Quasi-uniform codes and information inequalities using group theory

Eldho Kuppamala Puthenpurayil Thomas

2015

Eldho Kuppamala Puthenpurayil Thomas. (2014). Quasi-uniform codes and information inequalities using group theory. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/62207>

<https://doi.org/10.32657/10356/62207>

QUASI-UNIFORM CODES AND INFORMATION INEQUALITIES USING GROUP THEORY

ELDHO KUPPAMALA PUTHENPURAYIL THOMAS

**DIVISION OF MATHEMATICAL SCIENCES
SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES**

*A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy*

2015

Acknowledgements

I would like to express my sincere gratitude to my PhD advisor Prof. Frédérique Oggier for giving me an opportunity to work with her. I appreciate her for all the support and guidance throughout the last four years. I believe, without her invaluable comments and ideas, this work would not have been a success. I have always been inspired by her profound knowledge, unparalleled insight, and passion for learning. Her patient approach to students is one of the key qualities that I want to adopt in my career.

I am deeply grateful to Dr. Nadya Markin, who is my co-author. She came up with crucial ideas and results that helped me to overcome many hurdles that I had during this research. I express my thanks to anonymous reviewers of my papers and thesis for their valuable feedback. I appreciate Prof. Sinai Robins, Prof. Wang Huaxiong and Prof. Axel Poschmann for being in my qualifying examination and giving me useful advices.

I wish to express my special thanks to Basu, who was the first source of help whenever I faced a new problem that I could not unlock myself. Also I am very happy to have helpful colleagues and friends Fuchun, Soon Sheng, Jerome, Su Le, Reeto, Huang Tao and many others.

I appreciate all my teachers and mentors I had in my life who helped and supported me to reach upto this stage. Special thanks to my Master thesis advisors Dr. Jonathan Woolf and Dr. Alexey Gorinov from Liverpool. Also I thank Dr. Sunil C Mathew who was my master thesis advisor in India and my inspiration.

I thank my parents for their endless love, prayers and caring. They have unconditionally put forth anything that they could for my success and progress. They are my motivation without any doubt. Love you Pappa and Mummy.

I thank my sisters, family and friends for their support and encouragement. Also I appreciate my friends in Singapore who stayed with me in struggles and difficulties to keep me relaxed and happy. Again I am extremely grateful to everyone who loved me, cared me or supported me in any manner to reach this milestone. Above all, I praise the Almighty God for doing wonders in my life.

Eldho K. Thomas
NTU-Singapore.

Contents

Acknowledgements	ii
Contents	iii
Publications	vi
List of Figures	vii
List of Tables	viii
Symbols	ix
Abstract	x
1 Introduction	1
2 Entropy and Information Measures	5
2.1 Information Measures	5
2.1.1 Probability and Independence	6
2.1.2 Shannon's Information Measures	6
2.1.3 Chain Rules for Information Measures	11
2.2 Basic Inequalities	13
3 Information Inequalities and Region of Entropic Vectors	17
3.1 Information Inequalities	17
3.1.1 Characterizing Information Inequalities	18
3.2 Entropic Vectors and their Region	19
3.2.1 Canonical Form and Elemental Inequalities	20
3.3 Attempts to Characterize Γ_n^*	24
4 Connection Between Groups and Entropy	27
4.1 Basics of Group Theory	27
4.1.1 Groups and Subgroups	27

4.1.2	Homomorphisms and Isomorphisms	28
4.1.3	Cyclic Groups	30
4.1.4	Cosets and Lagrange's Theorem	30
4.1.5	Normal Subgroups and Quotient Groups	31
4.1.6	Direct Product of Groups	33
4.2	Group Representable Entropy Function	34
4.3	$\overline{\Gamma}_n^*$ and Group Representability	36
4.4	Introduction to Quasi-Uniform Random Variables	37
4.4.1	Asymptotic Equipartition Property	38
4.4.2	Uniform Distribution	39
4.4.3	Quasi-Uniform Distributions	39
4.5	Region of Entropic Vectors from Quasi-Uniform Distributions	40
5	Abelian Group Representability of Finite Groups	43
5.1	Abelian Group Representability	43
5.2	Abelian Group Representability of Classes of 2 -Groups	45
5.2.1	Dihedral and Quasi-Dihedral 2 -Groups	46
5.2.1.1	Dihedral 2 -Groups	48
5.2.1.2	Quasi-dihedral 2 -Groups	50
5.2.2	Dicyclic 2 -Groups	53
5.3	Abelian Group Representability of p -Groups	54
5.4	Abelian Group Representability of Nilpotent Groups	58
5.5	Applications of Information Inequalities	62
6	Violations of Non-Shannon Inequalities	63
6.1	Information Inequalities and Group Inequalities	63
6.2	Ingleton Inequalities	67
6.2.1	Minimal Set of Ingleton Inequalities	68
6.2.2	Group Theoretic Formulation of Ingleton Inequalities	69
6.3	DFZ Inequalities on 5 Variables	70
6.4	Negative Conditions for DFZ Inequalities	74
6.4.1	Eliminating Classes of Subgroups	74
6.4.2	Negative Conditions of the Form $G_i \leq G_j$	80
6.5	Smallest Violations Using Groups	83
6.5.1	Smallest Violating Groups	84
7	Quasi-Uniform Codes	87
7.1	Quasi-Uniform Codes from Groups	89
7.1.1	Quasi-Uniform Codes from Abelian Groups	91
7.1.2	Quasi-Uniform Codes from Nonabelian Groups	95
7.1.2.1	The Case of Quotient Groups	95
7.1.2.2	Normal Subgroups of D_{2m}	95
7.1.2.3	Quasi-Uniform Codes from D_{2m} of Maximum Length	99
7.1.2.4	The Case of Nonnormal Subgroups	100
7.2	Some Classical Bounds for Quasi-Uniform Codes	100
7.2.1	Singleton Bound for Quasi-Uniform Codes	101
7.2.1.1	Examples of Quasi-Uniform Codes Satisfying the Above Bound	103

7.2.2	Gilbert-Varshamov Bound	106
7.2.3	Hamming Bound	107
7.2.4	Plotkin Bound	108
7.2.4.1	q -ary Plotkin bound	110
7.2.5	Shortening	111
7.2.6	Litsyn-Laihonen Bound	111
8	Applications of Quasi-Uniform Codes	115
8.1	Quasi-Uniform Codes from Dihedral 2 -Groups	116
8.1.1	Quasi-Uniform Codes from D_8	118
8.2	Storage Applications	120
8.2.1	Code Comparisons	121
8.2.2	A Storage Example	122
8.3	Bounds on the Minimum Distance	122
8.4	Quasi-Uniform Codes in Network Coding	125
8.5	Almost Affine Codes from Groups	127
9	Future Works	129
A	Normal Subgroups of Dihedral Groups	131
A.1	Conjugacy Classes of D_{2m}	132
A.2	Normal Subgroups	133
	Bibliography	137

Publications

Journal Paper

1. E. Thomas, N. Markin, and F. Oggier, *On Abelian Group Representability of Finite Groups*, Advances in Mathematics of Communications, 8(2):139-152, May 2014.

Conference Papers

1. E. Thomas and F. Oggier, *Applications of Quasi-uniform Codes to Storage*, International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, July 2014 (Invited Paper).
2. N. Markin, E. Thomas, and F. Oggier, *Groups and Information Inequalities in 5 Variables*, Fifty-first Annual Allerton Conference, October 2013.
3. E. Thomas and F. Oggier, *Explicit Constructions of Quasi-uniform Codes from Groups*, International Symposium on Information Theory (ISIT), Istanbul, Turkey, July 2013.
4. E. Thomas and F. Oggier, *A Note on Quasi-uniform Distributions and Abelian Group Representability*, International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, July 2012.

List of Figures

4.1	Quasi-uniform and non quasi-uniform distributions.	40
7.1	On the right, the dihedral group D_{12} , and on the left, the abelian group $C_3 \times C_2 \times C_2$, both with some of their subgroups.	98
8.1	The dihedral group D_8 and its lattice of subgroups.	118

List of Tables

7.1	Quasi-uniform code constructed from $C_3 \times C_3 \simeq \{0, 1, 2\} \times \{0, 1, 2\}$	94
7.2	Quasi-uniform code constructed from S_3 and some nonnormal subgroups	100
8.1	A $(8, C , 3)$ code constructed from D_8 , $ C = 8$. Pairs are elements in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$	120
8.2	Minimum distance comparison with known codes [29].	121

Symbols

$\mathcal{N}$	$\{1, \dots, n\}$
$\mathcal{A}$	Any subset of $\mathcal{N}$
$G_{\mathcal{A}}$	$\cap_{i \in \mathcal{A}} G_i$
$\mathcal{H}_n$	$2^n - 1$ Euclidean space
Γ_n^*	Entropic vector set
$\overline{\Gamma}_n^*$	Closure of Γ_n^*
Γ_n	Set of Shannon-type inequalities
Υ_n	Group representable vector set
Υ_n^{ab}	Abelian group representable vector set
ψ^*	Quasi-uniform vector set
$\overline{con}$	Convex closure of a given set
$\mathbb{R}$	Set of real numbers
$\mathbb{Z}$	Set of integers
$\mathbb{Q}$	Set of rationals
S_n	Symmetric group on n symbols
D_{2m}	Dihedral group of order $2m$

Abstract

This thesis is dedicated to the study of information inequalities and quasi-uniform codes using group theory.

Understanding the region of entropic vectors for dimension $n \geq 4$ is an open problem in network information theory. It can be studied using information inequalities and their violations. The connection between entropic vectors and finite groups, known as 'group representability', is a useful tool to compute these violations. In the first part of this thesis we address the problem of extracting 'abelian group representable' vectors out of the whole set of group representable vectors. We prove that certain classes of non-abelian groups are abelian group representable and non-nilpotent groups are not abelian group representable. We then address the question of finding linear inequality violators for $n = 5$ and obtain the smallest group violators of two linear inequalities.

Random variables which are uniformly distributed over their support are known as quasi-uniform. One way of getting quasi-uniform random variables is by using finite groups and subgroups. Codes are constructed in such a way that the associated random variables are quasi-uniform and group theory is used to construct such codes. In the second part of this thesis, we consider the construction of quasi-uniform codes coming from groups and their algebraic properties. We compute some coding parameters and bounds in terms of groups. Finally we propose some applications of quasi-uniform codes especially to distributed storage.

To my family

Introduction

Understanding the region of entropic vectors is of long term interest in information theory. An entropic vector is a $2^n - 1$ dimensional vector whose components are joint entropies of all the possible subsets of a collection of finite alphabet random variables $X_1, \dots, X_n$. The region formed by these vectors is denoted as Γ_n^* (see page 326 of [42]). Characterizing Γ_n^* is important because there is a fundamental connection between the entropy region and the capacity region of networks. In fact, many network information theory problems can be formulated as linear optimization problems over Γ_n^* (see[20, 41]). Thus, determining Γ_n^* can lead to the solution of a plethora of information-theoretic problems. Moreover, many proofs of the converse coding theorems involve the so-called information inequalities (inequalities involving functions of information measures such as entropy etc.), the complete set of which can be found as a result of characterizing Γ_n^* (see[42]).

There were many approaches to find inner and outer bounds of the region Γ_n^* . One of them is by understanding the region of vectors satisfying 'Shannon-type inequalities'. They are information inequalities formed by the non-negative linear combinations of 'Shannon information measures' such as entropy, conditional entropy, mutual information and conditional mutual information. The space of all $2^n - 1$ dimensional vectors which only satisfy the Shannon-type inequalities is denoted by Γ_n . In fact, there is not much difference between Γ_n and Γ_n^* when $n = 2, 3$ since $\Gamma_2 = \Gamma_2^*$ and $\Gamma_3 = \bar{\Gamma}_3^*$, the closure of Γ_3^* (see[42, 44]). It is also proven that $\bar{\Gamma}_n^*$ is a convex cone, however the complete characterization of Γ_n^* for $n \geq 4$ is still unknown, because there exist non-Shannon type inequalities for $n \geq 4$ (see[14, 44]).

Surprisingly, there exists a connection between entropic vectors and finite groups using the notion of 'group representability'. A set of random variables $X_1, \dots, X_n$ is said to be group representable, if there exists a finite group G and subgroups $G_1, \dots, G_n$ such that the joint

entropy $H(X_{\mathcal{A}}) = \log[G : G_{\mathcal{A}}]$ for all $\mathcal{A} \subseteq \mathcal{N} = \{1, \dots, n\}$, where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$. The space of all group representable entropic vectors is denoted by Υ_n and the smallest closed convex set containing Υ_n , $\overline{\text{con}}(\Upsilon_n) = \bar{\Gamma}_n^*$ [11]. That is, a good understanding of the space of group representable entropic vectors is enough to characterize Γ_n^* . Considering the space Υ_n^{ab} using only abelian groups yields a non-trivial inner bound for $\bar{\Gamma}_n^*$.

Conversely, associated to a finite group and n subgroups, there exist n random variables $X_1, \dots, X_n$ which are uniformly distributed over their support and the joint entropy $H(X_{\mathcal{A}}) = \log[G : G_{\mathcal{A}}]$ for all $\mathcal{A} \subseteq \mathcal{N}$ [12]. Such random variables which are uniformly distributed over their support are known as 'quasi-uniform' and we denote the space of entropic vectors formed by quasi-uniform random variables as Ψ^* . It is proven that a group representable entropic vector is quasi-uniform and the space of all entropic vectors is the minimal closed convex set containing Ψ^* . That is, $\Upsilon_n \subset \Psi^*$ and $\overline{\text{con}}(\Psi^*) = \bar{\Gamma}_n^*$ [12]. Therefore Ψ^* is another sufficient class of vectors to understand the entropic region Γ_n^* .

The above analysis provides us a hierarchy of entropic regions as follows:

$$\Upsilon_n^{ab} \subset \Upsilon_n \subset \Psi^* \subset \Gamma_n^* \subset \bar{\Gamma}_n^*$$

and

$$\overline{\text{con}}(\Upsilon_n) = \overline{\text{con}}(\Psi^*) = \bar{\Gamma}_n^*.$$

However, $\overline{\text{con}}(\Upsilon_n^{ab}) \neq \bar{\Gamma}_n^*$ for $n \geq 4$ [12] and hence it is a non-trivial inner bound for $\bar{\Gamma}_n^*$.

As discussed above, for $n \geq 4$, with the existence of non-Shannon type inequalities, the classification of Γ_n^* gets more complicated. However, it is proven in [19] that, the set of Shannon-type inequalities together with the minimal set of Ingleton inequalities (see Equation (5.2)) are sufficient to characterize the set of all linear rank inequalities (inequalities involving rank of subspaces of vector spaces) for $n = 4$ and are not sufficient for $n \geq 5$. Any linear information inequality that always holds is also a linear rank inequality which always holds for finite dimensional vector spaces over some field. The Ingleton inequality always holds for ranks of subspaces, but does not always hold for random variables and hence the converse is false. So finding entropic vectors violating the Ingleton inequality is a good way to understand Γ_n^* when $n = 4$. Some of the violations are given in [28]. It is also known that entropic vectors which are abelian group representable do not violate the Ingleton inequality [12], which means we have to take an entropic vector which is not abelian representable to get violations of the Ingleton inequality.

However, information theory does not provide us much insight as how to find such violators. But its explicit connection with group theory using quasi-uniform random variables opens a path to look for violators which can be done as follows: rewrite all information

inequalities and in particular the Ingleton inequality as a group inequality and look for finite groups and subgroups that violate this group inequality. Mao et al. proved in [26] that S_5 , the symmetric group of order 120 is the smallest violator of the Ingleton inequality for $n = 4$ (we denote it as 4-Ingleton). When $n = 5$, the Shannon inequalities, Ingleton inequalities for 5 variables (5-Ingleton, see Equation (6.3)) together with 24 other inequalities known as 'DFZ' inequalities determine the space of linear rank inequalities [14].

In the first part of the thesis, we focus on finding violators of the linear rank inequalities for 5 random variables using the corresponding group inequalities. We concentrate on the DFZ inequalities because a group violator of the 5-Ingleton inequalities violates the 4-Ingleton inequality and vice versa [27]. We propose some negative conditions that help us identifying small violators.

We also work on certain classes of non-abelian finite groups to see whether they are abelian group representable. That is, the entropic vector coming from the associated random variables of the former and its subgroups can be obtained from an abelian group and subgroups. We found that dihedral 2-groups and some other well known families of non-abelian groups are abelian group representable. This study is important to understand the gap between regions Υ_n^{ab} and Υ_n . Also it helps to propose some negative conditions when looking at the violators of the DFZ inequalities apart from its applications in coding theory.

Finding violations of information inequalities are also useful in network coding. The fact that entropic vectors which are abelian group representable do not violate the Ingleton inequality [12] leads to the conclusion that linear network codes also do not since they are abelian groups. This explains why they can not achieve the capacity of some networks [13]. To resolve this issue, non-linear codes are required and one way of looking for them is by getting non-abelian violators of the Ingleton inequality. How to translate these violators into network codes is not yet fully understood.

Quasi-uniform codes were introduced in [9]. They are called so because if we associate its components to a set of random variables, they form a quasi-uniform distribution. Each component of a quasi-uniform code lies possibly in a different alphabet. We can construct a quasi-uniform code as the support of the joint distribution of n quasi-uniform random variables. In fact, all linear codes are coming under this class and so quasi-uniform codes are much more general than linear codes. One way of constructing these codes is by using finite groups and subgroups.

Many information theoretic properties of these codes are already studied in [9]. We focus more on the algebraic properties of these codes. In fact, non-linear codes may be constructed in this way. Therefore, this coding scheme possibly allows us to make use of the above violators in network coding. However the encoding and decoding of these codes as

network codes is still unknown and therefore we start by looking at them as classical codes even though the initial motivation is to apply them as network codes, which is the second part of the study.

Applications of quasi-uniform codes are many [9], but we focus more on the storage aspect of it initially. We can construct quasi-uniform storage codes on small alphabets to increase the local and global reparability of nodes involved.

In summary, this thesis proposes a way to understand information inequalities in higher dimensions using group theory and a study of quasi-uniform codes coming from groups and their algebraic properties. It opens many questions in the direction of network coding and distributed storage.

Structure of the Thesis: The required background on information theory to develop the subject is given in Chapter 2 and 3. The concepts of entropy, information measures, information inequalities and the region of entropic vectors are all described in these chapters. We use group theory as a major tool to unlock the information theoretic problem we face and therefore their connections are given in Chapter 4.

New results are presented from Chapter 5 onwards. Abelian group representability of finite groups and the details on some classes of finite groups which are abelian representable are given in Chapter 5. Most of the results from this chapter are published in [37]. Chapter 6 mainly focuses on finding some violators of the DFZ inequalities and is based on [27].

The explicit construction of quasi-uniform codes and their algebraic properties are discussed in Chapter 7. Initial results of this chapter are published in [39]. Some bounds satisfied by these codes are also computed there. Finally in Chapter 8, some applications of quasi-uniform codes in storage (published in [40]) are proposed with the help of abelian representability.

Entropy and Information Measures

This chapter introduces information measures, their basic properties and some relations among different information measures.

2.1 Information Measures

It is not easy to quantify information, which is not a physical entity. Claude E. Shannon (1916-2001) introduced two concepts to capture information from a communication point of view. Firstly, information is *uncertainty* and secondly, information to be transmitted over a communication channel is *digital* [34].

To be more specific, if we are interested in a piece of information which is deterministic, then it has no value, since it is known already without any uncertainty. Consequently, any information source can be modeled as a random variable and probability can be used to develop the theory of information.

Secondly, the information that is to be transmitted is digital means that the information should be converted to symbols 0s and 1s called bits which are then transmitted over a communication channel, without any reference to the actual meaning.

In this chapter we recall the basics of some of the most important measures of information: entropy, conditional entropy, mutual information and conditional mutual information, known as *Shannon's information measures*. This chapter is based on the book [42].

2.1.1 Probability and Independence

We start by introducing some basic concepts of probability. All the random variables we consider are discrete unless otherwise stated. Let X be a random variable with alphabet $\mathcal{X}$. Denote the probability distribution of X as $P_X(x) = Pr(X = x)$; $x \in \mathcal{X}$. We omit the subscript X and denote the probability as $P(x)$ if there is no ambiguity.

Define the *support* of X as $\lambda(X) = \{x \in \mathcal{X} : P(x) > 0\}$.

Definition 1. Two random variables X and Y are said to be *independent* denoted as $X \perp Y$, if $P(x, y) = P(x)P(y)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Next we define the conditional independence:

Definition 2. For random variables X, Y and Z , X is independent of Z conditioning on Y , denoted by $X \perp Z | Y$, if

$$P(x, y, z)P(y) = P(x, y)P(y, z) \text{ for all } x, y, z$$

or equivalently,

$$P(x, y, z) = \begin{cases} \frac{P(x, y)P(y, z)}{P(y)} = P(x, y)P(z|y); & P(y) > 0 \\ 0; & \text{otherwise} \end{cases}$$

2.1.2 Shannon's Information Measures

The entropy associated with a random variable is a measure of uncertainty associated with it, and is defined as:

Definition 3. The *entropy* $H(X)$ of a random variable X is defined as

$$H(X) = - \sum_{x \in \lambda(X)} P(x) \log P(x).$$

In the above definition of entropy, the base of the logarithm can be taken as any positive number greater than 1. When the base is 2, the unit of entropy is *bit* and *nat* when the base is e . In information theory, the entropy is measured in bits and hence we use the base 2. All entropies we consider are finite unless specified otherwise.

The entropy $H(X)$ of a random variable X measures the average amount of information contained in X , or equivalently, entropy is the average amount of uncertainty removed upon revealing the outcome of X and is a function of the probability distribution $P(x)$.

Similarly, we define the joint entropy associated with more than one random variable.

Definition 4. The *joint entropy* $H(X, Y)$ associated with a set of random variables X and Y is defined as

$$H(X, Y) = - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x, y).$$

Next we define the conditional entropy of two random variables where one of them is already given.

Definition 5. Let X and Y be two random variables. The *conditional entropy* of Y given X is defined as

$$H(Y|X) = - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(y|x).$$

Since $P(y|x) = P(x, y)/P(x)$, we have

$$H(Y|X) = \sum_{x \in \lambda(X)} P(x) \left[- \sum_{y \in \lambda(Y)} P(y|x) \log P(y|x) \right]$$

where $-\sum_{y \in \lambda(Y)} P(y|x) \log P(y|x) = H(Y|X = x)$, the entropy of Y conditioning on a particular $x \in \lambda(X)$. That is,

$$H(Y|X) = \sum_{x \in \lambda(X)} P(x) H(Y|X = x).$$

Similarly, for random variables X, Y and Z ,

$$H(Y|X, Z) = \sum_{z \in \lambda(Z)} P(z) H(Y|X, Z = z),$$

where

$$H(Y|X, Z = z) = - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y|z) \log P(y|x, z).$$

Suppose the outcomes of two random variables X and Y are revealed in two steps: first the outcome of X and then of Y .

The following Proposition says that the total amount of uncertainty removed upon revealing the outcomes of both X and Y together is equal to the sum of the uncertainty removed upon revealing X and the uncertainty removed upon revealing Y given that X is already known. That is,

Proposition 1.

$$H(X, Y) = H(X) + H(Y|X)$$

and

$$H(X, Y) = H(Y) + H(X|Y).$$

Proof. Consider

$$\begin{aligned}
H(X, Y) &= - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x, y) \\
&= - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x) P(y|x) \\
&= - \left[\sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x) + \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(y|x) \right] \\
&= - \left[\sum_{x \in \lambda(X)} P(x) \log P(x) + \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(y|x) \right] \\
&\quad \text{where } P(x) = \sum_{y \in \lambda(Y)} P(x, y) \\
&= H(X) + H(Y|X).
\end{aligned}$$

The second part follows by symmetry. $\square$

Next we define the mutual information between two random variables.

Definition 6. The *mutual information* between random variables X and Y is defined as

$$I(X; Y) = \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log \frac{P(x, y)}{P(x)P(y)}.$$

Note that $I(X; Y)$ is symmetrical in X and Y .

Intuitively, it is a measurement of information that both X and Y share between them or the amount of information about X provided by Y and vice versa.

For example, if X and Y are independent, the knowledge of one does not give any information about the other. This implies that the mutual information is zero, which is clear from the definition:

$$\begin{aligned}
I(X; Y) &= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \\
&= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log \frac{P(x)P(y)}{P(x)P(y)} \\
&= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log 1 = 0.
\end{aligned}$$

If Y is a function of X , then by knowing X , we know everything about Y . That is, $P(y|x) = 1$ implies $P(x, y) = P(x)$ and therefore, $I(X; Y) = H(X)$, the entropy of X .

Proposition 2. The mutual information between a random variable X and itself is equal to the entropy of X , i.e., $I(X; X) = H(X)$.

