N its number of $\mathbb{F}_p$-rational points, and n_∞ its number of $\mathbb{F}_p$-rational point at ∞. Then $N = \sum_{a\in\mathbb{F}_p} \psi(f(a)) - p + n_\infty$, and

- $g = (d-2)/2$ and $n_\infty = \psi(f_d) + 1$, when d is even
- $g = (d-1)/2$ and $n_\infty = 1$, when d is odd.

See for instance [Sti93] for a proof.

In our situation, we have the projective nonsingular hyperelliptic curve $\mathcal{C}_S$ whose function field is $\mathbb{F}_p(x, y), y^2 = f_S(x)$. Thus the genus of $\mathcal{C}_S$ is $(|S| - 2)/2$, and the following are equal:

- $\sum_{a\in\mathbb{F}_p} \psi(f_S(a))$
- $-p$ + number of (x, y) solutions in $\mathbb{F}_p^2$ of the equation $y^2 = f_S(x)$
- $-(p+2)$ + number of $\mathbb{F}_p$-rational points on $\mathcal{C}_S$.

Thus our problem can be reduced to bounding the character sums, or equivalently bounding the deviation from p of the number of rational points on the corresponding curves.

Weil's bound [Wei48] on the number N of $\mathbb{F}_q$-rational points on a nonsingular projective curve over any finite field $\mathbb{F}_q$ says that

$$|N - (q+1)| \leq 2g\sqrt{q}$$

where g is the genus of the curve. See also [Mor94, Sti93, LN83]. The strength of this bound is in the fact that it holds for any nonsingular projective curve over any finite field $\mathbb{F}_q$. In our situation it says that

$$|\sum_{a\in\mathbb{F}_p} \psi(f_S(a))| \leq (|S|-2)\sqrt{q}+1.$$

This bound is good for small $|S|$, but it becomes trivial when $|S|$ is large, i.e., when the genus of the corresponding curve is large. We cannot conclude from Weil's bound any more than a $\sqrt{p}$ lower bound on the minimum distance of the code. The problem we are dealing with requires using the special features of the curves under consideration to obtain a better estimate. The reason why we believe that a better bound exists is that the QQR code appears to be a very nondegenerate code in an ensemble of codes where almost all the codes achieve the GV bound. The bound should be independent of $|S|$. The asymptotic goodness of the QQR code will follow if we can show that there exists $\beta > 0$ such that

$$|\sum_{a\in\mathbb{F}_p} \psi(f_S(a))| \leq (1-2\beta)p,$$

when p is large enough. The GV bound will follow from $\beta = \beta_{GV} \stackrel{\text{def}}{=} h^{-1}(1/2) \sim 0.110$. So far, the only families of curves we were able to find where β goes below β_{GV} are when -1 or 2 are quadratic residues mod p. One of the worst recorded values of β is ~ 0.031 for $f_S(x) = x^{300} - 1$ with $p = 4801$ corresponding to the case when -1 is a quadratic residue. We know that 2 should be a non-quadratic residue since otherwise the code and the curves are not related. One explanation of why -1 being quadratic residue is bad is that in this case the quadratic residue string is symmetric around $(p-1)/2$, and hence is degenerate in some sense.

What is special about the family of curves we have? First we note that up to an $O(1)$ additive term, the even assumption on the size of S can be dropped.

Theorem 4.3.4 *Let p be a prime, then*

$$\max_{\emptyset \neq S \subset \mathbb{F}_p} |\sum_a \psi(f_S(a))| \leq \max_{\emptyset \neq S \subset \mathbb{F}_p \ s.t. \ |S| \ even} |\sum_a \psi(f_S(a))| + 1$$

Proof. See Section 4.3.3. ■

The additive 1 term results form the displacement of some points at infinity.

Corollary 4.3.5 *Let p be a prime such that $p = 3 \pmod 8$, then the minimum relative distance of the block-length-$2p$ QQR code is the maximum of $\beta > 0$ such that for any nonconstant square free polynomial $f(x)$ that splits completely over $\mathbb{F}_p$, the number of $\mathbb{F}_p$-rational points on the hyperelliptic curve $y^2 = f(x)$ is smaller than $2(1-\beta)p$.*

4.3.2 The hyperelliptic curves conjectures

The special features that we are left with are that $f(x)$ splits completely over $\mathbb{F}_p$, and that the field of constants $\mathbb{F}_p$ is prime, and probably a special prime .

Conjecture 4.3.6 *There exists $\beta > 0$ such that for any prime p (or for infinitely many primes p) and for any nonconstant square free polynomial $f(x)$ that splits completely over $\mathbb{F}_p$, the number of $\mathbb{F}_p$-rational points on the hyperelliptic curve $y^2 = f(x)$ is smaller than $2(1-\beta)p$.*

Note that, it is necessary for p to be a prime, or at least not a square, since in the later case we know that the conjecture is not true. We will elaborate on this point in Section 4.3.4.

The splitting condition may be only needed to handle the high genus cases. A stronger statement might be true:

Conjecture 4.3.7 *There exists $\beta > 0$ such that for any prime p (or for infinitely many primes p) and for any nonconstant square free polynomial $f(x)$ whose degree is sublinear in p, the number of $\mathbb{F}_p$-rational points on the hyperelliptic curve $y^2 = f(x)$ is smaller than $2(1-\beta)p$?*

This basically means improving Weil's bound in the setting of hyperelliptic curves over prime fields.

The correctness of any of these conjectures implies the asymptotic goodness of the QQR code.

We were not able to establish any of those claims, or find a counter example. We leave the questions open.

4.3.3 Proof of Theorem 4.3.4

If we look at the projective version of the curves, we can argue that fractional linear transformations over $\mathbb{F}_p$ induce isomorphisms between the corresponding function fields. We can establish the claim by sending a point of the even cardinality set to infinity via a fractional linear transformation to make its cardinality odd.

Lemma 4.3.8 *Let S be a subset of $\mathbb{F}_p$, α a fractional linear transformation in $PSL_2(\mathbb{F}_p)$, $w \in \mathbb{F}_p^\times$, and*

$$T = \begin{cases} \alpha^{-1}(S)\backslash\{\infty\} & \text{if } |S| \text{ even or } \alpha(\infty) = \infty \\ \alpha^{-1}(S \cup \{\infty\})\backslash\{\infty\} & \text{if } |S| \text{ odd and } \alpha(\infty) \neq \infty \end{cases}$$

Then $\mathbb{F}_p(x, \sqrt{w f_S(x)}) \cong \mathbb{F}_p(x, \sqrt{vw f_T(x)})$, where

$$v = \begin{cases} u^{|S|} & \text{when } \alpha(\infty) = \infty \text{ ,with } \alpha(x) = ux + v \\ f_S(\alpha(\infty)) & \text{when } \infty \neq \alpha(\infty) \notin S \\ -det(\alpha) f_{S\backslash\{\alpha(\infty)\}}(\alpha(\infty)) & \text{when } \alpha(\infty) \in S. \end{cases}$$

Proof. Say that

$$\alpha(x) = \frac{\alpha_{11}x + \alpha_{12}}{\alpha_{21}x + \alpha_{22}} \quad \text{thus} \quad \alpha^{-1}(x) = \frac{\alpha_{22}x - \alpha_{12}}{-\alpha_{21}x + \alpha_{11}},$$

and recall that $f_S(x) = \prod_{t \in S}(x - t)$. Direct computations show the following. Let t be an element of $\mathbb{F}_p$, then:

- when $\alpha^{-1}(t) \neq \infty$ and $\alpha(\infty) \neq \infty$, we have

$$\alpha(x) - t = \frac{\alpha(\infty) - t}{x - \alpha^{-1}(\infty)}(x - \alpha^{-1}(t)), \tag{4.10}$$

- when $\alpha(\infty) = \infty$ (hence $\alpha^{-1}(t) \neq \infty$), we have

$$\alpha(x) - t = \frac{\alpha_{11}}{\alpha_{22}}(x - \alpha^{-1}(t)), \tag{4.11}$$

- and when $\alpha^{-1}(t) = \infty$ (hence $\alpha(\infty) \neq \infty$), we have

$$\alpha(x) - t = \frac{-det(\alpha)/\alpha_{21}^2}{x - \alpha^{-1}(\infty)}. \tag{4.12}$$

Let K and K' be the function fields and write them as the fraction fields of $\mathbb{F}_p[x, y]/(y^2 - wf_S(x))$ and $\mathbb{F}_p[x', y']/(y'^2 - vwf_T(x'))$ respectively. Form (4.10), (4.11), and (4.12) we can see that

$$wf_S(\alpha(x)) = g^2(x)vwf_T(x),$$

where $g(x)$ is a rational function. Thus the map $K \to K'$ that takes x to $\alpha(x')$ and y to $g(x')y'$ is an isomorphism since saying $(g(x')y')^2 = wf_S(\alpha(x'))$ is the same thing as saying $y'^2 = vwf_T(x')$. ■

Let $N_1(\mathcal{C}_{w,S})$ be the number of $\mathbb{F}_p$-rational points on the abstract nonsingular projective curve $\mathcal{C}_{w,S}$ whose function field is $K_{w,S} \stackrel{\text{def}}{=} \mathbb{F}_p(x, \sqrt{wf_S(x)})$ and for simplicity denote

$$\psi_\Sigma(S) \stackrel{\text{def}}{=} \sum_{a \in \mathbb{F}_p} \psi(f_S(a)).$$

We have

$$N_1(\mathcal{C}_{w,S}) = \psi(w)\psi_\Sigma(S) - p + n_\infty(\mathcal{C}_{w,S}),$$

where $n_\infty(\mathcal{C}_{w,S})$ is the number of $\mathbb{F}_p$-rational point at ∞ and is given by

$$n_\infty(\mathcal{C}_{w,S}) = \begin{cases} 2 & \text{if } \psi(w) = 1 \text{ and } S \text{ even} \\ 1 & \text{if } S \text{ odd} \\ 0 & \text{if } \psi(w) = -1 \text{ and } S \text{ even .} \end{cases}$$

Now, with S, T and v as in Lemma 4.3.8 in the setting when $w = 1$, we have $K_{1,S} \cong K_{v,T}$, and thus $N_1(\mathcal{C}_{1,S}) = N_1(\mathcal{C}_{v,T})$. Therefore,

$$\psi_\Sigma(S) + n_\infty(\mathcal{C}_{w,S}) = \psi(v)\psi_\Sigma(T) + n_\infty(\mathcal{C}_{v,T}). \tag{4.13}$$

If T is an odd cardinality subset of $\mathbb{F}_p$ such that $\psi_\Sigma(T) \neq 0$, let a be such that $f_T(a) \neq 0$, and let α be a fractional linear transformation such that $\alpha(a) = \infty$. Then use the even cardinality set $S = \alpha(T) \cup \{\alpha(\infty)\}$. Therefore

$$\psi_\Sigma(S) + 2 = \psi(v)\psi_\Sigma(T) + 1$$

since $n_\infty(\mathcal{C}_{1,S}) = 2$ and $n_\infty(\mathcal{C}_{v,T}) = 1$. It follows that $|\psi_\Sigma(S)| \geq |\psi_\Sigma(T)| - 1$.

4.3.4 A note on the prime field setting

Conjecture 4.3.6 does not hold when the size of the base field is a square. It is known that Weil's bound is tight in that setting. Classical examples are over $\mathbb{F}_{p^2}$ where the hyperelliptic curve $y^2 = f_{\mathbb{F}_p}(x) = x^p - x$ has $p^2 - p$ $\mathbb{F}_{p^2}$-rational point when -1 is a nonquadratic residue mod p.

We can argue that similar cases cannot happen over a prime field. By similar cases, we mean curves of the form $y^2 = f(x)$ where the square free $f(x)$ has two nonzero coefficients and splits completely over $\mathbb{F}_p$.

Theorem 4.3.9 *Let p be a prime such that -1 is a non-quadratic residue (i.e., $p = -1 \pmod 4$). If $f(x)$ is a nonconstant square-free polynomial over $\mathbb{F}_p$ that splits completely over $\mathbb{F}_p$ and has only two nonzero coefficient, then $|\sum_a \psi(f(a))| \leq (p+1)/2$.*

Proof. Without loss of generality, say that $f(x)$ is a monic. Thus $f(x) = f_S(x)$ for some subset S of $\mathbb{F}_p$. Since $f_S(x)$ has weight 2, it should of the form $f_S(x) = x^k - x_0^k$ or $f_S(x) = x(x^k - x_0^k)$ where $k|(p-1)$ and x_0 nonzero. Thus S is a coset of some multiplicative subgroup of order k in addition to possibly zero. Without loss of generality we can assume that $x_0 = 1$ since we can sum over x/x_0 after taking x_0 out. This will only affect the character sum by a $\psi(x_0^{|S|})$ factor. The idea of the proof is to note that $\psi(f_S(a))$ and $\psi(f_S(a^{-1}))$ are opposite in sign for many a's. When a is nonzero, in the first case we have

$$\psi(a^{-k} - 1) = \psi(-1)\psi(a^{-k})\psi(a^k - 1),$$

and in the second case we have

$$\psi(a^{-1}(a^{-k} - 1)) = \psi(-1)\psi(a^{-k-2})\psi(a(a^k - 1)).$$

So if we assume that -1 is a non-quadratic residue, we get $\psi(f_S(a^{-1})) = -\psi(f_S(a))$ for each $a \in Q$, where Q is the set of quadratic residues mod p. In other words, the sum vanishes on Q. It follows that $|\sum_a \psi(f_S(a))| \leq (p-1)/2 + 1$. ■

Note that the proof uses a special case of (4.13), but we could not go very far with similar arguments alone.

4.3.5 Relation to cyclic quadratic residue codes over $\mathbb{F}_4$

The explicit binary codes we are talking about, i.e., the QQR code, can be related, after minor modifications, to a special class of non-binary classical cyclic quadratic residue codes over $\mathbb{F}_4$ when the prime is special enough. The distinguishing features of this special case comes from a code ensemble where the random is good.

When 2 is a nonquadratic residue, we can relate a rate 1/2 subcode EQQR of the QQR code to a special class of nonbinary cyclic quadratic residue codes over $\mathbb{F}_4$. The relation is a bijection that preserves weight in a suitable sense. The identification argument is similar to the argument in [HKSS94] that relates nonlinear binary codes

to codes over $\mathbb{Z}/4\mathbb{Z}$. Let R^* be the ideal of even weight vectors in $R = \mathbb{F}_2[\mathbb{Z}/p\mathbb{Z}]$. Consider the R-submodule $C_Q^* = R^* C_Q$ of the QQR code C_Q, and call it the EQQR (Even QQR) code. Let I_4 be the cyclic quadratic residue code over $\mathbb{F}_4$ generated in $\mathbb{F}_4[\mathbb{Z}/p\mathbb{Z}] = \mathbb{F}_4[x]/(x^p - 1)$ by the the polynomial $(x-1)\prod_{i \in Q}(x - \beta^i)$, where β is a primitive p'th root of 1 in an extension of $\mathbb{F}_2$. We can argue that when 2 is a non-quadratic residue, there is a bijection between the EQQR code and I_4 that preserves the minimum distance if we measure the weight of a vector in $\mathbb{F}_4^p$ by counting each occurrence of 1 twice. The choice of 1 is arbitrary; any of the three nonzero elements of $\mathbb{F}_4$ will do. The bijection is essentially an inverse concatenation given by

$$C_Q^* \to I_4 : (r(x), r_Q(x)r(x)) \mapsto \gamma r(x) + r_Q(x) f(x),$$

where $\gamma = r_Q(\beta)$ generates $\mathbb{F}_4^\times$. Note that here we are viewing $\mathbb{F}_2[\mathbb{Z}/p\mathbb{Z}]$ as $\mathbb{F}_2[x]/(x^p - 1)$, so here $r_A = \sum_{g \in A} g$ means $r_A(x) = \sum_{i \in A} x^i$. The proof is not hard. The reason why we need 2 to be a nonquadratic residue is essentially that otherwise we get $r_Q^2 = r_Q$ and consequently $\gamma = 1$. It is worth mentioning that when 2 is a nonquadratic residue, binary quadratic residue codes of block length p do not exist.

4.4 The dihedral group randomized construction

In this section, we establish the claim of Section 4.1.3.1. We will argue in Theorem 4.4.3 that for infinitely many block lengths, a random ideal in the binary group algebra $\mathbb{F}_2[D_m]$ of the dihedral group D_m is an asymptotically good rate 1/2 binary code. The condition we require on m is satisfied by almost half the primes, namely all primes p such that 2 is a nonquadratic residues mod p (i.e., $p = \pm 5 \pmod 8$), and such that the size of the multiplicative group generated by 2 in $\mathbb{F}_p^\times$ grows asymptotically faster than $\log p$. By random here we mean according to some specific distribution based on the $\mathbb{F}_2$-representations of D_m in Theorem 4.4.2. The implicit bound on the relative minimum distance is $h^{-1}(1/4)$, where h is the binary entropy function.

Let m be odd, and consider the dihedral group

$$D_m = \langle \alpha, \beta | \alpha^2 = 1, \beta^m = 1, \alpha\beta = \beta^{-1}\alpha \rangle,$$

i.e., D_m is generated by α and β subject to the above relations. So D_m has $2m$ elements: the $\alpha^i\beta^j$'s.

We are interested in the structure of $\mathbb{F}_2[D_m]$ in terms of its ideals. Note that since the characteristic 2 of $\mathbb{F}_2$ divides the even order of D_m, the ring $\mathbb{F}_2[D_m]$ is not semisimple, i.e., its radical is nonzero.

Let N be the subgroup generated by β, and H that generated by α. Note that N is normal. Let

$$Q = \mathbb{F}_2[N].$$

Any element r of $\mathbb{F}_2[D_m]$ can be represented uniquely as $r = q + \alpha q'$, where $q, q' \in Q$. From the relation $\alpha\beta^i = \beta^{-i}\alpha$ we see that $\alpha q = \tilde{q}\alpha$ when $q \in Q$, where for $q = \sum_{g\in S} g$ in Q, $\tilde{q} \stackrel{\text{def}}{=} \sum_{g\in S^{-1}} g$. Note that $\tilde{}$ as a map from Q to Q is a ring homomorphism.

Since Q is commutative and semisimple (because m is odd), let

$$Q = \oplus_{i=0}^{w} Q_i$$

be the unique decomposition of Q into irreducible ideals, and let e_i be the idempotent of Q_i. Since the Q_i's are finite, each Q_i is a field with e_i as its unit element. Assume that the Q_i's are ordered so that $Q_0 = (\sum_{g\in N} g)$, $\tilde{Q}_i = Q_i$ for $i = 1, \ldots, t$, and $\tilde{Q}_i = Q_{i+v}$, $v = s - t$, for $i = t+1, \ldots, s$.

Terminology 4.4.1 *By an ideal, unless otherwise specified, we mean a left ideal.*

If F is a field, by $F^\times$ we mean the multiplicative group of F. More generally, if A is a commutative ring with identity, $A^\times$ will denote the multiplicative group of the units of A.

In addition to a direct sum of ideals in a group algebra, $\oplus$ will be used also in the following context. If $T \subset \{1, \ldots, w\}$, by $\oplus_{i\in T} Q_i^\times$ we mean the multiplicative group defined as the direct product of the multiplicative groups $\{Q_i^\times\}_{i\in T}$ and realized in Q

as $\{\sum_{i\in T} q_i | q_i \in Q_i^\times, \forall i \in T\}$.

Finally, $M_k(F)$ *means the set of* $k \times k$ *matrices over the field* F.

Theorem 4.4.2 *Let* $R = \mathbb{F}_2[D_m]$, *where* D_m *is the dihedral group, and* m *is odd. The unique decomposition of* R *into 2-sided ideals is*

$$R = \oplus_{i=0}^{s} R_i,$$

where the structure of the R_i*'s is as follows.*

1) $dim R_0 = 2$. *The ideals of* R_0 *are*

$$(0) \subset J_0 = (\sum_{g\in D_m} g) \subset R_0.$$

The ideal J_0 *is the radical of* R.

2) For $i = 1, \ldots, t$, *we have*

$$R_i = Q_i \oplus \alpha Q_i.$$

Each such R_i *is simple as a ring and isomorphic as a ring to* $M_2(\mathbb{F}_{2^{l_i/2}})$, $l_i = dim\ Q_i$. *Moreover,* R_i *contains* $2^{l_i/2} + 1$ *nontrivial irreducible ideals all isomorphic and each of dimension* l_i. *They are given by*

$$I^i_{[b]} = \{q(1+\alpha)b | q \in Q_i\},\ for\ [b] \in Q_i^\times / Z_i^\times,$$

where $Z_i = \{q \in Q_i | q = \tilde{q}\}$, *a subfield of* Q_i.

3) For $i = t+1, \ldots, s$, *we have*

$$R_i = Q_i \oplus \alpha Q_i \oplus \tilde{Q}_i \oplus \alpha \tilde{Q}_i.$$

Each such R_i *is simple as a ring and isomorphic as a ring to* $M_2(\mathbb{F}_{2^{l_i}})$, $l_i = dim\ Q_i$. *Moreover,* R_i *contains* $2^{l_i} + 1$ *nontrivial irreducible ideals all isomorphic and each of*

dimension $2l_i$. *They are given by*

$$I^i_{[b]} = \{q(1+\alpha)b | q \in Q_i \oplus \tilde{Q}_i\}, \text{ for } [b] \in Q_i^\times \oplus \tilde{Q}_i^\times / T_i,$$

where $T_i = \{q \in Q_i^\times \oplus \tilde{Q}_i^\times | q = \tilde{q}\}$, *a subgroup of the multiplicative group* $Q_i^\times \oplus \tilde{Q}_i^\times$.

Proof (when $t = s$)**.** The representations are essentially similar to the semisimple case corresponding to the situation when instead of $\mathbb{F}_2$ there is a field F whose characteristic does not divide the order 2 of D_m (see for instance [Bur65, CR62]). We need however to worry about the fact that the ring is not semisimple and furthermore list all the irreducible ideals. This is not hard since the group is easy to deal with. We outline the main steps in the simpler case when $\tilde{Q}_i = Q_i$ for all i, i.e., we do not have to worry about (3). The situation in (3) follows by a similar argument that we skip without affecting the completeness of this chapter since we are going to exclude this situation later on.

Note first that $J_0^2 = (0)$ and that R_0 consists of $\sum_{g\in N} g$ and $\alpha \sum_{g\in N} g$ in addition to $\sum_{g\in D_m} g$. Observe next that $\alpha q(1+\alpha) = \tilde{q}(1+\alpha)$, for $q \in Q$, so $\alpha Q(1+\alpha)b = \tilde{Q}(1+\alpha)b$, for $b \in Q$, and hence the distinction between cases (2) and (3). Moreover, the decomposition is clearly in terms of 2-sided ideals since for each i, $qR_i = R_iq = R_i$ for each $q \in R$, and $\alpha R_i = R_i\alpha = R_i$. The claimed structure will essentially follow once we show that: $\forall i \neq 0$,

i) R_i contains no other 2-sided ideal, and is thus simple as a ring,

ii) each $I^i_{[b]}$ is irreducible, $I^i_{[b_1]} \neq I^i_{[b_2]}$ if and only if $[b_1] \neq [b_2]$, and any nonzero left ideal in R_i must contain one of the $I^i_{[b]}$.

Assume in what follows that $i \neq 0$. From (i) we get also that J_0 is the radical of R since none of the R_i's can be nilpotent because Q_i is a field inside R_i. Moreover, the fact that R_i is simple implies that all the nonzero irreducible ideals of R_i are isomorphic and that R_i is isomorphic to $M_d(K)$ for some finite field K, where d is such that $R_i = \oplus_{i=1}^d R_{i,j}$, with the $R_{i,j}$'s irreducible and the decomposition not unique unless $d = 1$. Combining with (ii), we see, from dimensional considerations,

that $d = 2$ and hence $|K| = 2^{l_i/2}$. The number of the nonzero irreducible ideals then follows from the fact that in general the number of nonzero irreducible ideals in $M_2(K)$ in $|K| + 1$.

Proof of (ii): Any $I^i_{[b]}$ is irreducible since Q_i is irreducible in Q. Moreover, if $b_1, b_2 \in Q_i^\times$ and $q_1, q_2 \in Q_i$, then $(q_1 + q_2\alpha)(1 + \alpha)b_2 = (1 + \alpha)b_2$ if and only if $q_1 + q_2 = b_2 b_1^{-1} = \widetilde{b_2 b_1^{-1}}$ where inversion is in Q_i as a field, thus $I^i_{[b_1]} = I^i_{[b_2]}$ if and only if $[b_1] = [b_2]$. Finally, if I is a nonzero left ideal in Q_i, then I must contain one of the $I^i_{[b]}$'s because I must contain a nonzero $r' = (1 + \alpha)b$ for some $b \in Q_i$. This is the case since if $r' = b_1 + \alpha b_2$ is any nonzero element in I, for some $b_1, b_2 \in Q_i$, we can use $r = r'$ if $b_1 = b_2$, or $r = r' + \alpha r' = (1 + \alpha)(b_1 + b_2)$ if $b_1 \neq b_2$.

Proof of (i): Let $r = r_1 + \alpha r_2$ be a nonzero element of R_i for some $r_1, r_2 \in Q_i$, and consider the 2-sided ideal I generated by r. First we note that I must contain an element $q = q_1 + \alpha q_2$, with $q_1, q_2 \in Q_i$, $q_1 \neq 0$, and $q_1 \neq q_2$. (If $r_1 = 0$, use $q = \alpha r$. If $r = r_1 + \alpha r_1$, try $q = gr$ for $g \in N$. Thus $gr = gr_1 + \alpha g^{-1} r_1$. It cannot be the case that $gr_1 = g^{-1} r_1$, i.e., $g^2 r_1 = r_1$, for all g in N. The reason is that this together with the fact that the square map in N is onto (because m is odd), mean that $r_1 = \sum_{g \in N} g$, i.e., $r_1 \in R_0$, which is not true). Thus $q + \alpha q q_1^{-1} q_2 = q_1^{-1}(q_1^2 + q_2^2)$ is a nonzero element in Q_i, where inversion is in Q_i as a field. Note that $q_1^{-1}(q_1^2 + q_2^2) \neq 0$ since $q_1 \neq q_2$ and the characteristic of Q_i is 2. Thus I contains Q_i, and hence R_i since the 2-sided ideal generated by Q_i is R_i.

This completes the proof of Theorem 4.4.2. ■

So we know all the left ideals in R: they are the direct sums of the left ideals in the R_i's.

Theorem 4.4.3 *Let m be an odd integer, and consider the dihedral group D_m. Let $R = \mathbb{F}_2[D_m]$, and consider the unique decomposition*

$$R = \oplus_{i=0}^{s} R_i,$$

of R into 2-sided ideals in Theorem 4.4.2.

Assume that representations of type (3) in Theorem 4.4.2 do not occur, i.e., $\tilde{Q}_i =$

Q_i, $i = 1, \ldots, t = s = w$.

Consider the following randomized code construction: generate a rate-$(\frac{1}{2} - \frac{1}{2m})$ *random ideal* I *of* $\mathbb{F}_2[D_m]$ *as*

$$I = \oplus_{i=1}^{s} I^i,$$

where each I^i *is selected uniformly at random from one of the* $2^{l_i/2} + 1$ *nonzero irreducible left ideals of* R_i.

If $\delta > 0$ *is such that* $h(\delta) \leq \frac{1}{4} - \frac{\log m}{2l(m)}$, *then the probability that the minimum relative distance of* I *is below* δ *is at most* $2^{-2l(m)(1/4-h(\delta))+5\log m}$, *where* h *is the binary entropy function.*

Moreover, there are infinitely many such m*'s such that* $l(m)$ *grows asymptotically faster than* $\log m$, *and representations of type (3) in Theorem 4.4.2 do not occur, for instance almost all the primes* $p = \pm 5 \pmod 8$.

Therefore, there are infinitely many integers m *such that the left ideal* I *of* $\mathbb{F}_2[D_m]$ *is an asymptotically good rate* $1/2$ *binary code with a high probability.*

We concentrate on the case when representations of type (3) in Theorem 4.4.2 do not occur for simplicity.

Proof. First we note that this construction is equivalent to the following: pick a random ideal

$$I_{[b]} = \{q(1+\alpha)b | q \in Q\},$$

where $[b]$ is selected uniformly at random from $Q^{*\times}/T$, $Q^* = \oplus_{i=1}^{w} Q_i$, and T is the subgroup of $Q^{*\times}$ given by $T = \{q \in Q^{*\times} | q = \tilde{q}\}$. Note that $Q^{*\times}$ is, in the sense of Terminology 4.4.1, the multiplicative group of units of of Q^*, thus $Q^{*\times} = \oplus_{i=1}^{w} Q_i^{\times}$.

From Section 4.2, we know that there are infinitely many integers m with $l(m) \gg \log m$, including almost all the primes. To show that representations of type (3) can be avoided when m is a prime $p = \pm 5 \pmod 8$, i.e., when 2 is a nonquadratic residue, assume for the moment that m is a prime p. Following the classical direction, realize Q as $Q = \mathbb{F}_2[x]/(x^p - 1)$, and let β be a primitive p'th root of 1 in an extension of $\mathbb{F}_2$,

thus the irreducible decomposition of $x^p - 1$ over $\mathbb{F}_2$ is

$$x^p - 1 = (x-1) \prod_{A \in \mathbb{F}_p^\times / \langle 2 \rangle} g_A(x), \text{ where } g_A(x) = \prod_{i \in A} (x - \beta^i).$$

In these terms rewrite $Q^* = \oplus_{i=1}^w Q_i$ as $Q^* = \oplus_A Q_A$, where Q_A is the ideal in Q generated by $g'_A(x) = (x-1) \prod_{B \neq A} g_B(x)$. Thus $\tilde{Q}_A$ is generated by $\tilde{g}'_A(x) = g'_{-A}(x)$. Hence $\tilde{Q}_A = Q_A$ if and only if $A = -A$. This holds for all $A \in \mathbb{F}_p / \langle 2 \rangle$ if and only if $-1 \in \langle 2 \rangle$, which can be guaranteed when 2 is a nonquadratic-residue since in such a case $2^{(p-1)/2} = -1 \pmod{p}$.

Now to establish the distance bound when in general representations of type (3) do not occur, we follow the argument in Theorem 4.2.1 while using the structure of the dihedral group representations from Theorem 4.4.2. Observe the relation between this randomized construction and the half randomized construction in Theorem 4.2.1; this ensemble of codes is, in a suitable sense, a subfamily of that ensemble.

For any b in $Q^{*\times}$,

$$I_{[b]} = Q(q + \alpha)b = aQ^*(1 + \alpha)b,$$

for all a in $Q^{*\times}$. Thus the probability P that the minimum distance of $I_{[b]}$ is below $2\delta m$, when $[b]$ is selected uniformly at random from $Q^{*\times}/T$, is the same as the probability that $aQ^*(1+\alpha)b$ has a minimum distance below $2\delta m$, when a and b are selected uniformly at random from $Q^{*\times}$.

Now we proceed as in Theorem 4.2.1. P is the probability that there is an $f \in Q^*$, $f \neq 0$, such that $0 \leq w(af(1+\alpha)b) < 2m\delta$. Thus P is at most

$$\sum_{f \in Q^*; f \neq 0} Pr_{a,b \in Q^{*\times}}[0 \leq w(af(1+\alpha)b) < 2m\delta],$$

and this is at most

$$\sum_{l=l(m)}^{m} |D_l^*| \max_{f \in D_l^*} Pr_{a,b \in Q^{*\times}}[0 \leq w(af(1+\alpha)b) < 2m\delta],$$

where $D_l^* = D_l \cap Q^*$, and $D_l = \{f \in Q |$ dim $fQ = l\}$. As before, we have

$$|D_l^*| \leq |D_l| \leq 2^l |\Omega_l|,$$

where Ω_l is the set of ideals in R of dimension l.

Consider any l, and any $f \in D_l^*$. We have

$$Pr_{a,b \in Q^{*\times}}[0 \leq w(af(1+\alpha)b) < 2m\delta] = \sum_{r \in U \text{ s.t. } 0 \leq w(r) < 2m\delta} Pr_{a,b \in Q^{*\times}}[af(1+\alpha)b = r],$$

where $U = Q^{*\times} f(1+\alpha) Q^{*\times}$, and this is at most

$$\max_{r \in U} Pr_{a,b \in Q^{*\times}}[fa(1+\alpha)b = r] \sum_{w_1, w_2 \geq 0; w_1 + w_2 < 2\delta m} |I^{(w_1)}||I^{(w_2)}|,$$

where $I = fQ$, and $I^{(v)}$ is the set of elements in I of weight v. Fix l, and any f in D_l^*, and any r in U. We will argue at the end that

$$Pr_{a,b \in Q^{*\times}}[fa(1+\alpha)b = r] \leq 2^{-3l/2}. \tag{4.14}$$

We have from Lemmas 4.2.2 and Lemma 4.2.3 that $|\Omega_l| \leq m^{l/l(m)+1}$, $|I^{(w_1)}| \leq 2^{lh(w_1/m)}$, and $|I^{(w_2)}| \leq 2^{lh(w_2/m)}$. Thus, modulo (4.14), we are done since by arguing as in Theorem 4.2.1, we get

$$P \leq \sum_{l=l(m)}^{m} 2^{-2l\left(\frac{1}{4} - h(\delta) - \frac{\log m}{2l(m)}\right) + 3\log m} \leq 2^{-2l(m)(\frac{1}{4} - h(\delta)) + 5\log m},$$

where the last bound holds when $\frac{1}{4} - h(\delta) - \frac{\log m}{2l(m)} \geq 0$. The difference is that now we have $1/4$ instead of $1/2$. The reason is that before we had 2^{-2l} instead of $2^{-3l/2}$.