Proof. Consider

$$\begin{aligned} I(X; X) &= \sum_{x \in \lambda(X)} P(x) \log \frac{P(x)}{P(x)P(x)} \\ &= \sum_{x \in \lambda(X)} P(x) \log \frac{1}{P(x)} = H(X). \end{aligned}$$

□

In the context of the above Proposition, the entropy of a random variable X is also called *self-information* of X .

Proposition 3. If X and Y are two random variables, then

$$I(X; Y) = H(X) - H(X|Y),$$

$$I(X; Y) = H(Y) - H(Y|X), \text{ and}$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y).$$

Proof. Consider

$$\begin{aligned} I(X; Y) &= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \\ &= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log \frac{P(x|y)}{P(x)} \\ &= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) [\log P(x|y) - \log P(x)] \\ &= - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x) + \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x|y) \\ &= - \sum_{x \in \lambda(X)} P(x) \log P(x) + \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x|y) \\ &= H(X) - H(X|Y). \end{aligned}$$

Similarly,

$$I(X; Y) = H(Y) - H(Y|X).$$

Now

$$\begin{aligned}
I(X; Y) &= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \\
&= \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) [\log P(x, y) - \log P(x) - \log P(y)] \\
&= - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x) - \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(y) \\
&\quad + \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x, y) \\
&= - \sum_{x \in \lambda(X)} P(x) \log P(x) - \sum_{y \in \lambda(Y)} P(y) \log P(y) \\
&\quad + \sum_{(x,y) \in \lambda(X \times Y)} P(x, y) \log P(x|y) \\
&= H(X) + H(Y) - H(X, Y).
\end{aligned}$$

□

From the above proposition, the mutual information $I(X; Y)$ is the reduction in uncertainty about X when Y is given or the reduction in uncertainty about Y when X is given.

From Propositions 1 and 3, we can see the connections between entropies, conditional entropies and mutual information for two random variables X and Y .

There is a conditional version of mutual information analogous to entropy.

Definition 7. For random variables X, Y and Z , the mutual information between X and Y conditioning on Z is defined as

$$I(X; Y|Z) = \sum_{(x,y,z) \in \lambda(X \times Y \times Z)} P(x, y, z) \log \frac{P(x, y|z)}{P(x|z)P(y|z)}.$$

Clearly, it is symmetrical in X and Y from the definition.

Analogous to Propositions 2 and 3, there are relations satisfied by conditional mutual information as well.

Proposition 4. The mutual information between a random variable X and itself conditioning on a random variable Z is equal to the conditional entropy of X given Z , i.e., $I(X; X|Z) = H(X|Z)$.

Proof. Consider

$$\begin{aligned} I(X; X|Z) &= \sum_{(x,z) \in \lambda(X \times Z)} P(x, z) \log \frac{P(x|z)}{P(x|z)P(x|z)} \\ &= \sum_{(x,z) \in \lambda(X \times Z)} P(x, z) \log \frac{1}{P(x|z)} = H(X|Z). \end{aligned}$$

□

Proposition 5. If X, Y and Z are random variables, then

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z),$$

$$I(X; Y|Z) = H(Y|Z) - H(Y|X, Z), \text{ and}$$

$$I(X; Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z).$$

The proof is similar to the proof of Proposition 3 except for the conditioning on Z .

Remark 1. All information measures we have seen so far namely, entropy, joint entropy, conditional entropy, mutual information and conditional mutual information are known as *Shannon's information measures*. That is, $H(X), H(X, Y), H(X|Y), I(X; Y), I(X; Y|Z)$ are all Shannon's information measures.

Remark 2. All Shannon's information measures are special cases of conditional mutual information. To see this, let Φ denote a new random variable that takes a constant value. Then $H(X) = I(X; X|\Phi)$, $H(X, Y) = I(X, Y; X, Y|\Phi)$, $H(X|Z) = I(X; X|Z)$ and $I(X; Y) = I(X; Y|\Phi)$.

2.1.3 Chain Rules for Information Measures

In this section, we prove some information identities known as chain rules which we use later when discussing information inequalities.

Proposition 6. If $X_1, \dots, X_n$ are random variables, then

$$H(X_1, \dots, X_n) = \sum_{i=1}^n H(X_i | X_1, \dots, X_{i-1}).$$

Proof. When $n = 2$, the result is true from Proposition 1. Now assume that the result is true for $n = m$; $m \geq 2$.

Consider

$$\begin{aligned}
 H(X_1, \dots, X_{m+1}) &= H(X_1, \dots, X_m) + H(X_{m+1}|X_1, \dots, X_m) \\
 &\quad \text{(by taking } X = (X_1, \dots, X_m), Y = X_{m+1} \text{ in Proposition 1)} \\
 &= \sum_{i=1}^m H(X_i|X_1, \dots, X_{i-1}) + H(X_{m+1}|X_1, \dots, X_m) \\
 &\quad \text{(using induction hypothesis for } n = m) \\
 &= \sum_{i=1}^{m+1} H(X_i|X_1, \dots, X_{i-1}).
 \end{aligned}$$

Therefore the result is proved by induction. $\square$

Next we prove the chain rule for conditional entropy.

Proposition 7. For random variables $X_1, \dots, X_n$ and Y ,

$$H(X_1, \dots, X_n|Y) = \sum_{i=1}^n H(X_i|X_1, \dots, X_{i-1}, Y).$$

Proof.

$$\begin{aligned}
 H(X_1, \dots, X_n|Y) &= H(X_1, \dots, X_n, Y) - H(Y) \\
 &\quad \text{(from Proposition 1)} \\
 &= H(X_1, Y, X_2, \dots, X_n) - H(Y) \\
 &= H(X_1, Y) + \sum_{i=2}^n H(X_i|X_1, \dots, X_{i-1}, Y) - H(Y) \\
 &\quad \text{(by Proposition 6)} \\
 &= H(X_1|Y) + \sum_{i=2}^n H(X_i|X_1, \dots, X_{i-1}, Y) \\
 &= \sum_{i=1}^n H(X_i|X_1, \dots, X_{i-1}, Y).
 \end{aligned}$$

$\square$

Analogous to the chain rules for entropy and conditional entropy, there are chain rules for mutual information and conditional mutual information.

Proposition 8.

$$I(X_1, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y|X_1, \dots, X_{i-1}).$$

Proof. From Proposition 3,

$$\begin{aligned}
 I(X_1, \dots, X_n; Y) &= H(X_1, \dots, X_n) - H(X_1, \dots, X_n|Y) \\
 &\quad \text{(by taking } X = (X_1, \dots, X_n)\text{)} \\
 &= \sum_{i=1}^n H(X_i|X_1, \dots, X_{i-1}) - \sum_{i=1}^n H(X_i|X_1, \dots, X_{i-1}, Y) \\
 &\quad \text{(using Propositions 6 and 7)} \\
 &= \sum_{i=1}^n I(X_i; Y|X_1, \dots, X_{i-1}) \\
 &\quad \text{(by Proposition 5.)}
 \end{aligned}$$

□

Finally, we prove the chain rule for conditional mutual information.

Proposition 9.

$$I(X_1, \dots, X_n; Y|Z) = \sum_{i=1}^n I(X_i; Y|X_1, \dots, X_{i-1}, Z).$$

Proof. From Proposition 3,

$$\begin{aligned}
 I(X_1, \dots, X_n; Y|Z) &= H(X_1, \dots, X_n|Z) - H(X_1, \dots, X_n|Y, Z) \\
 &\quad \text{(from Proposition 5)} \\
 &= \sum_{i=1}^n [H(X_i|X_1, \dots, X_{i-1}, Z) - H(X_i|X_1, \dots, X_{i-1}, Y, Z)] \\
 &\quad \text{(by Proposition 7)} \\
 &= \sum_{i=1}^n I(X_i; Y|X_1, \dots, X_{i-1}, Z).
 \end{aligned}$$

□

2.2 Basic Inequalities

In this section, we will see that all Shannon's information measures defined in Section 2.1 are non-negative, for which we need the concept of *relative entropy* or *informational divergence*.

Definition 8. The *informational divergence* between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p \parallel q) = \sum_{x \in \lambda_p} p(x) \log \frac{p(x)}{q(x)}.$$

In the above definition we assume that $c \log \frac{c}{0} = \infty$ for $c > 0$. Because of this, if $D(p \parallel q) < \infty$ and $q(x) = 0$, then $p(x) = 0$, implies that $\lambda_p \subset \lambda_q$, where λ_p is the support of p and λ_q is the support of q .

Informational divergence is a non-symmetric measure of the difference between two probability distributions p and q . Because of the non-symmetry, it is not a true distance or metric.

We state the following theorem which says that the informational divergence is always non-negative.

Theorem 10. [42] For any two probability distributions p and q on a common alphabet $\mathcal{X}$,

$$D(p \parallel q) \geq 0$$

with equality if and only if $p = q$.

In the next proposition, we see that the conditional mutual information is always nonnegative.

Proposition 11. For random variables X, Y and Z ,

$$I(X; Y|Z) \geq 0,$$

with equality if and only if X and Y are independent when conditioning on Z .

Proof. Consider

$$\begin{aligned} I(X; Y|Z) &= \sum_{x,y,z} P(x, y, z) \log \frac{P(x, y|z)}{P(x|z)P(y|z)} \\ &= \sum_z P(z) \sum_{x,y} P(x, y|z) \log \frac{P(x, y|z)}{P(x|z)P(y|z)} \\ &= \sum_z P(z) D(P_{XY|z} \parallel P_{X|z} P_{Y|z}), \end{aligned}$$

where $P_{XY|z} = \{P(x, y|z) : (x, y) \in \mathcal{X} \times \mathcal{Y}\}$, $P_{X|z} = \{P(x|z) : x \in \mathcal{X}\}$ and $P_{Y|z} = \{P(y|z) : y \in \mathcal{Y}\}$.

For a fixed z , both $P_{XY|z}$ and $P_{X|z}P_{Y|z}$ are distributions on $\mathcal{X} \times \mathcal{Y}$ implies

$$D(P_{XY|z} \parallel P_{X|z}P_{Y|z}) \geq 0$$

and hence $I(X; Y|Z) \geq 0$.

From Theorem 10, we have $D(P_{XY|z} \parallel P_{X|z}P_{Y|z}) = 0$ if and only if $P(x, y|z) = P(x|z)P(y|z)$ for all $z \in \lambda(Z)$ and for all x and y . Therefore, X and Y are independent conditioning on Z if and only if $I(X; Y|Z) = 0$. $\square$

Corollary 12. All Shannon's information measures are always non-negative.

Proof. Since all Shannon's information measures $H(X)$, $H(X, Y)$, $H(X|Y)$, $I(X; Y)$ and $I(X; Y|Z)$ are particular cases of the conditional mutual information (Remark 2), the result follows from Proposition 11. $\square$

Next result says that a random variable has zero entropy if and only if it is deterministic.

Proposition 13. $H(X) = 0$ if and only if X is deterministic.

Proof. Suppose that the random variable X is deterministic. Then $\exists x' \in X$ with $P(x') = 1$ and $P(x) = 0$ for all $x \neq x'$. Therefore, $H(X) = -P(x') \log P(x') = 0$.

If X is non-deterministic, $\exists x' \in X$ with $0 < P(x') < 1$ and then $H(X) \geq -P(x') \log P(x') > 0$, concludes the proof. $\square$

There is an analogous result for conditional entropy as well.

Corollary 14. $H(Y|X) = 0$ if and only if Y is a function of X .

Proof. We have

$$H(Y|X) = \sum_x P(x)H(Y|X = x).$$

That is, $H(Y|X) = 0$ if and only if $H(Y|X = x) = 0$ for all $x \in \lambda(X)$, if and only if Y is deterministic for each x (Proposition 13) and this means that Y is a function of X . $\square$

Corollary 15. $I(X; Y) = 0$ if and only if X and Y are independent.

The proof follows from Proposition 11 by assuming Z is a random variable which takes a constant value.

Definition 9. The non-negativity of all Shannon's information measures yields the so-called *basic inequalities*. i.e., $H(X) \geq 0$, $H(Y|X) \geq 0$, $H(X, Y) \geq 0$, $I(X; Y) \geq 0$, $I(X; Y|Z) \geq 0$ are all basic inequalities.

However, this set is not minimal, since some of the basic inequalities are implied by others. For example, the basic inequality

$$H(X) = H(X|Y) + I(X; Y) \geq 0$$

is implied by basic inequalities $H(X|Y) \geq 0$ and $I(X; Y) \geq 0$.

Definition 10. Inequalities involving only Shannon's information measures are known as information inequalities.

From Propositions 1, 3 and 5, we can say that all Shannon's information measures can be expressed only in terms of entropies. (By entropies, we mean entropies of single random variables and joint entropies of many variables). More precisely,

Corollary 16.

$$\begin{aligned} H(Y|X) &= H(X, Y) - H(X) \\ I(X; Y) &= H(X) + H(Y) - H(X, Y) \text{ and} \\ I(X; Y|Z) &= H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z). \end{aligned}$$

Hence using identities in Corollary 16, all information inequalities involving three random variables can be expressed in terms of entropies only. This result can be generalized to the case of n random variables, thanks to the chain rules.

Information Inequalities and Region of Entropic Vectors

In Chapter 2, we have seen the basics of information measures such as entropy, conditional entropy, mutual information and conditional mutual information involving one or more random variables. Towards the end, we defined information inequalities involving these information measures. In this chapter, we survey what is known about the characterization of the region of entropic vectors .

3.1 Information Inequalities

In a formal way, an *information expression* f is a linear combination of Shannon's information measures involving a finite number of random variables. For example,

$$H(X, Y) + 2.5I(X; Z) + 1.5H(X, Y|Z),$$

$$I(X; Y|Z) - H(Y) + 2I(A; B|C, D)$$

are valid information expressions.

An information expression f becomes an *information inequality* if $f \geq 0$ or $f \leq 0$. Two information expressions f and g form an information inequality if $f \leq g$ or $f \geq g$.

It is not required to state the equality explicitly, since $f = g$ is equivalent to the pair of inequalities $f \geq g$ and $f \leq g$.

Information expressions are functions of information measures, which themselves are functions of probability distributions. Therefore, an information inequality is said to be true or

to *always hold*, if it holds for all probability distributions of the random variables involved. For example,

$$I(X; Y) \geq 0$$

always holds because it is true for any joint distribution $P(x, y)$.

Conversely, we say that an information inequality does *not always hold* if there exists a joint distribution for which it is violated.

Example 1.

$$I(X; Y) \leq 0$$

does not always hold. This is because, $I(X; Y) \geq 0$ always holds and if $I(X; Y) \leq 0$ holds, then $I(X; Y) = 0$ which implies X and Y are independent (Corollary 15). In other words, if X and Y are not independent, $I(X; Y) \leq 0$ is not true. Therefore, $I(X; Y) \leq 0$ does not always hold.

3.1.1 Characterizing Information Inequalities

It is a natural question to ask whether it is possible to characterize all information inequalities. In [42], Yeung proposed a method to characterize a type of information inequalities called *Shannon-type inequalities*, which we define here.

Definition 11. Information inequalities which can be expressed as non-negative linear combinations of *basic inequalities* (Definition 9) are defined as *Shannon-type inequalities*.

That is; the inequalities formed by non-negative linear combinations of any of $H(X) \geq 0$, $H(X, Y) \geq 0$, $H(X, Y|Z) \geq 0$, $I(X; Y) \geq 0$ and $I(X; Y|Z) \geq 0$ are Shannon-type. Almost all information inequalities known to date are Shannon-type.

Is there any information inequality which is not implied by the basic inequalities? In fact, yes. There are information inequalities which are not implied by basic inequalities [42] and are known as *non-Shannon-type inequalities*. While the framework in [42] verifies and solves all Shannon-type inequalities, there is no such method to find and verify non-Shannon-type inequalities. But a good understanding of the non-Shannon-type inequalities is necessary to study the set of all entropic vectors for higher dimensions. That is, there exist entropic vectors given by some set of random variables violating non-Shannon-type inequalities and a complete list of such vectors is unknown since the non-Shannon-type inequalities are not fully characterized.

Before going into further details of information inequalities, we go through the following section where we discuss the region of entropic vectors.

3.2 Entropic Vectors and their Region

Let $X_1, \dots, X_n$ be a collection of n jointly distributed discrete random variables over some alphabet of size N . We denote by $\mathcal{A}$ a subset of indices from $\mathcal{N} = \{1, \dots, n\}$, and $X_{\mathcal{A}} = \{X_i, i \in \mathcal{A}\}$.

Definition 12. The *entropic vector* corresponding to $X_1, \dots, X_n$ is the vector

$$h = (H(X_1), \dots, H(X_1, X_2), \dots, H(X_1, \dots, X_n))$$

which collects all the joint entropies $H(X_{\mathcal{A}})$, $\mathcal{A} \subseteq \mathcal{N}$.

For example, when $n = 3$,

$$h = (H(X_1), H(X_2), H(X_3), H(X_1, X_2), H(X_1, X_3), H(X_2, X_3), H(X_1, X_2, X_3)).$$

Consider the $(2^n - 1)$ -dimensional Euclidean space $\mathcal{H}_n$ known as *entropy space*, with coordinates labeled by $h_{\mathcal{A}}$, for all non-empty subsets $\mathcal{A} \subseteq \mathcal{N}$. The $(2^n - 1)$ -tuple corresponding to the joint entropies of set of n random variables is then a column vector in $\mathcal{H}_n$.

For $n = 3$, the coordinates of $\mathcal{H}_3$ are labeled by

$$h_1, h_2, h_3, h_{12}, h_{13}, h_{23}, h_{123}$$

where $h_{i,j} = h_{ij}$.

A column vector $h \in \mathcal{H}_n$ is said to be *entropic* if the $(2^n - 1)$ -tuple representing h are joint entropies of a valid set of n random variables. In other words, when the vector h contains elements corresponding to the joint entropies of any valid set of random variables, then h is entropic.

Example 2. Consider $h = (1, 0.5, 0.25)^\top \in \mathcal{H}_2$ where $(h_1, h_2, h_{12})^\top$ denotes a vector.

If $H(X_1) = 1$, $H(X_2) = 0.5$ and $H(X_1, X_2) = 0.25$, then

$$H(X_2|X_1) = H(X_1, X_2) - H(X_1) = 0.25 - 1 < 0,$$

a contradiction to the non-negativity of $H(X_2|X_1)$. Therefore, there exist no such random variables X_1 and X_2 and hence h is not entropic.

The region in $\mathcal{H}_n$ formed by all entropic vectors h is denoted by Γ_n^* . That is,

$$\Gamma_n^* = \{h \in \mathcal{H}_n : h \text{ is entropic}\}.$$

Immediate properties of Γ_n^* are:

1. Γ_n^* contains the origin: The joint entropies corresponding to n deterministic random variables form a 0-vector.
2. Γ_n^* is in the non-negative octant of $\mathcal{H}_n$: This is because all entropy measures are non-negative and hence so are the coefficients of all entropic vectors.
3. The closure of Γ_n^* , denoted by $\bar{\Gamma}_n^*$, is a convex cone (see Theorem 20).

3.2.1 Canonical Form and Elemental Inequalities

From Corollary 16, it is clear that all Shannon's information measures can be expressed in terms of joint entropies by applying the following identities:

$$\begin{aligned} H(X|Y) &= H(X, Y) - H(Y) \\ H(Y|X) &= H(X, Y) - H(X) \\ I(X; Y) &= H(X) + H(Y) - H(X, Y) \text{ and} \\ I(X; Y|Z) &= H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z) \end{aligned}$$

and the chain rules. We call this representation as the *canonical form* of an information expression. If an expression f is represented as $f(h)$, in terms of joint entropies, it means that f is in canonical form.

Since an information expression in canonical form is the linear combination of joint entropies we can write it as

$$f(h) = b^\top h, \text{ where } b^\top \text{ is the transpose of a column vector of real constants.}$$

It is proven in Corollary 13.3 of [42] that the canonical form representation of an information expression is unique.

We have already seen that the Shannon's information measures are non-negative and their non-negativity forms a set of inequalities called *basic inequalities*. However, the basic inequalities are not unique since some Shannon's information measures can be written as the linear combination of others.

For example, consider

$$\begin{aligned}
 H(X|Y) &= H(X, Y) - H(Y) \\
 &= H(X, Y, Z) - H(Y, Z) + H(X, Y) + H(Y, Z) - H(Y) - H(X, Y, Z) \\
 &= H(X|Y, Z) + I(X; Z|Y)
 \end{aligned}$$

Note that the Shannon's information measure $H(X|Y)$ is written as the sum of two other Shannon's information measures.

An information measure in the form of entropy, conditional entropy, mutual information or conditional mutual information is known as *elemental information measure*. They are of one of the following general forms:

- $H(X_i|X_{\mathcal{N}-\{i\}}); i \in \mathcal{N}$
- $I(X_i; X_j|X_K); i \neq j, K \subset \mathcal{N} - \{i, j\}$.

Proposition 17. The total number m of two elemental forms of Shannon's information measures for n random variables is equal to

$$m = n + \binom{n}{2} 2^{n-2}.$$

Proof. The total number of information measures of the form $H(X_i|X_{\mathcal{N}-\{i\}}); i \in \mathcal{N}$ is n since i varies from 1 to n .

The total number of information measures of the form $I(X_i; X_j|X_K); i \neq j, K \subset \mathcal{N} - \{i, j\}$ is

$$\binom{n}{2} \left[\binom{n-2}{0} + \binom{n-2}{1} + \dots + \binom{n-2}{n-2} \right] = \binom{n}{2} 2^{n-2}.$$

Therefore the total number $m = n + \binom{n}{2} 2^{n-2}$ as required. $\square$

With the aid of the following identities

$$\begin{aligned}
H(X) &= H(X|Y) + I(X; Y) \\
H(X, Y) &= H(X) + H(Y|X) \\
H(X|Z) &= H(X|Y, Z) + I(X; Y|Z) \\
H(X, Y|Z) &= H(X|Z) + H(Y|X, Z) \\
I(X; Y, Z) &= I(X; Y) + I(X; Z|Y) \\
I(X; Y, Z|T) &= I(X; Y|T) + I(X; Z|Y, T)
\end{aligned}$$

and the chain rules, any Shannon's information measure can be expressed as the sum of elemental forms above.

Example 3. [42] Consider

$$\begin{aligned}
H(X_1, X_2) &= H(X_1) + H(X_2|X_1) \\
&= H(X_1|X_2, X_3) + I(X_1; X_2, X_3) + H(X_2|X_1, X_3) + I(X_2; X_3|X_1) \\
&= H(X_1|X_2, X_3) + I(X_1; X_2) + I(X_1; X_3|X_2) + H(X_2|X_1, X_3) \\
&\quad + I(X_2; X_3|X_1).
\end{aligned}$$

In the above example, we saw that the basic inequality $H(X_1, X_2) \geq 0$ can be expressed as the sum of five elemental inequalities which are non-negative.

Recall that all information expressions can be expressed uniquely in canonical form; as the linear combination of $2^n - 1$ joint entropies involving all or some of the random variables $X_1, \dots, X_n$.

If elemental inequalities are expressed in canonical form, they become linear inequalities in the space $\mathcal{H}_n$. Denote this set of inequalities by $Gh \geq 0$ where G is an $m \times k$ matrix where $m = n + \binom{n}{2}2^{n-2}$ and $k = 2^n - 1$ and define the set

$$\Gamma_n = \{h : Gh \geq 0\}.$$

Example 4. Let $n = 2$, then there are $m = n + \binom{n}{2}2^{n-2} = 3$ elemental forms,

$$I(X_1; X_2), H(X_1|X_2) \text{ and } H(X_2|X_1).$$

They can be written in canonical form as:

$$\begin{aligned} I(X_1; X_2) &= H(X_1) + H(X_2) - H(X_1, X_2) \geq 0 \\ H(X_1|X_2) &= 0 - H(X_2) + H(X_1, X_2) \geq 0 \\ H(X_2|X_1) &= -H(X_1) + 0 + H(X_1, X_2) \geq 0. \end{aligned}$$

In matrix form

$$\begin{bmatrix} I(X_1, X_2) \\ H(X_1|X_2) \\ H(X_2|X_1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} H(X_1) \\ H(X_2) \\ H(X_1, X_2) \end{bmatrix} = Gh \geq 0,$$

where

$$G = \begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \text{ and } h = \begin{bmatrix} H(X_1) \\ H(X_2) \\ H(X_1, X_2) \end{bmatrix}.$$

Therefore the region Γ_2 is given by

$$\Gamma_2 = \{h : Gh \geq 0\}.$$

The region Γ_n is a *pyramid* (an n -dimensional geometric object formed by taking the union of all line segments joining points in the hyperplane spanned by an $(n - 1)$ -dimensional base B and a point P outside that hyperplane) in the non-negative orthant of $\mathcal{H}_n$. To see this, note that

- 1) the origin is in Γ_n and
- 2) the constraints $Gh \geq 0$ are linear.