We still have to establish (4.14). The first thing to note is that when a and b are selected uniformly at random from $Q^{*\times}$, each $r \in U$ is equally likely to occur. The reason is that if $r = a'f(1+\alpha)b'$, $a', b' \in Q^{*\times}$, then the event $af(1+\alpha)b = r$ can be expressed as

$$a''af(1+\alpha)b''b = e^* f(1+\alpha)e^* = f(1+\alpha),$$

where a'' (respectively b'') is the inverse of a' (respectively b') in the multiplicative group $Q^{*\times}$, and where e^* is the unit element of $Q^{*\times}$ and the idempotent for Q^* (recall that $f \in Q^*$). Hence, since the uniform selection of a and b is equivalent to the uniform selection of $a''a$ and $b''b$, we get that the probability that r occurs is equal to the probability that $f(1+\alpha)$ occurs, i.e., all the elements of U are equally likely to occur. In other words, for each $r \in U$,

$$Pr_{a,b\in Q^{*\times}}[fa(1+\alpha)b = r] = \frac{1}{|U|}. \tag{4.15}$$

Now, decompose f uniquely as $f = \sum_{i=1}^{s} f_i$, where each $f_i \in Q_i$, and let S be the set of i's such that $f_i \neq 0$, thus $l = \sum_{i\in S} l_i$. We can express U as

$$U = \{\sum_{i\in S} u_i | u_i \in Q_i^{\times}(1+\alpha)Q_i^{\times}, \forall i \in S\}.$$

Note that this is the first time we used the assumption $\tilde{Q}_i = Q_i$. Using this one more time, in terms of the expression of type (2) ideals in Theorem 4.4.2, we have

$$Q_i^{\times}(1+\alpha)Q_i^{\times} = \bigcup_{[b]\in Q_i^{\times}/Z_i^{\times}} I^i_{[b]}\backslash\{0\},$$

where the union is a disjoint union. Thus

$$|U| = \prod_{\in S} \sum_{[b]\in Q_i^{\times}/Z_i} |I^i_{[b]}\backslash\{0\}| = \prod_{i\in S}(2^{l_i/2}+1)(2^{l_i}-1) \geq 2^{3/2\sum_{i\in S} l_i} = 2^{3l/2},$$

and hence (4.14) via (4.15).

This completes the proof of Theorem 4.4.3. ■

It is important to note that the $h^{-1}(1/4)$ bound we obtained on the minimum relative distance is unlikely to be tight. We ended up with this bound because our argument is based on counting, and the construction does not have enough randomness in such a way that a counting argument can go up to the GV bound, i.e., up to $h^{-1}(1/2)$.

4.5 Open questions

We conclude with the resulting open questions:

- **Conjectures 4.3.6 and 4.3.7**.
- Decoding the codes in the PQC ensemble, or specifically the QQR code in Section 4.3.
- Improving the bound on the minimum distance in the dihedral group construction.
- Extending the statement of Theorem 4.4.3 by allowing type (3) representations.
- Studying randomized constructions of codes that are ideals in the group algebras of other nonabelian groups.

Chapter 5

On the pseudorandomness based on minimum distance

We study in this chapter the derandomization capabilities of probability measures on the hamming cube having the k-wise independence property, the δ-bias property, or the δ-almost k-wise independence property. Classical explicit constructions of such probability measures are based on linear codes with good distance properties. In general understanding the power and limitations of such pseudorandomness properties is of fundamental importance due to their basic nature.

Mostly the questions we consider are about statements that hold for any probability measure having one of those properties. The exceptions are when we focus on the k-wise independent probability measures that are based on linear codes.

We note first that linear-programming duality can be used to get a purely analytical characterization of the class of boolean function that can be fooled by the δ-almost k-wise independence property. The characterization is necessary and sufficient and is in terms of tight average sandwichability between real valued functions with low degree and small L_1-norm in the Fourier domain.

Then we characterize the location of classical linear-codes-based constructions of k-wise independent probability measures in the convex polytope of all such measures, and its subpolytope consisting of those measures whose Fourier transform is nonnegative.

On the negative side, we prove that the exponentially-small-bias property is not sufficient to fool small log-depth circuits nor the weakest branching programs.

From a concrete viewpoint, we prove first that any sufficiently log-wise independent probability measure looks random to all polynomially small read-once DNF formulas. The setting is naturally extendable to almost k-wise independent probability measures. We give an application related to the distribution of quadratic-residues.

Then we establish a very sharp upper bound on the probability that a random binary string generated according to a k-wise independent probability measure has any given weight. The setting is naturally extendable to almost k-wise independent probability measures. We give applications related to the distribution of quadratic-residues and the weight distribution of linear codes.

We consider also the problem of derandomizability of AC_0 by arbitrary k-wise independent probability measures, when k is made polylogaritmically large enough. We reduce this problem to a conjecture about the the symmetry of the optimum of some symmetric optimization problem with linear constraints and a nonlinear objective function.

Finally, we study of the problem of approximability of high-degree parity functions on high-dual-distance binary linear codes by low-degree polynomials with coefficients in fields of odd characteristics. This problem has applications to the ability of binary linear codes with sufficiently large dual distance to derandomize AC_0, or low-degree polynomial equations on binary input variables with coefficients in small finite fields of odd order. Among other results, we relax this problem into essentially a single low-dimensional low-complexity linear program in terms of Krawtchouk polynomials. The problem of bounding the optimum of the linear program remains open.

5.1 Introduction

A PseudoRandom Generator (PRG) is an efficient algorithm that takes as input a short string called a seed of purely random bits and outputs a much longer string of bits with some desirable properties. The notion of a PRG was introduced by Blum

and Micali [BM82] and Yao [Yao82] in a setting where the ultimate objective is to construct PRG's in such a way that the seed length is logarithmic and there is no polynomial size boolean circuit that can distinguish between the PRG output and a purely random string of the same length. The indistinguishablity is in the sense that the probability that the circuit outputs one does not change significantly between the two environments. The existence of nonuniform PRG's follows easily by counting. The difficulty of the problem is in the construction part. Constructing such PRG's implies the correctness of the conjecture $P = BPP$.

Known answers to this problem rely on unproven hardness assumptions. Impagliazzo and Wigderson [IW97] proved that $P = BPP$ under the worst-case hardness assumption that there exists a language in EXPTIME that is not decidable by (any nonuniform family of) subexponential-size circuits. The hardness versus randomness approach was started by Nisan and Wigderson [NW88] in the setting of average-case hardness assumptions. This approach is a generalization of the unconditional quasipolynomial complexity PRG of Nisan [Nis91] for AC_0, which is based on Hastad's lower bound [Has86] on the hardness of approximation of parity by AC_0 circuits. Under the umbrella of worst-case hardness assumptions, Sudan, Trevisan, and Vadhan [STV99] gave different proofs of the Impagliazzo-Wigderson Theorem. One of the proofs is based purely on using locally-list-decodable error correcting codes to turn average-case hardness into worst-case hardness. A solution of this problem seems currently far from being reachable unconditionally.

A more modest goal is to construct PRG's for classes of computations that are less general than arbitrary polynomial size boolean circuits, for instance, the class AC_0 of polynomial size constant-depth unbounded-fanin AND-OR circuits [AW85, Nis91], and the class RL of logarithmic-space randomized algorithms [Nis90]. Nisan [Nis91, Nis90] constructed PRG's for these classes with quasipolynomial complexity.

It is essentially not known as to how to construct polynomial complexity PRG's for any relatively-general computational model. This is the case if we exclude models such as polynomial size decision trees, and DNF formulas where the number of inputs per clause is bounded by a constant.

In this chapter we consider an even more basic question which is about studying the derandomization capabilities of probability measures on the Hamming cube having the k-wise independence property, the δ-almost k-wise independence property, or the δ-bias property as introduced by Vazirani [Vaz86] and Naor and Naor [NN93].

These pseudorandomness notions are special purpose generators that were originally introduced to derandomize some randomized algorithms whose analyses can be made to work when only limited independence is assumed. A probability measure μ on $\{0,1\}^n$ is said to have the *δ-almost k-wise independence* property if μ can $\delta/2$-fool all parity functions on k or fewer of the n bits. The δ-almost n-wise independence is called the *δ-bias* property. The 0-almost k-wise independence property is called the *k-wise independence* property. Saying that μ has the k-wise independence property is equivalent to saying that any k or fewer of the n binary random variables are statistically independent, and each of those random variables is equally likely to be 0 or 1.

Classical constructions of such probability measures are based on linear codes with good distance properties [Vaz86, NN93, AGHP92]. For instance, if C is a block-length-n binary linear code whose dual has minimum distance above k, then the uniform distribution on the codewords of C is k-wise independent as a probability measure on $\{0,1\}^n$. See Section 5.1.1.5 for the other constructions.

The δ-almost k-wise independence property is the weakest of these properties, and it is necessarily satisfied by any pseudorandom generator for suitable values of k and δ. The k-wise independence property is stronger, but when k is relatively small, the two notions are loosely speaking equivalent in the sense that statements about foolability by the k-wise independence property can be translated to statements about foolability by the δ-almost k-wise independence property. The δ-bias property is stronger than the δ-almost k-wise independence property, and it is necessarily satisfied by any PRG for NC_1 or RL for suitable values δ.

Thus, in general, understanding the power and the limitations of such pseudorandomness properties is of fundamental importance due to their basic nature.

Mostly the questions we consider are about statements that hold for any proba-

bility measure having one of those properties. The exceptions are when we focus on linear-codes-based k-wise independent probability measures.

5.1.1 Preliminaries

5.1.1.1 Basic terminologies

The group $(\mathbb{Z}/2\mathbb{Z})^n$ is denoted by $\mathbb{Z}_2^n$. The complex *characters* of the abelian group $\mathbb{Z}_2^n$ are

$$\mathcal{X}_z(x) \stackrel{\text{def}}{=} \{(-1)^{xz}\}_{z\in\mathbb{Z}_2^n},$$

where $xz \stackrel{\text{def}}{=} \sum_i x_i z_i$. The dual group of $\mathbb{Z}_2^n$ (i.e., its group of characters) is identified with $\mathbb{Z}_2^n$ by identifying $\mathcal{X}_z$ with z. If $z \in \mathbb{Z}_2^n$, by $w(z)$ we mean the *weight* of z, i.e., the number of nonzero coordinates.

If $g : \mathbb{Z}_2^n \to \mathbb{C}$, we let $\hat{g} : \mathbb{Z}_2^n \to \mathbb{C}$ denote the *Fourier transform* of g with respect to the complex characters of the abelian group $\mathbb{Z}_2^n$, i.e.,

$$g(x) = \sum_z \hat{g}(z)\mathcal{X}_z(x),$$

or equivalently

$$\hat{g}(z) = \frac{1}{2^n}\sum_x g(x)\mathcal{X}_z(x).$$

The *degree* of g is defined to be the weight of the largest z such that $\hat{g}(z) \neq 0$.

In what follows, by a *circuit* we implicitly mean a boolean circuit, i.e., a single-sink multi-source directed acyclic graph, where the non-source nodes are associated with logic-gates (e.g. AND, OR, ...), and the edges are labeled with NOT gates, in such a way that the circuit computes some boolean function $\{0,1\}^n \to \{0,1\}$, n being the number of sources. The number of incoming edges to a node (i.e., gate) is called the *fanin* of the node or the gate. The *size* of the circuit is the total number of nodes in the graph, and its *depth* is the graph depth. The class AC_0 is the class of AND-OR constant-depth unbounded-fanin polynomial-size boolean circuits. A DNF (Disjunctive Normal Form) formula is an unbounded-fanin depth-2 circuit realized as an OR of AND gates. The AND gates are called *clauses*. A *decision tree* is a special

type of DNF formulas where no two distinct clauses can be simultaneously satisfied. It is called a decision tree because it can be realized as a binary decision tree where the clauses correspond to the leafs.

If q is power of a prime, the finite field of size q will be denoted by $\mathbb{F}_q$.

When q is odd, the *quadratic character* of $\mathbb{F}_q^\times$ will be denoted by ψ, i.e.,

$$\psi(a) = \begin{cases} 1 & \text{if } a \in Q \\ -1 & \text{if } a \in \mathbb{F}_q^\times \backslash Q \\ 0 & \text{if } a = 0, \end{cases}$$

where $Q = \{a^2 : a \in \mathbb{F}_q^\times\}$ is the set of *quadratic residues* in $\mathbb{F}_q$, i.e., the non-zero squares.

By a *code* we mean a *binary linear code*, i.e., an $\mathbb{F}_2$-linear subspace C of $\mathbb{F}_2^n$ whose elements are called *codewords*. The *minimum distance* of C is the minimum weight of a nonzero codeword. By the *dual* of C we mean the block-length-n linear code denoted by $C^\perp$ and defined as

$$C^\perp \stackrel{\text{def}}{=} \left\{ y \in \mathbb{Z}_2^n : \sum_i x_i y_i = 0 \pmod 2 \ , \forall y \in C \right\}.$$

Finally, if n_1, n_2 are integers and n is a positive integer, we will use the terminologies $[n_1 : n_2] \stackrel{\text{def}}{=} \{n_1, \ldots, n_2\}$ and $[n] \stackrel{\text{def}}{=} [1 : n]$.

5.1.1.2 Indistinguishablity

Definition 5.1.1 [BM82, Yao82] *If μ is a probability measure on $\{0,1\}^n$, and $f : \{0,1\}^n \to \{0,1\}$, we say that μ can ϵ-*fool *f if*

$$|Pr_{x\sim\mu}[f(x) = 1] - P_{x\sim\mu_0}[f(x) = 1]| \leq \epsilon,$$

where μ_0 is the the uniform probability measure on $\{0,1\}^n$.

5.1.1.3 Limited independence and small bias

Definition 5.1.2 [Vaz86, NN93] *Let μ be a probability measure on $\{0,1\}^n$. We say that μ has*

- The δ-almost k-wise independence property*: if μ can $\delta/2$-fool all parity functions on k or fewer of the n bits, or equivalently if $|E_\mu \mathcal{X}_z| \leq \delta$ for each nonzero z in $\mathbb{Z}_2^n$ whose weight is less than or equal to k.*

- The δ-bias property*: if μ has the δ-almost n-wise independence property, i.e., if $|E_\mu \mathcal{X}_z| \leq \delta$ for each nonzero z in $\mathbb{Z}_2^n$.*

- The k-wise independence property*: if any k or fewer of the n binary random variables are statistically independent and each of the random variables is equally likely to be 0 or 1, or equivalently if μ has the 0-almost k-wise independence property, i.e., if $E_\mu \mathcal{X}_z = 0$ for each nonzero z in $\mathbb{Z}_2^n$ whose weight is less than or equal to k.*

5.1.1.4 Relations

Theorem 5.1.3 *a)* [Vaz86] *If μ is a δ-almost k-wise independent probability measure on $\{0,1\}^n$, then the projection of μ on any k or fewer of the n coordinates is δ-close in the L_∞-norm sense to the uniform probability measure on those coordinates.*

b) [Gol92] *Any δ-almost k-wise independent probability measure μ_1 on $\{0,1\}^n$ is $n^k\delta$-close to a k-wise independent probability measure μ_2 on $\{0,1\}^n$ in the sense that $|\mu_1(A) - \mu_2(A)| \leq n^k\delta$, $\forall A \subset \{0,1\}^n$.*

Thus, if the k-wise independence property can ϵ-fool a function $f : \{0,1\}^n \to \{0,1\}$, then the δ-almost k-wise independence property can $(\epsilon + \delta n^k)$-fool f.

Note that in the setting of Section 5.2, (a) follows immediately from the fact that the L_1-norm of the Fourier transform of an AND gate is 1.

5.1.1.5 Classical explicit constructions from codes

1) [NN93, AGHP92] If $C \subset \{0,1\}^q$ is an n-dimensional linear code such that the weight of any codeword is $q\delta/2$-close to $q/2$, then the uniform distribution on the rows of any matrix whose n columns are linearly independent codewords is δ-biased as a probability measure on $\{0,1\}^n$.

   Classical explicit constructions from concatenated Reed-Solomon codes or, more generally, concatenated algebraic geometric codes achieve $q = \left(\frac{n}{\delta}\right)^{\Theta(1)}$.

2) [Vaz86] If $C \subset \{0,1\}^n$ is a linear code whose dual has minimum distance above k, then the uniform distribution on the codewords of C is k-wise independent as a probability measure on $\{0,1\}^n$.

   Classical explicit codes constructions achieve $|C| = n^{\Theta(k)}$.

3) [NN93, AGHP92] If λ is a δ-biased probability measures on $\{0,1\}^d$ and $G_{d\times n}$ is a generator matrix of a d-dimensional block-length-n binary linear code whose dual has minimum distance above k, then the probability measure induced on $\{0,1\}^n$ by λ, via G as a linear map $\{0,1\}^d \to \{0,1\}^n$, is δ-almost k-wise independent.

   By constructing λ as in (1), this construction gives a δ-almost k-wise independent probability measure μ that is discrete on its support. Classical explicit codes constructions achieve an $O\left(\frac{k \log n}{\delta}\right)^{\Theta(1)}$ support size.

Note that the correctness of (3) follows from observing that the δ-bias property of probability measures on $\{0,1\}^d$ is invariant under nonsingular $\mathbb{F}_2$-linear maps from $\{0,1\}^d$ to $\{0,1\}^d$.

5.1.2 Related literature

5.1.2.1 Known polynomial approximations of AC_0

We review in this section the literature of the polynomial approximations of AC_0 for future reference.

Let $f : \{0,1\}^n \to \{0,1\}$ be computable by an AND-OR depth-d circuit of size M and unbounded fanin. Then:

0) [Has86] Hastad switching Lemma: Consider a random restriction ρ that independently keeps each input bit unset with a probability $p = \frac{1}{(20k)^d}$, sets it to 1 with a probability $1-\frac{p}{2}$, and to 0 with a probability $1-\frac{p}{2}$. Then the probability, over the choice of ρ, that f, when restricted to the values set by ρ, cannot be evaluated by a decision tree of depth k is at most $M2^{-2k}$.

1) [LMN89] $\sum_{z;w(z)>t} \hat{f}(z)^2 \leq M2^{-\frac{1}{4}t^{\frac{1}{d+3}}}$. The bound is based on (0).

2) [LMN89] Let $\tilde{f}_t = \sum_{z;w(z)\leq t} \hat{f}(z)\mathcal{X}_z$. Then it follows from (1) that

$$Pr[sign(\tilde{f}_t) \neq f] \leq \sum_{z;w(z)>t} \hat{f}(z)^2 \leq \theta$$

when $t \geq \log^{d+3}\left(\frac{M}{\theta}\right)^4$.

3) [BRS91] There is a polynomial $p(x,y)$ in $\mathbb{Z}[X,Y]$, $X = x_1,\ldots,x_n$, $Y = y_1,\ldots,y_r$, $r = O(\log\frac{1}{\theta}\log^2 n)$, of degree $O(\log\frac{1}{\theta}\log^2 n \log M)^d$ such that for each x in $\{0,1\}^n$, $Pr_{y\in\{0,1\}^r}[p(x,y) \neq f(x)] \leq \theta$

4) [ABFR94, BRS91] There exists a polynomial $p \in \mathbb{Z}[x_1,\ldots,x_n]$ of degree $O(\log\frac{M}{\theta}\log M)^d$ such that $Pr_{x\in\{0,1\}^n}[p(x) \neq f(x)] \leq \theta$.

5) [Raz87] For any finite field $\mathbb{F}_q$, there is a polynomial $p \in \mathbb{F}_q[x_1,\ldots,x_n]$ of degree $O(q^d \log^d\frac{M}{\theta})$ such that $Pr_{x\in\{0,1\}^n}[p(x) \neq f(x)] \leq \theta$.

We will use a variation of (4) in Sections 5.7 and 5.8, and a variation of (5) in Section 5.8.

For future reference in Section 5.7.1, we note that

Remark 5.1.4 In (2), (3), and (4), the polynomial p can take very large values when it disagrees with f, and $E|p-f|$ is potentially very large. For example, in (4), p can take values as large as $2^{O(\log\frac{M}{\theta}\log M)^d}$, and $E|p-f|$ is potentially as large as $\theta \times 2^{O(\log\frac{M}{\theta}\log M)^d}$.

5.1.2.2 Nisan generator for AC_0

For future reference in Section 5.8.2.3, we review in this section the proof technique of the Nisan generator for AC_0.

The PRG of Nisan for AC_0 is based on Hastad's lower bound on the hardness of approximating parity by AC_0 circuits.

Lemma 5.1.5 [Has86] *Let $f : \{0,1\}^n \to \{0,1\}$ be computable by an AND-OR depth-d circuit of size $M \leq 2^{n^{\frac{1}{d+1}}}$ and unbounded fanin, then $\left|Pr_x[f(x) = \oplus_i x_i] - \frac{1}{2}\right| \leq 2^{n^{\frac{1}{d+1}}}$.*

Hastad lower bound is one of the consequences of the Hastad Switching Lemma.

The generator of Nisan is defined as $G^N : \{0,1\}^r \to \{0,1\}^n$, $x \mapsto (\oplus_{j \in S_i} x_j)_{i=1}^n$, where $S_1, \ldots, S_n$ are subsets of $[r]$ that form a (v,l)-design in the sense that: 1) each subset has size l, and 2) no two distinct subsets share more than v elements. The setting of the values is $v = \log n$, $r = O(l^2)$, and $r = O(\log^{2d+6} n)$.

Theorem 5.1.6 [Nis91] *G^N can ϵ-fool any boolean function $f : \{0,1\}^n \to \{0,1\}$ that can be realized by an AND-OR depth-d circuit of size $M = n^{O(1)}$ and unbounded fanin, for all $\epsilon = n^{-O(1)}$.*

The argument of Nisan is the following. If G^N cannot ϵ-fool f, then by Yao's unpredictability argument, there is an $i \in \{0, \ldots, n\}$, a setting $b_{i+1}, \ldots, b_n$ of the input variables $x_{i+1}, \ldots, x_n$, and $b_0 \in \{0,1\}$, such that the function

$$f'(x_1, \ldots, x_{i-1}) = f(x_1, \ldots, x_{i-1}, b_{i+1}, \ldots, b_n) \oplus b_0 \oplus 1$$

can predict the value of i'th bit of G^N from its previous bits with a probability at least $\frac{1}{2} + \frac{\epsilon}{n}$. Since for each $j < i$, S_j can intersect S_i in at most v elements, we get a boolean function

$$f''(x) = f'(D_1(x|_{S_i \cap S_1}), \ldots, D_{i-1}(x|_{S_i \cap S_{i-1}}))$$

that can compute f correctly on at least $\frac{1}{2}+\frac{\epsilon}{n}$ fraction of the inputs, where the D_i's are DNF formulas. When the parameters are suitably selected, this contradicts Lemma 5.1.5 as f'' is computable by a circuit of depth $d+2$ and size at most $(i-1)2^v + M$.

5.1.2.3 The quadratic-residues PRG

When deriving general statements about δ-almost k-wise independent probability measures, we will give applications to the distribution of quadratic residues via the quadratic-residues PRG that we review in this section. See Sections 5.5, 5.6, and 5.7. Also, from a different perspective that we explain in Section 5.9.1, this PRG was the original motivation behind the start of the research in this chapter.

Consider the quadratic residues PRG which is defined as

$$G : \mathbb{F}_q \to \{0,1\}^n,\ G(a) = \{x(a+t)\}_{t \in I},$$

where q is an power of an odd prime, I a subset of $\mathbb{F}_q$ of size n, and $x : \mathbb{F}_q \to \{0,1\}$

$$x(a) = \begin{cases} 1 & \text{if } \psi(a) = 1 \\ 0 & \text{o.w. .} \end{cases}$$

Recall that ψ means the quadratic character of $\mathbb{F}_q^\times$, i.e.,

$$\psi(a) = \begin{cases} 1 & \text{if } a \in Q \\ -1 & \text{if } a \in \mathbb{F}_q^\times \backslash Q \\ 0 & \text{if } a = 0. \end{cases}$$

where Q is the set of quadratic residues in $\mathbb{F}_q$.

This PRG was introduced in [AGHP92] in the setting when q is prime and $I = \{0, \ldots, n-1\}$. It was shown in [AGHP92] that this PRG has the $2n/\sqrt{q}$-bias property. This is a direct consequence of Weil's theorem on the analog of the Riemann Hypothesis for projective nonsingular curves over finite fields, which implies the following bound in the hyperelliptic case.

Theorem 5.1.7 [Wei48] *Let q be a power of an odd prime, and $g \in \mathbb{F}_q[x]$ be a non-constant square free degree-d polynomial, then*

$$|\sum_{a\in\mathbb{F}_q} \psi(g(a))| \leq (d-1)\sqrt{q}+1.$$

See for instance [Mor94, Sti93, LN83]. By the same argument as in [AGHP92], it can be shown also that Weil's bound immediately implies also that $|E_{a\in\mathbb{F}_q}\mathcal{X}_z(G(a))| \leq 2k/\sqrt{q}$ for each $z \in \{0,1\}^n$ whose weight is at most k. The calculations are at the end of this section. Thus this PRG has the $2k/\sqrt{q}$-almost k-wise independence property.

More generally, if $p(x) \in \mathbb{F}_q[x]$ is a nonconstant square free polynomial of small degree l, we can define a quadratic-residues-like PRG with respect to p as follows. Let

$$G_p : \mathbb{F}_q \to \{0,1\}^n,\ G(a) = \{x(p(a+t))\}_{t\in I}.$$

Then it also follows immediately from Theorem 5.1.7 that G_p has the $2ln/\sqrt{q}$-bias property, and more specifically the $2lk/\sqrt{q}$-almost k-wise independence property.

Namely, if $z \in \{0,1\}^I$ has weight at most k, then with $S = \{t \in [n] : z_t = 1\}$, and $Z = \{a \in \mathbb{F}_q : p(a) = 0\}$, we have

$$E_{a\in\mathbb{F}_q}\mathcal{X}_z(G(a)) = \frac{1}{q}\sum_{a\in\mathbb{F}_q}\psi(\prod_{t\in S}p(a+t)) + \frac{1}{q}\sum_{a_1\in Z, a_2\in S}\mathcal{X}_z(G(a_1-a_2)) - \psi(\prod_{t\in S}p(a_1-a_2+t)),$$

hence

$$\left|E_{a\in\mathbb{F}_q}\mathcal{X}_z(G(a))\right| \leq \frac{(lk-1)\sqrt{q}+1}{q} + \frac{lk}{q} \leq \frac{2lk}{\sqrt{q}}.$$

Note that the bound becomes worse as l grows. The reason why we are interested in this generality is not for the purpose of constructing δ-almost k-wise independent probability measures, but because of the implications of statements that holds for any δ-almost k-wise independent probability measure to this general setting.

5.1.3 Summary of results

5.1.3.1 When are the basic pseudorandomness properties sufficient? the dual perspective

We note in Section 5.2 that linear programming duality gives a purely analytical characterization of the class of boolean function that can be fooled by the δ-almost k-wise independence property. The characterization is necessary and sufficient, and it is in terms of tight average sandwichability between real valued functions with low degree and small L_1-norm in the Fourier domain.

Namely, let $f : \{0,1\}^n \to \{0,1\}$. Then, any δ-almost k-wise independent probability measure on $\{0,1\}^n$ can $o(\epsilon)$-fool f if and only if there exists $f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ such that

a) $f_1 \leq f \leq f_2$,

b) $E(f_2 - f_1) = o(\epsilon)$,

c) $\|\widehat{f_1}\|_1, \|\widehat{f_2}\|_1 = o\left(\frac{\epsilon}{\delta}\right)$.

d) $deg(f_1), deg(f_2) \leq k$.

Thus, specifically, 1) any δ-biased independence probability measure on $\{0,1\}^n$ can $o(\epsilon)$-fool f if and only if there exists $f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ satisfying (a,b,c), and 2) any k-wise independent probability measure on $\{0,1\}^n$ can $o(\epsilon)$-fool f if and only if there exists $f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ satisfying (a,b,d).

We will use the characterizations in Sections 5.5 and 5.7.1.

Then, we consider the unpredictability perspective of derandomizability by basic pseudorandomness properties. We note that we can take advantage of the generality of the situation to improve upon the reduction one can get by using Yao's unpredictability lemma as a black box. We will use this relation in Sections 5.7.2, 5.8.2, and 5.8.2.2.

5.1.3.2 Linear codes versus general k-wise independent measures

We study in Section 5.3 the position of classical linear-codes-based constructions of k-wise independent probability measures in the convex polytope of all such measures, and its subpolytope consisting of those measures whose Fourier transform is nonnegative.

If $C \subset \{0,1\}^n$ is an $\mathbb{F}_2$-linear code whose dual $C^\perp$ has minimum distance above k, then the probability measure $\mu_C \stackrel{\text{def}}{=} \frac{1}{|C|} 1_C$ supported by C is k-wise independent. This is the classical construction of k-wise independent probability measures from linear codes.

Consider the convex polytope $Q_k \subset \mathbb{R}^{\{0,1\}^n}$ of k-wise independent probability measures μ on $\{0,1\}^n$. This polytope is specified by the constraints: $\mu \geq 0$, $\sum_x \mu(x) = 1$, and $\sum_x \mu(x) \mathcal{X}_z(x) = 0$, $\forall z \neq 0$ such that $w(z) \leq k$.

We show that the linear codes C (and their translations, i.e., cosets) that are minimal (with respect to inclusion) with the property the dual $C^\perp$ has minimum distance above k are extreme points of Q_k.

We note that they are not all the extreme points, which leaves us with the open problem of studying the other extreme points.

A very special property of μ_C is that its Fourier transform $\widehat{\mu_C}$ is nonnegative. Let $P_k \subset Q_k$ be the convex polytope of k-wise independent probability measures on $\{0,1\}^n$ whose Fourier transform is nonnegative.

We argue that the binary linear codes with dual distance above k are exactly the elements of P_k that are uniform on their support, and exactly the elements of P_k that are on the boundary of a specific radius-$\frac{1}{2}$ sphere containing P_k and centered at $\frac{1}{2}\mu_{\{0\}}$. Thus they are specifically extreme points of P_k.

Here again we note that they are not all the extreme points, which again leaves us with the open problem of studying the other extreme points.

Let L_k be the set of linear codes with dual distance above k. Relaxing the set L_k to P_k is one way to look at Delsarte LP (Linear Programming) coding bound [Del73] in the setting of linear codes. We will explore in Section 5.8.5 other relaxations based

on this approach.

5.1.3.3 Some limitations of the small bias property

We argue in Section 5.4 that there is a $2^{-\Omega(n)}$-biased probability measure μ on $\{0,1\}^n$ that cannot $o(1)$-fool a function that can be realized as an $O(\log n)$-depth circuit of linear size, and as an $O(1)$-width read-once oblivious branching program.

5.1.3.4 Log-wise independence versus read-once DNF formulas

We argue in Section 5.5 that any sufficiently log-wise independent probability measure looks random to all polynomially small read-once DNF formulas.

More specifically, we show that if $f : \{0,1\}^n \to \{0,1\}$ is computable by a read once (i.e., the clauses are disjoint) DNF formula with m clauses, then any k-wise independent probability measure μ on $\{0,1\}^n$ can ϵ-fool f, with

$$\epsilon = \min_{1 \le s \le n} 2^{-(s - \log m)} + 2^{-\frac{k-37}{11s}}.$$

Therefore $\epsilon = o(1)$, when for instance $k = \log m \log\log m$.

We conclude with a similar statement for δ-almost k-wise independent measures when k is relatively small, and we give an application to the distribution of quadratic residues by applying the result to the quadratic residues PRG.

After using the sandwiching approach of Section 5.2, we complete the result as a consequence of a more general result on weak probability measures, which is a notion that is naturally suggested by the problem. We say that a probability measure γ on $\{0,1\}^m$ is (k,ϵ)-weak if when γ' is another probability measure on $\{0,1\}^m$ that agrees with γ on all its projection on any k of the coordinates, then the L_∞-distance between γ and γ' is at most ϵ. We show that if $X_1, \ldots, X_m$ are independent binary random variables, then the corresponding probability measure on $\{0,1\}^m$ is $(k, 2^{-\frac{k-37}{11}})$-weak.

5.1.3.5 Limited independence versus weight probability

We establish in Section 5.6 a sharp $O(k^{-1/2})$ upper bound on the probability that a random binary string generated according to a k-wise independent measure has any given weight.

More precisely, we show that if μ is a k-wise independent probability measure on $\{0,1\}^n$. Then

$$\begin{aligned} \max_{a=0,1,\ldots,n} Pr_{x\sim\mu}[w(x)=a] &\leq \frac{1}{\sum_{0\leq l \text{ even } \leq \lfloor k/2\rfloor -1} \frac{1}{2^l}\binom{l}{l/2}} \\ &= \frac{\sqrt{\pi+o(1)}}{\sqrt{k}}, \end{aligned}$$

where the asymptotic statement is in terms of the growth of k.

The setting is naturally extendable to almost k-wise independent measures. We give applications related to the distribution of quadratic-residues and the weight distributions of linear codes.

We give another application of the bound in Section 5.7.3.

5.1.3.6 Poly-log-wise independence versus AC_0

We consider in Section 5.7 the problem of derandomizability of AC_0 by arbitrary k-wise independent probability measures, when k is made polylogaritmically large enough. We reduce this problem to a conjecture about the the symmetry of the optimum of some symmetric optimization problem with linear constraints and a nonlinear objective function.

We consider in Section 5.7 the following problem which was essentially proposed by Linial and Nisan [LN90].

How large should k be as a function of n, M, d, and ϵ so that the k-wise independence property is sufficient to ϵ-fool any AND-OR circuit on n-bits of size M, depth d, and unbounded fanin?

The generality of the problem has many potential applications. The setting is naturally extendable to δ-biased probability measures, and consequently has appli-

cations related to the distribution of quadratic residues. Note that the dual of the problem is asking for a new characterization of AC_0 by low-degree polynomials over the reals in the sense of Section 5.2.

First, we reduce this problem to the following question.

How large should k be in terms of h and n, so that if $X_1, \ldots, X_{n+1}$ are binary k-wise independent random variables, no degree $\leq h$ polynomial p over the reals on $X_1, \ldots, X_n$ can predict the value of X_{n+1} with a probability significantly better than $1/2$?

The reduction corresponds to the case when h is polylogarithmic in n, and is based on the approximability of AC_0 circuits by low-degree polynomials over the reals (Beigel, Reingold, and Spielman [BRS91], Aspnes, Beigel, Furst, and Rudich [ABFR94]), and the unpredictability perspective in Section 5.2.

Using the bound we established in Section 5.6, we establish a good bound in the restricted version of the problem corresponding to the case when p is a symmetric polynomial. We show that if $k \geq 16\pi h^2$, h is larger than some absolute constant, and $X_1, \ldots, X_{n+1}$ are binary k-wise independent random variables, then no symmetric degree-h polynomial over the reals on $X_1, \ldots, X_n$ can predict the value of X_{n+1} with a probability larger than $1/2$.

Due to the highly symmetric nature of the low-degree polynomials predictors problem, we conjecture that the symmetric case is a worst case.

Establishing this conjecture will pull the bound we established on the symmetric case to the more general setting of arbitrary low-degree polynomials, and consequently will resolve in a satisfactory way the problem of derandomizability of AC_0 by any polylog-wise independent probability measure. The correctness of the symmetric optimum conjecture implies that in order to guarantee that the k-wise independence property is sufficient to $M^{-\Theta(1)}$-fool any size-M depth-d circuit in AC_0, it is sufficient to make $k = \Theta(\log^{4d} M)$.

5.1.3.7 Parity with encrypted linear help

We study in Section 5.8 the problem of approximability of high-degree parity functions on high-dual-distance binary linear codes by low-degree polynomials with coefficients in fields of odd characteristics. This problem has applications to the ability of binary linear codes with sufficiently large dual distance to derandomize AC_0, or low-degree polynomial equations on binary input variables with coefficients in small finite fields of odd order. Among other results, we relax this problem into essentially a single low-dimensional low-complexity linear program in terms of Krawtchouk polynomials. The problem of bounding the optimum of the linear program remains open.

See Section 5.8.1 for a detailed summary.

5.2 When are the basic pseudorandomness properties sufficient? the dual perspective

Definition 5.2.1 *We say that a measure* property can ϵ-fool *a function* $f : \{0,1\}^n \to \{0,1\}$*, if any probability measure on* $\{0,1\}^n$ *with this property can* ϵ*-fool* f*.*

We note in Theorem 5.2.4 that linear programming duality gives a purely analytical characterization of the class of boolean function that can be fooled by the δ-almost k-wise independence property. The characterization is necessary and sufficient and it is in terms of tight average sandwichability between real valued functions with low degree and small L_1-norm in the Fourier domain. We conclude in Corollaries 5.2.5 and 5.2.6 similar statements for the δ-bias property and the k-wise independence properties.

We illustrate in Lemma 5.2.2 that allowing real values compared to binary values is essential.

In Section 5.2.1 we consider the unpredictability perspective of derandomizability by basic pseudorandomness properties. We note that we can take advantage of the generality of the situation to improve upon the reduction one can get by using Yao's unpredictability lemma as a black box.

We will use the sandwichability characterizations in Sections 5.5 and 5.7.1, and the unpredictability perspective in Sections 5.7.2, 5.8.2, and 5.8.2.2.

Consider first the δ-bias property. Let $f : \{0,1\}^n \to \{0,1\}$. When is the δ-bias property sufficient to $o(1)$-fool f?

Let μ be a probability measure on $\{0,1\}^n$ with the δ-bias property, i.e., $|E_\mu \mathcal{X}_z| < \delta$ for each nonzero z in $\mathbb{Z}_2^n$. Consider the Fourier transform $\hat{f} : \mathbb{Z}_2^n \to \mathbb{R}$ of f, i.e., $f(x) = \sum_z \hat{f}(z)\mathcal{X}_z(x)$. After taking the expectations and noting that $\hat{f}(0) = Ef$, we get

$$E_\mu f - Ef = \sum_{z \neq 0} \hat{f}(z) E_\mu \mathcal{X}_z,$$

thus

$$|E_\mu f - Ef| \leq \delta \|\hat{f}\|_1,$$

where $\|\hat{f}\|_1 = \sum_z |\hat{f}(z)|$. Thus, if we set $\delta = o\left(\frac{1}{|\hat{f}\|_1}\right)$, we can guarantee that μ can $o(1)$-fool f. Since the smallest size of the support of any μ with the δ-bias property is $\left(\frac{1}{\delta}\right)^{\Theta(1)}$, we need $\|\hat{f}\|_1$ to be small.

How large is this class of functions? It contains for instance small decision trees [KM91]. More generally

Lemma 5.2.2 *Let $f : \{0,1\}^n \to \{0,1\}$ be computable by a decision tree where each node is associated with a set of variables whose parity determines the next node. Then $\|\hat{f}\|_1 \leq M$, where M is the number of leafs.*

Proof. Since f is a disjoint OR of the leafs, it can be expressed as $f = \sum_{l=1}^{M} f_l \circ L$, where $L : \mathbb{Z}_2^n \to \mathbb{Z}_2^m$ is $\mathbb{F}_2$-linear and each $f_l : \mathbb{Z}_2^m \to \mathbb{R}$ is an AND gate on some of the variables $x_1, \ldots, x_m$ with possibly negated inputs. Thus $\|\hat{f}\|_1 \leq \sum_l \|\widehat{f_l \circ L}\|_1 \leq M$ because $\|\widehat{f_l \circ L}\|_1 \leq \|\widehat{f_l}\|_1 = 1$. ∎

But for an arbitrary binary function f, $\|\hat{f}\|_1$ can be as large as $2^{\frac{n}{2}}$. What about low complexity boolean functions? The AC_0 ones for instance?

Lemma 5.2.3 *Consider the read-once monotone depth-2 $O(n)$-size circuit $f : \{0,1\}^n \to \{0,1\}$ given as the AND of a NAND gates each on b bits, thus $n = ab$. Assume that $a = 2^b$, then*

a) $\|\hat{f}\|_1 = 2^{\Omega(\frac{n}{\log n})}$

b) $deg(f) = n$

c) $Ef = \Theta(1)$,

but $\exists f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ *such that*

d) $\|\widehat{f_1}\|_1, \|\widehat{f_2}\|_1 = 2^{O(\log n \log\log n)}$

e) $deg(f_1), deg(f_2) = O(\log n \log\log n)$

f) $f_1 \leq f \leq f_2$

g) $E(f_2 - f_1) \leq \frac{1}{\log n}$.

Note that a version of (a) appears in [BS92].

Proof. In Section 5.2.2 ■

Thus, we cannot hope to go far with the small L_1-norm requirement alone. But the existence of f_1 and f_2 resolves the problem for the function f in the above Lemma. Indeed, we have

$$(E_\mu f_1 - Ef_1) - E(f - f_1) \leq E_\mu f - Ef \leq (E_\mu f_2 - Ef_2) + E(f_2 - f) \tag{5.1}$$

Thus

$$\begin{aligned} |E_\mu f - Ef| &\leq \delta \max\{\|\widehat{f_1}\|_1, \|\widehat{f_2}\|_1\} + E(f_2 - f_1) \\ &\leq \delta 2^{O(\log n \log\log n)} + \frac{1}{\log n} \\ &\leq \frac{2}{\log n} \end{aligned}$$

when $\delta = 2^{-O(\log n \log\log n)}$ is sufficiently large. Compare this with the L_1-norm approach alone which requires setting δ to $2^{-\Omega(\frac{n}{\log n})}$.

Thus, those boolean functions that can be well trapped between two real valued functions whose L_1-norm in the Fourier domain is not very large can be well fooled

by the not too small bias property, and this class is provably larger than the class of boolean functions whose L_1-norm in the Fourier domain is not very large. It turns out this is a complete characterization. See Corollary 5.2.5 below. More generally,

Theorem 5.2.4 *Let $f : \{0,1\}^n \to \{0,1\}$, $\delta, \epsilon > 0$, and k a positive integer. Then the δ-almost k-wise independence property can ϵ-fool f if and only if $\exists f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{R}$ such that*

- $f_1 \leq f \leq f_2$
- $\delta \sum_{z \neq 0} |\widehat{f_1}(z)| + E(f - f_1) \leq \epsilon$
- $\delta \sum_{z \neq 0} |\widehat{f_2}(z)| + E(f_2 - f) \leq \epsilon$
- $deg(f_1), deg(f_2) \leq k$.

Therefore, asymptotically speaking, the δ-almost k-wise independence property can $o(\epsilon)$-fool a boolean function $f : \{0,1\}^n \to \{0,1\}$ if and only if $\exists f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ such that

- $f_1 \leq f \leq f_2$
- $E(f_2 - f_1) = o(\epsilon)$
- $\|\widehat{f_1}\|_1, \|\widehat{f_2}\|_1 = o\left(\frac{\epsilon}{\delta}\right)$
- $deg(f_1), deg(f_2) \leq k$.

Proof. By linear-programming duality, see Section 5.2.3 for the calculations. Note that the two primals are: $\max \sum_x \mu(x) f(x) - Ef$ and $\max Ef - \sum_x \mu(x) f(x)$ where we are optimizing on $\mu : \mathbb{Z}_2^n \to \mathbb{R}$ such that $\mu > 0$, $\sum_x \mu(x) = 1$, and $|\sum_x \mu(x) \mathcal{X}_z(x)| \leq \delta$ for each nonzero z in $\mathbb{Z}_2^n$ whose weight is less than or equal to k. ■

Corollary 5.2.5 *Let $f : \{0,1\}^n \to \{0,1\}$, and $\delta, \epsilon > 0$. Then the δ-bias property can ϵ-fool f if and only if $\exists f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{R}$ such that*

- $f_1 \leq f \leq f_2$

- $\delta \sum_{z \neq 0} |\widehat{f_1}(z)| + E(f - f_1) \leq \epsilon$
- $\delta \sum_{z \neq 0} |\widehat{f_2}(z)| + E(f_2 - f) \leq \epsilon$.

Therefore, asymptotically speaking, the δ-bias property can $o(\epsilon)$-fool a boolean function $f : \{0,1\}^n \to \{0,1\}$ if and only if $\exists f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ such that

- $f_1 \leq f \leq f_2$
- $E(f_2 - f_1) = o(\epsilon)$
- $\|\widehat{f_1}\|_1, \|\widehat{f_2}\|_1 = o\left(\frac{\epsilon}{\delta}\right)$.

Corollary 5.2.6 *Let $f : \{0,1\}^n \to \{0,1\}$, $\epsilon > 0$, and k a positive integer. Then the k-wise independence property can ϵ-fool f if and only if $\exists f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{R}$ such that*

- $f_1 \leq f \leq f_2$
- $E(f - f_1) \leq \epsilon$ *and* $E(f_2 - f) \leq \epsilon$
- $deg(f_1)$, $deg(f_2) \leq k$.

Therefore, asymptotically speaking, the k-wise independence property can $o(\epsilon)$-fool a boolean function $f : \{0,1\}^n \to \{0,1\}$ if and only if $\exists f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ such that

- $f_1 \leq f \leq f_2$
- $E(f_2 - f_1) = o(\epsilon)$
- $deg(f_1), deg(f_2) \leq k$.

5.2.1 The unpredictability perspective

Arguing by unpredictability will divide ϵ by n if we want to use Yao's Lemma as a black box. We note that we can take advantage of the generality of the situation and manage with dividing ϵ by 2 only.

Lemma 5.2.7 *I) Let $f : \{0,1\}^n \to \{0,1\}$, and μ a probability measure on $\{0,1\}^n$, and $\epsilon > 0$ such that*

$$|Pr_{x\sim\mu}[f(x) = 1] - Pr_x[f(x) = 1]| > \epsilon, \tag{5.2}$$

then there exists $b \in \{0,1\}$ and $z \in \{0,1\}^n$ such that

$$Pr_{x'\sim\mu'}[f(x'_1, \ldots, x'_n) \oplus b = x'_{n+1}] > \frac{1}{2} + \frac{\epsilon}{2}, \tag{5.3}$$

where μ' is the probability measure on $\{0,1\}^{n+1}$ given by

$$\mu'(x') = \begin{cases} \frac{1}{2}\mu(x'|_{[n]}) & \text{if } x'_{n+1} = 0 \\ \frac{1}{2}(\sigma_z\mu)(x'|_{[n]}) & \text{if } x'_{n+1} = 1, \end{cases} \tag{5.4}$$

and where $(\sigma_z\mu)(x) \stackrel{\text{def}}{=} \mu(x \oplus z)$.

II) In general, assume that μ and μ' are related via (5.4) for some z. If μ is a δ-biased, k-wise independent, or a discrete measure (i.e., uniform on its support) supported by a linear code, then so is μ'. Moreover, in the case when μ is k-wise independent, μ' has also the additional property that any $k+1$ bit including the last are independent.

See Sections 5.7.2 for an application.

Note that unlike the setting of Theorem 5.2.4 and Corollaries 5.2.5, and 5.2.6, the statement of the Lemma make sense only when f takes binary values.

Note also that μ' works by generating x'_{n+1} uniformly at random, and depending on the value of x'_{n+1}, generates $x'_1, \ldots, x'_n$ according to μ or $\sigma_z\mu$.

Proof. The main point is to substitute the nested sequence of PRG's in Yao's unpredictability lemma, which is the part responsible for dividing ϵ by n, by an expectation argument based on translations of the PRG.

Consider the translations $\sigma_z\mu$ of μ by elements z in $\{0,1\}^n$ with respect to addition $\oplus$ in $\mathbb{Z}_2^n$. We have $Pr_x[f(x)=1] = E_z Pr_{x\sim\sigma_z\mu}[f(x)=1]$, thus

$$Pr_{x\sim\mu}[f(x)=1] - Pr_x[f(x)=1] = E_z(Pr_{x\sim\mu}[f(x)=1] - Pr_{x\sim\sigma_z\mu}[f(x)=1]),$$

and hence Hypothesis (5.2) implies $\exists z \in \{0,1\}^n$ such that

$$|Pr_{x\sim\mu}[f(x)=1] - Pr_{x\sim\sigma_z\mu}[f(x)=1]| > \epsilon.$$

Fix such a z. Thus $\exists b \in \{0,1\}$ such that

$$Pr_{x\sim\sigma_z\mu}[f(x)\oplus b=1] - Pr_{x\sim\mu}[f(x)\oplus b=1] > \epsilon,$$

and hence (5.3) since

$$\begin{aligned} Pr_{x'=(x,x'_{n+1})\sim\mu'}[f(x)\oplus b = x'_{n+1}] &= \frac{1}{2}Pr_{x\sim\mu}[f(x)\oplus b=0] + \frac{1}{2}Pr_{x\sim\sigma_z\mu}[f(x)\oplus b=1] \\ &= \frac{1}{2} + \frac{1}{2}(Pr_{x\sim\sigma_z\mu}[f(x)\oplus b=1] - Pr_{x\sim\mu}[f(x)\oplus b=1]). \end{aligned}$$

Verifying (II) is straightforward. ■

We can elaborate on (II) in the special case of linear codes as follows.

Corollary 5.2.8 *If $f : \{0,1\}^n \to \{0,1\}$, and $C \subset \mathbb{Z}_2^n$ is a linear code whose dual distance is above k, and $\epsilon > 0$ are such that:*

$$|Pr_{x\in C}[f(x)=1] - Pr_{x\in\mathbb{Z}_2^n}[f(x)=1]| > \epsilon,$$

then

$$Pr_{x\in C''}[f(x)\oplus b = Parity_A(x)] > \frac{1}{2} + \frac{\epsilon}{2},$$

for some linear code $C'' \subset \mathbb{Z}_2^n$, $b \in \mathbb{Z}_2$, and $A \subset [n]$ such that:

b) the minimum dual distance of C'' is above k

c) $|A'| > k$, for all $A' \subset [n]$ such that $Parity_A|_{C''} = Parity_{A'}|_{C''}$.

Here $Parity_A : \{0,1\}^n \to \{0,1\}$ *means* $Parity_A(x) = \oplus_{i \in A} x_i$.

Proof. We are in the setting where μ is a discrete probability measure supported by a linear code $C \subset \mathbb{Z}_2^n$ whose dual distance is above k, and μ' is related to μ via (5.4). Let $C'' = C \cup (C + z)$, $C' \subset \mathbb{Z}_2^{n+1}$ be the support of μ', and A such that $C' = \{(x, \text{Parity}_A(x)) : x \in C''\}$. Note also that $z \notin C$ since $\epsilon > 0$. ∎

See Sections 5.8.2 and 5.8.2.2 for an application.

5.2.2 Proof of Lemma 5.2.3

Partition $[n]$ into a disjoint interval $I_1, \ldots, I_a$ each of size b, thus $f : \mathbb{Z}_2^n \to \{0,1\}$ is given by

$$f(x) = \Pi_{i=1}^a (1 - f_i(x)), \quad f_i(x) = \Pi_{i \in I_i} x_i. \tag{5.5}$$

1) We want to compute the Fourier transform $\hat{f}$ of f from that of the f_i's. To do this, we need the following straightforward lemma.

Lemma 5.2.9 *If* $G = G_1 \times \ldots \times G_m$ *be a direct product of finite abelian groups, and* $f_1, \ldots, f_m : G \to \mathbb{C}$ *are such that* $f_i(g)$ *depends on the* G_i*-component of* g*, then* $\widehat{f_i}(\mathcal{X}) = \widehat{f_i|_G}(\mathcal{X}_i)$*, where* $\mathcal{X}_i$ *is the* $\widehat{G_i}$*-component of* $\mathcal{X}$*, and* $\widehat{\prod_i f_i} = \prod_i \hat{f}_i$.

The proof is direct. Let us go back to our setting with $G = \mathbb{Z}_2^n$, and $G_i = \{x : x|_{[n]\setminus I_i} = 0\}$, we get $\hat{f} = \prod_i \widehat{1 - f_i}$. We have

$$1 - f_i = 1 - \frac{1}{2^b} \sum_{T \subset I_i} (-1)^{|T|} \mathcal{X}_T(x),$$

thus $\|\widehat{1 - f_i}\|_1 = 2 - \frac{2}{2^b}$, and hence

$$\|\hat{f}\|_1 = \|\widehat{1 - f}\|_1 = \prod_i \|\widehat{1 - f_i}\|_1 = \left(2 - \frac{2}{2^b}\right)^a = (Ef) 2^a \geq 2^{\frac{n}{\log n}}.$$

2) To construct f_1 and f_2, we proceed by an inclusion-exclusion mentality. Expand (5.5) as

$$f = \prod_{i=1}^a (1 - f_i) = \sum_{k=0}^n (-1)^k g_k, \quad g_k = \sum_{T \subset [a]; |T| = k} \prod_{i \in T} f_i, \tag{5.6}$$

and let $g_{\leq k} = \sum_{i=0}^{k} (-1)^i g_i$. By the same proof of inclusion-exclusion, or equivalently by a direct induction, we have

$$g_{\leq k-1} \leq f \leq g_{\leq k}$$

for each even k. Say that $k = \Theta(\log \log n)$ is even, and let $f_1 = g_{\leq k-1}$ and $f_2 = g_{\leq k}$. We have

$$E(f_2 - f_1) = Eg_k = \sum_{T \subset [a]; |T|=k} \prod_{i \in T} Ef_i = \binom{a}{k} 2^{-kb} \leq 2^{-k \log \frac{k}{e}} \leq \frac{1}{\log n}.$$

Note that the main point why we got a good bound from truncating the inclusion-exclusion is that the f_i's are independent as random variables and the probability of f is small enough. For an elaborate study of this approach, see Section 5.5. The degree of f_2 is at most $kb = O(\log n \log \log n)$, similarly for f_1. Finally

$$\|\widehat{f_2}\|_1 \leq \sum_{T \subset [a]; |T| \leq k} \|\widehat{\prod_{i \in T} f_i}\|_1 \leq k \binom{a}{k} = 2^{O(k \log n)} = 2^{O(\log n \log \log n)}$$

since $\|\widehat{\prod_{i \in T} f_i}\|_1 = 1$, and similarly for f_1.

5.2.3 Linear-programming duality calculations

Theorems 5.2.4 and Lemma 5.5.8 follow immediately from two applications of the following lemma.

Lemma 5.2.10 *Let X be a finite set, B be a collection of real valued functions on X, μ_0 a fixed probability measure on X, $\delta > 0$, and let $\mathcal{M}_\delta$ be the set of probability measures μ on X such that $|E_\mu \beta - E_{\mu_0} \beta| \leq \delta$, for each β in B. Let f be real valued function on X. Then*

$$\max_{\mu \in \mathcal{M}_\delta} E_\mu f = \min \left\{ E_{\mu_0} g + \delta \sum_{\beta \in B} |\hat{g}_\beta| : \begin{array}{c} (\hat{g}_0, (\hat{g}_\beta)_{\beta \in B}) \in \mathbb{R} \times \mathbb{R}^B \text{ s.t. with } g : X \to \mathbb{R} \\ \text{given by } g = \hat{g}_0 + \sum_{\beta \in B} \hat{g}_\beta \beta, \text{ we have } g \geq f \end{array} \right\}$$

Proof. Index the functions in B by $z \in I$, and let $\tilde{\beta}_z = \beta_z - E_{\mu_0}\beta_z$. We have a linear program: maximize $\sum_x \mu(x)f(x)$, $(\mu(x))_{x\in X} \in \mathbb{R}^X$, subject to the constraints

$$\begin{cases} \sum_x \mu(x) = 1 \\ \mu(x) \geq 0, \forall x \\ -\delta \leq \sum_x \mu(x)\tilde{\beta}_z(x) \leq \delta, \forall z. \end{cases}$$

Taking the dual of this feasible linear program, we get: minimize $\theta_0 + \delta \sum_z (\theta'_z + \theta''_z)$, $(\theta_0, (\theta'_z, \theta''_z)_z) \in \mathbb{R}^{2|I|+1}$, subject to the constraints

$$\begin{cases} \theta_0 + \sum_z (\theta'_z - \theta''_z)\tilde{\beta}_z(x) \geq f(x), \forall x \\ \theta'_z, \theta''_z \geq 0, \forall z, \end{cases}$$

or equivalently, minimize $\theta_0 + \delta \sum_z |\theta_z|$, $(\theta_0, (\theta_z)_z) \in \mathbb{R}^{|I|+1}$, subject to the constraints

$$\theta_0 + \sum_z \theta_z \tilde{\beta}_z(x) \geq f(x), \forall x,$$

since the minimum of $a + b$ subject to $a - b = c$ and $a, b \geq 0$ is $|c|$. ■

5.3 Linear codes versus general k-wise independent probability measures

We study in this section the position of classical linear-codes-based constructions of k-wise independent probability measures in the convex polytope of all such probability measures, and its subpolytope consisting of those whose Fourier transform is nonnegative.

Let $C \subset \mathbb{Z}_2^n$ be a binary linear code, i.e., an $\mathbb{F}_2$-linear space. The dual of $C^\perp$ of C is

$$C^\perp \stackrel{\text{def}}{=} \left\{ y \in \mathbb{Z}_2^n : \sum_i x_i y_i = 0 \pmod 2 \ , \forall y \in C \right\}.$$

Recall from Section 5.1.2 the classical construction: if the minimum distance of the dual $C^\perp$ of C is above k, then then $\mu_C \stackrel{\text{def}}{=} \frac{1}{|C|} 1_C$ is a k-wise independent probability

measure on $\mathbb{Z}_2^n$.

Consider the convex polytope $Q_k \subset \mathbb{R}^{\{0,1\}^n}$ of k-wise independent probability measures μ on $\{0,1\}^n$. This polytope is specified by the constraints: $\mu \geq 0$, $\sum_x \mu(x) = 1$, and $\sum_x \mu(x)\mathcal{X}_z(x) = 0$, $\forall z \neq 0$ such that $w(z) \leq k$.

We show in Section 5.3.1 that the linear codes C (and their translations, i.e., cosets) that are minimal (with respect to inclusion) with the property that the dual $C^\perp$ has minimum distance above k are extreme points of Q_k.

We note that they are not all the extreme points. The problem of studying the other extreme points remains open.

A very special property of μ_C is that its Fourier transform $\widehat{\mu_C}$ is nonnegative. Let $P_k \subset Q_k$ be the convex polytope of k-wise independent probability measures on $\{0,1\}^n$ whose Fourier transform is nonnegative.

We argue in Section 5.3.2 that the binary linear codes with dual distance above k are exactly the elements of P_k that are uniform on their support, and exactly the elements of P_k that are on the boundary of a specific radius-$\frac{1}{2}$ sphere containing P_k and centered at $\frac{1}{2}\mu_{\{0\}}$. Thus they are specifically extreme points of P_k.

Here again we note that they are not all the extreme points, and the problem of studying the other extreme points remains open.

Let L_k be the set of linear codes with dual distance greater than k. Relaxing the set L_k to P_k is one way to look at Delsarte LP (Linear-Programming) coding bound [Del73] in the setting of linear codes. We will explore in Section 5.8.5 other relaxations based on this approach.

5.3.1 Relation to general k-wise independent measures

Consider the convex polytope $Q_k \subset \mathbb{R}^{\mathbb{Z}_2^n}$ consisting of the k-wise independent probability measures μ on $\mathbb{Z}_2^n$, i.e., μ such that $E_\mu \mathcal{X}_z = 0$ for each nonzero z in $\mathbb{Z}_2^n$ whose weight is at least k.

Note that Q_k is given by the the linear constraints: $\sum_x \mu(x) = 1$, $\mu(x) \geq 0$ for each x in $\mathbb{Z}_2^n$, and $\sum_x \mu(x)\mathcal{X}_z(x) = 0$ for each nonzero z in $\mathbb{Z}_2^n$ whose weight is at least k.

Let C be a binary linear code in $\mathbb{Z}_2^n$ whose dual $C^\perp$ has minimum distance above k. Then the discrete measure $\mu_C \stackrel{\text{def}}{=} \frac{1}{|C|}1_C$ supported by C is an element of Q_k. More generally, for each translation D of C, i.e., for each coset $D \in \mathbb{Z}_2^n/C$, $\mu_D \stackrel{\text{def}}{=} \frac{1}{|D|}1_D$ is an element of Q_k.

Theorem 5.3.1 *Let $k \geq 0$. The translations (i.e., coset) of the binary linear codes in $\mathbb{Z}_2^n$ that are minimal (with respect to inclusion) with the property that their dual has distance greater than k are (i.e., corresponds to) extreme points of the convex polytope Q_k of k-wise independent probability measures on $\mathbb{Z}_2^n$.*

Proof. Let $\mu \in Q_k$. Let S be the support of μ, and consider the $\mathbb{R}$-vector space

$$I_S = \{\alpha : S \to \mathbb{R} \mid \textstyle\sum_{x \in S} \alpha(x)\mathcal{X}_z(x) = 0, \forall z \in \mathbb{Z}_2^n \text{ s.t. } w(z) \leq k\}.$$

Then μ is an extreme point of Q_k if and only if I_S is zero dimensional.

The equivalence follows from the basic characterization of an extreme point of a linear program whose constraints are in the canonical form $Ay = b, y \geq 0$. Namely, a feasible solution y is an extreme point if and only if the columns of A corresponding to the nonzero entries of y are linearly independent.

Note that this means that if μ is an extreme point, then μ is uniquely determined by S and hence must be of minimal support.

First, observe that for any $S \subset \mathbb{Z}_2^n$, we have $I_S \cong I_{S+a}$ for all $a \in \mathbb{Z}_2^n$. This follows from the multiplicativity of the characters, and namely because $\mathcal{X}_z(x+a) = \mathcal{X}_z(x)\mathcal{X}_z(a)$. Thus, rather than starting from a translation of a linear code, we can start without loss of generality from a linear code.

Suppose that the support of a measure μ in Q_k is a linear code $C \subset \mathbb{Z}_2^n$ that is minimal with the property that its dual $C^\perp$ has minimum distance above k. We want to argue that μ is an extreme point, or equivalently that I_C contains only the zero function. From the multiplicativity of the characters, we see that I_C is invariant under translation by elements in C, i.e., if α is a function in I_C, then so is $(\sigma_a\alpha)(x) \stackrel{\text{def}}{=} \alpha(x+a)$

for each $a \in C$. Thus I_C is an ideal in

$$\hat{A}(C) \stackrel{\text{def}}{=} \{\alpha : C \to \mathbb{R}\}$$

when viewed as a ring under convolution [1]

$$(\alpha * \beta)(x) = \sum_y \alpha(y)\beta(x+y).$$

Since C is abelian, the ring $\hat{A}(C)$ decomposes as a direct sum of 1-dimensional ideals. The decomposition is

$$\hat{A}(C) = \bigoplus_{\bar{z} \in \mathbb{Z}_2^n / C^\perp} (\mathcal{X}_{\bar{z}})$$

into 1-dimensional ideals generated by the characters

$$\{\mathcal{X}_{\bar{z}}(x) \stackrel{\text{def}}{=} (-1)^{xz}\}_{\bar{z} \in \mathbb{Z}_2^n / C^\perp}$$

of the abelian group C, where $xz \stackrel{\text{def}}{=} \sum_i x_i z_i$, and the definition is independent of the choice of $z \in \bar{z}$.

If I_C was nonzero, then it must be a direct sum of some of those ideals, so it is sufficient to argue that $\mathcal{X}_{\bar{z}} \notin I_C$, for each $\bar{z} \in \mathbb{Z}_2^n / C^\perp$.

Assume that there is a $\bar{z}_0 \in \mathbb{Z}_2^n / C^\perp$ such that $\mathcal{X}_{\bar{z}_0} \in I_C$, then we have

$$\sum_{x \in C} \mathcal{X}_{\bar{z}+\bar{z}_0}(x) = \sum_{x \in C} \mathcal{X}_{\bar{z}}(x)\mathcal{X}_{\bar{z}_0}(x) = 0,$$

for each z whose weight is at least k, i.e., $\bar{z}_0 + \bar{z} \neq \bar{0}$ for each such z, or equivalently $\bar{z}_0 \neq \bar{z}$ for each such z. Let $Q = \bar{0} \cup \bar{z}_0$ and $C' = Q^\perp$. Then:

[1] Note that $\hat{A}(C)$ is isomorphic as an $\mathbb{R}$-algebra to group algebra

$$\mathbb{R}[C] \stackrel{\text{def}}{=} \left\{ \sum_{z \in C} a_z z : \alpha_{\bar{z}} \in \mathbb{R}, \forall z \in C \right\},$$

i.e., formal sums of elements in C with coefficients in $\mathbb{R}$. The isomorphism maps $\alpha : C \to \mathbb{R}$ to $\sum_z \alpha(z)z$.

1) $C' \subsetneq C$ since $C = \bar{0}^{\perp}$ and $\bar{z}_0 \neq \bar{0}$, and

2) the minimum distance of Q is above k since $\bar{0} = C^{\perp}$ has minimum distance above k, and $\bar{z}_0 \neq \bar{z}$ for each z whose weight is at least k.

This contradicts the assumption that C is minimal with the property that its dual has minimum distance greater than k. ■

We conclude this section with some open questions.

Conjecture 5.3.2 *If an extreme point of Q_k is uniform on its support, then it must be supported by a binary linear code.*

Problem 5.3.3 *It is experimentally evident that in general not all the extreme points of Q_k are uniform on their support. Classify the other extreme points and study their algebraic structure.*

Judging from a small number of machine generated examples, it is tempting to speculate that they come from linear codes over other finite fields by some way of binarizing and assigning weights. Prove or disprove?

5.3.2 The nonnegative Fourier transform property

Let $C \subset \mathbb{Z}_2^n$ be a linear code. Let μ_C be the discrete measure on $\mathbb{Z}_2^n$ supported by C, i.e., $\mu_C(x) = \frac{1}{|C|} 1_C(x)$. Consider the Fourier transform $\widehat{\mu_C}$ of μ_C, i.e., $\widehat{\mu_C}(z) = \frac{1}{2^n} E_{\mu} \mathcal{X}_z$. We have

$$E_{\mu_C} \mathcal{X}_z = \sum_{x \in \mathbb{Z}_2^n} \mu_C(x) \mathcal{X}_z(x) = \frac{1}{|C|} \sum_{x \in C} \mathcal{X}_z(x) = 1_{C^{\perp}}(z), \tag{5.7}$$

where $C^{\perp}$ is the dual of C. So, a very special feature of μ_C is that $\widehat{\mu_C} \geq 0$.

Let $k \geq 0$. We show below where exactly the binary linear codes lie in the convex polytope P_k of k-wise independent probability measures on $\mathbb{Z}_2^n$ whose Fourier transform is nonnegative. We argue that the binary linear codes with dual distance above k are exactly the elements of P_k that are uniform on their support, and exactly

the elements of P_k that are on the boundary of a specific radius-$\frac{1}{2}$ sphere containing P_k and centered at $\frac{1}{2}\mu_{\{0\}}$. Thus they are all extreme points of P_k.