Now let $e_j : 1 \leq j \leq k$ be a column vector whose j -th component is 1 and all other components are zeros. Consider the inequality

$$e_j^\top h \geq 0.$$

It corresponds to the non-negativity of a joint entropy, a basic inequality. Since the set of all basic inequalities is equivalent to the set of elemental inequalities, if $h \in \Gamma_n$, h satisfies an elemental inequality, then h satisfies the basic inequality $e_j^\top h \geq 0$. That is,

$$\Gamma_n \subset \{h : e_j^\top h \geq 0\}$$

for all $1 \leq j \leq k$. This means that

- 3) Γ_n is in the non-negative orthant of $\mathcal{H}_n$.

From statements 1), 2) and 3) we conclude that Γ_n is a pyramid in the non-negative octant of $\mathcal{H}_n$.

Now assume that $h \in \Gamma_n^*$. Then the elemental inequalities are satisfied by the entropies associated with the entropic vector h of any n random variables $X_1, \dots, X_n$. That is, $h \in \Gamma_n$ implies that

$$\Gamma_n^* \subset \Gamma_n.$$

The above expression says that Γ_n is an outer bound of the space of entropic vectors Γ_n^* .

3.3 Attempts to Characterize Γ_n^*

After defining the regions Γ_n and Γ_n^* , both pyramids in the non-negative orthant of the space $\mathcal{H}_n$ of n random variables, one can decide whether an information inequality is always true. If an inequality is formed by vectors in Γ_n , it is implied by the basic inequalities and therefore it is a Shannon inequality. Since basic inequalities always hold, so do Shannon-type inequalities. The linear structure of Γ_n is helpful in proving all Shannon-type inequalities, which is done in Chapter 14 of [42], using optimization techniques.

If $\Gamma_n = \Gamma_n^*$, then all information inequalities which always hold are Shannon-type, and hence all of them can be completely characterized. It is already proven in [42] that for dimensions $n = 2$ and 3 , there is 'not much difference' between Γ_n and Γ_n^* (see Theorem 18 and 19 below). That is, non-Shannon inequalities do not exist for $n = 2, 3$.

Theorem 18. [42]

$$\Gamma_2^* = \Gamma_2$$

The proof is given in [42], Chapter 15.

The result says that all information inequalities involving joint entropies of two random variables are Shannon-type.

For $n = 3$, Γ_n^* and Γ_n are not the same (Theorem 15.2, [42]). However,

Theorem 19. The closure $\bar{\Gamma}_3^* = \Gamma_3$ (Theorem 15.6, [42], [44]).

Furthermore, it is proven in [42] (Theorem 15.5) that

Theorem 20. $\bar{\Gamma}_n^*$ is a convex cone.

Definition 13. Information inequalities with certain constraints on the joint distribution of the random variables involved are called *constrained information inequalities*.

Usually these constraints are expressed as linear constraints on the entropies.

Suppose there are q linear constraints on the entropies given by $Qh = 0$ where Q is a $q \times k$ matrix. Then the constraints confine h to a linear subspace ϕ in $\mathcal{H}_n$ where

$$\phi = \{h \in \mathcal{H}_n : Qh = 0\}.$$

Information inequalities having no such constraints are called *unconstrained information inequalities*.

The region $\bar{\Gamma}_n^*$ is not sufficient to characterize all information inequalities, however it is sufficient to characterize all unconstrained information inequalities, which can be explained as follows:

Consider an unconstrained information inequality $f \geq 0$ where $f(h) = b^\top h$. Then $f \geq 0$ corresponds to the set

$$\{h \in \mathcal{H}_n : b^\top h \geq 0\},$$

a half space containing the origin. More precisely, for any $h \in \mathcal{H}_n$, $f(h) \geq 0$ if and only if h lies in this half space.

An information inequality always holds if and only if it is satisfied by the entropy of any joint distribution of the random variables involved. Therefore, a geometric interpretation of an unconstrained inequality is:

$$f \geq 0 \text{ always holds if and only if } \Gamma_n^* \subset \{h \in \mathcal{H}_n : f(h) \geq 0\}.$$

In other words, if an entropic vector $h_0 \in \Gamma_n^*$ lies outside the half space $\{h \in \mathcal{H}_n : f(h) \geq 0\}$, then $f(h_0) < 0$, and hence $f \geq 0$ is not always true. This gives a complete characterization of all unconstrained inequalities in terms of Γ_n^* .

Since the set $\{h \in \mathcal{H}_n : f(h) \geq 0\}$ is closed and the smallest closed set containing Γ_n^* is $\bar{\Gamma}_n^*$, we have

$$\Gamma_n^* \subset \bar{\Gamma}_n^* \subset \{h \in \mathcal{H}_n : f(h) \geq 0\}.$$

Therefore, $\bar{\Gamma}_n^*$ is sufficient for characterizing all unconstrained information inequalities and from Theorem 19, it follows that there exists no unconstrained information inequalities involving three random random variables other than the Shannon-type inequalities.

Since $\bar{\Gamma}_3^* = \Gamma_3$, it is natural to ask whether it is possible to extend this result to $n \geq 4$. But it is not known even for $n = 4$. Also the existence of non-Shannon-type information inequalities for $n = 4$ are proven in [42, 44].

We look at non-Shannon-type inequalities involving more than three random variables in Chapter [5](#).

Connection Between Groups and Entropy

A *group* is a basic structure in abstract algebra while entropy is a basic information measure. At a first glance, these two structures seem to have no obvious relations. However, they are intimately related to each other [6, 11, 12, 42] which in turn bring forth a nice connection between two branches of mathematics, abstract algebra and information theory. This chapter explains these relations between entropies and finite groups and how these connections are useful to each other.

4.1 Basics of Group Theory

This section contains elementary results from group theory which are recalled for the sake of completeness. A reader familiar with group theory may skip this section.

4.1.1 Groups and Subgroups

A group is defined as follows:

Definition 14. A group is a non-empty set G together with a binary operation $(a, b) \mapsto ab$ which satisfies the following axioms:

- 1) $ab \in G$ if $a, b \in G$ (Closure)
- 2) $a(bc) = (ab)c$ for all $a, b, c \in G$ (Associativity)
- 3) There exists an element $e \in G$, called identity, such that $ae = ea$ for all $a \in G$ (Identity)
- 4) There exists an element a^{-1} , inverse of a , such that $aa^{-1} = a^{-1}a = e$ for all $a \in G$ (Inverse).

If for all $a, b \in G$, $ab = ba$, then G is called *abelian* (commutative).

Example 5. Some common groups are

- 1) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are groups under addition with $e = 0$ and $a^{-1} = -a$ for all a .
- 2) $\mathbb{Q} \setminus \{0\}$, $\mathbb{R} \setminus \{0\}$, $\mathbb{C} \setminus \{0\}$, $\mathbb{Q}^+$, $\mathbb{R}^+$ are groups under multiplication with $e = 1$ and $a^{-1} = 1/a$.
- 3) $\mathbb{Z} \setminus \{0\}$ is not a group under multiplication, since $2^{-1} = 1/2 \notin \mathbb{Z} \setminus \{0\}$.

Definition 15. The *order* of a group G is the number of elements of G and is denoted by $|G|$.

If $|G| = n < \infty$, we call G a *finite group*. All groups here are finite unless otherwise stated.

Example 6. 1) The trivial group $G = \{e\}$ containing the identity element is the only group of order 1.

2) The group $G = \{0, 1, \dots, n-1\}$ of integers under addition modulo n is a group of order n , denoted by $\mathbb{Z}_n$.

Remark 3. If the binary operation $(a, b) \mapsto ab$ is addition or multiplication, we denote $e = 0$ or $e = 1$ respectively.

A non-empty subset of a group G is said to be a *subgroup*, if it is a group itself under the binary operation defined on G . More precisely,

Definition 16. A non-empty subset H of a group G is a *subgroup*, if it is closed under product and inverse. That is, if $x, y \in H$, $xy \in H$ and $x^{-1} \in H$ for all $x \in H$.

$H \leq G$ denotes a subgroup H of G .

A subgroup $H \leq G$ is said to be *non-trivial*, if $H \neq \{e\}$, the trivial subgroup and H is said to be *proper*, if $H \neq G$, denoted by $H < G$.

Example 7. 1) $\mathbb{Z} < \mathbb{Q}$, $\mathbb{Q} < \mathbb{R}$, $\mathbb{R} < \mathbb{C}$, $\mathbb{Z} < \mathbb{C}$ are all subgroups under addition.

2) If $G = \mathbb{Z}_4 = \{0, 1, 2, 3\}$, then $H = \{0, 2\} \leq G$.

Proposition 21. Let G_1 and G_2 be subgroups of G . Then $G_{12} \leq G$ where $G_{12} = G_1 \cap G_2$.

Proof. Suppose $g_1, g_2 \in G_{12}$. Then $g_1, g_2 \in G_1$ and G_2 .

Since $G_1, G_2 \leq G$, $g_1 g_2 \in G_1$ and G_2 . Therefore, $g_1 g_2 \in G_{12}$.

Similarly, $g^{-1} \in G_{12}$ for all $g \in G_{12}$. Hence the axioms of subgroup are satisfied. □

Corollary 22. If $G_1, \dots, G_n$ are subgroups of G , then $\cap_{i=1}^n G_i$ is also a subgroup of G .

4.1.2 Homomorphisms and Isomorphisms

After defining groups and subgroups, let us now introduce a map that helps us to go from one group to another while preserving the group operations of respective groups.

Definition 17. Let G and H be two groups. A map $f : G \rightarrow H$ is said to be a *homomorphism* if $f(xy) = f(x)f(y)$ for all $x, y \in G$.

Note that the binary operation xy (written multiplicatively) on the left is computed in G and the binary operation $f(x)f(y)$ (written multiplicatively) on the right is computed in H .

Proposition 23. If $f : G \rightarrow H$ is a homomorphism then

- 1) $f(e_G) = e_H$; e_G and e_H are identity elements of G and H respectively
- 2) $f(x^{-1}) = (f(x))^{-1}$.

Proof. 1) Let f be a homomorphism from G to H . Then $f(e_G) \in H$.

Now $f(e_G) = f(e_G e_G) = f(e_G)f(e_G)$. Left multiplying with $(f(e_G))^{-1}$ on both sides implies $e_H = f(e_G)$.

2) $f(xx^{-1}) = f(e_G) = e_H = f(x)f(x^{-1})$. Therefore, $(f(x))^{-1} = f(x^{-1})$. □

Definition 18. The *kernel* of a group homomorphism $f : G \rightarrow H$ is defined as

$$\text{Ker}(f) = \{a \in G : f(a) = e_H\}$$

where e_H is the identity element in H .

Remark 4. The kernel of a homomorphism $f : G \rightarrow H$ is a normal subgroup of G .

Proof.

$$f(aba^{-1}) = f(a)f(b)f(a^{-1}) = f(a)f(b)[f(a)]^{-1} = f(a)[f(a)]^{-1} = e_H$$

for all $b \in \text{Ker}(f)$ and for all $a \in G$. □

Definition 19. A homomorphism $f : G \rightarrow H$ is said to be an *isomorphism* if it is a bijection.

If there exists an isomorphism from G to H , we say that G isomorphic to H , denoted by $G \cong H$. Intuitively speaking, G and H are the same except for the different representations of elements and operation.

Example 8. Let $G = (\mathbb{R}, +)$ and $H = (R^+, \times)$. Define $f : G \rightarrow H$ such that $f(x) = e^x$ where e is the base of the natural logarithm. Then $G \cong H$. Note that f is a bijection since it has an inverse namely $f^{-1} = \log_e$.

We have already seen the order of a group. Next we define the order of an element of a group G .

Definition 20. The *order* of an element $g \in G$ is the least positive integer n such that $g^n = e$ where e is the identity in G . We denote it by $|g|$. If such an integer n does not exist, the order is said to be infinity.

4.1.3 Cyclic Groups

Let G be a group and $g \in G$. If all the other elements of G are powers of g , then we say that g generates the group G . More precisely,

Definition 21. If there exists an element $g \in G$ such that $x = g^n$ for all $x \in G$ and for some integer n , then G is called a *cyclic group* and g , a *generator* of G . i.e., $G = \{g^n | n \in \mathbb{Z}\}$.

In additive notation, G is cyclic if $G = \{ng | n \in \mathbb{Z}\}$ and in both multiplicative and additive cases, we can write $G = \langle g \rangle$. Note that the generator need not be unique. For example, if $G = \langle g \rangle$, then $G = \langle g^{-1} \rangle$ since $(g^{-1})^n = g^{-n}$ where n runs over all integers and so does $-n$. We denote a cyclic group of order n by C_n .

Example 9. 1) The set of integer $\mathbb{Z}$ under addition is a cyclic group generated by either 1 or -1. i.e., $(\mathbb{Z}, +) = \langle 1 \rangle = \langle -1 \rangle$.

Proposition 24. If G is a cyclic group, then G is abelian.

Proof. The result obviously follows from the definition. □

Proposition 25. A subgroup of a cyclic group is cyclic.

Proof. Let $G = \langle g \rangle$ be a cyclic group of order n . Let H be a proper non-trivial subgroup of G . All elements of H are some powers of g . Let k be the smallest power of g such that $g^k \in H$. Then its inverse $g^{-k} \in H$.

Now let $h \in H$ be an arbitrary element of H . Since $h \in G$, $h = g^m$ for some integer m . Then $h = g^m = g^{kq+r} = g^{kq}g^r$ where q and r are integers such that $0 \leq r < k$ by division algorithm. This implies $g^r = g^m g^{-kq}$. Since $g^{-k} \in H$, $g^{-kq} \in H$ and hence $g^r \in H$. But k is the smallest integer with $g^k \in H$ implying $r = 0$ and H is cyclic. □

4.1.4 Cosets and Lagrange's Theorem

This section introduces cosets of a subgroup and how these structures are related to a given group.

Definition 22. Let H be a subgroup of a group G . If $g \in G$, the left coset of H generated by g is $gH = \{gh, h \in H\}$. Similarly, the right coset is given by $Hg = \{hg, h \in H\}$.

If G is abelian, there is no distinction between the left and right cosets.

Lemma 26. $aH = bH$ if and only if $a^{-1}b \in H$ for $a, b \in G$. Similarly $Ha = Hb$ if and only if $ab^{-1} \in H$.

Proof. Assume that $aH = bH$. Since $H \leq G$, $e \in H$, we see that $b = ah$ for some $h \in H$, which implies that $h = a^{-1}b \in H$.

Conversely, if $a^{-1}b = h \in H$, then $b = ah$ and hence $bH = ahH = aH$. $\square$

Corollary 27. Two cosets of a subgroup H of G are either the same or disjoint.

Proof. Suppose $k \in aH \cap bH$. Then $k = ah_1 = bh_2$ for some $h_1, h_2 \in H$ and hence $a^{-1}b = h_1h_2^{-1} \in H$. By Lemma 26, $aH = bH$. $\square$

Furthermore, since the map $h \mapsto ah, h \in H$, is one-one, each left(right) coset of H has $|H|$ elements. Next we define the index of a subgroup.

Definition 23. The *index* of a subgroup H of a group G is the number of left(right) cosets of H in G and it is denoted by $[G : H]$.

Even though we work with non-abelian groups, by coset we simply mean left coset unless otherwise specified.

Theorem 28 (Lagrange's Theorem). If $H \leq G$, then $|G| = |H|[G : H]$. In particular, if G is finite, then $|H|$ divides $|G|$ and $[G : H] = \frac{|G|}{|H|}$.

Proof. Since any two cosets of H in G are either the same or disjoint, cosets of H forms a partition of G . That is, $G = \sqcup gH$ where g runs through each coset of H .

Now $|G| = |\sqcup gH| = \sum |gH| = \sum |H|$ since we have a disjoint union of cosets and each coset has $|H|$ elements. Therefore, $|G| = [G : H]|H|$ since the summation is over all cosets. $\square$

4.1.5 Normal Subgroups and Quotient Groups

We know how to get cosets given a subgroup of a group. This set of cosets does not have a specific structure however. In order to give a group structure to the set of all cosets, we introduce normal subgroups and quotient groups.

Definition 24. Let $H \leq G$. If $gHg^{-1} = H$ for all $g \in G$, we say that H is a *normal subgroup* of G or H is normal in G , denoted by $H \trianglelefteq G$.

Example 4.1. 1) Let $G = GL_n(\mathbb{R})$ denote the set of invertible $n \times n$ real matrices, which is a group under matrix multiplication, called *general linear group*. Then the set of all invertible real matrices with determinant 1, $H = SL_n(\mathbb{R})$ is a subgroup of G , called *special linear group*. In fact H is normal in G . To see this, consider $A \in G$, $B \in H$. Now $\det(ABA^{-1}) = \det(B) = 1$ implies $ABA^{-1} \in H$ for all $A \in G$ and for all $B \in H$.

2) It is obvious that all subgroups of an abelian group are normal.

The normal structure of a given subgroup provides a group structure to its set of all cosets.

Proposition 29. If $H \trianglelefteq G$, then the set of all cosets of H forms a group.

Proof. Define a binary operation on cosets as follows:

$$(aH, bH) \mapsto (aH)(bH) = \{ahbh' : ah \in aH, bh' \in bH\}.$$

Now we check that all group axioms are satisfied.

1) Let aH, bH are two cosets of H , then $aHbH = a(bH)H = abH$, another coset since $gH = Hg$ for all $g \in G$.

2) Associativity of G implies the associativity of the set of cosets.

3) Identity element is given by $eH = H$.

4) Inverse of a coset aH is $a^{-1}H$. □

Definition 25. The group of cosets of a normal subgroup N of G is called the *quotient group* of G and is denoted by G/N .

Next we see that every normal subgroup of a given group can be seen as the kernel of a homomorphism.

Lemma 30. Let G be a group. Every normal subgroup of G is the kernel of a homomorphism.

Proof. Suppose that $N \trianglelefteq G$. Then G/N is the quotient group. Consider a mapping $\pi : G \rightarrow G/N$ defined by $\pi(a) = aN$. In fact, π is a homomorphism from G to G/N since $\pi(ab) = abN = (aN)(bN) = \pi(a)\pi(b)$.

Now $\ker(\pi) = \{a \in G : \pi(a) = N\} = \{a \in G : aN = N\} = N$. That is, N is the kernel of the homomorphism π . □

The mapping π above is known as *canonical mapping*.

Next we state the isomorphism theorems without proof. Some of them are useful in later chapters.

Theorem 31. (1st isomorphism theorem)[15] If $f : G \rightarrow H$ is a group homomorphism with kernel K , then the image of f is isomorphic to G/K :

$$\text{Im}(f) \simeq G/\text{Ker}(f).$$

Theorem 32. (2nd isomorphism theorem) [15] If H and N are subgroups of G , with N normal in G , then

$$H/(H \cap N) \simeq HN/N.$$

Theorem 33. (3rd isomorphism theorem) [15] If H and N are normal subgroups of G , with N contained in H , then

$$G/H \simeq (G/N)/(H/N).$$

Theorem 34. (4th or lattice isomorphism theorem) [15] Let G be a group and $N \trianglelefteq G$. Then there is a bijection from the set of subgroups A of G which contain N onto the set of subgroups $\bar{A} = A/N$ of $G/N = \bar{G}$. In particular, every subgroup of $\bar{G}$ is of the form A/N for some subgroup A of G containing N .

This bijection has the following properties:

1. $A \leq B$ if and only if $\bar{A} \leq \bar{B}$,
2. $A \leq B$, then $[B : A] = [\bar{B} : \bar{A}]$,
3. $A \trianglelefteq G$ if and only if $\bar{A} \trianglelefteq \bar{G}$,
4. $\overline{A \cap B} = \bar{A} \cap \bar{B}$.

4.1.6 Direct Product of Groups

It would be nice to understand ways of building larger new groups given a collection of groups. Start with groups H and K , and let $H \times K$ be their cartesian product. That is,

$$H \times K = \{(h, k) : h \in H, k \in K\}.$$

Defining componentwise binary operation as the overall binary operation, the set becomes a group. To see this,

- 1) Let $(h_1, k_1), (h_2, k_2) \in H \times K$. Then $(h_1 h_2, k_1 k_2) \in H \times K$, closure property.

- 2) Associativity follows from the associativity of H and K .
- 3) Identity element is given as $1 = (1_H, 1_K)$.
- 4) Inverse of (h, k) is (h^{-1}, k^{-1}) .

Definition 26. Let H and K be two groups. The group $G = H \times K$ with componentwise multiplication as the binary operation as defined above is called the *external direct product* of H and K .

The direct product can be extended to any number of terms.

Example 10. Let $\mathbb{Z}_2 = \{0, 1\}$ under addition modulo 2. Then $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, the Klein's four group.

4.2 Group Representable Entropy Function

From Chapter 3, Γ_n^* is the set of all entropic vectors in the space $\mathcal{H}_n$ for n random variables. To establish the connection between entropy and groups, we first discuss the entropic vectors in Γ_n^* which can be described by a finite group G and subgroups $G_1, \dots, G_n$. Such entropic vectors are called *group representable*.

We use the following lemma to prove the main results in this section.

Lemma 35. Let $G_1, \dots, G_n$ be subgroups of G . For any non-empty subset $\mathcal{A} \subseteq \mathcal{N}$, if $\cap_{i \in \mathcal{A}} g_i G_i$ is non-empty for $g_i \in G_i$, then there exists an element $a \in G$ such that $\cap_{i \in \mathcal{A}} g_i G_i = a G_{\mathcal{A}}$ where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$.

Proof. Since $\cap_{i \in \mathcal{A}} g_i G_i$ is non-empty, there exists an element $a \in g_i G_i$ for all i . Therefore $g_i G_i = a G_i$ for all $i \in \mathcal{A}$.

Now $G_{\mathcal{A}} \subseteq G_i$ for all $i \in \mathcal{A}$ implies $a G_{\mathcal{A}} \subseteq a G_i$ for all $i \in \mathcal{A}$. Then $a G_{\mathcal{A}} \subseteq \cap_{i \in \mathcal{A}} a G_i = \cap_{i \in \mathcal{A}} g_i G_i$.

Conversely, suppose that $b \in \cap_{i \in \mathcal{A}} g_i G_i$. That is, $b \in \cap_{i \in \mathcal{A}} a G_i$ implies $b \in a G_i \forall i$. Therefore, $a^{-1}b \in G_i \forall i$ and hence $b \in a G_{\mathcal{A}}$. $\square$

From the above Lemma, it is clear that

$$|\cap_{i \in \mathcal{A}} a_i G_i| = |G_{\mathcal{A}}|$$

if $\cap_{i \in \mathcal{A}} a_i G_i$ is non-empty.

Theorem 36. [6, 12] For any finite group G and any subgroups $G_1, \dots, G_n$ of G , there exists a set of n jointly distributed discrete random variables $X_1, \dots, X_n$ such that for all non-empty subsets $\mathcal{A}$ of $\mathcal{N}$, $H(X_{\mathcal{A}}) = \log [G : G_{\mathcal{A}}]$, where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$ and $[G : G_{\mathcal{A}}]$ is the index of $G_{\mathcal{A}}$ in G . i.e, the vector $h \in \mathcal{H}_n$ defined by $h_{\mathcal{A}} = \log [G : G_{\mathcal{A}}]$ is entropic, belongs to Γ_n^* .

Proof. Let X be a random variable uniformly distributed over G with probability $P(X = g) = 1/|G|$ for all $g \in G$. For each $i \in \{1, \dots, n\}$, define new random variables X_i such that $X_i = XG_i$. That is, if $X = g$, then $X_i = gG_i$, a left coset of G_i in G .

Since there are $[G : G_i]$ left cosets of G_i , the support of X_i is the set of all $[G : G_i]$ cosets of G_i in G . Also $gG_i = hG_i$ if and only if g and h are in same left coset of G_i . Then

$$P(X_i = gG_i) = \sum_{h \in gG_i} P(X = h) = \frac{|gG_i|}{|G|} = \frac{|G_i|}{|G|}.$$

The total probability is 1 since there are $\frac{|G|}{|G_i|}$ distinct cosets.

If $P(X_i = g_iG_i : i \in \mathcal{A}) > 0$, there exists a $a \in G$ such that $a \in \cap_{i \in \mathcal{A}} g_iG_i$. Therefore using the Lemma 35, we have

$$P(X_i = g_iG_i : i \in \mathcal{A}) = \frac{|\cap_{i \in \mathcal{A}} g_iG_i|}{|G|} = \frac{|aG_{\mathcal{A}}|}{|G|} = \frac{|G_{\mathcal{A}}|}{|G|}.$$

That is, $X_i; i \in \mathcal{A}$ distributes uniformly on its support whose cardinality is $\frac{|G|}{|G_{\mathcal{A}}|}$. Moreover,

$$H(X_{\mathcal{A}}) = \sum_{g_iG_i} P(X_i = g_iG_i, i \in \mathcal{A}) \log \frac{|G|}{|G_{\mathcal{A}}|} = \log \frac{|G|}{|G_{\mathcal{A}}|} \left[\frac{|G|}{|G_{\mathcal{A}}|} \frac{|G_{\mathcal{A}}|}{|G|} \right] = \log [G : G_{\mathcal{A}}].$$

□

This result leads to the following definition of *group representability*.