Theorem 5.3.4 *Let:*

- $k \geq 0$
- $P_k \subset \mathbb{R}^{\mathbb{Z}_2^n}$ *be the convex polytope of k-wise independent probability measures μ on $\mathbb{Z}_2^n$ satisfying $\hat{\mu} \geq 0$.*
- L_k *be the set of probability measure on $\mathbb{Z}_2^n$ corresponding to binary linear codes with dual distance above k, i.e.,*

$$L_k = \{\mu_C : C \subset \mathbb{Z}_2^n \text{ linear s.t. min-dist}(C^\perp) \geq k\}.$$

- U *be the set of probability measure on $\mathbb{Z}_2^n$ that are uniform on their support.*
- D *be the radius-$\frac{1}{2}$ sphere in $\mathbb{R}^{\mathbb{Z}_2^n}$ centered at $\frac{1}{2}\mu_0$, where $\mu_0 = 1_{\{0\}}$, i.e., $\mu_0 : \mathbb{Z}_2^n \to \mathbb{R}$ is given by*

$$\mu_0(x) = \begin{cases} 1 & x = 0 \\ 0 & o.w. \end{cases},$$

then

$$L_k = P_k \cap U = P_k \cap \partial D \text{ and } P \subset D,$$

where ∂D means the boundary of D. Thus, in particular, all the elements of L_k are extreme points of P_k.

We will explore in Section 5.8.5 some relaxations based on relaxing L_k to P_k.

Proof. Without loss of generality we can assume that $k = 0$ since the more general case follows by intersecting with the convex polytope Q_k of k-wise independent probability measures on $\mathbb{Z}_2^n$.

First, note that P_0 consists of the set of $\mu : \mathbb{Z}_2^n \to \mathbb{R}$ such that : $\mu \geq 0$, $\sum_x \mu(x) = 1$, and $\sum_x \mu(x)\mathcal{X}_z(x) \geq 0$ for each z in $\mathbb{Z}_2^n$.

The next thing to note is that $\alpha : \mathbb{Z}_2^n \to \mathbb{R}$ belongs to D if and only if

$$\left(\frac{1}{2}\right)^2 \geq \left(\alpha(0) - \frac{1}{2}\right)^2 + \sum_{x \neq 0} \alpha(x)^2,$$

which can be written as $\alpha(0) \geq \sum_x \alpha(x)^2$. Accordingly,

- $\alpha \in D$ if and only if $\alpha(0) \geq \sum_x \alpha(x)^2$, and
- $\alpha \in \partial D$ if and only if $\alpha(0) = \sum_x \alpha(x)^2$.

Let μ be a probability measure on $\mathbb{Z}_2^n$. Then it is sufficient to show that

A) (I) and (II) below are equivalent.

I) There is a linear code $C \subset \mathbb{Z}_2^n$ such that $\mu = \frac{1}{|C|} 1_C$.

II) a) $\hat{\mu} \geq 0$

b) $\mu(0) = \sum_x \mu(x)^2$

B) If (a) holds, i.e., if $\hat{\mu} \geq 0$, then:

1) $\mu(0) \leq \sum_x \mu(x)^2$

2) If μ is uniform on its support, then (b) holds.

The implication from (I) to (II) is immediate. We want to prove the other direction, and establish (B). First, some observations:

i) In general we have

$$2^{-n} \sum_z E_\mu \left(\mathcal{X}_z - (E_\mu \mathcal{X}_z)^2 \right) = \sum_z \hat{\mu}(z) - 2^n \sum_z \hat{\mu}^2(z) = \mu(0) - \sum_x \mu^2(x), \quad (5.8)$$

where we have used Parseval equality $2^n \sum_z \hat{\mu}^2(z) = \sum_x \mu^2(x)$ in the last step.

ii) If μ satisfies (a), then $E_\mu \mathcal{X}_z - (E_\mu \mathcal{X}_z)^2 \geq 0$ for each z.

This is the case because $0 \leq E_\mu \mathcal{X}_z$ by (a), and in general we have $E_\mu \mathcal{X}_z \leq 1$.

Proof of (B): Assume that μ satisfy (a). Then (5.8) and (ii) imply that $0 \leq \mu(0) - \sum_x \mu(x)^2$, and hence (1).

To establish (2), note first that this means that $\mu(0) \neq 0$. Thus the requirement that μ is uniform on its support together with $\mu(0) \neq 0$ imply that, for each x, either $\mu(x) = \mu(0)$ or $\mu(x) = 0$, hence

$$0 = \sum_x \mu(x)(\mu(0) - \mu(x)) = \mu(0) - \sum_x \mu(x)^2,$$

i.e., (b).

(II)⇒ (I): Now back to (A), using (b), we see that the sum on the right hand side of (5.8) is actually 0. This together with (ii) mean that the only possible scenario is that $E_\mu \mathcal{X}_z - (E_\mu \mathcal{X}_z)^2 = 0$, i.e., $E_\mu \mathcal{X}_z \in \{0, 1\}$ for each z.

Let

$$Q = \{z : E_\mu \mathcal{X}_z = 1\} = \{z : \mathcal{X}_z(x) = 1, \forall x \in \text{Support}(\mu)\}.$$

By the multiplicativity of the characters, we have $\mathcal{X}_{z_1+z_2}(x) = 1$ when $\mathcal{X}_{z_1}(x) = 1$ and $\mathcal{X}_{z_2}(x) = 1$, thus Q is linear. Let $C = Q^\perp$, the dual of Q. Then by (5.7), $E_{\mu_C} \mathcal{X}_z = 1_Q(z)$. But since $E_\mu \mathcal{X}_z \in \{0, 1\}$ for each z, we have $1_Q(z) = E_\mu \mathcal{X}_z$ by the definition of Q. Thus for each z, $E_\mu \mathcal{X}_z = E_{\mu_C} \mathcal{X}_z$. In other words $\widehat{\mu_C} = \hat{\mu}$, and hence $\mu = \mu_C$. ■

Here again

Problem 5.3.5 *It is experimentally evident that in general not all the extreme points of P_k or P are uniform on their support. Classify the other extreme points and study their algebraic structure.*

One way to look at Delsarte LP coding bound [Del73] in the special case of linear codes is as relaxing L_k to P_k. See Lemma 5.8.19. Since experimentally not all the extreme points of P_k come from linear codes, this suggests the following

Question 5.3.6 *Is Delsarte LP coding bound tight for linear codes?*

The answer is not clear. Note that whether Delsarte LP coding bound is tight for arbitrary binary code is open also. It is an old and famous open problem.

5.4 Some limitations of the small bias property

We show that the exponentially small bias property is not sufficient to fool small log-depth circuits, nor bounded-space computations as it cannot fool some of the weakest branching programs. More specifically

Theorem 5.4.1 *There is a $2^{-\Omega(n)}$-biased probability measure on $\{0,1\}^n$ that cannot $o(1)$-fool a function from $\{0,1\}^n$ to $\{0,1\}$ that can be realized as an $O(\log n)$-depth circuit of linear size, and as an $O(1)$-width read-once oblivious branching program.*

Proof. Assume n is even, and let $f : \{0,1\}^n \to \{0,1\}$ be the binary quadratic form

$$f(x) = \sum_{i=1}^{n/2} x_i x_{i+n/2} \pmod 2.$$

Thus f is the XOR of $\frac{n}{2}$ AND gates, hence computable by $O(\log n)$-depth circuit, and obviously computable by an $O(1)$-width read-once oblivious branching program. Let μ be the probability measure on $\{0,1\}^n$ given by

$$\mu(x) = \frac{1}{|\Omega|} f(x), \text{ where } \Omega = \{x : f(x) = 1\}.$$

We will argue that μ is $(1 - o(1))2^{-n/2}$-biased.

Since

$$E_\mu \mathcal{X}_z = 2^n \hat{\mu}(z) = \frac{2^n}{|\Omega|} \hat{f}(z) = \frac{\hat{f}(z)}{\hat{f}(0)},$$

we need to compute the Fourier transform of f. It is more natural to deal with $1 - 2f = (-1)^f$. Indeed, $(-1)^f$ is an eigenfunction of the Fourier transform with eigenvalue $2^{-n/2}$, i.e.,

$$\widehat{(-1)^f} = \frac{1}{2^{n/2}} (-1)^f. \tag{5.9}$$

One way to verify this is to note that, by Lemma 5.2.9, we only have to check it when $n = 2$, i.e., when $f(x) = x_1 x_2$, in which case (5.9) is the identity

$$(-1)^{x_1 x_2} = \frac{1}{2} + \frac{1}{2}(-1)^{x_1} + \frac{1}{2}(-1)^{x_2} - \frac{1}{2}(-1)^{x_1 + x_2}.$$

Thus, since

$$\hat{f} = \frac{1}{2}\widehat{1} - \frac{1}{2}\widehat{(-1)^f},$$

we get

$$\hat{f}(z) = \begin{cases} \frac{1}{2} + \frac{1}{2^{n/2+1}} & \text{when } z = 0 \\ (-1)^{f(z)}\frac{1}{2^{n/2+1}} & \text{when } z \neq 0 \end{cases}.$$

Therefore when z is nonzero

$$E_\mu \mathcal{X}_z = \frac{1}{2^{n/2}} \frac{(-1)^{f(z)}}{1 + 2^{-n/2}},$$

and consequently $|E_\mu \mathcal{X}_z| = 2^{-n/2}(1 - o(1))$. ■

Note that the measure constructed in the proof is uniform on its support Ω whose size is $2^{n-1}(1 + o(1))$. Note also that its bias δ can be shown to be optimal up to a $1 \pm o(1)$ factor in the class of measures that have the same support size and are uniform on their support. By picking random subsets of Ω, other support sizes can be achieved with very low bias also.

5.5 Log-wise independence versus read-once DNF formulas

We argue that any sufficiently log-wise independent probability measure looks random to all polynomially small read-once DNF formulas, more specifically

Theorem 5.5.1 *Let $f : \{0,1\}^n \to \{0,1\}$ be computable by a read-once (i.e., the clauses are disjoint) DNF formula with m clauses. Then any k-wise independent probability measure μ on $\{0,1\}^n$ can ϵ-fool f, with*

$$\epsilon = \min_{1 \leq s \leq n} 2^{-(s - \log m)} + 2^{-\frac{k-37}{11s}}.$$

So $\epsilon = o(1)$, when for instance $k = \log m \log\log m$.

Proof. See Corollary 5.5.12. ■

After using the sandwiching approach of Corollary 5.2.6 to reduce to a simpler case, the proof of Theorem 5.5.1 is partially by inclusion-exclusion, which is the argument we used to establish parts (d),(f), and (g) of Lemma 5.2.3. We discuss in Section 5.5.3 the intrinsic limitations of Inclusion-exclusion when used in the more general setting of arbitrary DNF formulas.

Starting with an arbitrary DNF formulas, we conclude Corollary 5.5.12 as a consequence of a more general result on weak probability measures in Lemma 5.5.10. The notion of weak probability measures is naturally suggested by the problem. See Definition 5.5.4.

Before going to weak probability measures, we derive a consequence of Theorem 5.5.1 and we give an application to the distribution of quadratic residues.

Corollary 5.5.2 *Let $f : \{0,1\}^n \to \{0,1\}$ be computable by a read-once DNF formula with m clauses, then any δ-almost k-wise independent probability measure μ on $\{0,1\}^n$ can ϵ-fool f, with*

$$\epsilon = \min_{1 \le s \le n} 2^{-(s - \log m)} + 2^{-\frac{k-37}{11s}} + \delta n^k.$$

Note that the bound is good only when k is relatively small.

Proof. This follows from Theorem 5.5.1 via part (b) in Theorem 5.1.3. ∎

Corollary 5.5.3 *Let $S_1, \ldots, S_m$ be disjoint subsets of $\mathbb{F}_q$, q a power of an odd prime. Let Q be the set of quadratic residues in $\mathbb{F}_q$. Let $n = |\cup_i S_i|$. Then*

$$\left| \frac{1}{q} \#\{a \in \mathbb{F}_q : S_i + a \not\subset Q, \forall i\} - \prod_i (1 - 2^{-|S_i|}) \right| \le \epsilon,$$

where

$$\epsilon = \min_{1 \le s,k \le n} 2^{-(s - \log m)} + 2^{-\frac{k-37}{11s}} + \frac{2kn^k}{\sqrt{q}}.$$

Thus $\epsilon = o(1)$, when for instance $q = 2^{2 \log n \log m}$.

Proof. Let $I = \cup_i S_i$. Consider the monotone read-once DNF on the variables $\{x_i\}_{i \in I}$ consisting of m clauses each corresponding to an S_i, and the quadratic residues PRG

$G : \mathbb{F}_q \to \{0,1\}^I$ as defined in Section 5.1.2.3. The claim follows from Corollary 5.5.2 since G has the $2k/\sqrt{q}$-almost k-wise independence property. ■

Similar statements can be derived for the quadratic-residues-like PRG's defined in Section 5.1.2.3.

5.5.1 Weak probability measures

Start with an arbitrary DNF formulas. We will restrict later on to the read-once case. Consider a DNF formula f on n bits with m clauses. Let μ be a k-wise independent probability measure on $\{0,1\}^n$. We want to make k large enough so that μ can ϵ-fool f.

First we note that, without loss of generality, we can assume that each clause has size at most s as long as we remember to add $m2^{-s} = 2^{-(s-\log m)}$ to ϵ. Here s is an integer that we will tune later on. In other words, if we can show that μ can ϵ-fool f under this restriction, then μ can $(\epsilon + 2^{-(s-\log m)})$-fool f without this restriction.

The reason is that if not all the clauses have size at most s, we can construct two new DNF formulas f' and f'', where f' is constructed by removing from f all the clauses whose size is above s, and f'' is constructed by removing from each clause in f whose size is above s as many variables as needed in an arbitrary way to make its size s. Thus $f' \leq f \leq f''$ and $E_{\mu_0}(f'' - f), E_{\mu_0}(f - f'') \leq m2^{-s}$, where μ_0 is the uniform measure on $\{0,1\}^n$. So establishing the claim for f' and for f'' implies the claim for f with an $m2^{-s}$ additive term to ϵ. This is the case because, with Corollary 5.2.4 in mind, we can sandwich f between the upper polynomial of f'' and the lower polynomial of f'.

Under this assumption, consider the map $F : \{0,1\}^n \to \{0,1\}^m$, where $F_t(x)$ is the value of the t'th clause on x. Let μ_0^* be the probability measure on $\{0,1\}^m$ induced by the uniform measure μ_0 on $\{0,1\}^n$ via F, and let μ^* be the one induced by μ. Since μ is k-wise independent, $\mu_A^* = \mu_{0A}^*$ for each $A \subset [m]$ such that $|A| \leq \frac{k}{s}$, where by μ_A we mean the probability measure on $\{0,1\}^A$ induced by μ

This suggests the following problem.

Definition 5.5.4 *Let γ be a probability measure on $\{0,1\}^m$. Say that γ is (k,ϵ)-weak if when γ' is another probability measure on $\{0,1\}^m$ that agrees with γ on all its projection on any k of the coordinates, then the L_∞-distance between γ and γ' is at most ϵ.*

In other words, γ is (k,ϵ)-weak if when γ' is another probability measure on $\{0,1\}^m$ such that $\gamma'_A = \gamma_A$ for each $A \subset [m]$ such that $|A| \leq k$, we must have $|\gamma(x) - \gamma'(x)| \leq \epsilon$, for each $x \in \{0,1\}^m$. Here by γ_A we mean the probability measure induced on $\{0,1\}^A$ by γ.

Problem 5.5.5 *Let $F : \{0,1\}^n \to \{0,1\}^m$, $x \mapsto y$, be such that each y_i is an AND on some of the variables $x_1, \ldots, x_n$ with possibly negated inputs. Let γ be the probability measure on $\{0,1\}^m$ induced by the uniform measure on $\{0,1\}^n$. Given ϵ, how large should k be so that γ is (k,ϵ)-weak for each such F?*

We will not resolve this problem for arbitrary DNF formulas. We will obtain a good bound in the read-once case, which corresponds to the situation when the induced random variables are statistically independent.

Before doing so, we make some observations about weak probability measures.

Theorem 5.5.6 [LN90] *Let $U_1, \ldots, U_m, V_1, \ldots, V_m$ be finite sets such that $|\cap_{i\in A} U_i| = |\cap_{i\in A} V_i|$ for each $A \subset [m]$, $|A| \leq k$, then*

$$\frac{|\cup_{i=1}^m U_i|}{|\cup_{i=1}^m V_i|} = 1 + O(e^{-\frac{2k}{\sqrt{m}}}),$$

when $k = \Omega(\sqrt{n})$, and

$$\frac{|\cup_{i=1}^m U_i|}{|\cup_{i=1}^m V_i|} = O\left(\frac{m}{k^2}\right),$$

when $k = O(\sqrt{m})$, for some global absolute constants. Moreover, the bound is tight in the worst case when $k \leq \sqrt{m}$.

Corollary 5.5.7 *Any probability measure on $\{0,1\}^m$ is $(k, O(e^{-\frac{2k}{\sqrt{m}}}))$-weak, for some global absolute constants. Moreover, the bound is tight in the sense that there is a probability measure on $\{0,1\}^m$ that is not $(\sqrt{m}, o(1))$-weak.*

Lemma 5.5.8 *A probability measure γ on $\{0,1\}^m$ is (k,ϵ)-weak if and only if $\forall x_0 \in \{0,1\}^m$, $\exists f_1, f_2 : \{0,1\}^m \to \mathbb{R}$ such that with*

$$\delta_{x_0}(x) = \begin{cases} 1 & \text{if } x = x_0 \\ 0 & \text{o.w.} \end{cases},$$

we have

a) $deg(f_1), deg(f_2) \leq k$

b) $f_1 \leq \delta_{x_0} \leq f_2$

c) $E_\gamma(f_2 - \delta_{x_0}), E_\gamma(\delta_{x_0} - f_1) \leq \epsilon$

Proof. By linear-programming duality, see Section 5.2.3 for the calculations. Note that the two primals are: max $E_\gamma \delta_{x_0} - E_{\gamma'} \delta_{x_0}$ and max $E_{\gamma'} \delta_{x_0} - E_\gamma \delta_{x_0}$, where we are optimizing on the probability measures γ' on $\mathbb{Z}_2^n$ such that $E_{\gamma'} \mathcal{X}_z = E_\gamma \mathcal{X}_z$ for each nonzero z in $\mathbb{Z}_2^n$ whose weight is at most k. ■

Lemma 5.5.9 *If γ is a probability measure on $\{0,1\}^m$ supported by a set of size N, and*

$$k \geq 2\frac{\log(N+1)}{\log \frac{m}{k}} + 3,$$

then γ is $(k,0)$-weak.

Proof. By suitably negating some of the variables if needed we can assume without loss of generality that $x_0 = 0$. Let $g : \mathbb{Z}_2^m \to \mathbb{R}$ be a polynomial of degree k_0 such that $g(0) = 1$ and $g(x) = 0$ for each nonzero x in the support of γ. Since all the characters of $\mathbb{Z}_2^m$ are linearly independent, and specifically the $V \stackrel{\text{def}}{=} \sum_{l=0}^{k_0} \binom{m}{l}$ characters $\mathcal{X}_z$ where the weight of z is at most k_0, we can set k_0 to the smallest integer satisfying $V \geq N+1$. Then, set f_1 and f_2 to $f_1(x) = (1 - \sum_{i=1}^m x_i) g^2(x)$ and $f_2 = g^2$. So (b) is satisfied. Regarding (c), f_1 and f_2 are zero on any nonzero x in the support of γ which means that $E_\gamma f_1 = E_\gamma f_2 = E_\gamma \delta_0$. As for (a), we have $k = 2k_0 + 1$. So it sufficient to make k large enough so that $\sum_{l=0}^{\lfloor \frac{k-1}{2} \rfloor} \binom{m}{l} \geq N+1$. ■

Now, we come to the main result in this section.

Lemma 5.5.10 *If $X_1, \ldots, X_m$ are statistically independent binary random variables, then the corresponding probability measure γ on $\{0,1\}^m$ is $(k, 2^{-\frac{k-37}{11}})$-weak.*

Proof. See Section 5.5.2. ■

Problem 5.5.11 *The bound is quite good compared to the general case. The constants are definitely not tight. What is the best exponent? Also, is this bound asymptotically tight in terms of k for small values of k such as $k = \log^{O(1)} m$ or $k = m^{o(1)}$?*

Corollary 5.5.12 *Let $f : \{0,1\}^n \to \{0,1\}$ be computable by a read-once (i.e., the clauses are disjoint) DNF formula with m clauses. Then any k-wise independent probability measure μ on $\{0,1\}^m$ can ϵ-fool f, with*

$$\epsilon = \min_{1 \le s \le n} 2^{-(s - \log m)} + 2^{-\frac{k-37}{11s}}.$$

So $\epsilon = o(1)$, when for instance $k = \log m \log\log m$.

Proof. As we noted earlier in this section we can assume without loss of generality that each clause has size at most s as long we remember to add $m2^{-s} = 2^{-(s-\log m)}$ to the error.

Under this assumption, consider the map $F : \{0,1\}^n \to \{0,1\}^m$, where $F_t(x)$ is the value of the t'th clause on x. Let μ_0^* be the probability measure on $\{0,1\}^m$ induced by the uniform measure μ_0 on $\{0,1\}^n$ via F, and let μ^* be the one induced by μ. Since μ is k-wise independent, $\mu^*|_A = \mu_0^*|_A$ for each $A \subset [m]$ such that $|A| \le \frac{k}{s}$. The claim then follows from Lemma 5.5.10 which is applicable since $X_1, \ldots, X_m$ are independent because the clauses are disjoint. ■

5.5.2 Proof of Lemma 5.5.10

Without loss of generality, we can assume that $x_0 = 0$. Consider the weight function $w : \mathbb{Z}_2^m \to [0 : n]$ given by $w(x) = \sum_{i=1}^m x_i$. Since the X_i's are independent, the distribution of $w(X_1, \ldots, X_n)$ has the shape of a bell. Let γ^* be the corresponding

probability measure, i.e., the one induced on $[0:n]$ by μ via the weight map w. It is sufficient to trap the delta function $\delta'_0 : [0:m] \to \{0,1\}$,

$$\delta'_0(w) = \begin{cases} 1 & \text{if } w = 0 \\ 0 & \text{o.w.} \end{cases},$$

between two low-degree polynomials $f'_1, f'_2 \in \mathbb{R}[w]$, i.e., $f'_1 \leq \delta'_0 \leq f'_2$, in such a way that $E_{\gamma^*}(f'_2 - \delta'_0), E_{\gamma^*}(\delta'_0 - f'_1) \leq \epsilon$. Then we can pull back f'_1 and f'_2 by w to construct f_1 and f_2, i.e., we set $f_1(x) = f'_1(w(x))$ and $f_2(x) = f'_2(w(x))$. Note that this preserves the degrees, i.e., $deg f_1 = deg f'_1$ and $deg f_2 = deg f'_2$.

The natural thing to do is to select the zeros of f'_1 and f''_2 so that they are at the places where γ^*_0 is large. This is indeed the only thing to do, given that we want to work in the projected framework, regardless of whether the X_i's are independent or not. What is special about the case we are considering now is that we can do this by exploiting the strong structure of independent Bernoulli random variables. Naturally, there are two cases to consider depending on the shape of the bell $\gamma^*(w)$, the first when the mean $u \stackrel{\text{def}}{=} E_{w \sim \gamma^*_0} w$ is large, and the second when the mean is small. Let $s > 0$ be some threshold that we will specify as we proceed.

- **Case 1:** When the mean $u > s$.

  Set

  $$f'_1(w) = 0 \text{ and } f'_2(w) = \left(1 - \frac{w}{u}\right)^{k'}$$

  where $k' \leq k$ is an even integer. We will restrict k' further and fix it as we proceed. This choice is experimentally optimal when u is sufficiently large. It works in general because, when the mean is large, 0 is in the tail of the bell, which means that $E_{\gamma^*}\delta'_0$ is already small. On the other hand f'_2 takes very small values on a relatively wide region around the center of the bell, which makes $E_{\gamma^*}(f'_2 - \delta_0)$ small.

  Let $0 < \delta < 1$. We will fix δ as we proceed. We have

  $$E_{\gamma^*}(f'_2 - f'_1) \quad \leq \quad Pr_{w \sim \gamma^*_0}[w \leq (1-\delta)u] + \delta^{k'} + Pr_{w \sim \gamma^*_0}[(1+\delta)u \leq w \leq 2u]$$

$$
\begin{aligned}
&+ \sum_{u \le d \le m-u} Pr_{w \sim \gamma_0^*}[w = u + d] \left(\frac{d}{u}\right)^{k'} \\
\le\ & Pr_{w \sim \gamma_0^*}[w \le (1-\delta)u] + \delta^{k'} + Pr_{w \sim \gamma_0^*}[w \ge (1+\delta)u] \\
&+ \sum_{u \le d \le m-u} Pr_{w \sim \gamma_0^*}\left[w \ge \left(1 + \frac{d}{u}\right) u\right] \left(\frac{d}{u}\right)^{k'}.
\end{aligned}
$$

Now, we use Chernoff bound[2] to obtain

$$
\begin{aligned}
E_{\gamma^*}(f_2' - f_1') \ \le\ & \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^u + \delta^{k'} + \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^u \\
&+ \sum_{u \le d \le m-u} \left(\frac{e^{\frac{d}{u}}}{(1+\frac{d}{u})^{1+\frac{d}{u}}}\right)^u \left(\frac{d}{u}\right)^{k'}.
\end{aligned} \tag{5.10}
$$

Assume that $k'/u \le \theta$, where $\theta < 1/2$ is a parameter that we specify later. Noting that $\frac{e^x}{(1+x)^{1+x}} \le 2^{-(\frac{x}{3}+\frac{1}{5})}$ for all $x \ge 1$, and that $x \le 2^{\frac{2}{3}x}$ for all $x \ge 0$, we can upper bound the last summation as follows.

$$
\begin{aligned}
\sum_{u \le d \le m-u} \left(\frac{e^{\frac{d}{u}}}{(1+\frac{d}{u})^{1+\frac{d}{u}}}\right)^u \left(\frac{d}{u}\right)^{k'} &\le \sum_{u \le d \le m-u} \left(\frac{e^{\frac{d}{u}}}{(1+\frac{d}{u})^{1+\frac{d}{u}}}\right)^u \left(\frac{d}{u}\right)^{\theta u} \\
&\le \sum_{u \le d \le m-u} 2^{-(\frac{1}{3}\frac{d}{u}+\frac{1}{5})u} 2^{\frac{2}{3}\frac{d}{u}\theta u} \\
&= 2^{-\frac{u}{5}} \sum_{u \le d \le m-u} 2^{-\frac{2}{3}(\frac{1}{2}-\theta)d} \\
&\le \frac{2^{-(\frac{2}{3}(\frac{1}{2}-\theta)+\frac{1}{5})u}}{1 - 2^{-\frac{2}{3}(\frac{1}{2}-\theta)}}.
\end{aligned} \tag{5.11}
$$

All the terms in (5.10) and (5.11) are nonincreasing as u grows. Thus we can use the lower bound s we have on u. Moreover the condition $k'/u \le \theta$ can be guaranteed by requiring that $k' \le \theta s$. The other conditions we have on k' are: $k' \le k$, and k' odd. If we assume that $\theta s \le k$, we can set k' to the largest odd integer less than or equal to θs, and use $\theta s - 2$ as an upper bound on k'. It

[2]The version of Chernoff bound we are using is the following (See [MR95]): If $z_1, \ldots, z_m$ are independent Poisson trials, and $z = \sum_i z_i$, then, for each $0 < \delta < 1$, the probability that $z < (1-\delta)Ez$ is at most $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{Ez} \le e^{-(Ez)\delta^2/2}$. On the other hand, for each $\alpha > 0$, the probability that $z > (1-\alpha)Ez$ is at most $\left(\frac{e^{\alpha}}{(1+\alpha)^{1+\alpha}}\right)^{Ez}$.

follows that

$$E_{\gamma^*}(f_2' - f_1') \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^s + \delta^{\theta s - 2} + \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^s + \frac{2^{-(\frac{2}{3}(\frac{1}{2}-\theta)+\frac{1}{5})s}}{1 - 2^{-\frac{2}{3}(\frac{1}{2}-\theta)}}, \tag{5.12}$$

for all settings of reals δ, θ satisfying: $0 < \delta < 1$, $0 < \theta < 1/2$, and $\theta s \leq k$.

- **Case 2:** When the mean $u \leq s$.

  When the mean is sufficiently small, the above choice is not good neither for f_1' nor for f_2'. In this case, judging from simulations, the right choice is to construct f_1' and f_2' so that they have relatively many distinct zeros distributed relatively sparsely, but not consecutively, and all close to zero. It is not clear however what is the optimal distribution of the zeros. To establish the claimed bound, we use a suboptimal choice that we can analyze. The first thing to try is to set

$$f_1'(w) = \prod_{t=1}^{k}\left(1 - \frac{w}{t}\right) \text{ and } f_2'(w) = \prod_{t=1}^{k-1}\left(1 - \frac{w}{t}\right)$$

  when k is odd. Note that since k is odd we have $f_1' \leq \delta_0' \leq f_2'$. This will give us the claimed bound, which seems asymptotically optimal.

  Note that this choice of f_1' and f_2' is equivalent to truncating the binomial representation of δ_0' in the sense that $\delta_0'(w) = \sum_{l=0}^{m}(-1)^l\binom{w}{l}$, $f_1'(w) = \sum_{l=0}^{k}(-1)^l\binom{w}{l}$, and $f_2'(w) = \sum_{l=0}^{k-1}(-1)^l\binom{w}{l}$. This is equivalent also to truncating the inclusion-exclusion formula (5.6) as in Section 5.2.2. Since we are doing a plain truncation, it makes sense that in general when we have dependencies among the random variables this construction is not going to work. This is indeed provably the case as we explain in Section 5.5.3.

  For the purpose of this proof, it is sufficient to note that

$$f_2'(w) - f_1'(w) = \frac{w}{k}\prod_{t=1}^{k-1}\left(1 - \frac{w}{t}\right) = \binom{w}{k}.$$

Thus by the independence of $X_1, \ldots, X_m$,

$$E_{\gamma^*}(f_2' - f_1') = \sum_{A \subset [m]; |A|=k} \prod_{i \in A} p_i,$$

where p_i is the probability that $X_i = 1$. Recall that we are in the situation where $E_{w \sim \gamma^*} w \leq s$, i.e., $\sum_{i=1}^m p_i \leq s$.

Claim: The maximum of

$$\sum_{A \subset [m]; |A|=k} \prod_{i \in A} p_i,$$

over $p_1, \ldots, p_m \geq 0$ such that $\sum_{i=1}^m p_i \leq s$, occurs when all the p_i's are equal.

Proof of claim. We use a local perturbation argument based on the symmetry of the objective function. Note that the objective function is not convex.

Consider any $p_1, \ldots, p_m$ satisfying the constraints, and any $i \neq j$. Assume that $p_i \neq p_j$. We will make $p_i = p_j$ while keeping $c \stackrel{\text{def}}{=} p_i + p_j$ and $(p_t)_{t \neq i,j}$ unchanged, and while increasing the value of the objective function. To do this it is sufficient to note that

$$\sum_{A \subset [m]; |A|=k} \prod_{t \in A} p_t = a(p_i + p_j) + b p_i p_j = ac + b p_i p_j,$$

where

$$a = \sum_{A \subset [m] \setminus \{i,j\}; |A|=k} \prod_{t \in A} p_t, \text{ and } b = \sum_{A \subset [m] \setminus \{i,j\}; |A|=k-1} \prod_{t \in A} p_t.$$

The claim then follows since $b \geq 0$, and the maximum of $p_i p_j$ subject to $p_i, p_j \geq 0$ and $p_i + p_j = c$ occurs when $p_i = p_j$. ▼

Thus

$$E_{\gamma^*}(f_2' - f_1') \leq \binom{m}{k} \left(\frac{s}{m}\right)^k \leq \left(\frac{em}{k}\right)^k \left(\frac{s}{m}\right)^k = 2^{-k \log \frac{k}{es}}.$$

This is assuming that k is odd. Thus, in general, for any integer $k \geq 2$, we can construct f_1' and f_2' with

$$E_{\gamma^*}(f_2' - f_1') \leq 2^{-(k-1) \log \frac{(k-1)}{es}}. \tag{5.13}$$

Combining (5.12) and (5.13), we get that in general, if $k \geq 2$ is an integer, then γ_0^* is $(k, \epsilon(k))$-weak, where

$$\begin{aligned}\epsilon(k) &= \min_{s,\delta,\theta} \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^s + \delta^{\theta s - 2} + \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^s \\ &\quad + \frac{2^{-(\frac{2}{3}(\frac{1}{2}-\theta)+\frac{1}{5})s}}{1-2^{-\frac{2}{3}(\frac{1}{2}-\theta)}} + 2^{-(k-1)\log\frac{(k-1)}{es}}, \end{aligned} \tag{5.14}$$

and we are optimizing on the reals s, δ, and θ satisfying: $s > 0$, $0 < \delta < 1$, $0 < \theta < 1/2$, and $\theta s \leq k$.

Setting $\theta = \frac{12}{25}$, $\delta = \frac{2}{3}$, and $s = \frac{k-1}{2^{1/10}e}$, we obtain

$$\begin{aligned}\epsilon(k) &< \left(\frac{e^{-\frac{2}{3}}}{(\frac{1}{3})^{\frac{1}{3}}}\right)^{\frac{k-1}{2^{1/10}e}} + \left(\frac{2}{3}\right)^{\frac{12}{25}\frac{k-1}{2^{1/10}e}-2} + \left(\frac{e^{\frac{2}{3}}}{(\frac{4}{3})^{\frac{4}{3}}}\right)^{\frac{k-1}{2^{1/10}e}} + \frac{2^{-(\frac{1}{75}+\frac{1}{5})\frac{k-1}{2^{1/10}e}}}{1-2^{-\frac{1}{75}}} + 2^{-\frac{k-1}{10}} \\ &< 2^{-\frac{k-37}{11}}, \end{aligned}$$

where the last inequality holds for all values of k where the claimed bound is nontrivial, i.e., for all $k > 37$. Note that the above choice of s, δ, and θ is not optimal, but it is sufficient to get a concrete bound.