Definition 27. Let $X_1, \dots, X_n$ be n jointly distributed discrete random variables. The corresponding entropic vector (Definition 12) h is said to be *group representable*, if there exists a finite group G , with subgroups $G_1, \dots, G_n$ such that $H(X_{\mathcal{A}}) = \log [G : G_{\mathcal{A}}]$ for all $\mathcal{A}$. If in addition, the group G is abelian, we say that h is *abelian group representable*.

In this context, Theorem 36 asserts that some entropic vectors in Γ_n^* have a group representation. Such entropic vectors are called group representable entropic vectors, and are useful in determining the region $\bar{\Gamma}_n^*$.

Example 11. [42] Let h be a vector in $\mathcal{H}_3$ as follows:

$$h_{\mathcal{A}} = \min(|\mathcal{A}|, 2).$$

That is,

$$h = (1, 1, 1, 2, 2, 2, 2)^\top.$$

Now consider the Klein's four group $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ and subgroups $G_1 = \{(0, 0), (1, 0)\}$, $G_2 = \{(0, 0), (0, 1)\}$ and $G_3 = \{(0, 0), (1, 1)\}$. Then (G, G_1, G_2, G_3) is a group representation of h .

Example 12. [42] Let $\mathcal{B}$ be a non-empty subset of $\mathcal{N}$ and define $h \in \mathcal{H}_n$ as:

$$h_{\mathcal{A}} = \begin{cases} \log 2 & \text{if } \mathcal{A} \cap \mathcal{B} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

Then $(G, G_1, \dots, G_n)$ is a group representation of h where $G = \mathbb{Z}_2$ and

$$G_i = \begin{cases} \{0\} & \text{if } i \in \mathcal{B} \\ G & \text{otherwise.} \end{cases}$$

The origin, $h = 0$, has a group representation $(G, G_1, \dots, G_n)$ with $G = G_1 = \dots = G_n$ by setting $\mathcal{B} = \emptyset$.

Remark 5. It is easy to see that the group representation is not unique, since it depends only on the index of the subgroups chosen.

4.3 $\overline{\Gamma}_n^*$ and Group Representability

We have introduced group representable entropic vectors in the last section. Denote this set by Υ_n . i.e.,

Definition 28. The region of all entropic vectors in $\mathcal{H}_n$ which are group representable is defined as

$$\Upsilon_n = \{h \in \mathcal{H}_n : h \text{ has a group representation}\}.$$

From Theorem 36, if $h \in \Upsilon_n$, then $h \in \Gamma_n^*$. Therefore, $\Upsilon_n \subset \Gamma_n^*$.

But this inclusion is strict, which can be seen as follows:

Suppose $h \in \Gamma_n^*$. Then there exists a collection of random variables $X_1, \dots, X_n$ such that

$$h_{\mathcal{A}} = H(X_{\mathcal{A}}), \forall \text{ non-empty subset } \mathcal{A} \subseteq \mathcal{N}.$$

If h is groups representable, then there exists finite groups G and subgroups $G_1, \dots, G_n$ such that

$$H(X_{\mathcal{A}}) = \log \frac{|G|}{|G_{\mathcal{A}}|}.$$

Since $|G|$ and $|G_{\mathcal{A}}|$ are integers, the entropy $H(X_{\mathcal{A}})$ is the logarithm of a rational number. But the joint entropy of a set of random variables is not necessarily the logarithm of a rational number (Corollary 2.44 [42]). Therefore, it is possible to construct an entropy function which has no group representation.

In the following part we can see that the set of all group representable entropic vectors is good enough to characterize the region Γ_n^* .

Theorem 37. (Theorem 16.22,[42], [11]) For any $h \in \Gamma_n^*$, there exists a sequence $\{f^{(r)}\} \in \Upsilon_n$ such that $\lim_{r \rightarrow \infty} \frac{1}{r} f^{(r)} = h$.

This theorem asserts that the region of entropic vectors lies in the convex closure (smallest closed convex set) of Υ_n , denoted by $\overline{\text{con}}(\Upsilon_n)$.

Corollary 38. We have $\overline{\text{con}}(\Upsilon_n) = \bar{\Gamma}_n^*$.

Proof. We have $\Upsilon_n \subset \Gamma_n^*$. Taking convex closure on both sides yields $\overline{\text{con}}(\Upsilon_n) \subset \overline{\text{con}}(\Gamma_n^*)$. But from Theorem 20, $\bar{\Gamma}_n^*$ is convex and therefore $\overline{\text{con}}(\Gamma_n^*) = \bar{\Gamma}_n^*$. Then $\overline{\text{con}}(\Upsilon_n) \subset \bar{\Gamma}_n^*$.

From Example 12, the origin has a group representation and belongs to Υ_n . Then it is true from Theorem 37 that $\bar{\Gamma}_n^* \subset \overline{\text{con}}(\Upsilon_n)$. Hence we have proved that $\overline{\text{con}}(\Upsilon_n) = \bar{\Gamma}_n^*$. $\square$

Therefore from the corollary above, to study the region of entropic vectors, we just need to focus on the region of group representable entropic vectors.

Using the group representation of entropies we can rewrite information inequalities in terms of groups. More results towards that direction follow in Chapter 6.

4.4 Introduction to Quasi-Uniform Random Variables

We have seen in Theorem 36 that the random variables $X_1, \dots, X_n$ associated with a finite group G and subgroups $G_1, \dots, G_n$ are uniformly distributed over their support for all non-empty subset $\mathcal{A} \subset \mathcal{N}$. We usually refer those kind of random variable as quasi-uniform. That is, associated to n subgroups of a finite group, we can find n quasi-uniform random variables such that $H(X_{\mathcal{A}}) = \log [G : G_{\mathcal{A}}]$, for all $\mathcal{A}$.

Quasi-uniform random variables were introduced in [12] when studying the connection between information inequalities and combinatorics. The concept of such random variables

in fact originated from the Asymptotic Equipartition Property (AEP) which we discuss here in order to understand the background.

4.4.1 Asymptotic Equipartition Property

[42] Asymptotic equipartition property is analogous to the weak law of large numbers, which says that for independent and identically distributed (i.i.d) random variables X_i defined according to $P(x)$, (the probability mass function of a random variable X), $\frac{1}{n} \sum_{i=1}^n X_i$ is close to its expectation $E(X)$ for large n where as AEP states that when n goes larger $\frac{1}{n} \log \frac{1}{P(X_1, \dots, X_n)}$ is closer to the entropy $H(X)$, where $P(X_1, \dots, X_n)$ is the probability of the sequence $(X_1, \dots, X_n)$ occurring. Formally,

Theorem 39. If $X_1, \dots, X_n$ are i.i.d random variables defined according to $P(x)$, then

$$-\frac{1}{n} \log P(X_1, \dots, X_n) \rightarrow H(X) \text{ in probability.}$$

That is, for all $\epsilon > 0$, $Pr[|\frac{1}{n} \log \frac{1}{P(X_1, \dots, X_n)} - H(X)| > \epsilon] \rightarrow 0$. This allows us to divide the set of all sequences into two sets, the *typical set* where the sample entropy is close to the true entropy and other set which is the complement of the typical set.

Definition 29. The *typical set* $\mathcal{A}_\epsilon^{(n)}$ with respect to $P(X)$ is the set of sequences $(x_1, \dots, x_n)$ with probability

$$2^{-n(H(X)+\epsilon)} \leq P(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}.$$

The typical set defined above has the following properties:

Theorem 40. [42] If $(x_1, \dots, x_n) \in \mathcal{A}_\epsilon^{(n)}$, then (1) $H(X) - \epsilon \leq -\frac{1}{n} \log P(x_1, \dots, x_n) \leq H(X) + \epsilon$ and (2) $Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1 - \epsilon$ for sufficiently large n .

Proof. The first part is immediate from the definition of the typical set which says that $|\frac{1}{n} \log \frac{1}{P(X_1, \dots, X_n)} - H(X)| \leq \epsilon$. If a sequence $(X_1, \dots, X_n)$ belongs to the typical set, then the second part follows from the first part and Theorem 39. $\square$

That is, for sufficiently long sequences of independent and identically distributed random variables, almost all the probability concentrates on the typical set. Moreover, each element of the typical set has approximately same probability. In particular, the typical set has an approximate uniform distribution with total probability close to one.

Quasi-uniform random variables provide an extreme case for AEP where the total probability is uniformly distributed over the support. Before introducing quasi-uniform random variables, we recall what is a uniform distribution.

4.4.2 Uniform Distribution

A uniform distribution is a probability distribution in which all the possible outcomes are equally likely. That is, the probability of drawing any of the outcomes in the sample space is same. For example, tossing an unbiased coin is uniform since getting a head or tail is equally likely. Similarly, a deck of cards has uniform distribution because drawing a diamond, heart, club or spade has the same probability.

There are two types of uniform distributions: continuous and discrete. If the sample space contains only finitely many values, then it is discrete and continuous otherwise. Formally, we define the uniform distribution associated with a discrete random variable as follows.

Definition 30. Let X be a discrete random variable taking n values, $n \in \mathbb{N}$. If the probability, $P_X(x) = 1/n$ for all x in the sample space, then X is called a *discrete uniform random variable*.

The support of a probability distribution is the set of points in the sample space where the probability is non-zero. Obviously the support of a uniform distribution is its sample space.

4.4.3 Quasi-Uniform Distributions

As the name itself suggests, quasi-uniform distributions resemble uniform distributions, but in fact they are not.

Definition 31. A probability distribution defined over a random variable X is said to be quasi-uniform, if it is uniformly distributed over the support. That is,

$$P_X(x) = \begin{cases} 1/|\lambda(X)| & \text{if } x \in \lambda(X), \\ 0 & \text{otherwise.} \end{cases}$$

We generalize the above definition for a joint-distribution of n different random variables. Consider a set of jointly distributed discrete random variables $X_1, \dots, X_n$ over an alphabet of size N . If $\mathcal{A} \subseteq \mathcal{N} = \{1, \dots, n\}$, $X_{\mathcal{A}} = \{X_i : i \in \mathcal{A}\}$ denotes the joint distribution and $\lambda(X_{\mathcal{A}})$ is its support.

Definition 32. A probability distribution over a set of random variable $X_1, \dots, X_n$ is said to be *quasi-uniform*, if for any $\mathcal{A} \subseteq \mathcal{N}$, $X_{\mathcal{A}}$ is uniformly distributed over its support $\lambda(X_{\mathcal{A}})$:

$$P((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \begin{cases} 1/|\lambda(X_{\mathcal{A}})| & \text{if } (x_1, \dots, x_n) \in \lambda(X_{\mathcal{A}}), \\ 0 & \text{otherwise.} \end{cases}$$

FIGURE 4.1: Quasi-uniform and non quasi-uniform distributions.

Remark 6. By definition, the entropy of a quasi-uniform distribution is

$$H(X_{\mathcal{A}}) = - \sum_{x_{\mathcal{A}} \in \lambda(X_{\mathcal{A}})} P(X_{\mathcal{A}} = x_{\mathcal{A}}) \log P(X_{\mathcal{A}} = x_{\mathcal{A}}) = \log |\lambda(X_{\mathcal{A}})|.$$

The following diagram shows examples of quasi-uniform and non quasi-uniform random variables. One way of constructing quasi-uniform random variables algebraically is using groups and subgroups (Theorem 36). It can also be seen that given a group G and n subgroups $G_1, \dots, G_n$ one obtains n quasi-uniform random variables and the subgroup structure G determines the correlation among random variables. We restate the theorem including the quasi-uniformity.

Theorem 41. [5, 6, 12] For any finite group G and any subgroups $G_1, \dots, G_n$ of G , there exists a set of n jointly distributed quasi-uniform discrete random variables $X_1, \dots, X_n$ such that for all non-empty subsets $\mathcal{A}$ of $\mathcal{N}$, $H(X_{\mathcal{A}}) = \log [G : G_{\mathcal{A}}]$, where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$ and $[G : G_{\mathcal{A}}]$ is the index of $G_{\mathcal{A}}$ in G .

4.5 Region of Entropic Vectors from Quasi-Uniform Distributions

Consider the region of entropic vectors coming from quasi-uniform random variables:

$$\Psi^* = \{h \in \mathcal{H}_n : h \text{ is quasi-uniform}\}. \quad (4.1)$$

Since all group representable entropic vectors are quasi-uniform by Theorem 41, we have

$$\Upsilon_n \subset \Psi^*.$$

From Corollary 38, it follows from [12] that

$$\overline{\text{con}}(\Psi^*) = \bar{\Gamma}_n^*.$$

Therefore, the quasi-uniform random variables form another class of entropic vectors good enough to characterize the region of entropic vectors.

We know from Section 3.3 that, an unconstrained information inequality $b^\top h \geq 0$ always holds if and only if

$$\bar{\Gamma}_n^* \subset \{h \in \mathcal{H}_n : b^\top h \geq 0\}.$$

Now

$$\Upsilon_n \subset \Psi^* \subset \Gamma_n^* \subset \bar{\Gamma}_n^* \subset \{h \in \mathcal{H}_n : b^\top h \geq 0\}.$$

Since $\overline{\text{con}}(\Upsilon_n) = \overline{\text{con}}(\Psi^*) = \bar{\Gamma}_n^*$ and $\{h \in \mathcal{H}_n : b^\top h \geq 0\}$ is closed, by taking the convex closure on both sides of the above relation, we have

$$\bar{\Gamma}_n^* = \overline{\text{con}}(\Upsilon_n) = \overline{\text{con}}(\Psi^*) \subset \{h \in \mathcal{H}_n : b^\top h \geq 0\}.$$

Therefore an unconstrained information inequality is valid if and only if it is satisfied by all quasi-uniform or group representable random variables.

In this chapter, we have seen the relation between entropies and groups and how do both branches complement each other in analyzing the region of entropic vectors Γ_n^* , which plays an important role in information theory, especially in proving information inequalities. In general $\Gamma_n^* \subset \mathcal{H}_n$, and it has a complicated structure. It is already mentioned that Γ_n^* is not even closed for $n \geq 3$. But the closure $\bar{\Gamma}_n^*$ is a closed convex cone and therefore it is more manageable than Γ_n^* . However the set of all group representable entropic functions Υ_n is sufficient to study $\bar{\Gamma}_n^*$ as seen earlier and it has some nice algebraic properties. The study on group representability leads us to another class of entropic vectors called quasi-uniform, which are also sufficient to understand the region $\bar{\Gamma}_n^*$.

That is;

$$\Upsilon_n \subset \Psi^* \subset \Gamma_n^* \subset \bar{\Gamma}_n^* \tag{4.2}$$

and

$$\overline{\text{con}}(\Upsilon_n) = \overline{\text{con}}(\Psi^*) = \bar{\Gamma}_n^*. \tag{4.3}$$

In the next chapter we will see a non-trivial inner bound for Γ_n^* which is the set of all abelian group representable (Definition 27) entropic vectors, Υ_n^{ab} and we study more on the gap between Υ_n^{ab} and Υ_n .

Abelian Group Representability of Finite Groups

In Chapter 4, we have seen that the set Υ_n of group representable entropic vectors are sufficient to understand the whole region Γ_n^* of entropic vectors, since $\overline{\text{con}}(\Upsilon_n) = \overline{\Gamma}_n^*$. This chapter is based on [37] and provides more details of a non-trivial inner bound for $\overline{\Gamma}_n^*$ given by the set of all abelian group representable entropic vectors.

5.1 Abelian Group Representability

Let Υ_n^{ab} denote the set of all abelian group representable entropic vectors. i.e.,

$$\Upsilon_n^{ab} = \{h \in \mathcal{H}_n : h \text{ abelian group representable}\}. \quad (5.1)$$

We may use *abelian representable* instead of abelian group representable for brevity.

Consider the following inequality known as *Ingleton inequality* [12],

$$\begin{aligned} & h(\mathcal{A}_1) + h(\mathcal{A}_2) + h(\mathcal{A}_3 \cup \mathcal{A}_4) + h(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3) + h(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_4) \\ & \leq h(\mathcal{A}_1 \cup \mathcal{A}_2) + h(\mathcal{A}_1 \cup \mathcal{A}_3) + h(\mathcal{A}_1 \cup \mathcal{A}_4) + h(\mathcal{A}_2 \cup \mathcal{A}_3) + h(\mathcal{A}_2 \cup \mathcal{A}_4) \end{aligned} \quad (5.2)$$

where $h(\mathcal{A}_i) = H(X_{\mathcal{A}_i})$, $\mathcal{A} \subseteq \mathcal{N} = \{1, \dots, n\}$. The following Theorem states that all abelian representable entropic vectors satisfy the Ingleton inequality.

Theorem 42. [12] If an entropic vector $h \in \mathcal{H}_n$ is abelian representable, then for any non-empty subsets $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ and $\mathcal{A}_4$ of $\mathcal{N}$, h satisfies the Ingleton inequality.

Since

$$\Upsilon_n^{ab} \subset \Upsilon_n \subset \Gamma_n^* \subset \bar{\Gamma}_n^*$$

and

$$\overline{\text{con}}(\Upsilon_n) = \bar{\Gamma}_n^*,$$

by taking the convex closure, we can see that $\overline{\text{con}}(\Upsilon_n^{ab}) \subseteq \bar{\Gamma}_n^*$, an inner bound.

The next result says that $\overline{\text{con}}(\Upsilon_n^{ab})$ is a non-trivial inner bound for $\bar{\Gamma}_n^*$ for $n \geq 4$.

Theorem 43. [12] For $n \leq 3$, $\overline{\text{con}}(\Upsilon_n^{ab}) = \overline{\text{con}}(\Upsilon_n) = \bar{\Gamma}_n^*$ and for $n \geq 4$, $\overline{\text{con}}(\Upsilon_n^{ab}) \neq \bar{\Gamma}_n^*$.

Thus entropic vectors which are abelian group representable are a proper subset of entropic vectors coming from quasi-uniform random variables using Equations (4.2) and (4.3). This addresses a natural question of classifying groups with respect to the entropic vectors which they induce. In particular, we want to understand which groups belong to the same class as abelian groups with respect to this classification. We make this precise in the definitions below.

Definition 33. Let G be a group and let $G_1, \dots, G_n$ be fixed subgroups of G . Suppose there exists an abelian group A with subgroups $A_1, \dots, A_n$ such that for every non-empty $\mathcal{A} \subseteq \mathcal{N}$, $[G : G_{\mathcal{A}}] = [A : A_{\mathcal{A}}]$, where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$, $A_{\mathcal{A}} = \cap_{i \in \mathcal{A}} A_i$. Then we say that $(A, A_1, \dots, A_n)$ *represents* $(G, G_1, \dots, G_n)$.

Definition 34. If for every choice of subgroups $G_1, \dots, G_n$ of G , there exists an abelian group A such that $(A, A_1, \dots, A_n)$ represents $(G, G_1, \dots, G_n)$, we say that G is *abelian (group) representable for n* .

Note that the abelian group A may vary for different choices of subgroups $G_1, \dots, G_n$. However, it may possible to find an abelian group that works for all choices $G_1, \dots, G_n$.

Definition 35. Suppose there exists an abelian group A such that for every choice of subgroups $G_1, \dots, G_n \leq G$, there exist subgroups $A_1, \dots, A_n \leq A$ such that $(A, A_1, \dots, A_n)$ represents $(G, G_1, \dots, G_n)$. Then we will say that G is *uniformly abelian (group) representable for n* . (Alternatively, A *uniformly represents G* .)

When G is abelian group representable, the quasi-uniform random variables $X_1, \dots, X_n$ corresponding to subgroups G_i can also be obtained using an abelian group A and its subgroups $A_1, \dots, A_n$. If we choose the subgroup $G_i = G$, then $\log[G : G] = 0$, that is $H(X_i) = 0$, which implies X_i is actually taking values deterministically. If $G_i = \{1\}$, then $\log[G : \{1\}] = \log |G|$. Thus the entropy chain rule yields

$$H(X_i, X_{\mathcal{A}}) = H(X_i) + H(X_{\mathcal{A}} | X_i)$$

for every $\mathcal{A}$ such that $i \notin \mathcal{A}$. Since $H(X_i) = \log |G|$ and $H(X_i, X_{\mathcal{A}}) = \log[G : \{1\} \cap G_{\mathcal{A}}] = \log |G|$, we conclude that

$$H(X_{\mathcal{A}}|X_i) = 0.$$

That is, given X_i , all the $n - 1$ other random variables are functions of X_i and we will consequently require that each subgroup $G_1, \dots, G_n$ is non-trivial and proper. Hence n is at most the number of non-trivial proper subgroups of G .

The contributions of this chapter are to introduce the notion of abelian group representability for an arbitrary finite group G , and to characterize the abelian group representability of several classes of groups (the definitions of these groups will be recalled in the respective sections):

- Dihedral, quasi-dihedral and dicyclic groups are shown to be abelian group representable for every $n \geq 2$ if (Section 5.2) and only if (Section 5.4) they are 2-groups. When they are abelian group representable, they are furthermore uniformly abelian group representable for every n (Section 5.2).
- p -groups: p -groups are shown to be uniformly abelian group representable for $n = 2$ in Section 5.3.
- Nilpotent groups: in Section 5.4 we show that representability of nilpotent groups is completely determined by representability of p -groups. The set of nilpotent groups is shown to contain the set of abelian representable groups for any n in Section 5.4, the two coincide for $n = 2$.

5.2 Abelian Group Representability of Classes of 2-Groups

In this section we establish uniform abelian group representability of dihedral, quasi-dihedral and dicyclic 2-groups for any n . We begin with a general lemma showing how abelian group representability of a group H may imply abelian group representability of a group G .

Lemma 44. Let

$$\psi : G \rightarrow H$$

be a bijective map, which is additionally *subgroup preserving*, i.e., for any subgroup $G_i \leq G$, the set $\psi(G_i)$ is a subgroup of H . Suppose that H is abelian (resp. uniformly abelian) group representable. Then so is G . In particular, if H itself is abelian, then G is abelian group representable.

Proof. We want to show that given subgroups $G_1, \dots, G_n \leq G$, there exists an abelian group A with subgroups $A_1, \dots, A_n$ such that for any subset $\mathcal{A} \subseteq \{1, \dots, n\}$ the intersection subgroup $G_{\mathcal{A}}$ has index $[G : G_{\mathcal{A}}] = [A : A_{\mathcal{A}}]$.

Since H is abelian group representable, $(H, \psi(G_1), \dots, \psi(G_n))$ can be represented by some $(A, A_1, \dots, A_n)$. We claim that $(A, A_1, \dots, A_n)$ represents $(G, G_1, \dots, G_n)$.

Since ψ is bijective, for any $G_i \leq G$, the subgroup $\psi(G_i)$ has the same size and index in H as G_i has in G . In particular, $[A : A_i] = [H : \psi(G_i)] = [G : G_i]$. This takes care of 1-intersection, i.e., when $|\mathcal{A}| = 1$.

Now we want to show that in fact $[G : G_{\mathcal{A}}] = [A : A_{\mathcal{A}}]$ for any $\mathcal{A} \subseteq \mathcal{N}$. When considering intersections $G_{\mathcal{A}}$, let us first consider 2-intersections $G_{12} = G_1 \cap G_2 \leq G$. First observe that $\psi(G_1 \cap G_2) = \psi(G_1) \cap \psi(G_2)$. The containment $\psi(G_1 \cap G_2) \subseteq \psi(G_1) \cap \psi(G_2)$ is immediate. To see the containment $\psi(G_1 \cap G_2) \supseteq \psi(G_1) \cap \psi(G_2)$ observe that if $y = \psi(g_1) = \psi(g_2)$ for some $g_1 \in G_1, g_2 \in G_2$ then bijectivity of ψ implies that $g_1 = g_2$ and $y \in \psi(G_1 \cap G_2)$.

Now, recalling that $|G| = |H|$, we see that

$$[G : G_1 \cap G_2] = [H : \psi(G_1) \cap \psi(G_2)] = [A : A_1 \cap A_2].$$

More generally for arbitrary intersection $G_{\mathcal{A}}$, we have $[G : G_{\mathcal{A}}] = [A : A_{\mathcal{A}}]$: this follows by induction on the number of subgroups involved in the intersection. We conclude that $(A, A_i)_{i \in \mathcal{N}}$ represents $(G, G_i)_{i \in \mathcal{N}}$. If H was uniformly abelian group representable, then A was chosen independently of subgroups $\psi(G_1), \dots, \psi(G_n)$ and it follows that G is also uniformly abelian group representable.

□

5.2.1 Dihedral and Quasi-Dihedral 2-Groups

Before establishing the abelian representability for dihedral and quasi-dihedral 2-groups, we recall the definition of a finite p -group.

Definition 36. Let p be a prime. Then a group G is said to be a p -group if the order of each element of G is a power of p .

If there exists a prime $q \neq p$ such that $q \mid |G|$, then by Cauchy's theorem [15], there exists an element $g \in G$ whose order is q , a contradiction to the definition of a p -group. Hence the order of G is also a power of p .

When $p = 2$, G becomes a 2-group.