5.5.3 The intrinsic limitations of Inclusion-Exclusion in the DNF case

Consider a DNF formula with m clauses $X_1, \ldots, X_m$ on n variables. The t'th term of the inclusion-exclusion formula is

$$T_t = \sum_{A \subset [m]; |A| = t} E \prod_{i \in A} X_i.$$

See the proof of Lemma 5.5.10 for other interpretations of this term in terms of the sandwiching polynomials in the projected setting. In general T_t does not vanish as t increases.

One trivial instance is when all the clauses are the same. This can however be resolved by excluding all but one of the clauses.

Another trivial case is when we have a very large number of clauses. This case can be resolved by removing as may clauses as needed to get a new DNF formula that bounds the original one from below, and use the constant 1 function to upper bound the old DNF formula.

Less trivial and hybrid cases that make T_t large are for instance when we have many clauses with some common variables (possibly none) such that, given that the common variables are correctly set, the probability that none of the clauses containing these variables is satisfied is very small (e.g. sunflowers with many pedals). Such cases can be also resolved. The motivation is the Razborov sunflowers trimming technique [Raz85]. Indeed, we can tightly sandwich the DNF between two new DNF's where such cases do not occur.

There is however an intrinsic limitation to this attempt. One example where nothing can be done to the DNF formula to make T_t converge is the following symmetric case.

Consider the case when we have $m = n$ clauses, the first on the variables $y_1, \ldots, y_s$ and the others are cyclic shifts of this clause modulo n. If we suitably select s in the order of $\Theta(\log n)$, so that the probability that the DNF is 1 is $\Theta(1)$, we get a symmetric situation where T_t diverges and no trick can be made to make it converge.

This example can however be resolved by other means not based on inclusion-exclusion. One trivial way to resolve it is to partition the clauses into $\Theta(\log n)$ collection where they are disjoint. This gives us an $O(\log^3 n)$ bound on the needed to k so that the k-wise independence property can fool this DNF.

5.6 Limited independence versus weight probability

We establish an $O(k^{-1/2})$ sharp upper bound on the probability that a random binary string generated according to a k-wise independent measure has any given weight.

Theorem 5.6.1 "Limited independence versus weight probability lemma": *Let* $2 \leq k \leq n$. *Let* μ *be a* k*-wise independent probability measure on* $\{0,1\}^n$. *Then*

$$\begin{aligned} \max_{a=0,1,\ldots,n} Pr_{x\sim\mu}[w(x)=a] &\leq \frac{1}{\sum_{0\leq l \ even \ \leq \lfloor k/2 \rfloor -1} \frac{1}{2^l}\binom{l}{l/2}} \\ &= \frac{\sqrt{\pi+o(1)}}{\sqrt{k}}, \end{aligned}$$

where the asymptotic statement is in terms of the growth of k.

Note that the bound is experimentally quite good for low values of k. Moreover, in the extreme case when k is very large it is also very good. For instance, when $k=n$, it is only off by the constant factor $\frac{\pi}{\sqrt{2}}(1+o(1))$ from the actual value $\sqrt{\frac{2}{\pi n}}(1-o(1))$.

We will give an application of this bound in Section 5.7.3.

First some direct consequences, and an application to the distribution of quadratic residues.

Corollary 5.6.2 *Let* $C \subset \{0,1\}^n$ *be a linear code whose minimum dual distance is above* k, $2 \leq k \leq n$. *Then*

$$\begin{aligned} \max_{a=0,1,\ldots,n} Pr_{x\in C}[w(x)=a] &\leq \frac{1}{\sum_{0\leq l \ even \ \leq \lfloor k/2 \rfloor -1} \frac{1}{2^l}\binom{l}{l/2}} \\ &= \frac{\sqrt{\pi+o(1)}}{\sqrt{k}}. \end{aligned}$$

Corollary 5.6.3 *Let* $2 \leq k \leq n$. *Let* μ *be a* δ*-almost* k*-wise independent probability measure on* $\{0,1\}^n$. *Then*

$$\begin{aligned} \max_{a=0,1,\ldots,n} Pr_{x\sim\mu}[w(x)=a] &\leq \frac{1}{\sum_{0\leq l \ even \ \leq \lfloor k/2 \rfloor -1} \frac{1}{2^l}\binom{l}{l/2}} + \delta n^k \\ &= \frac{\sqrt{\pi+o(1)}}{\sqrt{k}} + \delta n^k. \end{aligned}$$

Note that the bound is good only when k is relatively small.

Proof. This follows from Theorem 5.6.1 via part (b) in Theorem 5.1.3. ■

Corollary 5.6.4 *Let q be a power of an odd prime, Q the set of quadratic residues in $\mathbb{F}_q$, I any subset of $\mathbb{F}_q$, and a any integer. Let $n = |I|$. Then*

$$\frac{1}{q}\#\{b \in \mathbb{F}_q : |(I+b) \cap Q| = a\} \leq \epsilon$$

where

$$\epsilon = \min_{4 \leq k \leq n} \frac{1}{\sum_{0 \leq l \ even \ \leq \lfloor k/2 \rfloor - 1} \frac{1}{2^l}\binom{l}{l/2}} + \frac{2k}{\sqrt{q}} n^k.$$

Thus when for instance $q = 2^{\log^2 n}$, we get $\epsilon \leq \sqrt{\frac{3\pi}{\log n}}$ when n is large enough.

Proof. Consider the quadratic residues PRG $G : \mathbb{F}_q \to \{0,1\}^I$ as defined in Section 5.1.2.3. The claim follows from Corollary 5.6.3 since G has the $2k/\sqrt{q}$-almost k-wise independence property. ■

Similar statements can be derived for the quadratic-residues-like PRG's defined in Section 5.1.2.3.

5.6.1 Proof of Theorem 5.6.1

The proof is based on Krawtchouk polynomials $\{\mathcal{K}_l^{(n)}(w)\}_{l=0}^n$,

$$\mathcal{K}_l^{(n)}(x) \stackrel{\text{def}}{=} \sum_{z \in \mathbb{Z}_2^n; w(z)=l} \mathcal{X}_z(x) = \sum_{i=0}^{l} (-1)^i \binom{w}{i}\binom{n-w}{n-i} \stackrel{\text{def}}{=} \mathcal{K}_l^{(n)}(w), w = w(x).$$

Note that the value of $\mathcal{K}_l^{(n)} : \mathbb{Z}_2^n \to \mathbb{R}$ on $x \in \mathbb{Z}_2^n$ depends only on the weight $w(x)$ of x, thus by abuse of notation $\mathcal{K}_l^{(n)} \in \mathbb{R}[w]$, $deg(\mathcal{K}_l^{(n)}) = l$. We compile first some elementary properties of Krawtchouk polynomials that we will be using. See for instance [Sze75, MRRW77].

- Let $\beta^{(n)}$ is the binomial measure on $[0:n]$, i.e

$$\beta^{(n)}(w) = \frac{1}{2^n}\binom{n}{w},$$

and let

$$\mathfrak{p}_l^{(n)} = \frac{\mathcal{K}_l^{(n)}}{\sqrt{\binom{n}{l}}},\ l = 0, \ldots, n.$$

Then $\{\mathfrak{p}_l^{(n)}\}_l$ are orthonormal w.r.t $\beta^{(n)}$, i.e., $E_{\beta^{(n)}}\mathfrak{p}_l^{(n)}\mathfrak{p}_s^{(n)} = \delta_{l,s}$. Note that this follows immediately from the orthogonality of the characters $\{\mathcal{X}_z\}_z$ of $\mathbb{Z}_2^n$.

- Being orthogonal, they satisfy a recurrence relation that takes the form

$$(l+1)\mathcal{K}_{l+1}^{(n)}(w) - (n-2w)\mathcal{K}_l^{(n)}(w) + (n-(l-1))\mathcal{K}_{l-1}^{n}(w) = 0, \tag{5.15}$$

for $l \geq 1$. Moreover, we have

$$\mathcal{K}_0^{(n)}(w) = 1 \tag{5.16}$$

$$\mathcal{K}_1^{(n)}(w) = n - 2w. \tag{5.17}$$

Lemma 5.6.5 *Let μ be a k-wise independent probability measure on $\mathbb{Z}_2^n$, $k \leq n$. Let $a \in \{0, 1, \ldots, n\}$. Then*

$$Pr_{x\sim\mu}[w(x) = a] \leq \frac{1}{S_{\lfloor k/2 \rfloor}^{(n)}(a)},$$

where

$$S_t^{(n)}(a) \stackrel{\text{def}}{=} \sum_{0\leq l\leq t} \frac{\mathcal{K}_l^{(n)}(a)^2}{\binom{n}{l}},\ t = 0, \ldots, n.$$

Proof. We have a linear program:

$$\max_{\mu} Pr_{x\sim\mu}[w(x) = a],$$

where we are maximizing over the k-wise independent probability measures μ on $\mathbb{Z}_2^n$. The dual is

$$\min_{g} E_{\mu_0} g,$$

where μ_0 is the uniform measure on $\mathbb{Z}_2^n$, and where we are minimizing over the polynomials $g : \mathbb{Z}_2^n \to \mathbb{R}$ such that $\deg(g) \leq k$, $g \geq 0$, and $g(x) \geq 1, \forall x$ such that $w(x) = a$. Taking advantage of the problem symmetry, we see that the dual is equivalent to the projected dual:

$$\min_f E_{\beta^{(n)}} f,$$

were we are minimizing over the polynomials $f \in \mathbb{R}[w]$ such that $\deg(f) \leq k$, $f \geq 0$, and $f(a) \geq 1$. We only need to note that the projected dual is an upper bound on the original dual, which is immediate. The fact that the projected dual is equal to the original dual follows form the symmetry of the problem by an averaging argument that takes advantage of the concavity (linearity in our case) of the objective function.

We will get the bound by setting $f(w) = f_1(w)^2$, where $f_1 \in \mathbb{R}[x]$ is such that : $f_1(a) = 1$ and $deg(f_1) \leq \lfloor k/2 \rfloor$. This will reduce the linear program to a least square problem, whose exact solution is the claimed bound. Express f_1 as $f_1 = \sum_{l=0}^{\lfloor k/2 \rfloor} c_l \mathfrak{p}_l$. Using the orthonormality of $\{\mathfrak{p}_l\}_l$, we get

$$E_{\beta^{(n)}} f = \sum_{l=0}^{\lfloor k/2 \rfloor} c_l^2.$$

Taking the constraint $f_1(a) = 1$ into consideration, we get the following upper bound on the projected dual

$$\min\{\textstyle\sum_l c_l^2 : \{c_p\}_{p=0}^{\lfloor k/2 \rfloor} \text{ s.t. } \sum_l c_l \mathfrak{p}_l(a) = 1\} = \frac{1}{\sum_{l=0}^{\lfloor k/2 \rfloor} \mathfrak{p}_l^2(a)},$$

since in general the minimum of $\|y\|_2^2$, $y \in \mathbb{R}^N$ such that $\langle a, x \rangle = 1$ is $\frac{1}{\|a\|_2^2}$. ■

The rest of the proof is about estimating the minimum of $S_t^{(n)}(a)$ over the choice of a in $\{0, \ldots, n\}$. Experimentally, it is evident that the minimum occurs at $\lfloor \frac{n}{2} \rfloor$. This is also very intuitive since this value carries the maximal weigh of $\beta^{(n)}$, but proving that the minimum occurs here is very tricky. Evaluating $S_t^{(n)}(a)$ at $a = \frac{n}{2}$, when n is even, can be done in a systematic way using the recurrence relation (5.15), which greatly simplifies under these conditions.

Lemma 5.6.6 *1) If n is even, and $0 \leq t \leq n$,*

$$S_t^{(n)}\left(\frac{n}{2}\right) \geq \sum_{0 \leq l \ even \ \leq t} \frac{1}{2^l}\binom{l}{l/2}.$$

2) As t increase,

$$\sum_{0 \leq l \ even \ \leq t} \frac{1}{2^l}\binom{l}{l/2} = \sqrt{\frac{2}{\pi + o(1)}}\sqrt{t}.$$

We take care of the case when n is odd by showing that

Lemma 5.6.7 *Assume that n is odd, then for any $1 \leq t \leq n$,*

$$S_t^{(n)}\left(\frac{n-1}{2}\right) \geq S_{t-1}^{(n-1)}\left(\frac{n-1}{2}\right).$$

The tricky part, is

Lemma 5.6.8 *For any n and t such that $0 \leq t \leq n$, the minimum of $S_t^{(n)}(a)$ occurs at $a = \lfloor \frac{n}{2} \rfloor$.*

Combining these lemmas, we get:

- when n is even

$$S_t^{(n)}(a) \geq S_t^{(n)}\left(\frac{n}{2}\right) \geq \sum_{0 \leq l \ \text{even} \ \leq t} \frac{1}{2^l}\binom{l}{l/2}$$

- when n is odd

$$S_t^{(n)}(a) \geq S_t^{(n)}\left(\frac{n-1}{2}\right) \geq S_{t-1}^{(n-1)}\left(\frac{n-1}{2}\right) \geq \sum_{0 \leq l \ \text{even} \ \leq t-1} \frac{1}{2^l}\binom{l}{l/2}.$$

Hence in general: for each n, each $1 \leq t \leq n$, and all $a \in \{0, \ldots, n\}$,

$$S_t^{(n)}(a) \geq \sum_{0 \leq l \ \text{even} \ \leq t-1} \frac{1}{2^l}\binom{l}{l/2}.$$

We still have to establish the lemmas. We start in the reverse order.

5.6.1.1 Proof of Lemma 5.6.8

The numerical experiments suggest the correctness of the following stronger statement

Conjecture 5.6.9 *Let γ be probability measure on $[0:m]$, and let $\{p_l\}_{l=0}^m$, $p_l \in \mathbb{R}[w]$, $deg(p_l) = l$, be the corresponding family of orthonormal polynomials. Assume that γ is nondecreasing on $[0:\lfloor \frac{n}{2} \rfloor]$ and that $\gamma(w) = \gamma(n-w)$. Assume further that t is odd. Then $\sum_{l=0}^t p_l(w)^2$ attains its minimum at $w = \lfloor \frac{n}{2} \rfloor$.*

But it is not clear at the moment how to prove this statement in this generality.

We will get the lemma by establishing something that is apparently more specific to Krawtchouk polynomials.

We will argue that

Lemma 5.6.10 *Let $0 \leq l \leq n-1$. Then*

$$\frac{\mathcal{K}_l^{(n)}(w)^2}{\binom{n}{l}} + \frac{\mathcal{K}_{l+1}^{(n)}(w)^2}{\binom{n}{l+1}}$$

attains its minimum at $w = \lfloor \frac{n}{2} \rfloor$.

This implies Lemma 5.6.8. Indeed, if t is odd, we can group the $t+1$ terms in the expression of $S_t^{(n)}(w)$ into $\frac{t+1}{2}$ term each as in Lemma 5.6.10. If t is even, noting that $S_0^{(n)}(w) = 1$ (via (5.16)), we can group the last t terms in the expression of $S_t^{(n)}(w)$ into $\frac{t}{2}$ term each as in Lemma 5.6.10.

To establish Lemma 5.6.10, it is instructive to prove it first for Hermite polynomials which are in a suitable setting (that is not sufficient for our purposes) limits of Krawtchouk polynomials. Then, we will imitate the proof in the Krawtchouk polynomials setting. Normalized Hermite polynomials $\{\bar{H}_l(x) = \frac{1}{\sqrt{2^l l!}} H_l(x)\}_{l=0}^\infty$ are orthonormal polynomials w.r.t to the Gaussian density $\frac{1}{\sqrt{\pi}} e^{-x^2/2}$ on $\mathbb{R}$, and they satisfy

$$H_l(x) = 2xH_{l-1}(x) - 2(l-1)H_{l-2}(x) \tag{5.18}$$

$$\frac{d}{dx} H_l(x) = 2lH_{l-1}(x). \tag{5.19}$$

See for instance [Sze75].

We show that

Lemma 5.6.11 *Let $l \geq 0$. Then*

$$\bar{H}_l(x)^2 + \bar{H}_{l+1}(x)^2$$

attains its minimum at $x = 0$.

Proof. If $l = 0$, the proof is straight forward, so assume $l \geq 1$. Let

$$V_l(x) \stackrel{\text{def}}{=} 2^l l!(\bar{H}_l(x)^2 + \bar{H}_{l+1}(x)^2) = H_l(x)^2 + \frac{1}{2(l+1)} H_{l+1}(x)^2.$$

Then

$$\frac{d}{dx} V_l(x) = 2l H_l(x) \frac{d}{dx} H_l(x) + \frac{1}{l+1} H_{l+1}(x) \frac{d}{dx} H_{l+1}(x).$$

Using (5.19), and then (5.18), we get

$$\frac{d}{dx} V_l(x) = 2H_l(x)(2l H_{l-1}(x) + H_{l+1}(x)) = 4x H_l(x)^2,$$

and hence the lemma. ■

We have the recurrence relation (5.15) for Krawtchouk polynomials, i.e., an analog to (5.18). To adapt this proof we need to find an analog to (5.19).

Lemma 5.6.12 *Let $0 \leq w \leq n-1$ be an integer.*

a) If $1 \leq l \leq n-1$,

$$\mathcal{K}_l^{(n)}(w+1) - \mathcal{K}_l^{(n)}(w) = -2\mathcal{K}_{l-1}^{(n-1)}(w).$$

b) If $0 \leq l \leq n-1$,

$$\mathcal{K}_l^{(n)}(w+1) + \mathcal{K}_l^{(n)}(w) = 2\mathcal{K}_l^{(n-1)}(w).$$

Proof. Define $x^{(w,n)} \in \mathbb{Z}_2^n$ by

$$x_i^{(w,n)} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } i \le w, \\ 0 & \text{if } i > w. \end{cases}$$

Thus

$$\mathcal{K}_l^{(n)}(w) = \sum_{z \in \mathbb{Z}_2^n; w(z)=l} \mathcal{X}_z(x^{(w,n)}).$$

a) Accordingly,

$$\begin{aligned} \mathcal{K}_l^{(n)}(w+1) - \mathcal{K}_l^{(n)}(w) &= \sum_{z \in \mathbb{Z}_2^n; w(z)=l} \left(\mathcal{X}_z(x^{(w+1,n)}) - \mathcal{X}_z(x^{(w,n)}) \right) \\ &= \sum_{z \in \mathbb{Z}_2^n; w(z)=l} \mathcal{X}_z(x^{(w,n)}) \left((-1)^{z_{w+1}} - 1 \right) \\ &= -2 \sum_{z \in \mathbb{Z}_2^n; w(z)=l \text{ and } z_{w+1}=1} \mathcal{X}_z(x^{(w,n)}) \\ &= -2 \sum_{z \in \mathbb{Z}_2^{n-1}; w(z)=l-1} \mathcal{X}_z(x^{(w,n-1)}). \end{aligned}$$

b) And similarly,

$$\begin{aligned} \mathcal{K}_l^{(n)}(w+1) + \mathcal{K}_l^{(n)}(w) &= \sum_{z \in \mathbb{Z}_2^n; w(z)=l} \mathcal{X}_z(x^{(w,n)}) \left((-1)^{z_{w+1}} + 1 \right) \\ &= 2 \sum_{z \in \mathbb{Z}_2^n; w(z)=l \text{ and } z_{w+1}=0} \mathcal{X}_z(x^{(w,n)}) \\ &= 2 \sum_{z \in \mathbb{Z}_2^{n-1}; w(z)=l} \mathcal{X}_z(x^{(w,n-1)}). \end{aligned}$$

■

Proof of Lemma 5.6.10. For integer values of w, let

$$U_l(w) \stackrel{\text{def}}{=} \binom{n}{l} \left(\frac{\mathcal{K}_l^{(n)}(w)^2}{\binom{n}{l}} + \frac{\mathcal{K}_{l+1}^{(n)}(w)^2}{\binom{n}{l+1}} \right) = \mathcal{K}_l^{(n)}(w)^2 + \frac{l+1}{n-l} \mathcal{K}_{l+1}^{(n)}(w)^2.$$

If $l = 0$, we know from (5.16) and (5.17), that $U_0(w) = 1 + \frac{1}{n}(n - 2w)$, and hence the claim is obvious in this case. Assume that $l \geq 1$. We have

$$\begin{aligned}
U_l(w) - U_l(w+1) &= \mathcal{K}_l^{(n)}(w)^2 - \mathcal{K}_l^{(n)}(w+1)^2 + \frac{l+1}{n-l}\left(\mathcal{K}_{l+1}^{(n)}(w)^2 - \mathcal{K}_{l+1}^{(n)}(w+1)^2\right) \\
&= \left(\mathcal{K}_l^{(n)}(w) - \mathcal{K}_l^{(n)}(w+1)\right)\left(\mathcal{K}_l^{(n)}(w) + \mathcal{K}_l^{(n)}(w+1)\right) \\
&\quad + \frac{l+1}{n-l}\left(\mathcal{K}_{l+1}^{(n)}(w) - \mathcal{K}_{l+1}^{(n)}(w+1)\right)\left(\mathcal{K}_{l+1}^{(n)}(w) + \mathcal{K}_{l+1}^{(n)}(w+1)\right) \\
&= 4\mathcal{K}_{l-1}^{(n-1)}(w)\mathcal{K}_l^{(n-1)}(w) + 4\frac{l+1}{n-l}\mathcal{K}_l^{(n-1)}(w)\mathcal{K}_{l+1}^{(n-1)}(w) \\
&= \frac{4}{n-l}\mathcal{K}_l^{(n-1)}(w)\left((l+1)\mathcal{K}_{l+1}^{(n-1)}(w) + (n-l)\mathcal{K}_{l-1}^{(n-1)}(w)\right) \\
&= \frac{4}{n-l}\mathcal{K}_l^{(n-1)}(w)(n-1-2w)\mathcal{K}_l^{(n-1)}(w) \\
&= \left(\frac{n-1}{2} - w\right)\frac{8}{n-l}\mathcal{K}_l^{(n-1)}(w)^2, \qquad (5.20)
\end{aligned}$$

where the third equality is from Lemma 5.6.12, and the one before the last is from the recurrence relation (5.15), which at $n-1$ becomes

$$(l+1)\mathcal{K}_{l+1}^{(n-1)}(w) - (n-1-2w)\mathcal{K}_{l-1}^{(n-1)}(w) + (n-l)\mathcal{K}_{l-1}^{(n-1)}(w) = 0.$$

This completes the proof of Lemma 5.6.10. Note that when n is odd, we have $U_l(\frac{n-1}{2}) = U_l(\frac{n+1}{2})$.

■

5.6.1.2 Proof of Lemma 5.6.7

We need the following relation along the lines of Lemma 5.6.12.

Lemma 5.6.13 *Let $0 \leq w \leq n-1$ and $l \geq 1$ be integers, then*

$$\mathcal{K}_l^{(n)}(w) = \mathcal{K}_l^{(n-1)}(w) + \mathcal{K}_{l-1}^{(n-1)}(w).$$

Proof. In the notations of Lemma 5.6.12, we have

$$\mathcal{K}_l^{(n)}(w) = \sum_{z \in \mathbb{Z}_2^n; w(z)=l} \mathcal{X}_z(x^{(w,n)})$$

$$
\begin{aligned}
&= \sum_{z\in\mathbb{Z}_2^n; w(z)=l \text{ and } z_n=0} \mathcal{X}_z(x^{(w,n)}) + \sum_{z\in\mathbb{Z}_2^n; w(z)=l \text{ and } z_n=1} \mathcal{X}_z(x^{(w,n)}) \\
&= \sum_{z\in\mathbb{Z}_2^{n-1}; w(z)=l} \mathcal{X}_z(x^{(w,n-1)}) + \sum_{z\in\mathbb{Z}_2^{n-1}; w(z)=l-1} \mathcal{X}_z(x^{(w,n-1)}).
\end{aligned}
$$

■

Accordingly, when n is odd, we have

$$
\begin{aligned}
\mathcal{K}_l^{(n)}\left(\frac{n-1}{2}\right)^2 &= \mathcal{K}_l^{(n-1)}\left(\frac{n-1}{2}\right)^2 + \mathcal{K}_{l-1}^{(n-1)}\left(\frac{n-1}{2}\right)^2 \\
&\quad +2\mathcal{K}_l^{(n-1)}\left(\frac{n-1}{2}\right)\mathcal{K}_{l-1}^{(n-1)}\left(\frac{n-1}{2}\right) \\
&= \mathcal{K}_l^{(n-1)}\left(\frac{n-1}{2}\right)^2 + \mathcal{K}_{l-1}^{(n-1)}\left(\frac{n-1}{2}\right)^2,
\end{aligned}
$$

since $\mathcal{K}_v^{(n-1)}\left(\frac{n-1}{2}\right) = 0$ when v is odd (this follows from (5.15) and (5.17)). Therefore,

$$
\begin{aligned}
S_t^{(n)}\left(\frac{n-1}{2}\right) &= \sum_{l=0}^{t} \frac{1}{\binom{n}{l}} \mathcal{K}_l^{(n)}\left(\frac{n-1}{2}\right)^2 \\
&= 1 + \sum_{l=1}^{t} \frac{1}{\binom{n}{l}} \left(\mathcal{K}_l^{(n-1)}\left(\frac{n-1}{2}\right)^2 + \mathcal{K}_{l-1}^{(n-1)}\left(\frac{n-1}{2}\right)^2\right) \\
&= \sum_{l=0}^{t-1} \left(\frac{1}{\binom{n}{l}} + \frac{1}{\binom{n}{l+1}}\right) \mathcal{K}_l^{(n-1)}\left(\frac{n-1}{2}\right)^2 + \frac{1}{\binom{n}{t}} \mathcal{K}_t^{(n-1)}\left(\frac{n-1}{2}\right)^2.
\end{aligned}
$$

Noting that

$$
\frac{1}{\binom{n}{l}} + \frac{1}{\binom{n}{l+1}} = \frac{1}{\frac{n}{n-l}\binom{n-1}{l}} + \frac{1}{\frac{n}{l+1}\binom{n-1}{l}} = \left(1+\frac{1}{n}\right)\frac{1}{\binom{n-1}{l}},
$$

we get

$$
\begin{aligned}
S_t^{(n)}\left(\frac{n-1}{2}\right) &= \left(1+\frac{1}{n}\right) S_{t-1}^{(n-1)}\left(\frac{n-1}{2}\right) + \frac{1}{\binom{n}{t}} \mathcal{K}_t^{(n-1)}\left(\frac{n-1}{2}\right)^2 \\
&\geq S_{t-1}^{(n-1)}\left(\frac{n-1}{2}\right),
\end{aligned}
$$

for each $t \geq 1$.

5.6.1.3 Proof of Lemma 5.6.6

1) When n is even and $w = \frac{n}{2}$, (5.15), (5.17), and (5.16) simplify to

$$l\mathcal{K}_l^{(n)}\left(\frac{n}{2}\right) + (n-l+2)\mathcal{K}_{l-2}^{(n)}\left(\frac{n}{2}\right) = 0 \ , \ \mathcal{K}_1^{(n)}\left(\frac{n}{2}\right) = 0 \ , \text{ and } \mathcal{K}_0^{(n)}\left(\frac{n}{2}\right) = 1.$$

Therefore, when l is odd,

$$\mathcal{K}_l^{(n)}\left(\frac{n}{2}\right) = 0,$$

and, when $l \geq 2$ is even,

$$\mathcal{K}_l^{(n)}\left(\frac{n}{2}\right) = \pm\frac{n-(l-2)}{l} \times \frac{n-(l-4)}{l-2} \times \ldots \times \frac{n-2}{4} \times \frac{n}{2}.$$

Thus, when $l \geq 2$ is even,

$$\begin{aligned}
\frac{\mathcal{K}_l^{(n)}\left(\frac{n}{2}\right)^2}{\binom{n}{l}} &= \frac{\left(\frac{n\times(n-2)\times(n-4)\times\ldots\times(n-(l-2))}{2\times4\times\ldots\times l}\right)^2}{\frac{n\times(n-1)\times\ldots\times(n-(l-1))}{l\times(l-1)\times\ldots\times2\times1}} \\
&= \frac{(l-1)\times(l-3)\times(l-3)\times\ldots\times3\times1}{l\times(l-2)\times\times(l-4)\times\ldots\times2} \times \frac{n}{n-1}\frac{n-2}{n-3}\cdots\frac{n-(l-2)}{n-(l-1)} \\
&\geq \frac{(l-1)\times(l-3)\times(l-3)\times\ldots\times3\times1}{l\times(l-2)\times\times(l-4)\times\ldots\times2} \\
&= \frac{\frac{l!}{2^{l/2}(l/2)!}}{2^{l/2}(l/2)!} \\
&= \frac{1}{2^l}\binom{l}{l/2}.
\end{aligned}$$

It follows that

$$S_t^{(n)}\left(\frac{n}{2}\right) \geq \sum_{0\leq l \text{ even } \leq t} \frac{1}{2^l}\binom{l}{l/2}.$$

2) From Sterling approximation

$$\sqrt{2\pi}m^{m+\frac{1}{2}}e^{-m+\frac{1}{12m+1}} < m! < \sqrt{2\pi}m^{m+\frac{1}{2}}e^{-m+\frac{1}{12m}},$$

we have

$$\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{l}}e^{-(\frac{1}{6l}-\frac{1}{12l+1})} < \frac{1}{2^l}\binom{l}{l/2} < \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{l}}e^{-(\frac{2}{12l+1}-\frac{1}{12l})},$$

when $l \geq 2$. Moreover, for each even $t_0 \geq 2$, and each even $t > t_0$,

$$2 \sum_{t_0+2 \leq l \text{ even } \leq t} \frac{1}{\sqrt{l}} < \int_{t_0}^{t+2} \frac{dl}{\sqrt{l}} < 2 \sum_{t_0 \leq l \text{ even } \leq t} \frac{1}{\sqrt{l}}$$

It follows that, for each t,

$$\sum_{0 \leq l \text{ even } \leq t} \frac{1}{2^l} \binom{l}{l/2} = (1 - o(1)) \sqrt{\frac{2}{\pi}} \frac{1}{2} \int_1^t \frac{dl}{\sqrt{l}} = \sqrt{\frac{2t}{\pi + o(1)}},$$

where the asymptotic statement is in terms of the growth of t.

5.7 Poly-log-wise independence versus AC_0

We consider in this section the problem of derandomizability of AC_0 by arbitrary k-wise independent probability measures, when k is made polylogaritmically large enough. We reduce this problem to a conjecture about the the symmetry of the optimum of some symmetric optimization problem with linear constraints and a nonlinear objective function.

Consider the following problem which was essentially proposed by Linial and Nisan [LN90].

Problem 5.7.1 "*k-wise independent versus AC_0*": How large should k be as a function of n, M, d, and ϵ so that the k-wise independence property is sufficient to ϵ-fool any AND-OR circuit on n bits of size M, depth d, and unbounded fanin?

The generality of the problem has many potential applications. We explain in Section 5.7.1 its relation to δ-biased probability measures, an application related to the distribution of quadratic residues, and its dual which is asking for a new characterization of AC_0 by low-degree polynomials over the reals.

First, we reduce Problem 5.7.1 in Section 5.7.2 to to the following question.

Problem 5.7.8 *"Low-degree polynomial predictors"*: How large should k be in terms of h and n, so that if $X_1, \ldots, X_{n+1}$ are binary k-wise independent random variables, no degree $\leq h$ polynomial p over the reals on $X_1, \ldots, X_n$ can predict the value of

X_{n+1} with a probability significantly better than 1/2?

The reduction corresponds to the case when h is polylogarithmic in n, and is based on the approximability of AC_0 circuits by low-degree polynomial over the reals (Beigel, Reingold, and Spielman [BRS91], Aspnes, Beigel, Furst, and Rudich [ABFR94]), and the unpredictability perspective in Section 5.2.

Using Theorem 5.6.1 (Limited independence versus weight probability lemma), we establish in Section 5.7.3 a good bound in the restricted setting of Problem 5.7.8 corresponding to the case when p is a symmetric polynomial.

We show that if $k \geq 16\pi h^2$, h is larger than some absolute constant, and $X_1, \ldots,$ X_{n+1} are binary k-wise independent random variables, then no symmetric degree-h polynomial over the reals on $X_1, \ldots, X_n$ can predict the value of X_{n+1} with a probability larger than 1/2.

Due to the highly symmetric nature of Problem 5.7.8, we conjecture in Section 5.7.4 the following.

Conjecture 5.7.14: The symmetric case is a worst case. In other words, for all $0 \leq h, k \leq n$, when maximizing the probability that a degree $\leq h$ polynomial on $x_1, \ldots, x_n$ successfully predicts the last bit of a k-wise independent probability measure on $\{0,1\}^{n+1}$, over the choice of the polynomial and the measure, the maximum is attained by a symmetric polynomial and (consequently) a symmetric measure.

Establishing this conjecture will pull the bound we established in the symmetric case to the more general setting of Problem 5.7.8, and consequently will resolve in a satisfactory way Problem 5.7.1. The correctness of the symmetric optimum conjecture implies that in order to guarantee that the k-wise independence property is sufficient to $M^{-\Theta(1)}$-fool any size-M depth-d circuit in AC_0, it sufficient to make $k = \Theta(\log^{4d} M)$.

5.7.1 The AC_0 conjectures

Problem 5.7.1 [LN90] "k-wise independence versus AC_0": *How large should k be so that the k-wise independence property is sufficient to ϵ-fool any circuit on n-bits of size M, depth d, and unbounded fanin?*

Denote the minimum such k by $K^{5.7.1}_{AC_0}(n, M, d, \epsilon)$.