We define the dihedral group D_{2m} for $m \geq 3$ to be the symmetry group of the regular m -sided polygon. The group D_{2m} is of order $2m$, with a well known description in terms of generators and relations:

$$D_{2m} = \langle r, s \mid r^m = s^2 = 1, rs = sr^{-1} \rangle.$$

Each element of D_{2m} is uniquely represented as $r^a s^j$ where $1 \leq a \leq m, j = 0, 1$.

Note that the generator s acts on r by conjugation sending r to an element $sr s^{-1} = r^{-1}$. More generally, consider other possibilities for $sr s^{-1} = r^z$. When we apply this map twice, we see that since s has order 2, it must be that $z^2 \equiv 1 \pmod{m}$.

In case $m = 2^k$, there are 4 such choices modulo 2^k for $z \in \mathbb{Z}_{2^k}^\times$, i.e., $z = \pm 1, 2^{k-1} \pm 1$.

All elements in the multiplicative group $\mathbb{Z}_{2^k}^\times$ are odd integers modulo 2^k . That is, they are of the form $z = 2n \pm 1$ for $n = 1, \dots, 2^{k-1}$. Then $z^2 = 4n(n \pm 1) + 1$. In order this to be a multiple of 2^k , n has to be 2^{k-2} or 2^{k-1} which leaves only 4 choices for z which are ± 1 and $2^{k-1} \pm 1$.

The choice $z = 1$ will result in an abelian group, $z = -1$ gives the dihedral group $D_{2^{k+1}}$ above. Now we cover the remaining two choices. In either case, the subgroup structure is similar to that of a dihedral group, which will eventually allow us to conclude that these groups are abelian representable via Lemma 44.

Define two quasi-dihedral groups, each of order 2^{k+1} :

$$\mathcal{QD}_{2^{k+1}}^{-1} = \langle r, s \mid r^{2^k} = s^2 = 1, rs = sr^{2^{k-1}-1} \rangle,$$

$$\mathcal{QD}_{2^{k+1}}^{+1} = \langle r, s \mid r^{2^k} = s^2 = 1, rs = sr^{2^{k-1}+1} \rangle.$$

Let us make some brief observations about the structure of subgroups of

$$D_{2^{k+1}}, \mathcal{QD}_{2^{k+1}}^{-1}, \mathcal{QD}_{2^{k+1}}^{+1}.$$

First of all, note that an element is of one of the forms $r^j, r^j s$ for some j .

Proposition 45. Any subgroup of the above groups can be expressed in terms of at most two generators of the form $r^{2^i}, r^j s$.

Proof. A subgroup generated by $\langle r^{j_1} s, r^{j_2} s \rangle$ can also be generated by $\langle r^{j_1} s, r^{j_2} s (r^{j_1} s)^{-1} \rangle = \langle r^{j_1} s, r^{j_2-j_1} \rangle$ which in turn can be expressed as $\langle r^{2^i}, r^{j_1} s \rangle$ for some i since $j_2 - j_1$ is an integer mod 2^k and $r^{j_2-j_1}$ generates r^{2^i} for some $i = 0, 1, \dots, k$. Thus we conclude that only one generator of the form $r^j s$ is required. Only one generator of the form r^j is required as

well, since the cyclic subgroup $\langle r \rangle$ contains only cyclic subgroups $\langle r^{2^i} \rangle$. Hence any subgroup can be expressed in terms of at most two generators of the form $r^{2^i}, r^j s$. $\square$

The following Proposition characterizes the subgroup types of dihedral and quasi-dihedral groups.

Proposition 46. Subgroups of $D_{2^{k+1}}, \mathcal{QD}_{2^{k+1}}^{-1}, \mathcal{QD}_{2^{k+1}}^{+1}$ can be only of the following types:

1. $\langle r^{2^i} \rangle, 0 \leq i < k,$
2. $\langle r^a s \rangle, 0 \leq a < 2^k,$
3. $\langle r^{2^i}, r^c s \rangle, 0 \leq i < k, 0 \leq c < 2^i.$

Proof. From Proposition 45, all subgroups are generated by at most two elements $r^{2^i}, r^a s$. So the first and second category of subgroups are generated by elements of the form r^{2^i} and $r^a s$ respectively. When a subgroup is given by $\langle r^{2^i}, r^c s \rangle$, without loss of generality we can further assume that $c < 2^i$: this is achieved by premultiplying the second generator $r^c s$ repeatedly with r^{-2^i} . $\square$

5.2.1.1 Dihedral 2-Groups

We now establish abelian group representability of dihedral 2-groups for all n by applying the lemma above. We start by showing that subgroups of $D_{2^{k+1}}$ map to subgroups of an abelian group.

Proposition 47. Let $A = \bigoplus_{i=0}^k \mathbb{Z}_2$ be generated by the standard basis $\{e_0, \dots, e_k\}$ over integers $\mathbb{Z}_2$ modulo 2. There exists a subgroup preserving bijection from the dihedral group $D_{2^{k+1}}$ to A .

Proof. For each element $r^a s^j$ we use the base 2 representation of exponent $a = \sum_{i=0}^{k-1} a_i 2^i$ to define the map $\psi : D_{2^{k+1}} \rightarrow A$:

$$\psi : r^a s^j \mapsto \sum_{i=0}^{k-1} a_i e_i + j e_k.$$

We show that the map ψ is a subgroup preserving bijection.

First note that since each element is uniquely represented as $r^a s^j$ for some a, j , the map ψ is indeed well-defined and clearly bijective. Hence we only need to verify that ψ is subgroup preserving. In other words, if $H \leq D_{2^{k+1}}$ is a subgroup, then the image $\psi(H)$ is a subgroup

of A . Since each element in A is its own inverse, and $\psi(H)$ contains the identity, we only need to check the closure property of $\psi(H)$ in order to prove that it is a subgroup.

To that end, we investigate the subgroups of $D_{2^{k+1}}$. All subgroups of $D_{2^{k+1}}$ are of the form (see Proposition 46)

1. $\langle r^{2^i} \rangle = \{r^a : a \equiv 0 \pmod{2^i}\}, 0 \leq i < k,$
2. $\langle r^a s \rangle = \{1, r^a s\}, 0 \leq a < 2^k,$
3. $\langle r^{2^i}, r^c s \rangle = \{r^a, r^{a+c} s : a \equiv 0 \pmod{2^i}\} = \{r^a, r^b s : a \equiv 0 \pmod{2^i}, b \equiv c \pmod{2^i}\},$
 $0 \leq i < k, 0 \leq c < 2^k.$

Case 1: $H = \langle r^{2^i} \rangle = \{r^a : a \equiv 0 \pmod{2^i}\}$. Note that $a < 2^k$ is a multiple of 2^i if and only if $a = \sum_{j=i}^{k-1} a_j 2^j$, i.e., the terms $a_0, \dots, a_{i-1}$ of the binary expansion of a are 0. In other words, the image

$$\psi(H) = \left\{ \sum_{j=i}^{k-1} a_j e_j \right\}.$$

Clearly this set is closed under addition.

Case 2: $H = \langle r^a s \rangle = \{1, r^a s\}$. Then the image

$$\psi(H) = \left\{ \sum_{i=0}^{k-1} a_i e_i + e_k, 0 \right\}.$$

Clearly this set of size 2 is a subgroup of A .

Case 3: $H = \langle r^{2^i}, r^c s \rangle, 0 \leq c < 2^i$.

Indeed $H = \{r^{2^i h}, r^{c+2^i h} s : h \in \mathbb{Z}\}$.

We verify directly that $\psi(H)$ is closed under addition by showing that

$$x, y \in H \implies \psi(x) + \psi(y) \in H.$$

This involves considering 3 cases:

- $x, y \in \langle r^{2^i} \rangle$
- $x \in \langle r^{2^i} \rangle, y = r^{c+2^i h} s$
- both x, y are of the form $r^{c+2^i h} s$.

1. $x, y \in \langle r^{2^i} \rangle$ implies that $\psi(x) + \psi(y) \in \psi(H)$. This follows identically to **Case 1**.
2. $x \in \langle r^{2^i} \rangle, y = r^{c+2^i h'} s$ implies that $\psi(x) + \psi(y) \in \psi(H)$.

We show that $\psi(r^{2^i h}) + \psi(r^{2^i h' + c} s) = \psi(r^{2^i h'' + c} s) \in \psi(H)$. To see this, observe using the binary expansion of exponents, that when $c < 2^i$ we can actually factor $\psi(r^{2^i h' + c}) = \psi(r^{2^i h'}) + \psi(r^c)$. Hence

$$\begin{aligned} \psi(r^{2^i h}) + \psi(r^{2^i h' + c} s) &= \psi(r^{2^i h}) + \psi(r^{2^i h'}) + \psi(r^c) + \psi(s) \\ &= \psi(r^{2^i h''}) + \psi(r^c) + \psi(s) \\ &= \psi(r^{2^i h'' + c} s) \in \psi(H). \end{aligned}$$

3. Both x, y of the form $r^{c+2^i h} s$ implies that $\psi(x) + \psi(y) \in \psi(H)$. Now

$$\begin{aligned} \psi(r^{2^i h + c} s) + \psi(r^{2^i h' + c} s) &= \psi(r^{2^i h + c}) + \psi(s) + \psi(r^{2^i h' + c}) + \psi(s) \\ &= \psi(r^{2^i h}) + \psi(r^c) + \psi(r^{2^i h'}) + \psi(r^c) \\ &= \psi(r^{2^i h''}) \in \psi(H). \end{aligned}$$

□

Proposition 48. The dihedral group $D_{2^{k+1}}$ is uniformly abelian group representable for all n .

Proof. This follows from Proposition 47 and Lemma 44. The uniform part comes from the fact that A only depends on the size of $D_{2^{k+1}}$ and not on the choice of subgroups of $D_{2^{k+1}}$.

□

We next use abelian representability of dihedral groups to derive abelian representability of quasi-dihedral groups. The idea is that these two classes of groups have the same subgroups.

5.2.1.2 Quasi-dihedral 2-Groups

We already know by Remark 46 what subgroups of quasi-dihedral groups look like in terms of generators. However, for the purpose of defining a subgroup preserving map ψ (which will not be a homomorphism!), we want to know exactly what elements these subgroups consist of. The next lemma describes what the subgroups of quasi-dihedral groups look like.

Lemma 49. Let $G = \mathcal{QD}_{2^{k+1}}^{-1}$ or $\mathcal{QD}_{2^{k+1}}^{+1}$ be a quasi-dihedral group of size 2^{k+1} . Then all the subgroups of G are of the form

1. $\langle r^{2^i} \rangle = \{r^a : a \equiv 0 \pmod{2^i}, 0 \leq i \leq k-1\}$,
2. $\langle r^{2^i}, r^j s \rangle = \{r^a, r^b s : a \equiv 0 \pmod{2^i}, b \equiv j \pmod{2^i}, 0 \leq i \leq k-1, 0 \leq j \leq 2^i - 1\}$,
3.
 - When $G = \mathcal{QD}_{2^{k+1}}^{-1}$, then $\langle r^j s \rangle = \begin{cases} \{1, r^j s\} & j \equiv 0 \pmod{2} \\ \{1, r^j s, r^{2^{k-1}}, r^{2^{k-1}+j} s\} & j \equiv 1 \pmod{2} \end{cases}$
 - When $G = \mathcal{QD}_{2^{k+1}}^{+1}$, then $\langle r^j s \rangle = \begin{cases} \langle r^{2j}, r^j s \rangle \text{ is of type (2)} & j \neq 0 \\ \{1, s\} & j = 0. \end{cases}$

Proof. First assume that $G = \mathcal{QD}_{2^{k+1}}^{-1}$.

1. $\langle r^{2^i} \rangle = \{r^a : a \equiv 0 \pmod{2^i}\}$

This case is obvious.

2. $\langle r^{2^i}, r^j s \rangle = \{r^a, r^b s : a \equiv 0 \pmod{2^i}, b \equiv j \pmod{2^i}; 0 \leq i \leq k-1, 0 \leq j \leq 2^i - 1\}$.

For this we observe that $r^j s r^{2^i} = r^{j+2^i(2^{k-1}-1)} s$.

Additionally note that

$$r^j s r^j s = r^{j+j(2^{k-1}-1)} s^2 = r^{j(2^{k-1})} = \begin{cases} 1 & j \equiv 0 \pmod{2} \\ r^{2^{k-1}} & j \equiv 1 \pmod{2} \end{cases}$$

3. $\langle r^j s \rangle = \begin{cases} \{1, r^j s\} & j \equiv 0 \pmod{2} \\ \{1, r^j s, r^{2^{k-1}}, r^{2^{k-1}+j} s\} & j \equiv 1 \pmod{2} \end{cases}$

In particular, when j is odd, the subgroup $\langle r^j s \rangle$ can be expressed in form (2) as $\langle r^{2^{k-1}}, r^j s \rangle$.

Now let $G = \mathcal{QD}_{2^{k+1}}^{+1}$. The cases (1), (2) follow identically to $\mathcal{QD}_{2^{k+1}}^{-1}$.

For case (3), the case $j = 0$ is obvious, for $j > 0$, observe

$$r^j s r^j s = r^j r^{j(2^{k-1}+1)} = r^{j(2^{k-1}+2)} = r^{2j(2^{k-2}+1)}.$$

But $\langle r^{2j(2^{k-2}+1)} \rangle = \langle r^{2j} \rangle$ and hence we have

$$\langle r^j s \rangle = \langle r^{2j}, r^j s \rangle.$$

Setting $2^i || 2j$ be the highest power of 2 dividing j , this subgroup is in fact of type (2) and consists of elements $\{r^a, r^{j'} s : a \equiv 0 \pmod{2^i}, j' \equiv j \pmod{2^i}\}$. $\square$

Proposition 50. The quasi dihedral group $\mathcal{QD}_{2^{k+1}}^{-1}$ and $\mathcal{QD}_{2^{k+1}}^{+1}$ are uniformly abelian group representable for all n .

Proof. To prove the result we construct a subgroup-preserving bijection from quasi-dihedral to dihedral groups. By Lemma 44 and the fact that dihedral groups are uniformly abelian group representable, the result follows.

Let $G = \mathcal{QD}_{2^{k+1}}^{-1}$ or $\mathcal{QD}_{2^{k+1}}^{+1}$.

Elements of G can be uniquely written as

$$\{r^i s^j : 0 \leq i < 2^k, j = 0, 1\}.$$

Also all the elements in $D_{2^{k+1}}$ can be uniquely written as

$$\{r^i s^j : 0 \leq i < 2^k, j = 0, 1\},$$

keeping in mind that $r, s \in D_{2^{k+1}}$ are not the same as $r, s \in G$, as the group law for the groups G and $D_{2^{k+1}}$ is not the same.

Define a subgroup preserving bijection

$$\psi : G \rightarrow D_{2^{k+1}}$$

$$\psi : r^i s^j \mapsto r^i s^j.$$

The map ψ is well-defined and clearly bijective. It remains to show that ψ is subgroup preserving.

Now Lemma 49 describes which elements subgroups of G consist of. The proof of Proposition 47 gives a similar description for subgroups of $D_{2^{k+1}}$. Verifying that the two coincide via ψ gives the result. As an example, consider a subgroup of form (2). We have

$$\begin{aligned} \psi(\langle r^{2^i}, r^j s \rangle) &= \psi(\{r^a, r^b s : a \equiv 0 \pmod{2^i}, b \equiv j \pmod{2^i}\}) = \\ &= \{r^a, r^b s : a \equiv 0 \pmod{2^i}, b \equiv j \pmod{2^i}\} \leq D_{2^{k+1}}. \end{aligned}$$

Cases (1) and (3) follow immediately as well. Hence ψ is a subgroup preserving bijection and since $D_{2^{k+1}}$ is abelian group representable by Proposition 48, Lemma 44 implies that $\mathcal{QD}_{2^{k+1}}^{-1}$ and $\mathcal{QD}_{2^{k+1}}^{+1}$ are uniformly abelian group representable as well.

□

5.2.2 Dicyclic 2-Groups

Next we consider the case of dicyclic groups, another well studied class of non-abelian groups. The results are similar to that of dihedral groups. A dicyclic group DiC_m of order $4m$ is generated by two elements a, x as follows:

$$DiC_m = \langle a, x : a^{2m} = 1, x^2 = a^m, xa = a^{-1}x \rangle.$$

Every element of DiC_m can be uniquely presented as $a^i x^j$, where $0 \leq i < 2m, j = 0, 1$.

A *generalized quaternion group* is a dicyclic group with m a power of 2. We now study the subgroups of DiC_{2^k} . From the definition, we know that $|DiC_{2^k}| = 2^{k+2}$. All the elements of DiC_{2^k} are of the form $a^i, a^i x, 1 \leq i \leq 2^{k+1}$ where $x^2 = a^{2^k}$ and $x^3 = a^{2^k} x$.

As is the case with dihedral groups, any subgroup $\langle a^{j_1} x, a^{j_2} x \rangle = \langle a^{j_1 - j_2}, a^{j_2} x \rangle$, which in turn can be represented as $\langle a^{2^i}, a^{j_2} x \rangle$, since $\langle a^{j_1 - j_2} \rangle = \langle a^{2^i} \rangle$ for some i . Hence we need at most one generator of the form $a^j x$. Trivially, since $\langle a \rangle$ is cyclic, we only need one generator of the form a^{2^i} as well.

Now consider the subgroup $\langle a^j x \rangle$. Its elements are $\{a^j x, x^2 = a^{2^k}, a^{2^k+j} x, 1\}$, so it can be written in the form $\langle a^{2^k}, a^j x \rangle$.

We conclude that subgroups of DiC_{2^k} are of types

1. $\langle a^{2^i} \rangle, 0 \leq i \leq k$
2. $\langle a^{2^i}, a^j x \rangle, 0 \leq i \leq k, 0 \leq j \leq 2^i - 1$.

Proposition 51. The dicyclic group $DiC_{2^{k-1}}$ is uniformly abelian group representable for all n .

Proof. This proof is an application of Lemma 44, which allows us to use the fact that $D_{2^{k+1}}$ was already shown to be abelian group representable, to conclude that so is $DiC_{2^{k-1}}$.

To apply Lemma 44, we must define a subgroup preserving bijection

$$\psi : DiC_{2^{k-1}} \rightarrow D_{2^{k+1}}.$$

Since the elements in $DiC_{2^{k-1}}$ can be uniquely written as

$$\{a^i x^j : 0 \leq i < 2^k, j = 0, 1\}$$

while all the elements in $D_{2^{k+1}}$ can be uniquely written as

$$\{r^i s^j : 0 \leq i < 2^k, j = 0, 1\},$$

the map

$$\psi : a^i x^j \mapsto r^i s^j$$

is well-defined and clearly bijective. It remains to show that ψ is subgroup preserving. But from the discussion above, we see that the subgroups of $DiC_{2^{k-1}}$ are

1. $\langle a^{2^i} \rangle$, $0 \leq i < k$,
2. $\langle a^{2^i}, a^j x \rangle$, $0 \leq i < k$, $0 \leq j \leq 2^i - 1$.

The images of both kinds of subgroups under ψ do indeed form subgroups of $D_{2^{k+1}}$:

1. $\psi(\langle a^{2^i} \rangle) = \psi(\{a^{2^i h} : h \in \mathbb{Z}\}) = \{r^{2^i h} : h \in \mathbb{Z}\}$,
2. $\psi(\langle a^{2^i}, a^j x \rangle) = \psi(\{a^{2^i h}, a^{2^i h+j} x : h \in \mathbb{Z}\}) = \{r^{2^i h}, r^{2^i h+j} s : h \in \mathbb{Z}\}$.

Hence ψ is a subgroup preserving bijection and since by Proposition 48 the dihedral group $D_{2^{k+1}}$ is abelian group representable, Lemma 44 implies that $DiC_{2^{k-1}}$ is uniformly abelian group representable as well.

□

Propositions 48 and 51 generalize Proposition 3 of [38], which showed that the two non-abelian groups of order 8 are abelian group representable.

The results we have obtained on representability of 2-groups rely on the use of a subgroup preserving bijection ψ and Lemma 44. The map ψ defined for dihedral groups in Proposition 47 does not generalize to p -groups, as we can no longer trivially claim closure under inverses. We employ different methods to deal with abelian representability of odd order p -groups in the following section.

5.3 Abelian Group Representability of p -Groups

We start with a simple lemma which establishes a necessary condition for abelian representability. In the next section this lemma is used to exclude the class of all non-nilpotent groups.

Lemma 52. Let G be a group which is abelian group representable. Then for any subgroups G_1, G_2 with intersection G_{12} and corresponding indices i_1, i_2, i_{12} in G , it must be the case that

$$i_{12} \mid i_1 i_2.$$

Proof. Let (A, A_1, A_2) represent (G, G_1, G_2) . Since A is abelian, the subgroups A_1, A_2, A_{12} are all normal in A and the quotient group A/A_{12} is abelian of order i_{12} .

Next note that the subgroups A_1/A_{12} and A_2/A_{12} of A/A_{12} are disjoint and normal. Hence A/A_{12} contains the subgroup $(A_1/A_{12})(A_2/A_{12})$ whose order divides the order of A/A_{12} :

$$|A_1/A_{12}| |A_2/A_{12}| \mid i_{12}.$$

Observing that $|A_1/A_{12}| = \frac{|A/A_{12}|}{|A/A_1|}$ we obtain

$$\frac{|A/A_{12}|}{|A/A_1|} \frac{|A/A_{12}|}{|A/A_2|} = \frac{i_{12}}{i_1} \frac{i_{12}}{i_2} \mid i_{12}.$$

Equivalently, $i_{12} \mid i_1 i_2$. □

The following proposition proves abelian representability of p -groups for $n = 2$ by establishing a sufficient condition for $n = 2$ and showing that all p -groups in fact satisfy it.

Proposition 53. Let G be a p -group. Then G is uniformly abelian group representable for $n = 2$.

Proof. Let p^m be the order of G . Consider some subgroups $G_1, G_2, G_{12} = G_1 \cap G_2$ of orders p^i, p^j, p^k respectively. We show that the exponents i, j, k, m obey an inequality, which is sufficient to guarantee abelian representability of (G, G_1, G_2) .

Claim 1. Inequality $i + j - k \leq m$ holds for any p -group G .

To that end, consider the subset $G_1 G_2 = \{g_1 g_2 \mid g_1 \in G_1, g_2 \in G_2\}$. Note that $G_1 G_2$ is only a subgroup when one of G_1, G_2 is normal in G . Counting the number of elements in $G_1 G_2$, we have

$$|G_1 G_2| = \frac{|G_1| |G_2|}{|G_{12}|} = \frac{p^i p^j}{p^k} = p^{i+j-k}.$$

Now since $G_1 G_2 \subseteq G$, we conclude that $i + j - k \leq m$.

Claim 2. Sufficiency of condition $i + j - k \leq m$.

We show that $i + j - k \leq m$ implies that (G, G_1, G_2) can be represented by some abelian (A, A_1, A_2) . Define A to be the elementary abelian p -group $A = C_p^m$.

We can express A as the following direct product:

$$C_p^k \times C_p^{i-k} \times C_p^{j-k} \times C_p^{m-(i+j-k)}.$$

Note that we needed the inequality $i + j - k \leq m$ in order for the exponent $m - (i + j - k)$ to be nonnegative. Now we define subgroups

$$A_1 = C_p^k \times C_p^{i-k} \times \{1\} \times \{1\},$$

$$A_2 = C_p^k \times \{1\} \times C_p^{j-k} \times \{1\}.$$

As the orders of A, A_1, A_2, A_{12} are same as those of G, G_1, G_2, G_{12} , clearly (A, A_1, A_2) represents (G, G_1, G_2) .

Since the choice of G_1, G_2 was arbitrary, we conclude that G is abelian group representable for $n = 2$. Moreover, G is uniformly abelian group representable, since A was chosen independently of G_1, G_2 . $\square$

Claim 2 of the above proof establishes a sufficient condition for abelian group representability for $n = 2$, namely the exponents i, j, k, m of orders of G_1, G_2, G_{12}, G obeying the inequality $i + j - k \leq m$. Lemma 52 on the other hand implies that this condition is also necessary: If (G, G_1, G_2) is representable, then $i_{12} \mid i_1 i_2$. But

$$\begin{aligned} i_{12} \mid i_1 i_2 &\iff \frac{p^m}{p^k} \mid \frac{p^m}{p^i} \frac{p^m}{p^j} \iff \frac{p^i p^j}{p^k} \mid p^m \\ &\iff i + j - k \leq m. \end{aligned}$$

Hence we have a numerical necessary and sufficient condition for abelian group representability of (G, G_1, G_2) in terms of exponents i, j, k, m of orders of subgroups G_1, G_2, G_{12} , and G :

$$i + j - k \leq m.$$

Certainly we can use similar methods as in the proof of Claim 2, Proposition 53 to guarantee abelian representability whenever a similar "inclusion-exclusion" inequality holds for higher n . It will no longer, however, be a necessary condition.

More specifically for $G_1, G_2, G_3 \leq G$ of orders $|G_{\mathcal{A}}| = p^{j_{\mathcal{A}}} : \mathcal{A} \subseteq \{1, 2, 3\}$, $|G| = p^m$, the similar inequality on exponents of orders

$$j_1 + j_2 + j_3 - (j_{12} + j_{13} + j_{23}) + j_{123} \leq m$$

while guaranteeing representability of (G, G_1, G_2, G_3) , need no longer hold for groups which are abelian representable. For example, we take the group of quaternions Q_8 with subgroups $\langle i \rangle, \langle j \rangle, \langle k \rangle$, the exponents of whose orders violate the inequality for $n = 3$: $2 + 2 + 2 - (1 + 1 + 1) + 1 = 4 \not\leq 3$.

The following question arises:

Remark 7. Can we establish a necessary and sufficient numerical condition for abelian group representability for $n > 2$.

We finish the section by giving a class of uniformly abelian representable p -groups for $n = 3$.

Proposition 54. ¹ If p is a prime, a non-abelian group G of order p^3 is abelian group representable for $n = 3$.

Proof. We show that the group $C_{p^2} \times C_p$ uniformly represents G .

Since G is a non-trivial p -group, it has a non-trivial center, $Z(G)$.

If $|Z(G)| \geq p^2$, $G/Z(G)$ is cyclic and G would be abelian, which is a contradiction. Thus we have $|Z(G)| = p$.

If G_1, G_2 and G_3 are three subgroups of G of order p^2 , then they are normal in G having non-trivial intersection with $Z(G)$, implies that $Z(G) \subseteq G_1, G_2, G_3$ (Theorem 1, Chapter 6, [15]). So the pairwise intersection of all subgroups of G of index p is $Z(G)$. Then the possible combinations of indices of $G_1, G_2, G_3, G_{12}, G_{13}, G_{23}, G_{123}$ are

1. $p, p, p, p^2, p^2, p^2, p^2$
2. $p, p, p^2, p^2, p^2, p^2, p^2$
3. $p, p, p^2, p^2, p^2, p^3, p^3$
4. $p, p, p^2, p^2, p^3, p^3, p^3$
5. $p, p^2, p^2, p^2, p^2, p^3, p^3$
6. $p, p^2, p^2, p^2, p^3, p^3, p^3$

¹It was conjectured by a reviewer that that the statement holds more generally for any p -group when $n = 3$.

7. $p, p^2, p^2, p^3, p^3, p^3, p^3$
8. $p^2, p^2, p^2, p^3, p^3, p^3, p^3$

The corresponding abelian group representation is given uniformly by the group $A = C_{p^2} \times C_p$, where $C_{p^2} = \langle g | g^{p^2} = 1 \rangle$, $C_p = \langle r | r^p = 1 \rangle$. For each combination of the indices above, let the subgroups G_1, G_2, G_3 be represented by A_1, A_2, A_3 respectively:

1. $A_1 = \langle g^p \rangle \times \langle r \rangle$, $A_2 = \langle g \rangle \times 1$, $A_3 = \langle (g, r) \rangle$, $A_{12} = \langle g^p \rangle \times 1$, $A_{13} = \langle g^p \rangle \times 1$, $A_{23} = \langle g^p \rangle \times 1$, $A_{123} = \langle g^p \rangle \times 1$
2. $A_1 = \langle g^p \rangle \times \langle r \rangle$, $A_2 = \langle g \rangle \times 1$, $A_3 = \langle g^p \rangle \times 1$, $A_{12} = \langle g^p \rangle \times 1$, $A_{13} = \langle g^p \rangle \times 1$, $A_{23} = \langle g^p \rangle \times 1$, $A_{123} = \langle g^p \rangle \times 1$
3. $A_1 = \langle g^p \rangle \times \langle r \rangle$, $A_2 = \langle g \rangle \times 1$, $A_3 = 1 \times \langle r \rangle$, $A_{12} = \langle g^p \rangle \times 1$, $A_{13} = 1 \times \langle r \rangle$, $A_{23} = 1 \times 1$, $A_{123} = 1 \times 1$
4. $A_1 = \langle (g, r) \rangle$, $A_2 = \langle g \rangle \times 1$, $A_3 = 1 \times \langle r \rangle$, $A_{12} = \langle g^p \rangle \times 1$, $A_{13} = 1 \times 1$, $A_{23} = 1 \times 1$, $A_{123} = 1 \times 1$
5. $A_1 = \langle g^p \rangle \times \langle r \rangle$, $A_2 = \langle g^p \rangle \times 1$, $A_3 = 1 \times \langle r \rangle$, $A_{12} = \langle g^p \rangle \times 1$, $A_{13} = 1 \times \langle r \rangle$, $A_{23} = 1 \times 1$, $A_{123} = 1 \times 1$
6. $A_1 = \langle g \rangle \times 1$, $A_2 = \langle g^p \rangle \times 1$, $A_3 = \langle (g^p, r) \rangle$, $A_{12} = \langle g^p \rangle \times 1$, $A_{13} = 1 \times 1$, $A_{23} = 1 \times 1$, $A_{123} = 1 \times 1$
7. $A_1 = \langle (g, r) \rangle$, $A_2 = \langle (g^p, r) \rangle$, $A_3 = 1 \times \langle r \rangle$, $A_{12} = 1 \times 1$, $A_{13} = 1 \times 1$, $A_{23} = 1 \times 1$, $A_{123} = 1 \times 1$
8. $A_1 = \langle g^p \rangle \times 1$, $A_2 = 1 \times \langle r \rangle$, $A_3 = \langle (g^p, r) \rangle$, $A_{12} = 1 \times 1$, $A_{13} = 1 \times 1$, $A_{23} = 1 \times 1$, $A_{123} = 1 \times 1$.

We obtain that A, A_1, A_2, A_3 represents G, G_1, G_2, G_3 and since the group A is independent of subgroups G_1, G_2, G_3 , the representation of G is uniform.

□

5.4 Abelian Group Representability of Nilpotent Groups

Nilpotent groups

The following definition is useful to define nilpotent groups.

Definition 37. If $|G| = p^r m$, where p does not divide m , then a subgroup P of order p^r is called a *Sylow p -subgroup* of G . Thus P is a p -subgroup of G of maximum possible size.

We state the following lemma without proof.

Lemma 55. [15] The subgroup P is a normal Sylow p -subgroup of a group G if and only if P is the unique Sylow p -subgroup of G .

We introduce the notion of a nilpotent group. We skip the general definition, and consider only finite nilpotent groups, for which the following characterization is available.

Proposition 56. The following statements are equivalent:

- G is the direct product of its Sylow subgroups.
- Every Sylow subgroup of G is normal.

Proof. Suppose G is the direct product of its Sylow subgroups. Since the factors of a direct product are normal subgroups, every Sylow subgroup of G is normal.

Assume that every Sylow subgroup of G is normal, then by Lemma 55, we know that every normal Sylow p -subgroup is unique, thus there is a unique Sylow p_i -subgroup P_i for each prime divisor p_i of $|G|$, $i = 1, \dots, k$.

Now by Proposition 62, we have that $|P_1 P_2| = |P_1| |P_2|$ since $P_1 \cap P_2 = \{1\}$, and thus

$$|P_1 \dots P_k| = |P_1| \dots |P_k| = |G|$$

by definition of Sylow subgroups. Since we work with finite groups, we deduce that G is indeed the direct product of its Sylow subgroups, having that $G = P_1 \dots P_k$ and $P_i \cap \prod_{j \neq i} P_j$ is trivial. $\square$

Definition 38. A finite group G which is the product of its Sylow subgroups, or equivalently by Proposition 56, each of its Sylow subgroup is normal is called a *nilpotent group*.

Example 13. [15] A dihedral group D_{2m} is nilpotent if and only if m is a power of 2.

That is, a finite nilpotent group is a direct product of its Sylow p -subgroups. In this section we obtain a complete classification of abelian representable groups for $n = 2$, as well as show that non nilpotent groups are never abelian group representable for general n . We start with an easy proposition which allows us to build new abelian representable groups from existing ones.

Proposition 57. Let G and H be groups of coprime orders. Suppose G, H are abelian group representable, then their direct product $G \times H$ is abelian group representable.

Proof. Since the orders of G, H are coprime, any subgroup $K_i \leq G \times H$ is in fact a direct product $K_i = G_i \times H_i$, where $G_i \leq G$, $H_i \leq H$. To prove abelian group representability for n , consider n subgroups $G_1 \times H_1, \dots, G_n \times H_n$ of $G \times H$.

Since G is abelian group representable, there exists an abelian group A with subgroups $A_1, \dots, A_n$ which represents $(G, G_1, \dots, G_n)$. Similarly some $(B, B_1, \dots, B_n)$ represents $(H, H_1, \dots, H_n)$.

It is easy to see that $(A \times B, A_1 \times B_1, \dots, A_n \times B_n)$ represents $G_1 \times H_1, \dots, G_n \times H_n$. To this end, consider an arbitrary intersection group $G_{\mathcal{A}} \times H_{\mathcal{A}} \leq G \times H$ for $\mathcal{A} \subseteq \{1, \dots, n\}$:

$$\begin{aligned} [G \times H : (G_{\mathcal{A}} \times H_{\mathcal{A}})] &= [G : G_{\mathcal{A}}][H : H_{\mathcal{A}}] = \\ &= [A : A_{\mathcal{A}}][B : B_{\mathcal{A}}] = [A \times B : (A_{\mathcal{A}} \times B_{\mathcal{A}})]. \end{aligned}$$

□

The previous proposition allows us to prove the following theorem.

Theorem 58. All nilpotent groups are abelian group representable for $n = 2$.

Proof. We know that finite nilpotent groups are direct products of their Sylow p -subgroups. That is, if G is a finite nilpotent group, then $G \cong S_{p_1} \times S_{p_2} \dots \times S_{p_r}$ where S_{p_i} is the Sylow p_i -subgroup. Since each S_{p_i} is abelian group representable by Proposition 53, the previous result implies that G is abelian group representable. □

Conversely, we show that the class of abelian representable groups must be contained inside the class of nilpotent groups for all n .

Proposition 59. A non-nilpotent group G is not abelian group representable for all values of n .

Proof. Recall that, a group is nilpotent if and only if all of its Sylow subgroups are normal. Since G is by assumption non-nilpotent, at least one of its Sylow p -subgroup S_p is not normal for a prime $p \mid r = |G|$. We know that $|S_p| = p^a$, the highest power of p dividing r , that is, $r = p^a m$ with $p \nmid m$. Since S_p is not normal, G has another Sylow p -subgroup $S_p^x = \{xS_px^{-1} \mid x \in G, s \in S_p\}$ of order p^a with $S_p \neq S_p^x$ for some x .

Say their intersection $S_p \cap S_p^x$ is of order p^t . Since S_p and S_p^x are distinct, $a > t$ and thus $[G : S_p \cap S_p^x] = mp^{a-t} > m$. Thus we have subgroups $G_1 = S_p, G_2 = S_p^x, G_{12} = S_p \cap S_p^x$ of indices m, m, mp^{a-t} , respectively, in G . But this contradicts Lemma 52, which implies that $mp^{a-t} \mid m^2$.

□

In particular, we have completely classified groups which are abelian group representable for $n = 2$.

Theorem 60. A group G is abelian group representable for $n = 2$ if and only if it is nilpotent.

The following corollary generalizes Proposition 4 of [38].

Corollary 61. If m is not a power of 2, then the dihedral group D_m , the quasi-dihedral groups QD_m^{-1} and QD_m^{+1} , and the dicyclic group DiC_m are not abelian group representable for any $n > 1$.

Proof. Let G_m denote either $D_m, QD_m^{-1}, QD_m^{+1}$, or DiC_m . Since subgroups of nilpotent groups are nilpotent, it suffices to show that $G_p \leq G_m$ is not nilpotent for a prime $p \mid m$, $p \geq 3$.

In case when $G_m = D_m, QD_m^{-1}, QD_m^{+1}$ is (quasi-)dihedral, clearly G_p is not nilpotent, as its Sylow 2-subgroup $\{1, s\}$ is not normal. Similarly, a Sylow 2-subgroup $\langle x \rangle \leq DiC_p$ is not normal, which shows that DiC_p is not nilpotent.

By Proposition 59 the result follows.

□

To summarize the contributions of this chapter, we proposed a classification of finite groups with respect to the quasi-uniform variables induced by the subgroup structure. In particular, we studied which finite groups belong to the same class as abelian groups with respect to this classification, that is, which finite groups can be represented by abelian groups. We provided an answer to this question when the number n of quasi-uniform variables is 2: it is the class of nilpotent groups. For general n , we showed that nilpotent groups are abelian representable if and only if p -groups are, while non-nilpotent groups do not afford abelian representation. Hence the question of classifying abelian representable finite groups is completely reduced to answering the question for p -groups.

We demonstrated how some classes of p -groups afford abelian representation for all n , opening various interesting questions for further work. What other classes of p -groups can be shown to be abelian group representable? Is there a generalization of numerical

criterion given for $n = 2$ providing a necessary and sufficient condition for abelian representability? It would be extremely interesting to show whether p -groups are indeed abelian group representable. If not - what is the grading with respect to representability within p -groups (and, consequently, nilpotent groups)? Finally, beyond the nilpotent case, the classification of groups with respect to the quasi-uniform distributions is completely open, e.g. - what are the finite groups which induce the same quasi-uniform variables as solvable groups?

5.5 Applications of Information Inequalities

Understanding fundamental information inequalities is of long term importance to network coding and network information theory. Indeed, it was shown in [21, 41], that computing the capacity of arbitrary multi-source multi-sink acyclic networks consists of optimizing a linear function over the set of $(2^n - 1)$ -dimensional entropic vectors (Definition 12) where n is the number of random variables involved in the communication network. In particular, exhibiting random variables which violate linear information inequalities (information inequalities containing only linear combinations of information measures involved) helps to improve inner bounds on the region of entropic vectors, and to obtain points not achievable using linear network coding strategies.

More precisely, it is shown by Chan [12] that all abelian group representable entropic vectors satisfy the Ingleton inequality (Equation (5.2)) and hence all linear network codes (since they are abelian groups). But linear network codes do not achieve capacity in general network coding problems [13]. Therefore the knowledge of entropic vectors coming from non-abelian groups (non-abelian group representable) which violate the Ingleton inequality is helpful in constructing non-linear network codes to achieve the capacity that can not be reached using linear codes. Matúš [28] mentioned that there exist entropic vectors which violate the Ingleton inequality and Mao et al.[26] showed that the smallest group that violates the Ingleton inequality (this concept of violation using groups is discussed in Section 6.1) for $n = 4$ is S_5 , the symmetric group of order 120. In fact, S_5 is isomorphic to the projective general linear group $PGL(2, 5)$ and $PGL(2, p)$, $p \geq 5$ is the first family of groups known to violate the Ingleton inequality. More examples of groups violating this inequality were studied in [3, 4, 26].

In the following chapter, we define and investigate violations of Ingleton inequalities in 5 random variables.

Violations of Non-Shannon Inequalities

We know that an information inequality is a linear inequality involving entropies of jointly distributed random variables and an information inequality always holds, if it is true for any joint distribution of the random variables involved. We have already seen the two types of information inequalities namely: Shannon inequalities and non-Shannon inequalities.

The next section (based on Chapter 16 of [42]) describes how we can formulate an information inequality (Shannon or non-Shannon) in terms of group theory using the connection between entropy and groups.

6.1 Information Inequalities and Group Inequalities

An unconstrained information inequality $b^\top h \geq 0$ always holds if and only if it is satisfied by all group representable entropic vectors (Section 4.5). That is,

$$b^\top h \geq 0$$

always holds if and only if

$$\Upsilon_n \subset \{h \in \mathcal{H}_n : b^\top h \geq 0\}$$

which is equivalent to say that

$$b^\top h \geq 0 \text{ for all } h \in \Upsilon_n.$$

An entropic vector $h \in \Upsilon_n$ if and only if

$$h_{\mathcal{A}} = \log \frac{|G|}{|G_{\mathcal{A}}|}$$

for all non-empty subset $\mathcal{A}$ of $\mathcal{N}$ and for some finite group G and subgroups $G_1, \dots, G_n$ where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$.

Therefore, an information inequality $b^\top h \geq 0$ holds for all random variables $X_1, \dots, X_n$ if and only if the inequality obtained by replacing $h_{\mathcal{A}}$ in $b^\top h \geq 0$ by $\log \frac{|G|}{|G_{\mathcal{A}}|}$ for all non-empty subset $\mathcal{A}$ of $\mathcal{N}$ holds for all finite group G and subgroups $G_1, \dots, G_n$.

In other words, for every unconstrained information inequality, there is a corresponding group inequality, and vice versa. Hence we can prove information inequalities using group techniques.

That is, if we are able to find a finite G and subgroups $G_1, \dots, G_n$ such that a group inequality is violated, then the corresponding information inequality does not always hold. We use this principle later in this chapter to verify information inequalities.

Next we prove some basic group identities which are helpful to understand group inequalities well.

Definition 39. Let G_1 and G_2 are subgroups of a finite Group G . Define

$$G_1 G_2 = \{g_1 g_2 : g \in G_1 \text{ and } g' \in G_2\}.$$

$G_1 G_2$ is not a subgroup of G in general. However, if either G_1 or G_2 is normal in G , then $G_1 G_2 \leq G$ [15].

The following proposition computes the cardinality of the set $G_1 G_2$.

Proposition 62. Let G_1 and G_2 are subgroups of a finite Group G . Then

$$|G_1 G_2| = \frac{|G_1| |G_2|}{|G_1 \cap G_2|}.$$

Proof. Consider

$$f : G_1 \times G_2 \rightarrow G_1 G_2 \text{ with } (g, g') \mapsto gg'.$$

Since f is surjective, $|G_1 G_2| \leq |G_1 \times G_2| = |G_1| |G_2| < \infty$.

Let $g_1 g'_1, \dots, g_d g'_d$ are distinct elements of $G_1 G_2$. Then $G_1 \times G_2$ is the disjoint union of $f^{-1}(g_i g'_i) : i = 1, \dots, d$.

That is, $|G_1 \times G_2| = |G_1| |G_2| = d |f^{-1}(g_i g'_i)|$.

Now

$$f^{-1}(g_i g'_i) = \{(g_i h, h^{-1} g'_i) : h \in G_1 \cap G_2\}$$

and hence $|f^{-1}(g_i g'_i)| = |G_1 \cap G_2|$.

Therefore,

$$d = |G_1 G_2| = \frac{|G_1||G_2|}{|G_1 \cap G_2|}.$$

□

Next result is very important to prove many information inequalities in later sections.

Theorem 63. [42] Let G_1, G_2, G_3 be subgroups of a finite group G . Then

$$|G_{13}||G_{23}| \leq |G_3||G_{123}|.$$

Proof. Consider

$$G_{13} \cap G_{23} = (G_1 \cap G_3) \cap (G_2 \cap G_3) = G_1 \cap G_2 \cap G_3 = G_{123} \subset G_3.$$

From Proposition 62, $|G_{13}G_{23}| = \frac{|G_{13}||G_{23}|}{|G_{123}|} \leq |G_3|$. That is,

$$|G_{13}||G_{23}| \leq |G_3||G_{123}|.$$

□

The following corollary explains the information theoretic meaning of the above inequality.

Corollary 64. For random variables X_1, X_2 , and X_3 ,

$$I(X_1; X_2|X_3) \geq 0.$$

Proof. From the above theorem, we have

$$|G_{13}||G_{23}| \leq |G_3||G_{123}|$$

for subgroups G_1, G_2, G_3 of a finite group G .

That is,

$$\frac{|G|^2}{|G_3||G_{123}|} \leq \frac{|G|^2}{|G_{13}||G_{23}|},$$

which is equivalent to

$$\log \frac{|G|}{|G_3|} + \log \frac{|G|}{|G_{123}|} \leq \log \frac{|G|}{|G_{13}|} + \log \frac{|G|}{|G_{23}|}.$$

This group inequality corresponds to the information inequality

$$H(X_3) + H(X_1, X_2, X_3) \leq H(X_1, X_3) + H(X_2, X_3),$$

which implies

$$I(X_1; X_2|X_3) \geq 0.$$

□

Now let $G_3 = G$ in Theorem 63. Then we have,

$$|G_1||G_2| \leq |G||G_{12}|,$$

which is equivalent to

$$\log \frac{|G|}{|G_{12}|} \leq \log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_2|}.$$

This corresponds to the information inequality,

$$H(X_1, X_2) \leq H(X_1) + H(X_2)$$

or

$$I(X_1; X_2) \geq 0.$$

Similarly we can prove all basic inequalities using group inequalities and vice versa.

Now consider a non-Shannon inequality in four random variables $X_1, \dots, X_4$

$$2I(X_3; X_4) \leq I(X_1; X_2) + I(X_1; X_3, X_4) + 3I(X_3; X_4|X_1) + I(X_3; X_4|X_2)$$

given in Theorem 15.7 [42]. Its canonical form representation is

$$\begin{aligned} & H(X_1) + H(X_1, X_2) + 2H(X_3) + 2H(X_4) + 4H(X_1, X_3, X_4) + H(X_2, X_3, X_4) \\ & \leq 3H(X_1, X_3) + 3H(X_1, X_4) + 3H(X_3, X_4) + H(X_2, X_3) + H(X_2, X_4). \end{aligned}$$

The corresponding group inequality is given by

$$\begin{aligned} & \log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_{12}|} + 2 \log \frac{|G|}{|G_3|} + 2 \log \frac{|G|}{|G_4|} + 4 \log \frac{|G|}{|G_{134}|} + \log \frac{|G|}{|G_{234}|} \\ & \leq 3 \log \frac{|G|}{|G_{13}|} + 3 \log \frac{|G|}{|G_{14}|} + 3 \log \frac{|G|}{|G_{34}|} + \log \frac{|G|}{|G_{23}|} + \log \frac{|G|}{|G_{24}|} \end{aligned}$$

Rearranging the terms imply,

$$|G_{13}|^3 |G_{14}|^3 |G_{34}|^3 |G_{23}| |G_{24}| \leq |G_1| |G_{12}| |G_3|^2 |G_4|^2 |G_{134}|^4 |G_{234}|.$$

Even though the meaning and implications of the last inequality is still unknown, we can use group theoretic methods to understand the initial information inequality.

6.2 Ingleton Inequalities

Ingleton inequalities (defined in Equation (5.2)) were first introduced by Ingleton in [23] while studying the rank functions of *representable matroids*. Later they are proven to be useful in the study of network codes to understand the capacity region. Some of the applications of Ingleton inequality in network coding can be found in Section 5.5.

A *linear rank inequality* is a linear inequality involving the rank of subspaces of a (finite) vector space over some field $\mathbb{F}$.

Connections between linear rank and information inequalities that always hold have been classically studied [19]. Indeed, there is a natural translation of entropy and mutual information in terms of rank: for A, B, C either random variables or subspaces, $H(A)$ is the entropy of A , or the rank of A , $H(A, B)$ is the joint entropy of A and B , or the rank of the span $\langle A, B \rangle$ of A and B , the mutual information

$$I(A; B) = H(A) + H(B) - H(A, B)$$

is the rank of $A \cap B$, the conditional entropy $H(A|B) = H(A, B) - H(B)$ is the excess of the rank of A over that of $A \cap B$, and the conditional mutual information

$$I(A; B|C) = H(A, C) + H(B, C) - H(A, B, C) - H(C)$$

is the excess of the rank of $(A + C) \cap (B + C)$ over that of C . The non-negativity of the conditional mutual information holds in both interpretations, and so does the non-negativity of the mutual information and of the conditional entropy.

It was shown in [19] that any linear information inequality that always holds is also a linear rank inequality which always holds for finite dimensional vector spaces over $\mathbb{F}$.

The converse is not true: the Ingleton inequality always holds for ranks of subspaces, yet does not hold for random variables.

Examples of random variables whose joint entropies do not satisfy Ingleton were given by Matúš in [28]. Furthermore, it is known [19] that the Ingleton inequality, together with its permuted variable forms, and the Shannon-type inequalities fully characterize linear rank inequalities on 4 subspaces (or random variables).

For 5 subspaces (or random variables), the situation gets more complicated [14]: this time, the Shannon-type inequalities, 4 Ingleton inequalities (see (6.3) below for a description of what these are), and 24 new inequalities (which we will refer to as DFZ inequalities after the authors of [14]) are needed to fully characterize all the linear rank inequalities. Finding violations of these inequalities is important in network coding as explained in Section 5.5. But it is never an easy task.

For 4 random variables, examples of violators of Ingleton inequality has been found using group theory.[3, 4, 26]. In the sequel we extend the approach of [26] to the case of 5 random variables and exploits group theory in order to find examples of random variables violating linear rank inequalities.

First, we prove in Section 6.2.2 that random variables from finite groups violate the Ingleton inequality for 4 random variables if and only they violate the Ingleton inequalities for 5 random variables. Hence to find new violators of linear rank inequalities, we must find violators of one of the DFZ inequalities.

We study the first ten of these 24 inequalities which have some common characteristics in Section 6.3.