Problem 5.7.2 "δ-bias versus AC_0": *How small should δ be so that the δ-bias property is sufficient to ϵ-fool any AND-OR circuit on n-bits of size M, depth d, and unbounded fanin?*

Denote the minimum such δ by $\delta^{5.7.2}_{AC_0}(n, M, d, \epsilon)$.

Conjecture 5.7.3 "The poly-log-wise independence versus AC_0 conjecture":

$$K^{5.7.1}_{AC_0}(n, n^{O(1)}, d, o(1)) = O(\log^{O(d)} n),$$

when $d = O(1)$.

Conjecture 5.7.3 can be called the relaxed Linial-Nisan conjecture since Linial and Nisan originally conjectured in [LN90] that $K^{5.7.1}_{AC_0}(n, M, d, 0.1) \leq \log^{d-1} M$, but with these tight parameters this statement is apparently not correct as noted in Luby and Velickovic [LV96]. Using their result stated in Theorem 5.5.6, Linial and Nisan established the bound $K^{5.7.1}_{AC_0}(n, M, 2, 0.1) \leq \sqrt{M} \log M$. Note that this is only slightly better than the trivial bound.

The correctness of the symmetric optimum conjecture, that we state in Section 5.7.4, implies that $K^{5.7.1}_{AC_0}(n, M, d, M^{-\Theta(1)}) = O(\log^{4d} M)$.

Recall that Corollary 5.5.12 is a statement about the correctness of Conjecture 5.7.1 in the special case of read-once DNF formulas.

Conjecture 5.7.4 "The inverse-quasi-poly-bias versus AC_0 conjecture":

$$\delta^{5.7.2}_{AC_0}(n, n^{O(1)}, d, o(1)) = \Omega(2^{-\log^{O(d)} n}),$$

when $d = O(1)$.

It follows from (b) in Theorem 5.1.3 that

$$-\log \delta^{5.7.2}_{AC_0}(n, M, d, \epsilon) \leq K^{5.7.1}_{AC_0}(n, M, d, \frac{\epsilon}{2}) \log n + \log \frac{2}{\epsilon}. \tag{5.21}$$

Thus

Proposition 5.7.5 *Conjecture 5.7.3 implies Conjecture 5.7.4.*

From Corollaries 5.2.5 and 5.2.6, the dual perspective is as follows

Corollary 5.7.6 *1) Conjecture 5.7.4 is equivalent to: For any $f : \{0,1\}^n \to \{0,1\}$ that is computable by an AC_0 circuit of depth d, there exists $f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{R}$ such that*

a) $\|\widehat{f_1}\|_1, \|\widehat{f_2}\|_1 = O(2^{\log^{O(d)} n})$

b) $f_1 \leq f \leq f_2$

c) $E(f_2 - f_1) = o(1)$.

2) Conjecture 5.7.3 is equivalent to: For any $f : \{0,1\}^n \to \{0,1\}$ that is computable by an AC_0 circuit of depth d, there exists $f_1, f_2 : \{0,1\}^n \to \mathbb{R}$ such that

a) $deg(f_1), deg(f_2) = O(\log^{O(d)} n)$

b) $f_1 \leq f \leq f_2$

c) $E(f_2 - f_1) = o(1)$.

Compare this with the known approximation of AC_0 by low degree polynomials summarized in Section 5.1.2.1. See Remark 5.1.4.

Note that Nisan generator for AC_0 (see Section 5.1.2.2) is a special $O(\log^{O(d)} n)$-wise probability measure. It is also linear. See Section 5.8.2.3 for a comparison of Nisan generator with arbitrary linear k-wise independent measures.

Finally, observe that by arguing as in Corollary 5.5.3, we can derive some consequences of small bias conjecture to the distribution of quadratic residues as follows.

Let $S_1, \ldots, S_m$ be arbitrary subsets of $\mathbb{F}_q$. Let Q be the set of quadratic residues in $\mathbb{F}_q^\times$. Let $I = \cup_i S_i$, $\vec{1} \in \mathbb{Z}_2^I$ the all ones vector, and $n = |I|$. Then

$$\left| \frac{1}{q} \#\{a \in \mathbb{F}_q : S_i + a \not\subset Q, \forall i\} - \frac{1}{2^n} \#\{z \in \mathbb{Z}_2^I : z|_{S_i} \neq \vec{1}|_{S_i}, \forall i\} \right| \leq \epsilon,$$

where ϵ is such that $\delta_{AC_0}^{5.7.2}(n, m, 2, \epsilon) \geq 2n/\sqrt{q}$.

Compare with Corollary 5.5.3. Similar statements can be derived for the quadratic-residues-like PRG's defined in Section 5.1.2.3.

5.7.2 Low-degree polynomials predictors

Consider

Theorem 5.7.7 [BRS91, ABFR94] *Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be computable by an AND-OR depth-d circuit of size M with unbounded fanin. Then there exists a family of functions $\{p_\alpha\}_{\alpha \in I}$, $p_\alpha \in \mathbb{Z}[x_1, \ldots, x_n]$, and I some index set such that*

1) the degree of each p_α is at most $h = h_0^d$, $h_0 = O(\log \frac{M}{\theta} \log M)$

2) for each x in $\{0,1\}^n$, $Pr_{\alpha \in I}[p_\alpha(x) \neq f(x)] \leq \theta$,

where $\theta > 0$ is tunable.

Note that the maximum absolute value of each p_α on $\{0,1\}^n$ can be as large as $2^{O(\log \frac{M}{\theta} \log M)^d}$, and that $E|p_\alpha - f|$ is potentially as large as $\theta 2^{O(\log \frac{M}{\theta} \log M)^d}$.

Note also that the statement in [BRS91, ABFR94] is (4) in Section 5.1.2.1. But, we need the stronger statement in Theorem 5.7.7 which can be extracted from the proof in [BRS91, ABFR94] used to establish (4).

Using Theorem 5.7.7 and Lemma 5.2.7, we can reduce Problem 5.7.1 (k-wise independence versus AC_0) to

Problem 5.7.8 "Low-degree polynomials predictors": *Let μ be a k-wise independent probability measure on $\mathbb{Z}_2^{n+1}$. Let $p : \mathbb{Z}_2^n \rightarrow \mathbb{R}$ such that $deg(p) \leq h$. How large should k be with respect to n, h, and ϵ, so that*

$$Pr_{x \sim \mu}[p(x_1, \ldots, x_n) = x_{n+1}] \leq \frac{1}{2} + \epsilon? \tag{5.22}$$

Denote minimum such k by $K_{poly-pred}^{5.7.8}(n, h, \epsilon)$.

It is important to stress here that essence of Problem 5.7.8 is that that p can take values other than 0 and 1.

The relation is as follows.

Lemma 5.7.9 $K^{5.7.1}_{AC_0}(n, M, d, \epsilon) \leq K^{5.7.8}_{poly-pred}(n, h_0^d, (1-\alpha)\frac{\epsilon}{2})$, *where* $h_0 = O(\log \frac{2M}{\alpha\epsilon} \log M)$, *and* $0 < \alpha \leq 1$ *tunable.*

Proof. Using Lemma 5.2.7, there is $f' : \{0,1\}^n \to \{0,1\}$ of the same circuit complexity as f, and a measure μ' on $\{0,1\}^{n+1}$ in the same class of μ such that

$$Pr_{x\sim\mu'}[f'(x_1, \ldots, x_n) = x_{n+1}] > \frac{1}{2} + \frac{\epsilon}{2}. \tag{5.23}$$

Now consider the family of polynomials $\{p_\alpha\}_{\alpha\in I}$, $p_\alpha \in \mathbb{Z}[x_1, \ldots, x_n]$, with respect to f' in the setting of Theorem 5.7.7. Since for each x in $\mathbb{Z}_2^n$, $Pr_{\alpha\in I}[p_\alpha(x) \neq f'(x)] \leq \theta$, we get that there is $\alpha \in I$ such that $Pr_{x\sim\mu'}[p_\alpha(x) \neq f'(x)] \leq \theta$. Hence there exists a $p : \mathbb{Z}_2^n \to \mathbb{Z}$ such that

1) $deg(p) \leq h \stackrel{\text{def}}{=} h_0^d, h_0 = c \log \frac{M}{\theta} \log M$, $c > 0$ a constant

2) $Pr_{x\sim\mu'}[p(x|_{[n]}) = x_{n+1}] > \frac{1}{2} + \frac{\epsilon}{2} - \theta$,

where θ can be tuned. Set $\theta = \alpha\epsilon/2$. ■

Thus Conjecture 5.7.3 (The poly-log-wise versus AC_0 conjecture) follows from

Conjecture 5.7.10 $K^{5.7.8}_{poly-pred}(n, h, o(1)) = poly(\log n, h)$.

We will establish in the next section this conjecture in the special case when p is a symmetric polynomial, which will lead us to the more specific version

Conjecture 5.7.11 $K^{5.7.8}_{poly-pred}(n, h, 0) = O(h^2)$.

Remark 5.7.12 Note that we need to proceed in Problem 5.7.8 (Low-degree polynomials predictors) by unpredictability in the sense that we cannot hope to go far with the following question. Consider the k-wise independent situation. If p is a low degree polynomial on $\{0,1\}^n$, how large can

$$\epsilon = |Pr_{x\in\{0,1\}^n}[p(x) = 0] - Pr_{x\sim\mu}[p(x) = 0]|$$

be when μ is a k-wise independent probability measure on $\mathbb{Z}_2^n$ and k is large enough? The issue with this question is that ϵ cannot be made arbitrarily polynomially small. For instance, say that μ is the discrete probability measure supported by a linear code defined by the single constraint which requires that the parity of all the variables $x_1, \ldots, x_n$ is zero. Say also that n is even and that $p(x) = \sum_{i=1}^n x_i - n/2 - 1$. Then μ is $(n-1)$-independent, and the degree of p is 1. But $\epsilon = \Theta(\frac{1}{\sqrt{n}})$. So with this question we cannot hope to achieve an ϵ below $\Theta(\frac{1}{\sqrt{n}})$ (e.g. $\epsilon = n^{-3}$ is not achievable) regardless of how large we make k as long we are below $k = n$.

5.7.3 A good bound in the symmetric case

Assuming that the polynomial p in Problem 5.7.8 (Low-degree polynomials predictors) is symmetric, we establish the following bound. The bound is consequence of Theorem 5.6.1 (Limited independence versus weight probability lemma).

Theorem 5.7.13 *Consider Problem 5.7.8 (Low-degree polynomials predictors) in the special case when p is restricted to be a symmetric polynomial on the variables $x_1, \ldots, x_n$, i.e., assume that $p : \mathbb{Z}_2^n \to \mathbb{R}$ is such that the value of $p(x)$ on $x \in \mathbb{Z}_2^n$ depends only on the weight $w(x)$ of x.*

Let $K^{5.7.8}_{poly-pred-symm}(n, h, \epsilon)$ be the corresponding minimum.

Then

$$K^{5.7.8}_{poly-pred-symm}(n, h, 0) \leq 16\pi h^2$$

when h is larger than some absolute constant.

This means that under these conditions the chances of predicting x_{n+1} correctly cannot be better than the value $1/2$, which is achieved by the constant polynomials $p = 1$ or $p = 0$.

Proof. Let μ be a k-wise independent probability measure on $\mathbb{Z}_2^{n+1}$, $k \geq 2$. We will use the following consequences of the k-wise independence of μ:

- The projection μ' of μ on $\mathbb{Z}_2^n$, i.e.,

$$\mu'(x_1, \ldots, x_n) = \mu(x_1, \ldots, x_n, 0) + \mu(x_1, \ldots, x_n, 1),$$

is k-wise independent.

- The last bit x_{n+1} of μ is unbiased, i.e., $Pr_{x\sim\mu}[x_{n+1}=1]=\frac{1}{2}$.

Let $p:\mathbb{Z}_2^n\to\mathbb{R}$ be a symmetric polynomial on the variables $x_1,\ldots,x_n$. Assume that $deg(p)\le h$. We will show that if

$$\sum_{0\le l \text{ even } l\le\lfloor k/2\rfloor-1}\frac{1}{2^l}\binom{l}{l/2}\ge 4h, \tag{5.24}$$

or equivalently if,

$$k\ge 16\pi h^2 \text{ when } h\ge \text{ some absolute constant}, \tag{5.25}$$

then

$$Pr_{x\sim\mu}[p(x_1,\ldots,x_n)=x_{n+1}]\le\frac{1}{2}. \tag{5.26}$$

Let

$$\beta=Pr_{x\sim\mu}[p(x_1,\ldots,x_n)=x_{n+1}],$$

and let μ' be the projection of μ

If the degree of p is zero, then $\beta=\frac{1}{2}$ when $p=0$ or $p=1$, and $\beta=0$ otherwise. So assume that $deg(p)\ge 1$. Since $p(x)$ depends on the weight of x, let f be a univariate polynomial in $\mathbb{R}[a]$, $deg(f)\le h$, such that $p(x)=f(w(x))$ for each $x\in\mathbb{Z}_2^n$. Let S_0 be the set of solutions of in $\mathbb{R}$ of the equation $f(a)=0$, and S_1 those of the equation $f(a)=1$. So $|S_0|,|S_1|\le h$, and for each $x\in\mathbb{Z}_2^{n+1}$ satisfying $p(x_1,\ldots,x_n)=x_{n+1}$, we have $w(x_1,\ldots,x_n)\in S_0\cup S_1$. Therefore

$$\beta\le Pr_{x\sim\mu'}[w(x)\in S_0\cup S_1]\le|S_0\cup S_1|\max_{a\in\{0,\ldots,n\}}Pr_{x\sim\mu'}[w(x)=a].$$

Using Theorem 5.6.1, and the bound $|S_0\cup S_1|\le 2h$, we get

$$\beta\le\frac{2h}{\sum_{0\le l \text{ even } l\le\lfloor k/2\rfloor-1}\frac{1}{2^l}\binom{l}{l/2}}\le\frac{1}{2},$$

where the last inequality is (5.24). The asymptotic expression in Theorem 5.6.1 means that (5.24) is equivalent to $k \geq 16(\pi + o(1))h^2$, and hence (5.25) since since k and h are integers and π is irrational. ■

It is important to note that we partially used in the proof the fact that μ is k-wise independent. We only used the fact that μ' is k-wise independent, and that the last bit of μ is unbiased. In general when p is not symmetric, we know that we must use the limited independence properties of the last bit. However, the symmetry of the problem suggests the following conjecture: in the setting where μ is k-wise independent, the worst case is achievable by a symmetric polynomial.

5.7.4 The symmetric optimum conjecture

Accordingly, we can reduce Problem 5.7.8 (Low-degree polynomials predictors) to the following conjecture.

Conjecture 5.7.14 $K^{5.7.8}_{poly-pred}(n, h, 0) = K^{5.7.8}_{poly-pred-symm}(n, h, 0)$.

In other words, the worst case of Problem 5.7.8 is achievable by a symmetric polynomial.

Backtracking, we get from Theorem 5.7.13, Lemma 5.7.9, and Proposition 5.7.5 that

Proposition 5.7.15 *Conjecture 5.7.14 implies:*

- *Conjecture 5.7.11 (The $K^{5.7.8}_{poly-pred}(n, h, 0) = O(h^2)$ conjecture),*
- *Conjecture 5.7.4 (The inverse-quasi-poly-bias versus AC_0 conjecture), and*
- *Conjecture 5.7.3 (The poly-log-wise independence versus AC_0 conjecture).*

It specifically implies that $K^{5.7.1}_{AC_0}(n, M, d, M^{-\Theta(1)}) = O(\log^{4d} M)$, when $d = O(1)$.

Note that Problem 5.7.8 (Low-degree polynomials predictors) can be phrased as a symmetric optimization problem over linear constraints as follows.

We are interested in

$$\Phi(n,h,k) \stackrel{\text{def}}{=} \max_{p,\mu} Pr_{x\sim\mu}[p(x_1,\ldots,x_n) = x_{n+1}],$$

where we are optimizing over the polynomials $p : \mathbb{Z}_2^n \to \mathbb{R}$, $deg(p) \leq h$, and the k-wise independent probability measures μ on $\mathbb{Z}_2^{n+1}$.

Denote the $\mathbb{R}$-linear subspace $\{f : \mathbb{Z}_2^m \to \mathbb{R}\}$ by $L(\mathbb{Z}_2^m)$.

Let $M \subset L(\mathbb{Z}_2^{n+1})$ be the convex polytope consisting of $\mu \in L(\mathbb{Z}_2^{n+1})$ satisfying: $\mu \geq 0$, $\sum_x \mu(x) = 1$, and $\sum_x \mu(x)\mathcal{X}_z(x) = 0$ for each $z \neq 0 \in \mathbb{Z}_2^{n+1}$ such that $w(z) \leq k$.

Let P be the linear subspace of $L(\mathbb{Z}_2^n)$ consisting of all the functions p in $L(\mathbb{Z}_2^n)$ satisfying: $\sum_x p(x)\mathcal{X}_y(x) = 0$ for each $y \in \mathbb{Z}_2^n$ such that $w(y) > h$.

Finally, let $U : L(\mathbb{Z}_2^{n+1}) \times L(\mathbb{Z}_2^n) \to \mathbb{R}$ be given by

$$U(\mu,p) = \sum\{\mu(x) : x \in \mathbb{Z}_2^{n+1} \text{ s.t. } p(x_1,\ldots,x_n) = x_{n+1}\}$$

Then

$$\Phi(n,h,k) = \max_{(\mu,p)\in M\times P} U(\mu,p). \tag{5.27}$$

The symmetric group S_m of permutations π of $\{1,\ldots,m\}$ acts on the vector space $L(\mathbb{Z}_2^m) = \{f : \mathbb{Z}_2^n \to \mathbb{R}\}$ by permutations of the variables, i.e., via $(\rho_\pi f)(x) = f((x_{\pi(i)})_i)$. S_n acts also on $L(\mathbb{Z}_2^{n+1})$ in a natural way when S_n is identified with the subgroup of S_{n+1} consisting of the permutations that leave $n+1$ fixed. Thus S_n acts on $L(\mathbb{Z}_2^n) \times L(\mathbb{Z}_2^{n+1})$ naturally via $\rho_\pi(f,\mu) = (\rho_\pi f, \rho_\pi \mu)$. In this sense we have:

a) U is invariant under the action of S_n, i.e., $U(\rho_\pi(p,\mu)) = U(p,\mu)$, $\forall \pi \in S_n$.

b) P is symmetric under the action of S_n, i.e., $\rho_\pi(P) = P$, $\forall \pi \in S_n$.

c) M is symmetric under the action of S_{n+1}, i.e., $\rho_\pi(M) = M$, $\forall \pi \in S_{n+1}$.

Conjecture 5.7.14 can be stated as follows: For each $0 \leq h,k \leq n$, the maximum in (5.27) is achievable by a pair (p^*,μ^*) invariant under the action of S_n, i.e., $\rho_\pi p^* = p^*$ and $\rho_\pi \mu^* = \mu^*$, $\forall \pi \in S_n$, Or equivalently, by a a pair (p^*,μ^*) such that p^* is a symmetric polynomial on the variables $x_1,\ldots,x_n$.

A final remark is that U is obviously not convex, and the above general symmetry remarks are unlikely to be sufficient alone if we ignore the other special features of the problem. Resolving the conjecture should involve taking more advantage of the structure of U, P, and M.

5.8 Parity with encrypted linear help

We study in this section the problem of approximability of high-degree parity functions on high-dual-distance binary linear codes by low-degree polynomials with coefficients in fields of odd characteristics. This problem has applications to the derandomizability of AC_0 or low-degree polynomial equations on binary input variables with coefficients in small finite fields of odd order by binary linear codes with sufficiently large dual distance. Among other results, we relax this problem into essentially a single low-dimensional low-complexity linear program in terms of Krawtchouk polynomials. We leave the problem of bounding the optimum of the linear program open.

5.8.1 Summary

We consider the following special case of Problem 5.7.8 (Low-degree polynomials predictors).

Problem 5.8.2 *"Parity with encrypted linear help":* Let $F = \mathbb{Q}$ or $\mathbb{F}_q$, q an odd prime. Let C be a block-length n binary linear code and $\mathcal{X}$ be a parity function on C such that:

1) the dual distance of C is above k

2) any realization of $\mathcal{X}$ on C requires a XOR of more than k bits.

How large should k be with respect to n, h, and ϵ, so that $\mathcal{X}$ cannot be approximated on C with a probability larger than $1/2 + \epsilon$ by a degree $\leq h$ polynomial p on the n bits of C with coefficients in F?

Our interest in this problem is motivated by the following applications:

- When $F = \mathbb{F}_q$, $q = O(1)$, this problem is essentially equivalent to the usability of arbitrary binary linear codes with sufficiently large dual distance to fool low-degree polynomial equations over binary input variables with coefficients in F. See Lemma 5.8.4.

- When $F = \mathbb{Q}$ or $\mathbb{F}_q$, $q = O(1)$, the problem of derandomizability of AC_0 by arbitrary linear codes with sufficiently large dual distance can be efficiently reduced to Problem 5.8.2 with h polylogarithmic in n. See Section 5.8.2.2. In the rationals case, this relation is a special case of the reduction in Section 5.7.2. In the finite-fields case the reduction is based on the AC_0 versus low-degree polynomials theorem of Razborov [Raz87].

The classical parity approximability problem by low-degree polynomials corresponds to the case when $C = \{0,1\}^n$ and $\mathcal{X}$ is a parity on all the n bits of C. The classical problem was resolved to some extent by Smolensky [Smo87] who argued that parity on the $\{0,1\}^n$ cannot be approximated by a degree $h = o(\sqrt{n})$ polynomial with a probability larger than $\frac{1}{2} + \Omega(\frac{h}{\sqrt{n}})$. Note that the error term $\frac{h}{\sqrt{n}}$ is very unlikely to be tight. For instance, when $h = 1$, and $F = \mathbb{F}_3$, the error term appear experimentally to be $2^{-\Theta(n)}$. Moreover, assuming Conjecture 5.7.14 (The symmetric optimum conjecture), the error term is zero when $F = \mathbb{Q}$ and $k > 16\pi h^2$.

We argue in Section 5.8.3.4 that the argument of Smolensky provably fails to generalize to the setting of Problem 5.8.2 (Parity with encrypted linear help) in the typically case when the minimum distance of the code C (not the dual) grows linearly with the block length. The reason is that, in such a case, the graph constructed by modding out the Hamming cube by the dual of C is a good expander. Roughly, this invalidate the geometric argument of Smolensky since this argument is essentially based on the low expansion of some sets of probability $1/2$ in the Hamming cube.

The Nisan generator for AC_0 (see Section 5.1.2.2) is a linear code and hence falls in this category of PRG's. The argument behind Nisan generator is by reduction to the problem of approximability of parity on the whole Hamming cube. We argue in Section 5.8.2.3 that arguing by reduction fails to generalize to arbitrary linear code

with sufficiently large dual distance. The failure occurs provably in the typical case when – again – the minimum distance of the code C (not the dual) grows linearly with the block length. The meaning of the failure of the reduction in such a case is that the extra information given to the original circuit or the polynomial is written in a format which is too hard for a small constant-depth circuit or a low-degree polynomial to reproduce even partially.

We explain the algebraic interpretation of Problem 5.8.2 (Parity with encrypted linear help) in Section 5.8.3.3. Algebraically, this problem is about bounding the dimension of the space of solutions of a type of difference equations on Cayley graphs based on linear codes, or, equivalently, the problem of bounding the dimension of the vector space spanned by translations of a function on the graph whose support has a special geometry. We use the later formulation to conclude that Smolensky's argument does not generalize.

Then we proceed by doubly relaxing Problem 5.8.2.

5.8.1.1 The first relaxation

We know from Smolensky's work [Smo87] that parity on the Hamming cube cannot be approximated by a low-degree polynomial. The setting of Problem 5.8.2 (Parity with encrypted linear help) is parity on a linear code. Consider (1) and (2) in Problem 5.8.2. What (2) is saying is that all the realizations of this help are difficult on the low-degree polynomial, i.e., the help is in a suitable sense encrypted and potentially useless. What (1) is saying is that the code is k-wise independent. This is supposed to make the help even more useless, but (2) is already a relatively strong condition alone. We relax the problem by dropping (1). This drastically simplifies the problem. We note in Section 5.8.4 that, by dropping (1), we can reduce to the case when the polynomial p has degree 1. We can reduce further to the setting when the polynomial has all its coefficients equal to 1 in the small finite-fields case. In the rationals case, we can reduce to a similar setting where all the coefficients of the linear form are ± 1 under bounds on the coefficients of the original polynomial, which are consistent with the AC_0 situation. See Section 5.8.5.3.

5.8.1.2 The LP relaxation

Then motivated by Delsarte LP coding bound [Del73], we relax the resulting problem further into a low-dimensional low-complexity linear program. The relaxation corresponds to relaxing the problem of optimizing on linear codes to the problem of optimizing on probability measures whose Fourier transform is nonnegative.

In the special case of linear codes, we can look at Delsarte LP coding bound as a bound via this relaxation. See Lemma 5.8.19. Recall that we studied in Lemma 5.3.4 this relaxation carefully and we characterized the position of linear codes in the set of all probability measures whose Fourier transforms is nonnegative.

To test the goodness of the relaxation in the setting of Problem 5.8.2 (Parity with encrypted linear help), we test it on the following related problem. How small can a linear code C of block-length n be if we know that there is a parity function $\mathcal{X}$ on C s.t any realization of $\mathcal{X}$ on C requires a XOR of more than k bits. This is equivalent to asking how large can a coset of a linear code be if we know that it has no element of weight below $d = k + 1$ bit. Let M_d the maximum. By applying the relaxation to this problem, we argue in Lemma 5.8.16 that we get the following linear program.

Let $f : [0 : d] \times [0 : n - d] \to \mathbb{R}$,

$$f(w_1, w_2) = \begin{cases} 2^n & \text{if } w_1 + w_2 = 0 \\ 0 & o.w. \end{cases} .$$

Let $\tilde{M}_d = \min E_\beta g$, where we are optimizing on $g : [0 : d] \times [0 : n-d] \to \mathbb{R}$ such that with $\hat{g} : [0 : d] \times [0 : n - d] \to \mathbb{R}$ related to g via

$$g(w_1, w_2) = \sum_{l_1, l_2} \hat{g}(l_1, l_2) \mathcal{K}_{l_1}^{(d)}(w_1) \mathcal{K}_{l_2}^{(n-d)}(w_2),$$

where the $\{\mathcal{K}_t^{(m)}(v)\}_t$ are the Krawtchouk polynomials, we have

a) $g \geq f$

b) $\hat{g}(l_1, l_2) \leq 0$, when $l_1 \leq l_2$ and $l_1 + l_2 \neq 0$,

where β is the product binomial measure on $[0:d] \times [0:n-d]$. Here we are using the notation $[n_1 : n_2] \stackrel{\text{def}}{=} \{n_1, \ldots, n_2\}$.

By suitably constructing g, we argue in Lemma 5.8.17 that $\tilde{M}_d = M_d = 2^{n-d}$. This is an indication of the goodness of the relaxation.

By similarly relaxing the reduced version of Problem 5.8.2 (Parity with encrypted linear help), i.e., after the first relaxation summarized in Section 5.8.1.1, we show in Lemma 5.8.15 that we obtain the following linear program.

The finite fields case: Let $b_0 \in \mathbb{F}_q$, q a small odd prime (e.g. $q = 3$), and $f : [0 : k+1] \times [0 : n-k-1] \to \{0, 1\}$ be given by

$$f(w_1, w_2) = \begin{cases} 1 & \text{if } w_1 + w_2 + b_0 = (-1)^{w_1} \pmod q \\ 0 & o.w. \end{cases}.$$

How large should k be so that there exists $\hat{g} : [0 : k+1] \times [0 : n-k-1] \to \mathbb{R}$ such that with $g : [0 : k+1] \times [0 : n-k-1] \to \mathbb{R}$ given by

$$g(w_1, w_2) = \sum_{l_1, l_2} \hat{g}(l_1, l_2) \mathcal{K}_{l_1}^{(k+1)}(w_1) \mathcal{K}_{l_2}^{(n-k-1)}(w_2),$$

we have

i) $g \geq f$

ii) $E_\beta g \leq \frac{1}{2} + \epsilon$

iii) $\hat{g}(l_1, l_2) \leq 0$, when $l_1 \leq l_2$ and $l_1 + l_2 \neq 0$,

where β is the product binomial measure on $[0 : k+1] \times [0 : n-k-1]$?

In the bounded-coefficients rational case, we end up in Section 5.8.5.3 with the following linear program.

The rational case: Let s_1, s_2, s_3, s_4 be nonnegative integers such that $s_1 + s_2 = k+1$ and $s_1 + s_2 + s_3 + s_4 = n$. Let b_0 and u be integers such that u is nonzero.

Let $f : \prod_{i=1}^{4}[0 : s_i] \to \{0, 1\}$ be given by

$$f(w) = \begin{cases} 1 & \text{if } w_1 + w_3 - w_2 - w_4 + b_0 = u(-1)^{w_1+w_2} \\ 0 & o.w. \end{cases}.$$

How large should k be so that there exists $g : \prod_{i=1}^{4}[0 : s_i] \to \mathbb{R}$ such that with $g : \prod_{i=1}^{4}[0 : s_i] \to \mathbb{R}$ given by

$$g(w) = \sum_{l} \hat{g}(l) \prod_{i=1}^{4} \mathcal{K}_{l_i}^{(s_i)}(w_i),$$

we have

i) $g \geq f$

ii) $E_\beta g \leq \frac{1}{2} + \epsilon$

iii) $\hat{g}(l) \leq 0$, when $l_1 + l_2 \leq l_3 + l_4$ and $l_1 + l_2 + l_3 + l_4 \neq 0$,

where β is the product binomial measure on $\prod_{i=1}^{4}[0 : s_i]$?

What is interesting about this relaxation is that we have now a single low-dimensional low-complexity linear program parameterized by $b_0 \in [0 : q-1]$ in the finite fields case. In the rationals case, it is parameterized by the feasible settings of the parameters s_1, s_2, s_3, s_4, b_0, and u. Note that each of those parameters can take $O(n)$ nontrivial value only. The dimension is very low since we have $O(nk) = O(n^2)$ variables in the finite fields case, and $O(n^2k^2) = O(n^4)$ in the rationals case. The same estimates hold for the number of constraints. Unfortunately, we cannot get much insight from experiments since this linear program is numerically unstable due the wide spectrum of the values taken by Krawtchouk polynomials. The numerical simulations are breaking down before $n = 50$.

We leave the problem of bounding the optimum of the linear programs open.

5.8.2 The problem

In this section F will be a field whose characteristic is not equal to 2.

Terminology 5.8.1 *If $C \subset \mathbb{Z}_2^n$ is a linear code and $g : \mathbb{Z}_2^n \to F$, by $\hat{g} : \widehat{C} = \mathbb{Z}_2^n/C^\perp \to F$ we mean the Fourier transform of g with respect to the characters*

$$\{\mathcal{X}_{\bar{z}}(x) \stackrel{\text{def}}{=} (-1)^{xz}\}_{z \in \mathbb{Z}_2^n/C^\perp}$$

of the abelian group C over F, where $xz \stackrel{\text{def}}{=} \sum_i x_i z_i$ and the definition is independent of the choice of $z \in \bar{z}$. Thus

$$g(x) = \sum_{\bar{z}} \hat{g}(\bar{z}) \mathcal{X}_{\bar{z}}(x)$$

or equivalently

$$\hat{g}(\bar{z}) = \frac{1}{|C|} \sum_x g(x) \mathcal{X}_{\bar{z}}(x).$$

The weight of a $\bar{z} \in \mathbb{Z}_2^n/C^\perp$ is defined to be the minimum weight of a $z' \in \bar{z}$, or equivalent the weight of $\bar{z}$ is the distance of $\bar{z}$ from $\bar{0}$ in the Cayley graph $\mathbb{Z}_2^n/C^\perp$ as defined in Section 5.8.3.1 to be the quotient of the Hamming cube Cayley graph on $\mathbb{Z}_2^n$ by the subgroup $C^\perp$ of $\mathbb{Z}_2^n$.

The degree of g is defined to be the weight of the largest $\bar{z}$ such that $\hat{g}(\bar{z}) \neq 0$.

Problem 5.8.2 "Parity with encrypted linear help ": *Let $C \subset \mathbb{Z}_2^n$ be a linear code and $\bar{z}_1 \in \mathbb{Z}_2^n/C^\perp$ such that :*

1) *The minimum distance of the dual $C^\perp$ of C is above k, or equivalently the probability measure $\mu_C = \frac{1}{|C|} 1_C$ is k-wise independent.*

2) *The weight of $\bar{z}_1$ is above k, or equivalently all the equivalent representations of the parity $\mathcal{X}_{\bar{z}_1}$ on C require the XOR of more than k bit.*

Let $p : C \to F$ such that $deg(p) \leq h$. How large should k be so that

$$Pr_{x \in C}[p(x) = \mathcal{X}_{\bar{z}_1}(x)] \leq \frac{1}{2} + \epsilon? \tag{5.28}$$

Denote the minimum such k by $K_{parity}^{5.8.2}(n, h, \epsilon)_F$.

Our interest is in the cases when $F = \mathbb{F}_q$, q an odd prime, or $F = \mathbb{Q}$.

5.8.2.1 The finite fields case

When $F = \mathbb{F}_q$, $q = O(1)$, Problem 5.8.2 (Parity with encrypted linear help) is essentially equivalent to Problem 5.8.3 below.