6.2.1 Minimal Set of Ingleton Inequalities

We define the generic version of Ingleton inequality here even though it is defined in Section 5.1 in terms of entropy.

Definition 40. [18] An Ingleton inequality $J(H; \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4) \geq 0$ is a linear inequality defined in terms of four subsets $\mathcal{A}_1, \dots, \mathcal{A}_4 \subseteq \mathcal{N}$ where

$$\begin{aligned} & J(H; \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4) \\ = & H(X_{\mathcal{A}_1 \cup \mathcal{A}_2}) + H(X_{\mathcal{A}_1 \cup \mathcal{A}_3}) + H(X_{\mathcal{A}_1 \cup \mathcal{A}_4}) - H(X_{\mathcal{A}_1}) \\ & + H(X_{\mathcal{A}_2 \cup \mathcal{A}_4}) + H(X_{\mathcal{A}_2 \cup \mathcal{A}_3}) - H(X_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3}) \\ & - H(X_{\mathcal{A}_3 \cup \mathcal{A}_4}) - H(X_{\mathcal{A}_2}) - H(X_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_4}). \end{aligned} \tag{6.1}$$

The notation $J(H; \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4)$ suggests that an Ingleton inequality could be defined on a function different from H . This definition is the same as Equation (5.2), where $h(\mathcal{A}_i)$ is used instead of $H(X_{\mathcal{A}_i})$ for all $\mathcal{A} \subseteq \mathcal{N}$.

The subsets $\mathcal{A}_1, \dots, \mathcal{A}_4$ of $\mathcal{N}$ are arbitrary and hence some inequalities are redundant. However Guille et al. [18] computed the minimal set of Ingleton inequalities, that is, the (unique) set of Ingleton inequalities which describes all of them without redundancy.

Example 14. When $n = 4$, there is one Ingleton inequality, which can be rephrased in terms of conditional mutual information as:

$$I(X_1; X_2) \leq I(X_1; X_2|X_3) + I(X_1; X_2|X_4) + I(X_3; X_4). \quad (6.2)$$

and it will be referred to as the 4-Ingleton inequality for short.

Example 15. When $n = 5$, the following three 5-Ingleton inequalities expressed in terms of conditional mutual information form a minimal set of Ingleton inequalities [14], together with the 4-Ingleton inequality:

$$\begin{aligned} I(X_1; X_2) &\leq I(X_1; X_2|X_3) + I(X_3; X_4, X_5) \\ &\quad + I(X_1; X_2|X_4, X_5) \\ I(X_1; X_2, X_3) &\leq I(X_1; X_2, X_3|X_4) + I(X_4; X_5) \\ &\quad + I(X_1; X_2, X_3|X_5) \\ I(X_1, X_2; X_1, X_3) &\leq I(X_1, X_2; X_1, X_3|X_1, X_4) \\ &\quad + I(X_1, X_4; X_1, X_5) + I(X_1, X_2; X_1, X_3|X_1, X_5). \end{aligned} \quad (6.3)$$

6.2.2 Group Theoretic Formulation of Ingleton Inequalities

We have seen methods to rewrite an information inequality in terms of groups in Section 6.1. Similarly we can write Equation (6.1) in group terms as:

$$\begin{aligned} J(H; \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4) &\geq 0 \iff \\ |G_{\mathcal{A}_1 \cup \mathcal{A}_2}| |G_{\mathcal{A}_1 \cup \mathcal{A}_3}| |G_{\mathcal{A}_1 \cup \mathcal{A}_4}| |G_{\mathcal{A}_2 \cup \mathcal{A}_3}| |G_{\mathcal{A}_2 \cup \mathcal{A}_4}| \\ &\leq |G_{\mathcal{A}_1}| |G_{\mathcal{A}_2}| |G_{\mathcal{A}_3 \cup \mathcal{A}_4}| |G_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3}| |G_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_4}|. \end{aligned} \quad (6.4)$$

Definition 41. We say that a finite group G violates the n -Ingleton inequality (6.4) if it contains n subgroups $G_1, \dots, G_n$ such that (6.4) does not hold.

Example 16. We get the 4-Ingleton inequality

$$|G_{12}| |G_{13}| |G_{14}| |G_{23}| |G_{24}| \leq |G_1| |G_2| |G_{34}| |G_{123}| |G_{124}| \quad (6.5)$$

from Example 14. Examples of groups violating this inequality were studied in [3, 4, 26].

We are interested in groups violating the 5-Ingleton inequalities, which, from Example 15, are

$$|G_{12}| |G_{13}| |G_{145}| |G_{23}| |G_{245}| \leq |G_1| |G_2| |G_{345}| |G_{123}| |G_{1245}| \quad (6.6)$$

$$|G_{123}| |G_{14}| |G_{15}| |G_{235}| |G_{234}| \leq |G_1| |G_{23}| |G_{54}| |G_{1235}| |G_{1234}| \quad (6.7)$$

$$|G_{123}| |G_{124}| |G_{125}| |G_{134}| |G_{135}| \leq |G_{12}| |G_{13}| |G_{145}| |G_{1234}| |G_{1235}| \quad (6.8)$$

and the 4-Ingleton inequality (6.5). Note first that both (6.6) and (6.7) in isolation can be reduced to the 4-Ingleton inequality. Indeed, take (6.6) and set $G_6 = G_{45}$, then (6.6) becomes

$$|G_{12}||G_{13}||G_{16}||G_{23}||G_{26}| \leq |G_1||G_2||G_{36}||G_{123}||G_{126}|,$$

which is (6.5) with a change of labels. Similarly, set $G_6 = G_{23}$ in (6.7), which becomes

$$|G_{16}||G_{14}||G_{15}||G_{56}||G_{46}| \leq |G_1||G_6||G_{54}||G_{156}||G_{146}|,$$

(6.5) again with a change of labels.

Now when $G_1 = G$, (6.8) becomes the Ingleton inequality for 4 random variables with a change of labels. It is, however, easy to see that looking for a finite group which violates inequalities in (6.5)-(6.8) reduces to looking for a violator of the 4-Ingleton inequality. We detail this in the following lemma.

Lemma 65. [27] A group G does not violate the 4-Ingleton inequality (6.2) if and only if it does not violate any of the four 5-Ingleton inequalities in (6.5)-(6.8).

Proof. Assume G does not violate the 4-Ingleton inequality (14). Fix arbitrary subgroups $G_1, G_2, G_3, G_4, G_5 \leq G$. The inequalities in (6.5)-(6.8) correspond to the following 4-Ingleton inequalities on different subgroups

$$\begin{aligned} J(H; 1, 2, 3, 4) &\geq 0 \\ J(H; 1, 2, 3, \{45\}) &\geq 0 \\ J(H; 1, \{23\}, 5, 4) &\geq 0 \\ J(H; \{12\}, \{13\}, \{14\}, \{15\}) &\geq 0 \end{aligned}$$

Since by assumption the group G does not violate the 4-Ingleton inequality (6.5) for any choice of subgroups $A, B, C, D \leq G$ and each of these is just the 4-Ingleton inequality evaluated on different subgroups, G does not violate any of them.

The converse is clear since the first 5-Ingleton inequality is the 4-Ingleton inequality. $\square$

6.3 DFZ Inequalities on 5 Variables

It was shown in [14] that linear rank inequalities for $n = 5$ random variables are characterized by Shannon-type inequalities, the Ingleton inequalities (6.5)-(6.8), and 24 other (DFZ) inequalities. Since by Lemma 65, a group violates the 5-Ingleton inequalities if and only if it violates the 4-Ingleton inequality, we search for new violators of linear rank inequalities

by looking for violators of the DFZ inequalities, starting with the first ten, which are

$$I(X_1; X_2) \leq I(X_1; X_2|X_3) + I(X_1; X_2|X_4) + I(X_3; X_4|X_5) + I(X_1; X_5) \quad (6.9)$$

$$I(X_1; X_2) \leq I(X_1; X_2|X_3) + I(X_1; X_3|X_4) + I(X_1; X_4|X_5) + I(X_2; X_5) \quad (6.10)$$

$$I(X_1; X_2) \leq I(X_1; X_3) + I(X_1; X_2|X_4) + I(X_2; X_5|X_3) + I(X_1; X_4|X_3, X_5) \quad (6.11)$$

$$I(X_1; X_2) \leq I(X_1; X_3) + I(X_1; X_2|X_4, X_5) + I(X_2; X_4|X_3) + I(X_1; X_5|X_3, X_4) \quad (6.12)$$

$$I(X_1; X_2) \leq I(X_1; X_3) + I(X_2; X_4|X_3) + I(X_1; X_5|X_4) + I(X_1; X_2|X_3, X_5) + I(X_2; X_3|X_4, X_5) \quad (6.13)$$

$$I(X_1; X_2) \leq I(X_1; X_3) + I(X_2; X_4|X_5) + I(X_4; X_5|X_3) + I(X_1; X_2|X_3, X_4) + I(X_1; X_3|X_4, X_5) \quad (6.14)$$

$$I(X_1; X_2) \leq I(X_2; X_4) + I(X_1; X_3|X_4) + I(X_1; X_5|X_3) + I(X_2; X_4|X_3, X_5) + I(X_1; X_2|X_4, X_5) \quad (6.15)$$

$$2I(X_1; X_2) \leq I(X_3; X_4) + I(X_3, X_4; X_5) + I(X_1; X_2|X_3) + I(X_1; X_2|X_4) + I(X_1; X_2|X_5) \quad (6.16)$$

$$2I(X_1; X_2) \leq I(X_1; X_3) + I(X_4; X_5) + I(X_1; X_2|X_4) + I(X_1; X_2|X_5) + I(X_2; X_4, X_5|X_3) \quad (6.17)$$

$$2I(X_1; X_2) \leq I(X_3; X_4) + I(X_1; X_5) + I(X_1; X_2|X_3) + I(X_1; X_2|X_4) + I(X_2; X_4|X_5) + I(X_1; X_3|X_4, X_5) \quad (6.18)$$

Remark 8. We choose to look at these 10 inequalities because they have in common that if there exists a random variable Z , called common information, such that

$$H(Z|X_1) = 0, H(Z|X_2) = 0, H(Z) = I(X_1; X_2),$$

then they are deduced from Shannon inequalities [14].

The other 14 inequalities are also deduced from Shannon inequalities through the concept of common information, but it takes different expressions.

These inequalities can be translated into group inequalities, using the following equivalences:

$$\begin{aligned} I(X_1; X_2) &= H(X_1) + H(X_2) - H(X_1, X_2) \\ &= \log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_2|} - \log \frac{|G|}{|G_{21}|} \\ &= \log \frac{|G_{12}||G|}{|G_1||G_2|}, \end{aligned}$$

and similarly

$$I(X_1; X_2, X_3) = \log \frac{|G_{123}||G|}{|G_1||G_{23}|},$$

$$I(X_1, X_2; X_3, X_4) = \log \frac{|G_{1234}||G|}{|G_{12}||G_{34}|},$$

$$I(X_1; X_2|X_3) = \log \frac{|G_{123}||G_3|}{|G_{13}||G_{23}|},$$

$$I(X_1; X_2|X_3, X_4) = \log \frac{|G_{1234}||G_{34}|}{|G_{134}||G_{234}|}.$$

This yields accordingly 10 group inequalities:

$$|G_{12}||G_{13}||G_{23}||G_{14}||G_{24}||G_{35}||G_{45}| \leq |G_2||G_3||G_{123}||G_{15}||G_{124}||G_4||G_{345}|, \quad (6.19)$$

$$|G_{12}||G_{13}||G_{23}||G_{14}||G_{34}||G_{15}||G_{45}| \leq |G_1||G_3||G_{123}||G_{25}||G_{134}||G_4||G_{145}|, \quad (6.20)$$

$$|G_{12}||G_{14}||G_{24}||G_{23}||G_{135}||G_{345}| \leq |G_2||G_{13}||G_{124}||G_4||G_{253}||G_{1345}|, \quad (6.21)$$

$$|G_{12}||G_{145}||G_{245}||G_{23}||G_{134}||G_{345}| \leq |G_2||G_{13}||G_{1245}||G_{45}||G_{234}||G_{1345}|, \quad (6.22)$$

$$\begin{aligned} & |G_{12}||G_{23}||G_{34}||G_{14}||G_{135}||G_{235}||G_{245}||G_{345}| \\ \leq & |G_2||G_{13}||G_{234}||G_{145}||G_4||G_{1235}||G_{35}||G_{2345}|, \end{aligned} \quad (6.23)$$

$$\begin{aligned} & |G_{12}||G_{25}||G_{35}||G_{134}||G_{234}||G_{145}| \\ \leq & |G_2||G_{13}||G_{245}||G_5||G_{1234}||G_{1345}|, \end{aligned} \quad (6.24)$$

$$\begin{aligned} & |G_{12}||G_{14}||G_{34}||G_{13}||G_{235}||G_{345}||G_{145}||G_{245}| \\ \leq & |G_1||G_{24}||G_{134}||G_{135}||G_3||G_{2345}||G_{1245}||G_{45}|, \end{aligned} \quad (6.25)$$

$$\begin{aligned} & |G_{12}|^2|G_{13}||G_{23}||G_{14}||G_{24}||G_{15}||G_{25}| \\ \leq & |G_1|^2|G_2|^2|G_{123}||G_{124}||G_{125}||G_{345}|, \end{aligned} \quad (6.26)$$

$$\begin{aligned} & |G_{12}|^2|G_{14}||G_{24}||G_{15}||G_{25}||G_{23}||G_{345}| \\ \leq & |G_1||G_2|^2|G_{13}||G_{45}||G_{124}||G_{125}||G_{2345}|, \end{aligned} \quad (6.27)$$

$$\begin{aligned} & |G_{12}|^2|G_{13}||G_{23}||G_{14}||G_{24}||G_{25}||G_{145}||G_{345}| \\ \leq & |G_1||G_2|^2|G_{34}||G_{15}||G_{123}||G_{124}||G_{245}||G_{1345}|. \end{aligned} \quad (6.28)$$

To find violating groups, it is useful to have conditions which guarantee that a group (or its chosen subgroups) will not result in a violation. For the 4-Ingletton inequality, the following conditions are all sufficient to tell when chosen subgroups $G_1, \dots, G_4$ of G will not result in a violation of the 4-Ingletton inequality [24, 36]:

- All G_i are normal.
- $G_1G_2 = G_2G_1$.

- $G_i = \{1\}$ or $G_i = G$ for some i .
- $G_i = G_j$ for some $i \neq j$.
- $G_{12} = \{1\}$.
- $G_i \leq G_j$ for some $i \neq j$.

We will extend some of these conditions to the DFZ inequalities considered.

The following lemma (generic version of Theorem 63) is useful in proving many inequalities.

Lemma 66. Let G be a finite group with n subgroups $G_1, \dots, G_n$. For any choice of subsets $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ of $\mathcal{N}$

$$|G_{\mathcal{A}_1 \cup \mathcal{A}_2}| |G_{\mathcal{A}_1 \cup \mathcal{A}_3}| \leq |G_{\mathcal{A}_1}| |G_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3}| \leq |G_{\mathcal{A}_1}| |G_{\mathcal{A}_2 \cup \mathcal{A}_3}|$$

is always satisfied.

Proof. Since $G_{\mathcal{A}_1 \cup \mathcal{A}_2}$ and $G_{\mathcal{A}_1 \cup \mathcal{A}_3}$ are subgroups of $G_{\mathcal{A}_1}$, to prove the first inequality, recall from [36] that one can define a map

$$f : G_{\mathcal{A}_1 \cup \mathcal{A}_2} \times G_{\mathcal{A}_1 \cup \mathcal{A}_3} \longrightarrow G_{\mathcal{A}_1}$$

that sends $(h, k) \mapsto hk$.

We have that $hk = h'k'$ if and only if $h'^{-1}h = k'k^{-1} \in G_{\mathcal{A}_1 \cup \mathcal{A}_2} \cap G_{\mathcal{A}_1 \cup \mathcal{A}_3} = G_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3}$. Then

$$|f(G_{\mathcal{A}_1 \cup \mathcal{A}_2} \times G_{\mathcal{A}_1 \cup \mathcal{A}_3})| = \frac{|G_{\mathcal{A}_1 \cup \mathcal{A}_2}| |G_{\mathcal{A}_1 \cup \mathcal{A}_3}|}{|G_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3}|} \leq |G_{\mathcal{A}_1}|.$$

The second inequality follows since

$$G_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3} \leq G_{\mathcal{A}_2 \cup \mathcal{A}_3}.$$

□

This lemma rephrases the non-negativity of conditional mutual information as well as the non-negativity of conditional entropy.

That is; the first inequality corresponds to the basic information inequality:

$$H(X_{\mathcal{A}_1}, X_{\mathcal{A}_2}) + H(X_{\mathcal{A}_1}, X_{\mathcal{A}_3}) - H(X_{\mathcal{A}_1}) - H(X_{\mathcal{A}_1}, X_{\mathcal{A}_2}, X_{\mathcal{A}_3}) = I(X_{\mathcal{A}_2}; X_{\mathcal{A}_3} | X_{\mathcal{A}_1}) \geq 0$$

where $H(X_{\mathcal{A}_i}, X_{\mathcal{A}_j}) = \log \frac{|G|}{|G_{\mathcal{A}_i \cup \mathcal{A}_j}|}$ and the second inequality corresponds to

$$H(X_{\mathcal{A}_1}, X_{\mathcal{A}_2}, X_{\mathcal{A}_3}) - H(X_{\mathcal{A}_2}, X_{\mathcal{A}_3}) = H(X_{\mathcal{A}_1}, X_{\mathcal{A}_2}, X_{\mathcal{A}_3} | X_{\mathcal{A}_2}, X_{\mathcal{A}_3}) \geq 0.$$

It may be easier to understand these inequalities in the language of information theory, or in that of group theory depending on one's interest. However, to identify examples of random variables, group theory is needed, and thus we will adopt its language, and not wonder anymore whether the inequalities may be rephrased in terms of mutual information.

6.4 Negative Conditions for DFZ Inequalities

Before going to investigate the negative conditions for individual inequalities, it would be nice to have some conditions which eliminate certain classes of subgroups altogether.

6.4.1 Eliminating Classes of Subgroups

We begin the section with the definition of a simultaneous violator.

Definition 42. If a finite group G and subgroups $G_1, \dots, G_n$ violate two or more inequalities, then $(G, G_1, \dots, G_n)$ is called a *simultaneous violator* for those inequalities.

The following result says that there do not exist any simultaneous violator $(G, G_1, \dots, G_n)$ which violates all the 24 DFZ inequalities.

Proposition 67. There is no simultaneous violator $(G, G_1, \dots, G_n)$ for (6.19) and (6.21).

Proof. Suppose $(G, G_1, G_2, G_3, G_4, G_5)$ be a simultaneous violator of (6.19) and (6.21). That is,

$$|G_{12}||G_{13}||G_{23}||G_{14}||G_{24}||G_{35}||G_{45}| > |G_2||G_3||G_{123}||G_{15}||G_{124}||G_4||G_{345}|$$

and

$$|G_{12}||G_{14}||G_{24}||G_{23}||G_{135}||G_{345}| > |G_2||G_{13}||G_{124}||G_4||G_{235}||G_{1345}|.$$

Since all these quantities are positive the product of the inequalities yields:

$$\begin{aligned} & |G_{12}|^2 |G_{23}|^2 |G_{14}|^2 |G_{24}|^2 |G_{35}| |G_{45}| |G_{135}| \\ & > |G_2|^2 |G_3| |G_{123}| |G_{15}| |G_4|^2 |G_{124}|^2 |G_{235}| |G_{1345}|. \end{aligned} \quad (6.29)$$

But using the Lemma 66 repeatedly, we have

$$\begin{aligned}
& |G_{12}|^2 |G_{23}|^2 |G_{14}|^2 |G_{24}|^2 |G_{35}| |G_{45}| |G_{135}| \\
& \leq |G_2| |G_{123}| |G_2| |G_{124}| |G_{23}| |G_{14}|^2 |G_{24}| |G_{35}| |G_{45}| |G_{135}| \\
& \leq |G_2|^2 |G_{123}| |G_3| |G_{235}| |G_4| |G_{124}|^2 |G_{14}| |G_{45}| |G_{135}| \\
& \leq |G_2|^2 |G_{123}| |G_3| |G_{235}| |G_4|^2 |G_{124}|^2 |G_{145}| |G_{135}| \\
& \leq |G_2|^2 |G_3| |G_{123}| |G_{15}| |G_4|^2 |G_{124}|^2 |G_{235}| |G_{1345}|,
\end{aligned}$$

a contradiction to inequality (6.29) which concludes the proof. $\square$

Corollary 68. There do not exist any simultaneous violators for all DFZ inequalities.

Proof. If there exists a simultaneous violator $(G, G_1, G_2, G_3, G_4, G_5)$ for all DFZ inequalities, it violates (6.19) and (6.21) simultaneously, a contradiction to Proposition 67. $\square$

The next result is an extended version of a similar result in [27], which says that the existence of a particular subgroup gives a negative condition for all DFZ inequalities under consideration.

Proposition 69. If $G_1 G_2$ is a subgroup of G , then the DFZ inequalities above hold.

Proof. By Remark 8, DFZ inequalities (6.9-6.18) hold if there exists a random variable Z , called common information, such that

$$H(Z|X_1) = 0, H(Z|X_2) = 0, H(Z) = I(X_1; X_2),$$

We translate the above result into group terms::

$$H(Z|X_1) = 0 \iff \log \frac{|G|}{|G_1|} = \log \frac{|G|}{|G_Z \cap G_1|} \iff G_1 < G_Z,$$

similarly $H(Z|X_2) = 0 \iff G_2 < G_Z$ and

$$\begin{aligned}
H(Z) = I(X_1; X_2) & \iff \log \frac{|G|}{|G_Z|} = \log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_2|} - \log \frac{|G|}{|G_{12}|} \\
& \iff |G_Z| = \frac{|G_1| |G_2|}{|G_{12}|}.
\end{aligned}$$

If $G_1 G_2$ is a subgroup of G , take $G_Z = G_1 G_2$, then

$$|G_1 G_2| = \frac{|G_1| |G_2|}{|G_{12}|}, G_1 < G_1 G_2, G_2 < G_1 G_2$$

which shows the existence of common information and concludes the proof. $\square$

Corollary 70. The above inequalities hold in the following cases:

1. $G_1 < G_2$, or $G_2 < G_1$,
2. G_1 or G_2 is normal in G .
3. G is abelian.

Corollary 71. If G is abelian group representable, then the above inequalities hold.

Proof. Let G be an abelian representable group and $G_1, \dots, G_5$ are subgroups of G .

Since G is abelian group representable, there exists an abelian group A and subgroups $A_1, \dots, A_5$ such that $[G : G_{\mathcal{A}}] = [A : A_{\mathcal{A}}]; \mathcal{A} \subseteq \mathcal{N} = \{1, \dots, 5\}$. Then the induced random variables $X_1, \dots, X_5$ have entropies

$$H(X_{\mathcal{A}}) = \log \frac{|G|}{|G_{\mathcal{A}}|} = \log \frac{|A|}{|A_{\mathcal{A}}|}.$$

Therefore the group inequalities can be rephrased in terms of the abelian group H and its subgroups. Hence the result follows from Proposition 69 and Corollary 70. $\square$

The next result says that subgroups of order pq where p and q are distinct primes will not violates certain DFZ inequalities under consideration.

Proposition 72. Let G be a group with the property that any two of its distinct proper subgroups intersect trivially. Then G does not violate any of the inequalities (6.19)-(6.28).

Proof. The case of inequalities (6.19) and (6.20) is proven in [27], and we use similar arguments to prove the remaining ones.

Recall that (6.21) is

$$\begin{aligned} & |G_{12}||G_{14}||G_{24}||G_{23}||G_{135}||G_{345}| \\ & \leq |G_2||G_{13}||G_{124}||G_4||G_{253}||G_{1345}|. \end{aligned}$$

By assumption, either G_{35} is trivial, or $G_3 = G_5$. If G_{35} is trivial, then the above inequality simplifies to

$$|G_{12}||G_{14}||G_{24}||G_{23}| \leq |G_2||G_{13}||G_{124}||G_4|.$$

Since $|G_{12}||G_{23}| \leq |G_2||G_{123}| \leq |G_2||G_{13}|$, it is enough to check that

$$|G_{14}||G_{24}| \leq |G_{124}||G_4|$$

which is always true. If $G_3 = G_5$, then

$$|G_{12}||G_{14}||G_{24}||G_{34}| \leq |G_2||G_{124}||G_4||G_{134}|.$$

Since $|G_{12}||G_{24}| \leq |G_2||G_{124}|$, we are left to check that

$$|G_{14}||G_{34}| \leq |G_4||G_{134}|$$

which holds.