Problem 5.8.3 "Linear codes versus low-degree F-polynomial-equations on binary input variables": *Let $C \subset \mathbb{Z}_2^n$ be a linear code such that the minimum distance of the dual $C^\perp$ of C is at least k. Let $p : C \to F$ such that $deg(p) \leq h$. How large should k be so that*

$$|Pr_{x \in \mathbb{Z}_2^n}[p(x) = 0] - Pr_{x \in C}[p(x) = 0]| \leq \epsilon? \tag{5.29}$$

Denote the minimum such k by $K^{5.8.3}_{poly-eqn}(n, h, \epsilon)_F$.

Our interest is in the case when $F = \mathbb{F}_q$, $q = O(1)$ a small odd prime.

Note that when, for instance, $F = \mathbb{Q}$ we cannot hope to get a good bound in terms of ϵ in the setting of Problem 5.8.3. See Remark 5.7.12.

The relation follows from the unpredictability perspective in Corollary 5.2.8. Namely,

Lemma 5.8.4 *Let q be an odd prime, then*

$$K^{5.8.2}_{parity}\left(n, (q-1)h, \frac{\epsilon}{2}\right)_{\mathbb{F}_q} \leq K^{5.8.3}_{poly-eqn}(n, h, \epsilon)_{\mathbb{F}_q} \leq K^{5.8.2}_{parity}(n+1, h, \epsilon)_{\mathbb{F}_q}.$$

Proof. Let p and C be as in Problem 5.8.3. Fix k, and assume that (5.29) does not hold. Let $f : \mathbb{Z}_2^n \to \{0, 1\}$ be given by $f(x) = p(x)^{q-1}$, then apply Corollary 5.2.8. This proves the first inequality. The second is immediate. ■

5.8.2.2 AC_0 implications

Lemma 5.8.5 *There are some global absolute positive constants c_1, c_2 such that if $f : \{0, 1\}^n \to \{0, 1\}$ is computable by an AND-OR size-M depth-d unbounded-fanin circuit, C is a block-length-n binary linear code whose minimum dual distance is above k, and*

I) either

$$k \geq K^{5.8.2}_{parity}\left(n, h_0^d, (1-\alpha)\frac{\epsilon}{2}\right)_{\mathbb{Q}},$$

where $h_0 = c_1 \log \frac{2M}{\alpha\epsilon} \log M$, *and* $0 < \alpha \leq 1$ *tunable,*

II) or

$$k \geq K^{5.8.2}_{parity}\left(n, h_q^d, \frac{\epsilon}{4}\right)_{\mathbb{F}_q},$$

where $h_q = c_2 q \log \frac{4M}{\epsilon}$,

then μ_C *can* ϵ*-fool* C.

We get (I), via Corollary 5.2.8, by specializing Lemma 5.7.9 to the setting of k-wise independent measures coming from linear codes.

The relation in (II) follows by the same argument, but while using Razborov Theorem 5.8.6 below instead of Theorem 5.7.7. Note that in (II) we fixed α to $1/2$ since, unlike in the $\mathbb{Q}$-setting, we cannot hope here to get a good bound when $\alpha = 1$.

Theorem 5.8.6 [Raz87] *Let* $f : \{0,1\}^n \to \{0,1\}$ *be computable by an AND-OR depth-d circuit of size* M *with unbounded fanin. Let* q *be a power of any prime Then there exists a family of polynomials* $\{p_\alpha\}_{\alpha \in I}$, $p_\alpha \in \mathbb{F}_q[x_1, \ldots, x_n]$, *and* I *some index set such that*

1) the degree of each p_α *is at most* $h = h_q^d$, $h_q = O(q \log \frac{M}{\theta})$

2) for each x *in* $\{0,1\}^n$, $Pr_{\alpha \in I}[p_\alpha(x) \neq f(x)] \leq \theta$,

where $\theta > 0$ *is tunable.*

Note that the statement in [Raz87] is (5) in Section 5.1.2.1. But, we need the stronger statement in Theorem 5.8.6 which can be extracted from the proof in [Raz87] used to establish (4). Note also that the theorem is nontrivial only when q is small.

5.8.2.3 The nature of the problem

We point out in this section why Problem 5.8.2 (Parity with encrypted linear help) cannot be resolved by reducing it to the problem of approximability of parity on the

whole Hamming cube by low-degree polynomials. For the sake of comparison with the Nisan generator for AC_0, rather than working with low-degree polynomials, we will illustrate the issue in the setting of AC_0 circuits. The same argument applies to low-degree polynomials.

We want to show that when k is made large enough with respect to $M = n^{O(1)}$ and $d = O(1)$, there is no size-M depth-d unbounded-fanin AND-OR circuit $f : \{0,1\}^n \to \{0,1\}$, and no binary linear code $C \subset \{0,1\}^n$ such that

$$Pr_{x \in C}[f(x) = \text{Parity}_A(x)] > \frac{1}{2} + \epsilon,$$

where $A \subset [n]$ is such that $|A'| > k$, for all $A' \subset [n]$ such that $\text{Parity}_A|_C = \text{Parity}_{A'}|_C$. Here, $\text{Parity}_A : \{0,1\}^n \to \{0,1\}$ means $\text{Parity}_A(x) = \oplus_{i \in A} x_i$.

In other words, we want to show that such an f cannot approximate a parity function on a code where the parity size can be made large enough so that any f having the same complexity cannot perform the approximation if it is going to do the direct thing: concentrate on a smallest weight representation and try to approximate it.

The most direct thing to try is a reduction, i.e., try to argue that if there is such an f with a small circuit complexity, then there is a small circuit of constant depth that can approximate a plain regular parity function to get a contradiction. Unfortunately we can argue that if this is going to work then the minimum relative distance of C must be zero in the limit, which is not a typical case. By "if this is going to work", we mean the natural meaning of a reduction in this setting: find a linear encoding map of E of C, $E : \{0,1\}^u \to C$, i.e., a bijective $\mathbb{F}_2$-linear map, such that Parity_A pulls back by E to a parity function $\text{Parity}_B : \{0,1\}^u \to \{0,1\}$, i.e., $\text{Parity}_A \circ E = \text{Parity}_B$ for some $B \subset [u]$, satisfying

- $|B|$ is large, and
- for each i, $1 \leq i \leq n$, $E_i(r)$ is easy to compute as a function of $r|_B$ by a parity of weight s_i, when $r|_{[u]\setminus B}$ is fixed to some desirable value.

This means that the total number of ones in the columns of the generator matrix of E corresponding to B is at most $\sum_{i=1}^{n} s_i$, and hence at least one of those columns must have a (nonzero) weight less than or equal to $\sum_{i=1}^{n} s_i/|B|$. Therefore the minimum distance of C can be at most $\sum_{i=1}^{n} s_i/|B| = o(n)$ when the s_i's are small enough and $|B|$ is large enough to get a result. This reduction framework is essentially the mentality of Nisan generator for AC_0 (see Section 5.1.2.2) which is a linear code whose minimum dual distance is poly-logarithmic, but has also a zero minimum relative distance in the limit by the above counting argument.

The meaning of the failure of the reduction when the minimum distance of the code C grows linearly with the block length is that the extra information given to the original AC_0 circuit is written in a format which is too hard for a constant depth circuit of a small size to reproduce even partially.

The same reasoning applies to the case when we have a low-degree polynomial instead of an AC_0 circuit. The reduction fails when the minimum distance of the code grows linearly with the block length because the extra information given to the original polynomial is written in a format which is too hard for a low-degree polynomial to reproduce even partially.

Note that, typically, the minimum distance of C grows linearly with n since a random C, satisfying constraints (1) and (2) in Problem 5.8.2 (Parity with encrypted linear help), with for instance $k = \log^{O(1)} n$, has minimum distance $n/2 - o(n)$.

5.8.3 The algebraic setting

5.8.3.1 Cayley graphs based on linear codes

Let C be a binary linear code of block length n and dual distance d. Thus $C \subset \mathbb{F}_2^n$ is a linear space and the minimum distance of $Q \stackrel{\text{def}}{=} C^{\perp}$ is d. Consider the space $X \stackrel{\text{def}}{=} \mathbb{Z}_2^n/Q$ which is isomorphic to C as $\mathbb{F}_2$-vector spaces. The group $\mathbb{Z}_2^n$ when seen as the Hamming cube is a Cayley graph generated by $e_1, \ldots, e_n$, where e_i is the vector in $\mathbb{Z}_2^n$ with a single one at position i. Consider the quotient graph structure on the group X resulting from modding out the Cayley graph $\mathbb{Z}_2^n$ by the subgroup Q. This Cayley

graph is generated by $\bar{e}_1, \ldots, \bar{e}_n$, where if $x \in \mathbb{Z}_2^n$ we mean by $\bar{x}$ the corresponding element in $\mathbb{Z}_2^n/Q$. In other words, $\bar{x}$ and $\bar{y}$ are connected by an edge if $\bar{x} = \bar{y} + \bar{e}_i$ for some $\bar{e}_i$. A fundamental domain or a Dirichlet region (by borrowing from Riemann surfaces terminology) of X in $\mathbb{Z}_2^n$ is any subset D of the Voronoi cell of zero in $\mathbb{Z}_2^n$ with respect to Q such that D is maximal with the property that no two points in Q are equivalent by translations via elements in Q. Since any such D contains the $\lfloor d/2 \rfloor$-neighborhood of zero in $\mathbb{Z}_2^n$, we see that any $\lfloor d/2 \rfloor$-neighborhood of a node in X looks exactly like the $\lfloor d/2 \rfloor$-neighborhood of zero in the Hamming cube $\mathbb{Z}_2^n$, i.e., they are isomorphic as graphs. Thus, if d is large, X looks locally like the Hamming cube $\mathbb{Z}_2^n$. In other words, $\mathbb{Z}_2^n$ is in a suitable sense a combinatorial covering of X. Note that X will have multiple edges or equivalently have degree below n if and only if we are in the extreme case $d = 1$. Assume for the rest that $d > 1$.

The local properties of X such as the volumes of small neighborhoods and expansion of sets inside small neighborhoods are the same as those of $\mathbb{Z}_2^n$. To study other properties such as expansions of more spread sets, consider the Laplacian of the graph X. Let $L(X)$ be the space of complex (or real, it does not make a difference) valued functions on X. As a linear operator on $L(X)$, the Laplacian Δ of X is given by

$$(\Delta f)(\bar{x}) = \sum_{i=1}^{n} f(\bar{x}) - f(\bar{x} + \bar{e}_i), \quad f \in L(X).$$

The complex characters of X are $\{\mathcal{X}_z(\bar{x}) \stackrel{\text{def}}{=} (-1)^{xz}\}_{z \in C}$, where $C = Q^{\perp}$ is the dual of Q, i.e., the original code we started from. Note that $xz \stackrel{\text{def}}{=} \sum_i x_i z_i$ and is independent of the choice of z in $\bar{z}$. These characters form an orthogonal basis for $L(X)$.

It is not hard to show that the eigenvalues of the Laplacian Δ are exactly twice the weights of the codewords of in C. A proof can be found in [AR94]. Moreover, the eigenspace of the eigenvalue $\lambda \in 2w(C)$, w meaning the weight function, is spanned by the characters $\{\mathcal{X}_z\}_{z \in C \text{ s.t.} w(z) = \lambda/2}$.

Hence, specifically, the smallest nonzero eigenvalue of Δ is $\lambda_1(X) = 2d_{min}(C)$, where $d_{min}(C)$ is the minimum distance of C. Thus we can get a bound on the

Cheeger constant of X

$$h(X) \stackrel{\text{def}}{=} \min_{A,B \subset \mathbb{Z}_2^n/Q; A \cap B = \emptyset} \frac{e(|A|,|B|)}{min\{|A|,|B|\}},$$

with e meaning the number of edges between A and B. From the bound $h(X) \geq \lambda_1(X)/2$, we obtain

$$h(X) \geq d_{min}(C). \tag{5.30}$$

5.8.3.2 Equations on codes versus difference equations on graphs

Let F be a field whose characteristic is not equal to 2, and C and X as above. Let $L(C,F)$ be the F-vector space of F-valued functions on C, and $L(X,F)$ be the F-vector space of F-valued functions on X. We can view $L(C,F)$ as an F-algebra under pointwise multiplications in which case we denote it by $A(C,F)$, and we can view $L(X,F)$ as an F-algebra under convolution $*$

$$(a * b)(\bar{z}) = \sum_{\bar{y}} a(\bar{y}) b(\bar{z} + \bar{y}),$$

in which case we denote by $\hat{A}(X,F)$. Thus we have the F-algebras isomorphism: $A(C,F) \hat{\rightarrow} \hat{A}(X,F)$, where $\hat{}$ is the Fourier transform

$$\hat{f}(\bar{z}) = \frac{1}{|C|} \sum_x f(x) \mathcal{X}_{\bar{z}}.$$

Note that $\hat{A}(X,F)$ is isomorphic to the group algebra

$$F[X] \stackrel{\text{def}}{=} \{r = \textstyle\sum_{\bar{z}} \bar{z} a(\bar{z}) \mid a : X \rightarrow F\}$$

via $\hat{A}(X,F) \tilde{\rightarrow} F[X]$, $\tilde{a} = \sum_{\bar{z}} a(\bar{z}) \bar{z}$.

Now, let $f \in L(C,F)$, and consider the set $V(f)$ of zeros of f, i.e.,

$$V(f) = \{x \in C \mid f(x) = 0\}.$$

We are interested in the number of zeros. We have $|V(f)| = dim_F A(C,F)/(f)$, where (f) is the ideal generated by f in $A(C,F)$.

Thus by the isomorphism,

$$|V(f)| = |C| - dim_F(a)_* = dim_F(a)_*^{\perp},$$

where

$$a \stackrel{\text{def}}{=} \hat{f} \in \hat{A}(X,F),$$

$(a)_*$ is the ideal generated by a in $\hat{A}(X,F)$, and $(a)_*^{\perp}$ is the annihilator of $(a)_*$, i.e., the ideal in $\hat{A}(X,F)$ given by

$$(a)_*^{\perp} = \{r \in \hat{A}(X,F) : r * a = 0\}.$$

Note that $\hat{A}(X,F) = (a)_* \oplus_F (a)_*^{\perp}$.

We can think of $dim_F(a)_*$ directly as it is defined, i.e., the dimension of the F-vector space spanned by the translation $\{\sigma_{\bar{z}} a\}_{\bar{z} \in X}$ of the function $a : X \to F$. w Here $\sigma_{\bar{z}}$ means the translation operator on X given by $(\sigma_{\bar{z}} a)(\bar{y}) = a(\bar{y} + \bar{z})$. Note that this is the case because $\sigma_{\bar{z}} a = a * \delta_{\bar{z}}$, where

$$\delta_{\bar{z}}(\bar{y}) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \bar{y} = \bar{z} \\ 0 & \text{o.w.} \end{cases}.$$

Regarding $dim_F(a)_*^{\perp}$, in general, define, for $g \in L(X,F)$, the F-linear operator $T(g)$ on $L(X,F)$ by $T(g) = \sum_z g(z)\sigma_z$, i.e.,

$$(T(g)r)(\bar{z}) \stackrel{\text{def}}{=} \sum_{\bar{y}} g(\bar{y}) r(\bar{z} + \bar{y}) = (r * g)(z).$$

We can think of $T(g)$ as a Toplitz-like (or Laurent-like) operator with respect to the Fourier transform on C, and hence the terminology. When g is nonzero only in a small neighborhood $N_h(0)$ of zero, we can think of $T(g)$ also as a local averaging

operator,

$$(T(g)f)(\bar{z}) = \sum_{\bar{y} \in N_h(\bar{0})} g(\bar{y}) f(\bar{z} + \bar{y}),$$

i.e., as difference operator by analogy with a differential operator on a Lie group. Here, if $\bar{z} \in X$, $N_h(\bar{z})$ means the h-neighborhood of $\bar{z}$ in X, i.e.,

$$N_h(\bar{z}) = \{\bar{y} \in X : d(\bar{z}, \bar{y}) \leq h\}.$$

Back to $dim_F(a)^{\perp}_*$, we are interested in the case when $f = \mathcal{X}_{\bar{z}_1} - p$, where the degree of p is small and the weight of $\bar{z}_1$ is large. This means that $T(b)$, where $b = \hat{p}$, is actually a local averaging operator. Thus $dim_F(a)^{\perp}_*$ is the dimension of the space of solutions of the difference equation

$$T(b)g = \sigma_{\bar{z}_1} g$$

in g.

5.8.3.3 The algebraic formulation

Back to Problem 5.8.2 (Parity with encrypted linear help), in the language of Section 5.8.3.2, we have

Lemma 5.8.7 *Let $C \subset Z_2^n$ be a linear code. Consider the Cayley graph $X = \mathbb{Z}_2^n / C$. Let $\bar{z}_1 \in X$. Assume that:*

1) *The minimum distance of the dual $Q = C^{\perp}$ of C is above k, or, equivalently, X looks up to neighborhoods of radius $\lfloor (k+1)/2 \rfloor$ exactly like the Hamming cube.*

2) *The distance between $\bar{z}_1$ and zero is above k.*

Then each of (I) and (II) below is an equivalent formulation of Problem 5.8.2. Let $b \in L(X, F)$ such that b is nonzero only inside $N_h(\bar{0})$.

I) *Let $a = b - \delta_{\bar{z}_1}$, and I be the ideal of $\hat{A}(X, F)$ spanned by the translation $\{\sigma_{\bar{z}} a\}_{\bar{z} \in X}$*

of a. How large should k be so that

$$dim_F I \geq |X|(\frac{1}{2} - \epsilon)?$$

II) Let J be the ideal of $\hat{A}(X, F)$ consisting of the set of $f \in L(X, F)$ satisfying the equation $(T(b) - \sigma_{\bar{z}_1})f = 0$. How large should k be so that

$$dim_F J \leq |X|(\frac{1}{2} + \epsilon)?$$

5.8.3.4 Smolensky's argument is ungeneralizable

In this section we note that Smolensky's argument is ungeneralizable in the typical case when the minimum distance of the code C grows linearly with the block length n. The reason is that the smallest eigenvalue of the Laplacian of X is twice the minimum distance of C, and when the later is linear in n, X is, unlike the Hamming cube, a good expander for sets of relative size $1/2$.

Smolensky [Smo87] argued that parity on $\mathbb{Z}_2^n$ cannot be approximated by a degree $h = o(\sqrt{n})$ polynomial with a probability larger than $\frac{1}{2} + \Omega(\frac{h}{\sqrt{n}})$. Note that the error term $\frac{h}{\sqrt{n}}$ is very unlikely to be tight. For instance, when $h = 1$, and $F = \mathbb{F}_3$, the error term appear experimentally to be $2^{-\Theta(n)}$. Moreover, assuming Conjecture 5.7.14 (The symmetric optimum conjecture), the error term is zero when $F = \mathbb{Q}$ and $k > 16\pi h^2$.

In the setting of Lemma 5.8.7, the setting of Smolensky's result corresponds to the case when $C = \mathbb{Z}_2^n$, $X = \mathbb{Z}_2^n$ with the Hamming cube graph structure, and $\bar{z}_1$ is the all ones vector.

In the setting of (I) in Section 5.8.3.3, Smolensky's argument can be understood as follows. Find a large set of linearly independent $\sigma_{\bar{z}}a$'s. We can phrase Smolensky's approach as looking for a subset T of X such that $T \cap N_h(T + \bar{z}_1) = \emptyset$ which makes the $\{\sigma_{\bar{z}}a\}_{\bar{z} \in T}$ linearly independent simply because for each $z \in T$, $(\sigma_{\bar{z}}a)(\bar{z} + \bar{z}_1) = 1$ while $(\sigma_{\bar{z}'}a)(\bar{z} + \bar{z}_1) = 0$, $\forall \bar{z}' \neq \bar{z} \in T$.

In the Hamming cube, we can interpret the approach of Smolensky as setting $T = N_{n/2-h/2-1}(0)$ (Smolensky actually sets $T = N_{n/2-h}(0)$, but the above choice is

tighter). This satisfies $T \cap N_h(T + \bar{z}_1) = \emptyset$. Thus when $h = o(\sqrt{n})$, by Sterling's approximation, we get $\frac{|T|}{|X|} = \frac{1}{2} - \Theta(\frac{h}{\sqrt{n}})$, and hence $\frac{dim_F I}{|X|} = \frac{1}{2} - O(\frac{h}{\sqrt{n}})$.

In order for this approach to work, we need $\frac{|T|}{|X|} = \frac{1}{2} - o(1)$ and $T \cap N_h(T + \bar{z}_1) = \emptyset$. But, unfortunately, when the minimum distance of C is $\Theta(n)$, no such S exists in a useful way not even in the most trivial case. More precisely

Lemma 5.8.8 *When the minimum distance of C is $\Theta(n)$, there is no subset $T \subset X$, such that $T \cap N_h(T + \bar{z}_1) = \emptyset$ and $\frac{|T|}{|X|} = \frac{1}{2} - o(1)$, not even when $h = 1$.*

Proof. Assume that such a T exists when $h \geq 1$. We have, with ∂ meaning boundary, i.e., $\partial A \stackrel{\text{def}}{=} N_1(A) \backslash A$,

$$T \cup \partial T \subset N_h(T) \subset T^c + \bar{z}_1,$$

where the second inequality follows from the hypothesis $T \cap N_h(T + \bar{z}_1) = \emptyset$. Thus $|T| + |\partial T| \leq |T^c|$, i.e.,

$$|\partial T| \leq 2|T^c| - |X|. \tag{5.31}$$

Using (5.30) in Section 5.8.3.1, and then the hypothesis $\frac{|T|}{|X|} = \frac{1}{2} - o(1)$, we get

$$e(T^c, T) \geq d_{min}(C) \min\{|T^c|, |T|\} = d_{min}(C)|X|(\frac{1}{2} - o(1)),$$

hence

$$|\partial T| \geq \frac{e(T^c, T)}{n} \geq \delta_{min}(C)|X| \left(\frac{1}{2} - o(1)\right),$$

where $\delta_{min}(C) = d_{min}(C)/n$ is the minimum relative distance of C. Replacing in (5.31), we get

$$\frac{|T^c|}{|X|} \geq \frac{1}{2} + \delta_{min}(C)(\frac{1}{4} - o(1)),$$

a contradiction when $\delta_{min}(C) = \Theta(1)$. ∎

Note that, typically, the minimum distance of C grows linearly with n since a random C, satisfying constraints (1) and (2) in Problem 5.8.2 (Parity with encrypted linear help) with for instance $k = \log^{O(1)} n$, has minimum distance $n/2 - o(n)$.

5.8.4 The first relaxation: parity with moderately encrypted linear help

We know from Smolensky's work [Smo87] that parity on the Hamming cube cannot be approximated by a low-degree polynomial. The setting of Problem 5.8.2 (Parity with encrypted linear help) is parity on a linear code. Consider (1) and (2) in Problem 5.8.2. What (2) is saying is that all the realization of this help are difficult on the low-degree polynomial, i.e., the help is in a suitable sense encrypted and potentially useless. What (1) is saying is that the code is k-wise independent. This is supposed to make the help even more useless, but (2) is already a relatively strong condition alone. Consider dropping (1), and call the resulting problem *"parity with moderately encrypted linear help"*. Denote the corresponding minimum by $\tilde{K}^{5.8.2}_{parity}(n,h,\epsilon)_F$. Thus

$$\tilde{K}^{5.8.2}_{parity}(n,h,\epsilon)_F \geq K^{5.8.2}_{parity}(n,h,\epsilon)_F.$$

One advantage of dropping (1), is that this, essentially, reduces the problem to the case when the degree of the polynomial is 1, i.e., a linear form.

Lemma 5.8.9 $\tilde{K}^{5.8.2}_{parity}(n,1,\epsilon)_F \leq \tilde{K}^{5.8.2}_{parity}(n,h,\epsilon)_F \leq h\tilde{K}^{5.8.2}_{parity}(n^h,1,\epsilon)_F$.

Proof. The construction behind the second inequality is to use the monomials of the polynomial to build a linear map through which the original linear code maps to a new linear code where the original polynomial reduces to a linear form. The block length of the new code is at most the number of monomials which can be at most $\sum_{i=1}^{h}\binom{n}{i} \leq n^h$. Since the parity on the original code cannot be realized as a parity of k or fewer bits, and the each monomial is a parity of at most h bit, the parity on the new linear code will have no realization as a parity of k/h or fewer bits. ■

When the $F = \mathbb{F}_q$, $q = O(1)$, we can actually reduce further to the canonical case where the coefficients are all 1.

Problem 5.8.10 "Parity with moderately encrypted linear help: the symmetric mod-q-case": *Let $C \subset \mathbb{Z}_2^n$ be a linear code and let $\bar{z}_1 \in \mathbb{Z}_2^n/C^\perp$ such that the weight of*

$\bar{z}_1$ is above k. Let q be an odd prime, how large should k be so that for each $b_0 \in \mathbb{F}_q$,

$$Pr_{x \in C}[b_0 + \sum_{i=1}^{n} x_i = \mathcal{X}_{\bar{z}_1}(x)(\ mod\ q)] \leq \frac{1}{2} + \epsilon?$$

Denote the minimum such k by $K^{5.8.10}_{parity-symm}(n, \epsilon)_q$.

The reduction is as follows.

Lemma 5.8.11 $K^{5.8.10}_{parity-symm}(n, \epsilon)_q \leq \tilde{K}^{5.8.2}_{parity}(n, h, \epsilon)_{\mathbb{F}_q} \leq hK^{5.8.10}_{parity-symm}((q-1)n^h, \epsilon)_q$.

Proof. The additional component to the construction in Lemma 5.8.9 is to repeat each monomial as many times as its coefficient when the coefficient is regarded as an integer between 1 and $q-1$. ■

Observe that when $q = O(1)$ and $h = O(\log^{O(1)} n)$, the reduction is up to quasipolynomial factors an equivalence. Our interest in the setting $q = O(1)$ and $h = O(\log^{O(1)} n)$ is motivated by the AC_0 problem. See Section 5.8.2.2.

In the rationals case, i.e., when $F = \mathbb{Q}$, we can do something similar when we have bounds on the coefficients of the polynomial p. We can reduce the problem to a setting where all the coefficients are ± 1. See Section 5.8.5.3.

5.8.5 The linear-programming relaxation

Until Section 5.8.5.3, we will focus on the finite fields situation.

Consider Problem 5.8.10 (Parity with moderately encrypted linear help: the symmetric mod-q-case). We are optimizing on the linear codes $C \subset \mathbb{Z}_2^n$ and $\bar{z}_1 \in \mathbb{Z}_2^n / C^\perp$ such that the weight of $\bar{z}_1$ is above k, i.e., the weight of any $z_1 \in \bar{z}_1$ is above k. Equivalently, we are optimizing on μ, a probability measure on $\mathbb{Z}_2^n$, and $z_1 \in \mathbb{Z}_2^n$ such that:

1) there exists a linear code $C \subset \mathbb{Z}_2^n$ such that μ is the discrete probability measure supported by C, i.e., $\mu = \mu_C$, where $\mu_C \stackrel{\text{def}}{=} \frac{1}{|C|} 1_C$.

2) $w(z_1) \geq k + 1$

3) $E_\mu \mathcal{X}_{z+z_1} = 0$ for each $z \in \mathbb{Z}_2^n$ such that $w(z) \leq w(z_1) - 1$.

We can do this by using a $z_1 \in \bar{z}_1$ of minimal weight.

Consider relaxing (1) in the sense of Section 5.3.2 to

1') $\hat{\mu} \geq 0$, or, equivalently, $E_\mu \mathcal{X}_z \geq 0$, for each z in $\mathbb{Z}_2^n$.

Thus, Problem 5.8.10 (Parity with moderately encrypted linear help: the symmetric mod-q-case) relaxes to a linear program as follows.

Problem 5.8.12 "LP-relaxed parity with moderately encrypted linear help: the symmetric mod-q-case": *Let $z_1 \in \mathbb{Z}_2^n$ such that $w(z_1) \geq k+1$.*
Let q be an odd prime, $b_0 \in \mathbb{F}_q$, and $f : \mathbb{Z}_2^n \to \{0,1\}$ be given by

$$f(x) = \begin{cases} 1 & \text{if } w(x) + b_0 = \mathcal{X}_{z_1}(x) \pmod q \\ 0 & o.w. \end{cases}.$$

Primal question: How large should k be so that

$$\max_{\mu \in P} E_\mu f \leq \frac{1}{2} + \epsilon?$$

where P is the polytope of probability measures on $\mathbb{Z}_2^n$ such that

- $E_\mu \mathcal{X}_z \geq 0$ *for each z in $\mathbb{Z}_2^n$*
- $E_\mu \mathcal{X}_{z+z_1} = 0$ *for each $z \in \mathbb{Z}_2^n$ such that $w(z) \leq w(z_1) - 1$.*

Dual question: How large should k be so that there exists $g : \mathbb{Z}_2 \to \mathbb{R}$ such that

- $g \geq f$
- $E_{\mu_0} g \leq \frac{1}{2} + \epsilon$
- $\hat{g}(z) \leq 0, \forall z \notin N_k(z_1) \cup \{0\}$,

where μ_0 the uniform measure on $\mathbb{Z}_2^n$?

Note that dual calculations are along the lines of Section 5.2.3.

A natural question here is how small can $|C|$ be, or equivalently how large can $|C^{\perp}|$ be, subject to: (2) and (3) above with $\mu = \mu_C$? The other question is what bound can we conclude from the above relaxation, i.e., when we replace (1) with (1′)?

Problem 5.8.13 "LP-relaxed coset-size versus minimum-weigh problem": *Let* $f : \mathbb{Z}_2^n \to \{0,1\}$,

$$f(x) = \begin{cases} 2^n & \text{if } w(x) = 0 \\ 0 & o.w. \end{cases}.$$

Let $z_1 \in \mathbb{Z}_2^n$ *such that* $w(z_1) \geq d$, P *be the polytope of probability measures on* $\mathbb{Z}_2^n$ *such that*

- $E_\mu \mathcal{X}_z \geq 0$ *for each* z *in* $\mathbb{Z}_2^n$
- $E_\mu \mathcal{X}_{z+z_1} = 0$ *for each* $z \in \mathbb{Z}_2^n$ *such that* $w(z) \leq w(z_1) - 1$,

and consider

$$\tilde{M}_d = \max_{\mu \in P} E_\mu f.$$

(To be precise, we are maximizing also over z_1*, but by symmetry they are all the same when the weight is fixed).*
The dual is: $\tilde{M}_d = min\ E_{\mu_0} g$, $g : \mathbb{Z}_2^n \to \mathbb{R}$, *such that*

- $g \geq f$
- $\hat{g}(z) \leq 0, \forall z \notin N_{d-1}(z_1) \cup \{0\}$,

where μ_0 *the uniform measure on* $\mathbb{Z}_2^n$. *Thus* $M_d \leq \tilde{M}_d$, *where*

$$M_d = \max\{|Q| \ : \ Q \subset \mathbb{Z}_2^n \text{ linear and } z_1 \in \mathbb{Z}_2^n, \text{ s.t. min-weight}(Q + z_1) \geq d\}.$$

This is the case because with $Q = C^{\perp}$ as above and $d = k + 1$,

$$E_{\mu_C} f = 2^n \mu_C(0) = \frac{2^n}{|C|} = |Q|.$$

We will argue in Theorem 5.8.17 that $M_d = \tilde{M}_d = 2^{n-d}$, which is an indication of the goodness of the relaxation.

To compare the above linear program with Delsarte LP bound, we can phrase Delsarte LP bound as follows.

Problem 5.8.14 "Delsarte LP bound [Del73]*": Let* $f : \mathbb{Z}_2^n \to \{0,1\}$,

$$f(x) = \begin{cases} 2^n & \text{if } w(x) = 0 \\ 0 & o.w. \end{cases}.$$

Let d *be integer,* U *be the polytope of probability measures on* $\mathbb{Z}_2^n$ *such that*

- $E_\mu \mathcal{X}_z \geq 0$ *for each* z *in* $\mathbb{Z}_2^n$
- $E_\mu \mathcal{X}_z = 0$ *for each nonzero* $z \in \mathbb{Z}_2^n$ *such that* $w(z) \leq d-1$,

and consider

$$\tilde{N}_d = \max_{\mu \in U} E_\mu f.$$

The dual is: $\tilde{N}_d = min\ E_{\mu 0} g$, $g : \mathbb{Z}_2^n \to \mathbb{R}$ *such that*

a) $g \geq f$

b) $\hat{g}(z) \leq 0, \forall z \in \mathbb{Z}_2^n$ *such that* $w(z) \geq d$,

where μ_0 *the uniform measure on* $\mathbb{Z}_2^n$. *Thus* $N_d \leq \tilde{N}_d$, *where*

$$N_d = \max\{|Q| \ : \ Q \subset \mathbb{Z}_2^n \text{ linear s.t. min-dist}(Q) \geq d\}.$$

Note that Delsarte LP bound is in terms of Krawtchouk polynomials, but, by symmetry, Delsarte linear program (regardless of whether its origin is linear or non-linear codes) can be lifted to the above linear program. See Lemma 5.8.19. Note also that it is an old open question whether $N_d = \tilde{N}_d$.

A final remark is that the linear codes rate-versus-distance tradeoff problem corresponds to the maximal acceptance probability of an AND gate on n-bits when the input is randomly selected from a linear code satisfying a bound on its dual distance.

5.8.5.1 Fourier transform of weight-based functions

Let $S_1, \ldots, S_c$ be a partition of $[1:n]$, i.e., $S_i \cap S_j = \emptyset$ for each $i \neq j$ and $\cup_i S_i = [1:n]$. Here, and in what follows, we will be using the notation $[n_1 : n_2] \stackrel{\text{def}}{=} \{n_1, \ldots, n_2\}$. Let $s_i = |S_i|$, and consider $\pi : \mathbb{Z}_2^n \to \prod_{i=1}^c [0 : s_i]$, $x \mapsto (w(x|_{S_i}))_{i=1}^c$.

Let $f : \mathbb{Z}_2^n \to \mathbb{R}$ such that $f(x)$ depends only on $\pi(x)$. By abuse of notation, we can think of f as $f : \prod_{i=1}^c [0 : s_i] \to \mathbb{R}$. Consider the Fourier transform $\hat{f} : \mathbb{Z}_2^n \to \mathbb{R}$ of f. We have

$$\hat{f}(z) = \frac{1}{2^n} \sum_x f(x) \mathcal{X}_z(x) = \frac{1}{2^n} \sum_{w \in \prod_{i=1}^c [0:s_i]} f(w) \bar{\mathcal{K}}_w^{(s)}(z), \tag{5.32}$$

where $s \stackrel{\text{def}}{=} (s_1, \ldots, s_c)$, and $\bar{\mathcal{K}}_w^{(s)} : \mathbb{Z}_2^n \to \mathbb{R}$ is given by

$$\begin{aligned} \bar{\mathcal{K}}_w^{(s)}(z) &\stackrel{\text{def}}{=} \sum_{x;\pi(x)=w} \mathcal{X}_z(x) = \sum_{x;w(x|_{S_i})=w_i} \prod_{i=1}^c \mathcal{X}_{z|_{S_i}}(x|_{S_i}) \\ &= \prod_{i=1}^c \sum_{x \in \mathbb{Z}_2^{S_i}} \mathcal{X}_{z|_{S_i}}(x) = \prod_{i=1}^c \mathcal{K}_{w_i}^{(s_i)}(w(z|_{S_i})). \end{aligned}$$

Here, if a and b are integers, $\mathcal{K}_b^{(a)} : [0:a] \to \mathbb{R}$, and by abuse of notation $\mathcal{K}_b^{(a)} : \mathbb{Z}_2^a \to \mathbb{R}$, is the Krawtchouk polynomial given by

$$\mathcal{K}_b^{(a)}(y) \stackrel{\text{def}}{=} \sum_{x \in \mathbb{Z}_2^a; w(x)=b} \mathcal{X}_y(x) = \sum_t (-1)^t \binom{v}{t} \binom{a-v}{b-t} \stackrel{\text{def}}{=} \mathcal{K}_b^{(a)}(v), \tag{5.33}$$

for $y \in \mathbb{Z}_2^a$ with $v = w(y)$.

Thus, $\hat{f}(z) = \hat{f}(\pi(z))$ and $\hat{f} : \prod_{i=1}^c [0 : s_i] \to \mathbb{R}$ is given by

$$\hat{f}(l) = \frac{1}{2^n} \sum_w f(w) \bar{\mathcal{K}}_w^{(s)}(l) \text{ and } f(w) = \sum_l \hat{f}(l) \bar{\mathcal{K}}_l^{(s)}(w), \tag{5.34}$$

where $\bar{\mathcal{K}}_w^{(s)} : \mathbb{Z}_2^n \to \mathbb{R}$, and by abuse of notation $\bar{\mathcal{K}}_w^{(s)} : \prod_{i=1}^c [0 : s_i] \to \mathbb{R}$, is a product of

Krawtchouk polynomials

$$\bar{\mathcal{K}}_w^{(s)}(z) \stackrel{\text{def}}{=} \sum_{x;\pi(x)=w} \mathcal{X}_z(x) = \prod_{i=1}^{c} \mathcal{K}_{w_i}^{(s_i)}(l_i) \stackrel{\text{def}}{=} \bar{\mathcal{K}}_w^{(s)}(l), \tag{5.35}$$

for $z \in \mathbb{Z}_2^n$ with $l = \pi(z)$.

The direct way to see the many of the properties Krawtchouk polynomials is to start from the hypercube. Two such immediate properties are

$$E_\lambda \mathcal{K}_{b_1}^{(a)} \mathcal{K}_{b_2}^{(a)} = 2^n \lambda(b_1) \delta_{b_1,b_2} \text{ and } \lambda(b_1) \mathcal{K}_{b_2}^{(a)}(b_1) = \lambda(b_2) \mathcal{K}_{b_1}^{(a)}(b_2), \tag{5.36}$$

where λ is the probability measure on $[0:a]$ induced via w by the uniform probability measure on $\mathbb{Z}_2^a$, i.e., λ is the binominal measure given by $\lambda(b) \stackrel{\text{def}}{=} \frac{1}{2^a}\binom{a}{b}$. See for instance [MRRW77, Sze75] for other properties Krawtchouk polynomials.

Similar relations apply in the higher dimensional situation under consideration. Let β be the probability measure on $\prod_{i=1}^{c}[0:s_i]$ induced via π by the uniform probability measure on $\mathbb{Z}_2^n$, i.e.,

$$\beta(w) \stackrel{\text{def}}{=} \frac{1}{2^n} \prod_{i=1}^{c} \binom{s_i}{w_i} \tag{5.37}$$

is a product binomial measure. Then the 1-dimensional relations immediately generalize as

$$E_\beta \bar{\mathcal{K}}_{l_1}^{(s)} \bar{\mathcal{K}}_{l_2}^{(s)} = 2^n \beta(l_1) \delta_{l_1,l_2} \text{ and } \beta(w) \bar{\mathcal{K}}_l^{(s)}(w) = \beta(l) \bar{\mathcal{K}}_w^{(s)}(l). \tag{5.38}$$

Replacing in (5.34), we get

$$2^n \beta(l) \hat{f}(l) = E_\beta f \bar{\mathcal{K}}_l^{(s)} \text{ and } \beta(w) f(w) = E_\beta \hat{f} \bar{\mathcal{K}}_w^{(s)}. \tag{5.39}$$

5.8.5.2 The low dimensional equivalent problems

Accordingly,

Lemma 5.8.15 *Problem 5.8.12 (LP-relaxed parity with moderately encrypted linear help: the symmetric mod-q-case) is equivalent to the following.*

Let q be an odd prime, $b_0 \in \mathbb{F}_q$, and $f : [0:k+1] \times [0:n-k-1] \to \{0,1\}$ be

given by

$$f(w_1, w_2) = \begin{cases} 1 & \text{if } w_1 + w_2 + b_0 = (-1)^{w_1} \pmod q \\ 0 & o.w. \end{cases}.$$

Primal question: How large should k be so that

$$\max_{\mu \in P} E_\mu f \leq \frac{1}{2} + \epsilon?$$

where P is the polytope of probability measures on $[0 : k+1] \times [0 : n-k-1]$ such that

- $E_\mu \bar{\mathcal{K}}_l^{(s)} \geq 0$ *for each* l
- $E_\mu \bar{\mathcal{K}}_l^{(s)} = 0$, *when* $l_1 > l_2$,

and $s = (k+1, n-k-1)$.
Dual question: How large should k be so that there exists $g : [0 : k+1] \times [0 : n-k-1] \to \mathbb{R}$ such that

- $g \geq f$
- $E_\beta g \leq \frac{1}{2} + \epsilon$
- $\hat{g}(l_1, l_2) \leq 0$, *when* $l_1 \leq l_2$ *and* $l_1 + l_2 \neq 0$,

where β is the product binomial measure on $[0 : k+1] \times [0 : n-(k+1)]$?

Proof. First, without loss of generality , we can assume that z_1 in Problem 5.8.12 has weight exactly $k+1$ since the problem is about how large should the lower bound on the weight of z_1 be. Let S_1 be the support of z_1, and $S_2 = [n] \backslash S_1$. Thus $|S_1| = k+1$ and $|S_2| = n-k-1$. The equivalence follows from noting that the objective function is a linear function invariant under permutations of S_1 and S_2, and all the constraints are symmetric with respect to such permutations. To get the equivalence, we can start with an optimum of the original problem, apply all such permutations, then average to get an invariant optimum.

■

By the same argument, we get

Lemma 5.8.16 *Problem 5.8.13 (LP-relaxed coset-size versus minimum-weigh problem) is equivalent to the following. Let* $f : [0:d] \times [0:n-d] \to \mathbb{R}$,

$$f(w_1, w_2) = \begin{cases} 2^n & \text{if } w_1 + w_2 = 0 \\ 0 & o.w. \end{cases}.$$

Primal: $\tilde{M}_d = \max_{\mu \in P} E_\mu f$, *where* P *is the polytope of probability measures on* $[0 : k+1] \times [0 : n-k-1]$ *such that*

- $E_\mu \bar{\mathcal{K}}_l^{(s)} \geq 0$ *for each* l
- $E_\mu \bar{\mathcal{K}}_l^{(s)} = 0$, *when* $l_1 > l_2$,

and $s = (d, n-d)$.
Dual: $\tilde{M}_d = min\ E_\beta g$, $g : [0:d] \times [0:n-d] \to \mathbb{R}$ *such that*

a) $g \geq f$

b) $\hat{g}(l_1, l_2) \leq 0$, *when* $l_1 \leq l_2$ *and* $l_1 + l_2 \neq 0$,

where β *is the product binomial measure on* $[0:d] \times [0:n-d]$.

Theorem 5.8.17 $\tilde{M}_d = M_d = 2^{n-d}$

Proof. First we show that $\tilde{M}_d \leq 2^{n-d}$. Set

$$\hat{g}(l_1, l_2) = \begin{cases} 2^{n-d} & \text{if } l_2 = 0 \\ 0 & o.w. \end{cases}.$$

Hence (b) is satisfied. To see why (a) is satisfied, note first that

$$\hat{g}(w_1, w_2) = \sum_{l_1, l_2} \hat{g}(l_1, l_2) \mathcal{K}_{l_1}^{(d)}(w_1) \mathcal{K}_{l_2}^{(n-d)}(w_2) = 2^{n-d} \sum_{l_1} \mathcal{K}_{l_1}^{(d)}(w_1) \mathcal{K}_0^{(n-d)}(w_2).$$

But $\mathcal{K}_0^{(n-d)}(w_2) = 1$, and

$$\sum_{l_1} \mathcal{K}_{l_1}^{(d)}(w_1) = \begin{cases} 2^d & \text{if } w_1 = 0 \\ 0 & o.w. \end{cases}.$$

Thus (a) is satisfied also. The bound $\tilde{M}_d \leq 2^{n-d}$ then follows since $E_\beta g = \hat{g}(0,0) = 2^{n-d}$.

The fact that $M_d \geq 2^{n-d}$ follows from the first example to try. Namely, set $Q = \mathbb{Z}_2^{n-d}0^d$ and $z_1 = 0^{n-d}1^d$. ■

Question 5.8.18 *A purely combinatorial proof of $M_d = 2^{n-d}$?*

Similarly,

Lemma 5.8.19 *Problem 5.8.14 (Delsarte LP bound) is equivalent to the following. Let $f : [0:n] \to \mathbb{R}$,*

$$f(w) = \begin{cases} 2^n & \text{if } w = 0 \\ 0 & o.w. \end{cases}.$$

Primal: $\tilde{N}_d = \max_{\mu \in U} E_\mu f$, where U is the polytope of probability measures on $[0:n]$ such that

- $E_\mu \mathcal{K}_l^{(n)} \geq 0$ *for each* l
- $E_\mu \mathcal{K}_l^{(n)} = 0$, *when* $1 \leq l \leq d$.

Dual: $\tilde{N}_d = min\ E_\beta g$, $g : [0:n] \to \mathbb{R}$ such that

- $g \geq f$
- $\hat{g}(l) \leq 0$, *when* $l \geq d$,

where β is the binomial measure on $[0:n]$.

Note that this the classical statement of Delsarte linear program.

5.8.5.3 The characteristic-zero case

Consider Problem 5.8.2 (Parity with encrypted linear help) when the base field is $F = \mathbb{Q}$. In order to be able to reduce to a canonical linear program similar to the one in Lemma 5.8.15, we need a reduction to a canonical problem like Problem 5.8.10 (Parity with moderately encrypted linear help: the symmetric mod-q-case). We

partially have such a problem from Lemma 5.8.9, but this is not directly reducible to a low dimensional LP after relaxation.

We can however reduce to a problem similar to Problem 5.8.10 when we have bounds on the coefficients of the polynomial. If we can assume such bounds, we can naturally reduce to the situation of a linear form where all the coefficients are ± 1. As far as the AC_0 application is concerned, we can assume such bounds. We can extract from the construction in [BRS91, ABFR94] the stronger version of Theorem 5.7.7 where in addition to (1) and (2), we can assume that:

3) The sum of the absolute values of the coefficients of each p_α is $2^{O(h \log M)}$.

Consider Problem 5.8.2 (Parity with encrypted linear help) in the setting when the base field is $\mathbb{Q}$. Restrict the problem further to the situation where we are given a bound L in such a way that p satisfies:

3) $up = \sum_{\bar{z}} a(\bar{z}) \mathcal{X}_z$ where $u \neq 0 \in \mathbb{Z}$ and $a : \mathbb{Z}_2^n / C^\perp \to \mathbb{Z}$ is such that $\sum_{\bar{z}} |a(\bar{z})| \leq L$.

Let $K^{5.8.2}_{parity-bounded-coef}(n, h, L, \epsilon)_{\mathbb{Q}}$ be the corresponding minimum, and let $\tilde{K}^{5.8.2}_{parity-bounded-coef}(n, h, L, \epsilon)_{\mathbb{Q}}$ be the minimum corresponding to dropping (1) as in Section 5.8.4.

Then, we can reduce to the following variation of Problem 5.8.10 (Parity with moderately encrypted linear help: the symmetric mod-q-case).

Problem 5.8.20 "Parity with moderately encrypted linear help: the symmetric rational-case": *Let $C \subset \mathbb{Z}_2^n$ be a linear code and let $\bar{z}_1 \in \mathbb{Z}_2^n / C^\perp$ such that the weight of $\bar{z}_1$ is above k.*

How large should k be so that for each partition S', S'' of $[n]$, and for each $b_0 \in \mathbb{Z}$ and $u \neq 0 \in \mathbb{Z}$,

$$Pr_{x \in C}[b_0 + \sum_{i \in S'} x_i - \sum_{i \in S''} x_i = u \mathcal{X}_{\bar{z}_1}] \leq \frac{1}{2} + \epsilon?$$

Denote the minimum such k by $K^{5.8.20}_{parity-symm}(n, \epsilon)_0$.

The reduction is as follows

$$\tilde{K}^{5.8.2}_{parity-bounded-coef}(n, h, L, \epsilon)_{\mathbb{Q}} \leq h K^{5.8.20}_{parity-symm}(2L, \epsilon)_0.$$

This proof is along the lines of Lemma 5.8.11.

The linear-programming relaxation then translates naturally as follows. We only have to change the the definition of f in Problem 5.8.12 (LP-relaxed parity with moderately encrypted linear help: the symmetric mod-q-case) as follows.

We have a partition S', S'' of $[n]$, $b_0 \in \mathbb{Z}$, $u \neq 0 \in \mathbb{Z}$, and $f : \mathbb{Z}_2^n \to \{0,1\}$ is given by

$$f(x) = \begin{cases} 1 & \text{if } w(x|_{S'}) - w(x|_{S''}) + a_0 = u\mathcal{X}_{z_1}(x) \\ 0 & o.w. \end{cases}.$$

As for the low dimensional equivalent problem, we get the following.

Problem 5.8.21 "LP-relaxed parity with moderately encrypted linear help: the symmetric rational-case": *Let s_1, s_2, s_3, s_4 be nonnegative integers such that $s_1 + s_2 = k+1$ and $s_1 + s_2 + s_3 + s_4 = n$. Let b_0 and u be integers such that u is nonzero. Let $f : \prod_{i=1}^4 [0 : s_i] \to \{0,1\}$ be given by*

$$f(w) = \begin{cases} 1 & \text{if } w_1 + w_3 - w_2 - w_4 + b_0 = u(-1)^{w_1+w_2} \\ 0 & o.w. \end{cases}.$$

Primal question: How large should k be so that

$$\max_{\mu \in P} E_\mu f \leq \frac{1}{2} + \epsilon?$$

where P is the polytope of probability measures on $\prod_{i=1}^4 [0 : s_i]$ such that

- $E_\mu \bar{\mathcal{K}}_l^{(s)} \geq 0$ *for each* l
- $E_\mu \bar{\mathcal{K}}_l^{(s)} = 0$, *when* $l_1 + l_2 > l_3 + l_4$,

and $s = (s_1, s_2, s_3, s_4)$.
Dual question: How large should k be so that there exists $g : \prod_{i=1}^4 [0 : s_i] \to \mathbb{R}$ such that

- $g \geq f$
- $E_\beta g \leq \frac{1}{2} + \epsilon$

- $\hat{g}(l) \leq 0$, *when* $l_1 + l_2 \leq l_3 + l_4$ *and* $l_1 + l_2 + l_3 + l_4 \neq 0$,

and β *is the product binomial measure on* $\prod_{i=1}^{4}[0 : s_i]$*?*

Let $K^{5.8.21}_{LP-parity-symm}(n, \epsilon)_0$ *be the minimum such* k*.*

This relates to the AC_0 situation as follows.

Lemma 5.8.22 *There are some absolute positive constants* c, c'*, such that if* $f : \{0,1\}^n \to \{0,1\}$ *is computable by an AND-OR of size-*M *depth-*d *unbounded-fanin circuit,* C *is a block-length-*n *binary linear code whose dual has minimum distance above* k*, and*

$$k \geq h_0^d K^{5.8.21}_{LP-parity-symm}\left(2L, (1-\alpha)\frac{\epsilon}{2}\right)_0,$$

where $h_0 = c \log \frac{2M}{\alpha\epsilon} \log M$*,* $L = 2^{c' h_0^d \log M}$*, and* $0 < \alpha \leq 1$ *is tunable, then* μ_C *can* ϵ*-fool* f*.*

5.8.6 Some generalities

In the setting of finite fields, Problem 5.8.3 (Linear codes versus low-degree F-polynomial-equations on binary input variables) is a special case of the following problem, which is essentially about approximating the number of solutions of low-degree character equations on abelian groups over small finite fields.

Let G be a finite abelian group together with some grading on its group of characters $\hat{G}$. Let H be a subgroup of G such that all the defining characters of H have high degree. Let p be a low-degree function on G taking values in a small finite field F whose characteristic does not divide the order of G. The problem is about when can we guarantee that the fraction of elements g of G satisfying the equation $p(g) = 0$ does not change significantly when we restrict g to be an element of H.

So, Problem 5.8.3 corresponds to the case when $G = \mathbb{Z}_2^n$, $F = \mathbb{F}_q$ with $q = 3$ for instance, and the grading is coming from the Hamming cube.

A similar example is, for instance, when $G = \mathbb{Z}_5^n$, $F = \mathbb{F}_2$, and the grading is coming from the Cartesian product of n copies of the circle graph on $\mathbb{Z}_5$.

5.9 Open problems

We conclude with the resulting open questions:

- The symmetric optimum conjecture (Conjecture 5.7.14), and the original low-degree polynomial predictors problem (Problem 5.7.8).
- A bound on the linear programs in Lemma 5.8.15 and Section 5.8.5.3, and the original parity with encrypted linear help problem (Problem 5.8.2) from the algebraic perspective in Lemma 5.8.7.
- The k-wise independent versus AC_0 problem (Conjecture 5.7.3 and Problem 5.7.1), and the δ-biased versions in Conjecture 5.7.4 and Problem 5.7.2.
- Classify the nonlinear extreme points of the convex polyptopes Q_k and P_k, i.e., Problem 5.3.5, Conjecture 5.3.2, and Problem 5.3.3.
- Problem 5.5.5: how weak are probability measures induced by the uniform measure via arbitrary depth-1 maps?

5.9.1 The power of the quadratic residues PRG

One of the basic questions motivating the start of the research in this chapter is the quadratic residues PRG. See Section 5.1.2.3. Consider the quadratic residues PRG over a prime field $\mathbb{F}_q$ for nondegeneracy reasons.

The mystery of the way quadratic residues are oddly distributed for a given prime promises great derandomization capabilities and intrigued people long before complexity theory existed. It is very tempting to conjecture that they look random even to something as powerful as all polynomially sized circuits. A more modest start is the following:

- Show that the quadratic residues PRG can $o(1)$-fool all DNF formulas of size M on n bits when the prime q is made polynomially large enough in terms of n and M.

- Show that the quadratic residues PRG can $o(1)$-fool all read-once oblivious branching programs of size S on n bits when the prime q is made polynomially large enough in terms of n and $\log S$.

Note that constructing polynomial complexity PRG's for small DNF formulas or low-memory read-once oblivious branching programs is a fundamental open question because it is essentially not known how to (unconditionally) construct polynomial complexity PRG's for any relatively-general computational model (If we exclude models such as polynomial size decision trees, and DNF formulas where the number of inputs per clause is bounded above by a constant).

It follows from Weil's bound that this PRG has the $2n/\sqrt{q}$-bias property as noted in [AGHP92]. The quadratic-residues PRG seems to have much stronger properties, but in all the proof attempts we were trying, we were not using more than this property, which lead us to the following question: what can we conclude from this property alone? i.e., what kind of functions can be derandomized by this property alone? Is this property alone sufficient to derandomize AC_0 (At this point the transition from a polynomial complexity objective to a quasipolynomial complexity objective happened)? or small read-once oblivious branching programs? And, naturally, the following related problems appeared. What can we conclude from the k-wise independence property alone? i.e., what kind of functions can be derandomized by this property? Is it sufficient to derandomize AC_0 when k is poly-logarithmically large? What kind of functions can be derandomized by arbitrary linear-codes-based k-wise independent probability measures? The rest of the story is in the beginning of this chapter.

Bibliography

[Ajt99] Miklos Ajtai, "Determinism versus Non-Determinism for Linear Time RAMs", Proc. 31st Annual ACM Symposium on Theory of Computing (1999), pp. 632-641.

[AW85] M. Ajtai and A. Wigderson, "Deterministic Simulation of Probabilistic Constant Depth Circuits", Proc. 26th IEEE Symposium on Foundations of Computer Science (1985), pp. 11-19.

[ABN+92] N. Alon, J. Bruck, J. Naor, M. Naor, and R. Roth, "Construction of asymptotically good, low-rate error-correcting codes through pseudo-random graphs", IEEE Transactions on Information Theory, 38 (1992), pp. 509-516.

[AGHP92] N. Alon, O. Goldreich, J. Hastad, and R. Peralta, "Simple Constructions of Almost k-wise Independent Random Variables", Random Stuctures and Algorithms, Vol 3, No 3, 1992, pp 289-304.

[AR94] N. Alon and Y. Roichman, "Random Cayley Graphs and Expanders", Random Structures and Algorithms, 5 (2), pp. 271-284 (1994).

[ABFR94] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich, "The Expressive Power of Voting Polynomials", Combinatorica 14(2), pp. 135-148 (1994).

[Bar01] Alexander Barg, private correspondence, Dec. 2001.

[BMS03] L. Bazzi, M. Mahdiam, and D. Spielman, "The Minimum Distance of Turbo Like Codes", submitted for publication review, 2003.

[BM03a] L. Bazzi and S. Mitter, "Encoding Complexity versus Minimum Distance", submitted for publication review, 2003.

[BM03b] L. Bazzi and S. Mitter, "Some Constructions of Codes from Group Actions", submitted for publication review, 2003.

[BRS91] R. Beigel, N. Reingold, and D. Spielman, "The Perceptron Strikes Back", Proc. 6th Annual IEEE Conference on Structure in Complexity Theory, pp. 286-291 (1991).

[BDMP98] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding", IEEE Transactions on Information Theory, vol. 44 (3), pp. 909-926, May 1998.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajashima. "Near Shannon limit error-correction coding: Turbo Codes", Proc. 1993 IEEE International Conference on Communications, pp. 1064-1070, Geneva, Switzerland, May 1993.

[BM82] M. Blum and S. Micali. "How to generate Cryptographically Strong Sequences of Pseudo-Random Bits", SIAM journal on Computing, Vol 31, pp. 850-864 (1984).

[Bre01] M. Breiling, "A Logarithmic Upper Bound on the Minimum Distance of Turbo Codes", submitted to the IEEE Transactions on Information Theoy, 2001.

[BS92] J. Bruck and R. Smolensky, "Polynomial threshold functions, AC0 functions,. and spectral norms", SIAM J. Comp., 21, pp. 33-42 (1992).

[Bur65] M. Burrow, "Representation theory of finite groups", Academic Press, 1965.

[CPW69] C. L. Chen, W. W. Peterson, and E. J. Weldon, Jr., "Some results on quasi-cyclic codes", Information and Control, 15(5):407-423, November 1969.

[Che92] V. Chepyzhov, "New Lower Bounds for Minimum Distance of Linear Quasi-Cyclic and Almost Linear Cyclic Codes", Problemy Peredachi Informatsii, 28, pp. 33-44, January 1992.

[DCC01] DIMACS Workshop on Codes and Complexity, December 2001.

[CR62] C. Curtis and I. Reiner, "Representation Theory of Finite Groups and Associative Algebras", Wiley Interscience, 1962.

[Del73] P. Delsarte, "An algebraic approach to the association schemes of coding theory", Philips Res. Reports Suppl. 10 (1973).

[DJM98] D. Divsalar, H. Jin, and R.J. McEliece, "Coding theorems for "turbo-like" codes", Proc. 36th Annual Allerton Conference on Comm., Control, and Computing, pp. 210-210, Sept. 1998.

[Elk01] Noam D. Elkies, "Excellent nonlinear codes from modular curves", Proc. 33rd Annual ACM Symposium on Theory of Computing (2001), pp. 200-208.

[Gal63] R. G. Gallager, "Low Density Parity Check Codes", Monograph, M.I.T. Press, 1963.

[Gol92] Oded Goldreich 92, private corespondence between Oded Goldreich and Madhu Sudan, July 2002.

[GRS00] O. Goldreich, D. Ron, and M. Sudan, "Chinese remaindering with errors", IEEE Transactions on Information Theory, 46 (2000), pp. 1330-1338.

[Gop70] Goppa, V.D., "A new class of linear error-correcting codes", Problems of Info. Transmissions, 6, pp 207-212 (1970).

[Has86] J. Hastad, "Computational Limitations for Small Depth Circuits", PhD dissertation, MIT , Cambridge, MA, 1986.

[HKSS94] A. R. Hammons Jr., P. Vijay Kumar, A. R. Calderbank, N. J. A. Sloane and P. Sole, "The Z4-Linearity of Kerdock, Preparata, Goethals and Related Codes", IEEE Transactions on Information Theory, 40 (1994), pp. 301-319.

[IW97] R. Impagliazzo and A. Wigderson, "P = BPP if E Requires Exponential Circuits: Derandomizing the XOR Lemma", Proc. 29th Annual ACM Symposium on the Theory of Computing, (1997), pp. 220-229.

[JM99] H. Jin and R.J. McEliece, "RA Codes Achieve AWGN Channel Capacity", Proc. 13th International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes, volume 1719 of Lecture Notes in Computer Science, pp. 10-18, 1999.

[Jus72] J. Justesen, "A class of constructive asymptotically good algebraic codes", IEEE Transactions on Information Theory 18 (1972), pp. 652-656.

[KU97] N. Kahale and R. Urbanke, "On the Minimum Distance of Parallel and Serially Concatenated Codes", to appear in IEEE Transactions on Information Theory, 1997.

[Kas74] T. Kasami, "A Gilbert-Varshamov bound for quasi-cyclic codes of rate 1/2", IEEE Transactions on Information Theory (1974), p. 679.

[KM91] Eyal Kushilevitz and Yishay Mansour, "Learning Decision Trees Using the Fourier Sprectrum" Proc. 23rd Annual ACM Symposium on Theory of Computing (1991) , pp. 455-464.

[LN83] R. Lidl and H. Niederreiter, "Finite Fields. Number 20 in Encyclopedia of Mathematics and its Applications", Addison-Wesley, 1983.

[LMN89] N. Linial, Y. Mansour, and N. Nisan, "Constant depth circuits, Fourier transform, and learnability", Proc. 30th IEEE Symposium on Foundations of Computer Science (1989), pp. 574-579.

[LN90] N. Linial and N. Nisan. "Approximate inclusion-exclusion", Combinatorica 10(4): 349-365 (1990).

[Lin99] Lint, J.H. van, "Introduction to coding theory", Graduate texts in mathematics, Berlin ; New York : Springer, 1999.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak, "Ramanujan Graphs", Combinatorica 8(3), pp. 261-277, 1988.

[LV96] M. Luby and B. Velickovic, "On Deterministic Approximation of DNF", Algorithmica 16(4/5), pp. 415-433 (1996).

[Mac69] F.J. MacWilliams, "Codes and ideals in group algebras", Combinatorial Mathematics and its Applications, R.C. Bose and T.A. Dowling, eds., University of North Carolina Press, Chapel Hill, NC (1969), 317-328.

[MS92] J.F. MacWilliams and N.J.A. Sloane, "The Theory of Error-Correcting Codes", North-Holland 1992.

[Mar88] G.A. Margulis, "Explicit group theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators", Problems of Information Transmission, Vol 24, No. 1, pp. 39-46, July, 1988.

[McD74] Bernard R. McDonald, "Finite rings with identity", Marcel Dekker, Inc., New York 1974.

[MRRW77] R. McEliece, E. Rodemich, H. Rumsey Jr., and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliam inequalities", IEEE Transactions on Information Theory, 23:157-166, 1977.

[MW02] Meshulam and Wigderson, "Expanders from symmetric codes", Proc. 34th IEEE Symposium on Foundations of Computer Science, 2002, pp. 669-677.

[MW03] Meshulam and Wigderson, "Expanders in Group Algebras", to appear in Combinatorica, 2003.

[Mor94] Carlos Moreno, "Algebraic Curves over Finite Fields", Cambridge University Press, 1994.

[MR95] R. Motwani and P. Raghavan, "Randomized Algorithms", Cambridge University Press, 1995.

[NN93] J. Naor and M. Naor, "Small bias probability sapces: efficient constructions and applications", SIAM J. on Computing, 22:838-856 (1993).

[Nis90] N. Nisan, "Pseudorandom Generator for Space Bounded Computation", Proc. 22nd Annual ACM Symposium on Theory of Computing (1990), pp. 204-212.

[Nis91] N. Nisan, "Pseudorandom bits for constant depth circuits", Combinatorica, 12(4). pp. 63-70, 1991.

[NW88] N. Nisan and A. Wigderson, "Hardness vs. Randomness", Proc. 29th IEEE Symposium on Foundations of Computer Science (1988), pp. 2-11.

[PS99] H. D. Pfister and P. H. Siegel, "The Serial Concatenation of Rate-1 Codes Through Uniform Random Interleavers", Proc. 37th Allerton Conference on Communication, Control and Computing, pp. 260–269, Monticello, Illinois, Sep 1999.

[Pir85] Piret P.H., "An upper bound on the weight distribution of some codes", IEEE Transactions on Information Theory, 31 no 4 (1985), pp. 520-521.

[PHB98] V. S. Pless (Editor), W. C. Huffman (Editor), and R. A. Brualdi (Editor), "Handbook of coding theory", New York, Elsevier, 1998.

[Raz85] Alexander Razborov, "Lower bounds on the monotone complexity of some Boolean functions", Dokl. Akad. Nauk SSSR 281(4) (1985) 798 - 801 (In Russian); English translation in: Soviet Math. Dokl. 31 (1985) 354-57.

[Raz87] Alexander Razborov, "Lower bounds on the size of bounded depth networks over a complete basis with logical addition", Mathematicheskie Zametki, 41 (4)'598-607, 1987.

[Shp86] Shparlinsky I.E., "On weight enumerators of some codes", Problemy Peredechi Inform., 22 no 2 (1986), 43-48.

[SS96] M. Sipser and D. Spielman, "Expander Codes", IEEE Transactions on Information Theory, 1996, Vol 42, No 6, pp. 1710-1722.

[Smo87] Roman Smolensky, "Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity", Proc. 19th Annual ACM Symposium on Theory of Computing, (1987), pp. 77-82.

[Spi96] Daniel A. Spielman, "Linear-time encodable and decodable error-correcting codes", IEEE Transactions on Information Theory, 1996, Vol 42, No 6, pp. 1723-1732.

[Spi96n] Daniel A. Spielman, "Applied Extremal Combinatorics", lecture notes, http://www-math.mit.edu/ ~spielman/AEC/

[Sti93] Stichtenoth, H., "Algebraic Function Fields and Codes", Springer-Verlag, Heidelberg 1993.

[Sud01] Madhu Sudan , "Algorithmic Introduction to Coding Theory", lecture notes, http://theory.lcs.mit.edu/ ~madhu/FT01/

[STV99] M. Sudan, L. Trevisan, and S. Vadhan, "Pseudorandom generators without the XOR Lemma". Proc. 31st Annual ACM Symposium on Theory of Computing (1999), pp. 537-548.

[Sze75] G. Szego, "Orthogonal Polynomials", Fourth edition, Colloquium Publications, Vol. 23, Amer. Math. Soc. Providence R.I., 1975.

[TVZ82] Tsfasman M. A., Vladuts S. and Zink T, "Modular curves, Shimura curves, and Goppa codes, better than Varshamov Gilbert bound", Math. Nachr. 109, 21–28, 1982.

[Vaz86] U. Vazirani, "Randomness, adversaries, and computation", Ph.D. Thesis, University of California, Berkeley, 1986.

[VY00] B. Vucetic and J. Yuan, "Turbo Codes: Principles and Applications", Kluwer Academic Publishers, 2000.

[War74] Harold N. Ward, "Quadratic residue codes and symplectic groups", J. Algebra 29 (1974), 150-171.

[Wei48] A. Weil, "Sur les courbes algebriques et les varietes qui s'en deduisent", Herman, Paris, 1948.

[Yao82] A.C. Yao, "Theory and application of Trapdoor functions", Proc. 23rd IEEE Annual Symposium on Foundations of Computer Science (1982), pp. 80-91.