Consider next (6.22), given by

$$\begin{aligned} & |G_{12}||G_{145}||G_{245}||G_{23}||G_{134}||G_{345}| \\ & \leq |G_2||G_{13}||G_{1245}||G_{45}||G_{234}||G_{1345}|. \end{aligned}$$

By assumption, either G_{45} is trivial, or $G_4 = G_5$. If G_{45} is trivial, then it simplifies to verifying that

$$|G_{12}||G_{23}||G_{134}| \leq |G_2||G_{13}||G_{234}|,$$

where $|G_{12}||G_{23}| \leq |G_2||G_{123}|$. But $|G_{123}||G_{134}| \leq |G_{13}||G_{234}|$ is indeed true, since $|G_{123}||G_{134}| \leq |G_{13}||G_{1234}|$.

If $G_4 = G_5$, then

$$|G_{12}||G_{14}||G_{24}||G_{23}||G_{134}||G_{34}| \leq |G_2||G_{13}||G_{124}||G_4||G_{234}||G_{134}|$$

which is (6.21) with $G_4 = G_5$.

The next inequality (6.23) is

$$\begin{aligned} & |G_{12}||G_{23}||G_{34}||G_{14}||G_{135}||G_{235}||G_{245}||G_{345}| \\ & \leq |G_2||G_{13}||G_{234}||G_{145}||G_4||G_{1235}||G_{35}||G_{2345}|. \end{aligned}$$

If $G_3 = G_4$, then

$$|G_{12}||G_{235}| \leq |G_2||G_{1235}|$$

which always holds. If $G_{34} = 1$, then

$$|G_{12}||G_{23}||G_{14}||G_{135}||G_{235}||G_{245}| \leq |G_2||G_{13}||G_{145}||G_{35}||G_4||G_{1235}|,$$

true always since $|G_{135}||G_{235}| \leq |G_{35}||G_{1235}|$, $|G_{12}||G_{23}| \leq |G_2||G_{123}|$,

$$|G_{123}||G_{1235}| \leq |G_{13}||G_{1235}| \text{ and } |G_{14}||G_{245}| \leq |G_4||G_{1245}| \leq |G_4||G_{145}|.$$

Next consider (6.24)

$$\begin{aligned} & |G_{12}||G_{25}||G_{35}||G_{134}||G_{234}||G_{145}| \\ & \leq |G_2||G_{13}||G_{245}||G_5||G_{1234}||G_{1345}|. \end{aligned}$$

If $G_2 = G_5$, then

$$|G_{12}||G_{23}||G_{134}||G_{234}||G_{124}| \leq |G_{13}||G_{24}||G_2||G_{1234}||G_{1234}|$$

which is always true since $|G_{234}||G_{124}| \leq |G_{24}||G_{1234}|$, and after cancellation, we have

$$|G_{12}||G_{23}||G_{134}| \leq |G_2||G_{123}||G_{134}| \leq |G_2||G_{13}||G_{1234}|.$$

If $G_{25} = \{1\}$, then

$$|G_{12}||G_{35}||G_{134}||G_{234}||G_{145}| \leq |G_2||G_{13}||G_{1234}||G_5||G_{1345}|.$$

$$\text{Now } |G_{12}||G_{234}||G_{134}| \leq |G_2||G_{1234}||G_{134}| \leq |G_2||G_{13}||G_{1234}|$$

and $|G_{35}||G_{145}| \leq |G_5||G_{1345}|$ implies that the inequality is true always.

Consider the inequality (6.25)

$$\begin{aligned} & |G_{12}||G_{14}||G_{34}||G_{13}||G_{235}||G_{345}||G_{145}||G_{245}| \\ & \leq |G_1||G_{24}||G_{134}||G_{135}||G_3||G_{2345}||G_{1245}||G_{45}| \end{aligned}$$

If $G_3 = G_5$, the inequality becomes,

$$|G_{12}||G_{14}||G_{23}||G_{34}| \leq |G_1||G_{24}||G_3||G_{1234}|,$$

which is true, as $|G_{12}||G_{14}| \leq |G_1||G_{124}| \leq |G_1||G_{24}|$.

If G_{35} is trivial, we have

$$|G_{12}||G_{14}||G_{34}||G_{13}||G_{145}||G_{245}| \leq |G_1||G_{24}||G_{134}||G_3||G_{1245}||G_{45}|$$

The inequality $|G_{145}||G_{245}| \leq |G_{1245}||G_{45}|$ is always true while the inequality involving the first four terms follows from

$$|G_{12}||G_{14}| \leq |G_1||G_{124}|, \quad |G_{34}||G_{13}| \leq |G_{134}||G_3|.$$

The inequality (6.26) is given by

$$\begin{aligned} & |G_{12}|^2 |G_{13}| |G_{23}| |G_{14}| |G_{24}| |G_{15}| |G_{25}| \\ & \leq |G_1|^2 |G_2|^2 |G_{123}| |G_{124}| |G_{125}| |G_{345}|. \end{aligned}$$

If $G_{25} = \{1\}$, the above inequality becomes

$$|G_{12}|^2 |G_{13}| |G_{23}| |G_{14}| |G_{24}| |G_{15}| \leq |G_1|^2 |G_2|^2 |G_{123}| |G_{124}| |G_{345}|.$$

We verify that by applying $|G_{12}||G_{23}| \leq |G_2||G_{123}|$, $|G_{12}||G_{24}| \leq |G_2||G_{124}|$ so that after cancellation it remains to show that $|G_{13}||G_{14}||G_{15}| \leq |G_1|^2 |G_{345}|$:

$$|G_{13}||G_{14}||G_{15}| \leq |G_1||G_{134}||G_{15}| \leq |G_1||G_1||G_{1345}| \leq |G_1|^2 |G_{345}|.$$

If $G_2 = G_5$, we have

$$|G_{12}|^2 |G_{13}| |G_{23}| |G_{14}| |G_{24}| \leq |G_1|^2 |G_2| |G_{123}| |G_{124}| |G_{234}|,$$

always true because of true inequalities

$$|G_{12}||G_{13}| \leq |G_1||G_{123}|, \quad |G_{12}||G_{14}| \leq |G_1||G_{124}| \text{ and } |G_{23}||G_{24}| \leq |G_2||G_{234}|.$$

The inequality (6.27) is given by

$$\begin{aligned} & |G_{12}|^2 |G_{14}| |G_{24}| |G_{15}| |G_{25}| |G_{23}| |G_{345}| \\ & \leq |G_1| |G_2|^2 |G_{13}| |G_{45}| |G_{124}| |G_{125}| |G_{2345}|. \end{aligned}$$

When $G_{15} = \{1\}$, it becomes

$$|G_{12}|^2 |G_{14}| |G_{24}| |G_{25}| |G_{23}| |G_{345}| \leq |G_1| |G_2|^2 |G_{13}| |G_{45}| |G_{124}| |G_{2345}|,$$

which can be verified by applying $|G_{24}||G_{25}||G_{345}| \leq |G_2||G_{245}||G_{345}| \leq |G_2||G_{45}||G_{2345}|$, $|G_{12}||G_{23}| \leq |G_2||G_{123}| \leq |G_2||G_{13}|$ and $|G_{12}||G_{14}| \leq |G_1||G_{124}|$.

If $G_1 = G_5$, the above inequality becomes

$$|G_{12}|^2 |G_{24}| |G_{23}| |G_{134}| \leq |G_2|^2 |G_{13}| |G_{124}| |G_{1234}|,$$

which is always true because $|G_{12}||G_{24}| \leq |G_2||G_{124}|$ and $|G_{12}||G_{23}||G_{134}| \leq |G_2||G_{123}||G_{134}| \leq |G_2||G_{13}||G_{1234}|$.

Finally, the inequality (6.28) is

$$\begin{aligned} & |G_{12}|^2 |G_{13}| |G_{23}| |G_{14}| |G_{24}| |G_{25}| |G_{145}| |G_{345}| \\ & \leq |G_1| |G_2|^2 |G_{34}| |G_{15}| |G_{123}| |G_{124}| |G_{245}| |G_{1345}|. \end{aligned}$$

If G_{14} is trivial. Then we have

$$|G_{12}|^2 |G_{13}| |G_{23}| |G_{24}| |G_{25}| |G_{345}| \leq |G_1| |G_2|^2 |G_{34}| |G_{15}| |G_{123}| |G_{245}|.$$

which is true since

$$\begin{aligned} & |G_{12}| |G_{25}| \leq |G_2| |G_{125}| \leq |G_2| |G_{15}|, \\ & |G_{12}| |G_{13}| \leq |G_1| |G_{123}|, |G_{23}| |G_{24}| \leq |G_2| |G_{234}| \text{ and} \\ & |G_{234}| |G_{345}| \leq |G_{34}| |G_{2345}| \leq |G_{34}| |G_{245}|. \end{aligned}$$

If $G_1 = G_4$, the above inequality becomes,

$$|G_{12}|^2 |G_{23}| |G_{25}| \leq |G_2|^2 |G_{123}| |G_{125}|$$

which we verify to hold:

$$(|G_{12}| |G_{25}|) |G_{12}| |G_{23}| \leq |G_2| |G_{125}| (|G_{12}| |G_{23}|) \leq |G_2| |G_{125}| |G_2| |G_{123}|.$$

□

The following corollary is immediate from the above proposition.

Corollary 73. Groups of order pq , p, q two distinct primes, always satisfy the inequalities (6.19)-(6.28).

6.4.2 Negative Conditions of the Form $G_i \leq G_j$

Now we will see some negative conditions of the form $G_i \leq G_j$ for each inequalities.

Lemma 74. The inequality (6.19) holds if any of the following conditions hold

$$G_1 \leq G_{25}, G_2 \leq G_1, G_3 \leq G_4, G_4 \leq G_3, G_5 \leq G_1.$$

The inequality (6.20) holds if any one of the following hold:

$$G_1 \leq G_{34}, G_2 \leq G_5, G_3 \leq G_1, G_4 \leq G_1, G_5 \leq G_2.$$

Proof. **1. Conditions on G_1**

We break the inequality (6.19) into two inequalities, the first one containing terms involving G_1 and the second inequality not involving G_1 :

$$\begin{aligned} |G_{12}||G_{13}||G_{14}| &\leq |G_{123}||G_{15}||G_{124}|, \\ |G_{23}||G_{24}||G_{35}||G_{45}| &\leq |G_2||G_3||G_4||G_{345}|. \end{aligned}$$

If each of the inequalities above holds, then so does (6.19).

Consider the inequality not involving G_1 . By Lemma 66

$$\begin{aligned} (|G_{23}||G_{24}|)|G_{35}||G_{45}| &\leq |G_2|(|G_{234}||G_{35}|)|G_{45}| \\ &\leq |G_2||G_3|(|G_{2345}||G_{45}|) \\ &\leq |G_2||G_3||G_4||G_{2345}| \\ &\leq |G_2||G_3||G_4||G_{345}|, \end{aligned}$$

the RHS of (6.19) without terms in G_1 . Hence to satisfy (6.19), it is enough for the inequality with terms in G_1 to hold:

$$|G_{12}||G_{13}||G_{14}| \leq |G_{15}||G_{123}||G_{124}|.$$

Now if $G_1 \leq G_{25}$, then $G_1 \leq G_2$ and $G_1 \leq G_5$, and

$$|G_{12}||G_{13}||G_{14}| = |G_1||G_{13}||G_{14}| = |G_{15}||G_{123}||G_{124}|.$$

2. Conditions on G_2

The LHS of inequality (6.19) without terms having G_2 is

$$\begin{aligned} (|G_{13}||G_{35}|)(|G_{14}||G_{45}|) &\leq |G_3||G_4|(|G_{135}||G_{145}|) \\ &\leq |G_3||G_4|(|G_{15}||G_{1345}|) \\ &\leq |G_3||G_4||G_{15}||G_{345}|, \end{aligned}$$

the RHS without terms in G_2 . It is then enough that $|G_{12}||G_{23}||G_{24}| \leq |G_2||G_{123}||G_{124}|$ in order for (6.19) to be satisfied. By Lemma 66, we only need $|G_{24}| \leq |G_{124}|$, which holds when $G_2 \leq G_1$.

3. Conditions on G_3

By avoiding terms in G_3 from (6.19)

$$\begin{aligned} (|G_{12}||G_{24}|)(|G_{14}||G_{45}|) &\leq |G_2||G_4|(|G_{124}||G_{145}|) \\ &\leq |G_2||G_4||G_{124}||G_{15}|. \end{aligned}$$

It thus suffices to have $|G_{13}||G_{23}||G_{35}| \leq |G_3||G_{123}||G_{345}|$ for (6.19) to hold. Now Lemma 66 yields $G_3 \leq G_4$ as the sufficient condition.

4. Conditions on G_4

Similarly

$$\begin{aligned} (|G_{12}||G_{23}|)(|G_{13}||G_{35}|) &\leq |G_2||G_{123}||G_3||G_{135}| \\ &\leq |G_2||G_3||G_{123}||G_{15}|, \end{aligned}$$

after omitting terms in G_4 from the LHS, and $|G_{14}||G_{24}||G_{45}| \leq |G_4||G_{124}||G_{345}|$ is a sufficient condition for (6.19) to hold, which is true when $G_4 \leq G_3$.

5. Conditions on G_5

Rewrite (6.19) as

$$\frac{|G||G_{12}|}{|G_1||G_2|} \frac{|G_1||G_5|}{|G||G_{15}|} \leq \frac{|G_3||G_{123}|}{|G_{13}||G_{23}|} \frac{|G_4||G_{124}|}{|G_{14}||G_{24}|} \frac{|G_5||G_{345}|}{|G_{35}||G_{45}|}$$

The RHS of the above inequality is always greater than 1 using Lemma 66 where as the LHS is less than 1 when $G_5 \leq G_1$, which implies the result [27].

A similar proof shows the results for the second inequality. $\square$

Corollary 75. If any $G_i = \{1\}$, then both (6.19) and (6.20), and hence (6.9) and (6.10), always hold.

Proof. If $G_i = \{1\}$ for any $i = 1, \dots, 5$, then $G_i \leq G_j$ for all j , hence satisfying the condition of Lemma 74. The result follows. $\square$

The following negative conditions are straight from the proof of Proposition 72.

Remark 9. The inequalities under the following conditions are never violated:

1. $G_{35} = 1$, or $G_3 = G_5$, or $G_3 \leq G_5$, for (6.21),
2. $G_{45} = 1$ for (6.22),

3. $G_{34} = 1$, or $G_3 = G_4$ for (6.23),
4. $G_{25} = 1$, or $G_2 = G_5$ for (6.24),
5. $G_{35} = 1$, or $G_3 = G_5$ for (6.25),
6. $G_{25} = 1$, or $G_2 = G_5$ for (6.26)
7. $G_{15} = 1$, or $G_1 = G_5$ for (6.27) and
8. $G_{14} = 1$, or $G_1 = G_4$ for (6.28).

6.5 Smallest Violations Using Groups

We would like to find the smallest group which violates any of (6.19)- (6.28). The results of this section allow us to eliminate a lot of groups of small order.

By Corollary 70, abelian groups do not violate these inequalities. This allows us to eliminate groups of order p and p^2 , which are known to be abelian for any prime p . Additionally we can eliminate abelian-representable groups using Corollary 71. i.e., groups whose corresponding joint entropies are known to be achieved by abelian groups. Among these are all groups order 8. Then until order 24, we are left with groups of possible orders 6; 10; 12; 14; 15; 16; 18; 20; 21; 22; 24.

Together with Corollary 73, which eliminates groups of order pq for primes $p \neq q$, we conclude that up to order 24, violators of the inequalities may only be found among groups of order 12; 16; 18; 20; 24.

Having narrowed down the list of possible suspects, we verify using GAP that there exist no violators of order < 24 of any of the inequalities.

Note that the lemmas that were proven earlier are very helpful in speeding up the computations. For example, for groups of order 16, there are 14 groups to be tested. Among them, 5 are abelian, 3 are abelian-group representable. This leaves 6 groups to be tested, with number of subgroups respectively 23, 15, 11, 35, 19, and 23, which all a priori require an exhaustive on 5-tuples across all subgroups.

In comparison, the actual number of checks performed is much smaller: e.g. 2252713 instead of $5^2 3$, 144841 instead of $5^1 5$, and 17641 instead of $5^1 1$ [27]. There does, in fact, exist a violator of order 24 for inequalities (6.19) and (6.21).

6.5.1 Smallest Violating Groups

Proposition 76. The symmetric group S_4 of permutations over 4 elements violates (6.19) and (6.21).

Proof. Consider $G = S_4$. For inequality (6.19), take

$$\begin{aligned} G_1 &= \langle (3, 4), (2, 4, 3) \rangle, & G_3 &= \langle (1, 2)(3, 4), (3, 4) \rangle \\ G_2 &= \langle (1, 3), (1, 3, 2) \rangle, & G_4 &= \langle (1, 3)(2, 4), (2, 4) \rangle \\ G_5 &= \langle (1, 4)(2, 3), (1, 3)(2, 4) \rangle \end{aligned}$$

with $|G_1| = 6, |G_2| = 6, |G_3| = 4, |G_4| = 4, |G_5| = 4$. Then

$$\begin{aligned} G_1 \cap G_2 &= \langle (2, 3) \rangle, & G_1 \cap G_3 &= \langle (3, 4) \rangle \\ G_2 \cap G_3 &= \langle (1, 2) \rangle, & G_1 \cap G_4 &= \langle (2, 4) \rangle \\ G_2 \cap G_4 &= \langle (1, 3) \rangle, & G_1 \cap G_5 &= \{1\} \end{aligned}$$

$$\begin{aligned} G_4 \cap G_5 &= \langle (1, 3)(2, 4) \rangle, & G_3 \cap G_4 \cap G_5 &= \{1\} \\ G_3 \cap G_5 &= \langle (1, 2)(3, 4) \rangle, & G_1 \cap G_2 \cap G_3 &= \{1\}. \end{aligned}$$

But for the trivial group, all the other intersection subgroups are of order 2. The left hand side of (6.19) yields

$$|G_{12}||G_{13}||G_{23}||G_{14}||G_{24}||G_{35}||G_{45}| = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 128$$

while the right hand side is

$$|G_2||G_3||G_{123}||G_{15}||G_{124}||G_4||G_{345}| = 6 \cdot 4 \cdot 1 \cdot 1 \cdot 1 \cdot 4 \cdot 1 = 96.$$

For inequality (6.21), take

- $G_1 = \langle (3, 4), (2, 4, 3) \rangle, |G_1| = 6,$
- $G_2 = \langle (1, 2)(3, 4), (3, 4) \rangle, |G_2| = 4,$
- $G_3 = \langle (1, 2)(3, 4), (1, 4)(2, 3), (1, 3) \rangle, |G_3| = 8,$
- $G_4 = \langle (1, 3), (1, 3, 2) \rangle, |G_4| = 6$
- $G_5 = \langle (1, 3)(2, 4), (2, 4) \rangle, |G_5| = 4.$

so that

$$G_{12} = \langle (3, 4) \rangle, G_{14} = \langle (2, 3) \rangle, G_{13} = \langle (2, 4) \rangle, G_{24} = \langle (1, 2) \rangle, G_{23} = \langle (1, 2)(3, 4) \rangle,$$

$$G_{135} = \langle (2, 4) \rangle, G_{345} = \langle (1, 3) \rangle$$

and

$$|G_{12}||G_{14}||G_{24}||G_{23}||G_{135}||G_{345}| = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 64$$

which is strictly larger than

$$|G_2||G_{13}||G_{124}||G_4||G_{235}||G_{1345}| = 4 \cdot 2 \cdot 1 \cdot 6 \cdot 1 \cdot 1 = 48.$$

□

In this chapter, we looked at violators for DFZ inequalities in 5 variables using corresponding group inequalities. We treated the initial 10 DFZ inequalities out of the 24 which have some common properties. We succeeded in finding violations for 2 of them and the smallest violator obtained is the symmetric group S_4 (see Proposition 76). We suggested some techniques to eliminate groups of certain orders altogether (see the results of Propositions 69, 72 and of corollaries 70, 71, 73) to prove the existence of violators and find the smallest ones. Also some negative conditions are given to reduce the computation time (see Subsection 6.4.2) if one uses any computer searches for violations.

To find more violators, we need more algebraic intuition of the subgroup structure of the violators obtained and its impact on the inequalities, which is currently under investigation.

It is also interesting to see how these violations can be useful in coding theory. We already mentioned in Section 5.5 that entropic vectors violating the Ingleton inequality are helpful to construct non-linear network codes for $n = 4$. Similarly, for $n = 5$, the above set of violators might be useful for the construction of network codes which improves the throughput of a given network. In order to understand these ideas, we need a framework of constructing codes using quasi-uniform random variables coming from groups which we discuss next.

Quasi-Uniform Codes

Recall the assumptions that $X_1, \dots, X_n$ is a collection of n jointly distributed discrete random variables over some alphabet of size N , $\mathcal{A}$ a non-empty subset of $\mathcal{N} = \{1, \dots, n\}$, $X_{\mathcal{A}} = \{X_i, i \in \mathcal{A}\}$ and $\lambda(X_{\mathcal{A}}) = \{x_{\mathcal{A}} : P(X_{\mathcal{A}} = x_{\mathcal{A}}) > 0\}$, the support of $X_{\mathcal{A}}$.

From Definition 32, a set of n random variables $X_1, \dots, X_n$ is said to be *quasi-uniform* if for any $\mathcal{A} \subseteq \mathcal{N}$, $X_{\mathcal{A}}$ is uniformly distributed over its support.

In this chapter, we study codes obtained from quasi-uniform random variables.

Definition 43. A code C of length n is an arbitrary non-empty subset of $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$ where $\mathcal{X}_i$ is the alphabet for the i th codeword symbol, and each $\mathcal{X}_i$ might be different.

Observe that this definition is much more general than the definition of a linear code, which is defined as:

Definition 44. A linear code of length n and dimension k is a linear subspace C with dimension k of the vector space $\mathbb{F}_q^n$ where $\mathbb{F}_q$ is the finite field with q elements. Such a code is called a q -ary code. If $q = 2$ or $q = 3$, the code is described as a binary code, or a ternary code respectively. The vectors in C are called codewords and the size of C , $|C| = q^k$. The *weight (Hamming weight)* of a codeword c is the number of nonzero coefficients of c and the *distance (Hamming Distance)* between two codewords is the number of coefficients in which they differ.

In this chapter, by code we mean a code as in Definition 43 unless otherwise specified.

We can associate to every code C a set of random variables [10] by treating each codeword $(X_1, \dots, X_n) \in C$ as a random vector with probability

$$P(X_{\mathcal{N}} = x_{\mathcal{N}}) = \begin{cases} 1/|C| & \text{if } x_{\mathcal{N}} \in C, \\ 0 & \text{otherwise.} \end{cases}$$

To the i th codeword symbol then corresponds a *codeword symbol random variable* X_i induced by C .

Definition 45. [10] A code C is said to be *quasi-uniform* if the induced codeword symbol random variables are quasi-uniform.

Given a code, we explained above how to associate a set of random variables, which might or not end up being quasi-uniform. Conversely, given a set of quasi-uniform random variables, a quasi-uniform code is obtained as follows.

Let $X_1, \dots, X_n$ be a set of quasi-uniform random variables with probabilities

$$P(X_{\mathcal{A}} = x_{\mathcal{A}}) = 1/|\lambda(X_{\mathcal{A}})| \text{ for all } \mathcal{A} \subseteq \mathcal{N}.$$

The corresponding quasi-uniform code C of length n is given by $C = \lambda(X_{\mathcal{N}}) = \{x_{\mathcal{N}} = P(X_{\mathcal{N}} = x_{\mathcal{N}}) > 0\}$.

Quasi-uniform codes were defined in [10], where some of their properties were discussed, and importantly their *weight enumerator polynomial* which specifies the number of words of each possible Hamming weight was computed.

For a linear $[n, k]$ code C of dimension k and length n (over some finite field), the weight enumerator polynomial of C is

$$W_C(x, y) = \sum_{c \in C} x^{n-wt(c)} y^{wt(c)},$$

where $wt(c)$ is the weight of c (see Definition 44).

For arbitrary codes, rather than the weight of the codewords, the distance between two codewords is of interest. Let

$$A_r(c) = |\{c' \in C: |\{j \in \mathcal{N}, c_j \neq c'_j\}| = r\}|$$

be the distance profile of C centered at c .

Note that we avoid defining A_r using $wt(c - c')$, since this already assumes that the difference of two codewords makes sense.

It was shown in [10] that quasi-uniform codes are distance-invariant, meaning that the distance profile does not depend on the choice of c .

Theorem 77. [10] Let C be a quasi-uniform code of length n . Then its weight enumerator $W_C(x, y) = \sum_{j=0}^n A_j x^{n-j} y^j$ is

$$W_C(x, y) = \sum_{\mathcal{A} \subseteq \mathcal{N}} q^{H(X_{\mathcal{N}}) - H(X_{\mathcal{A}})} (x - y)^{|\mathcal{A}|} y^{n - |\mathcal{A}|}, \quad (7.1)$$

where $H(X_{\mathcal{A}}) = \log_q(|\lambda(X_{\mathcal{A}})|)$ is the joint entropy of the induced codeword symbol quasi-uniform random variables.

The formula for the weight enumerator shows that it only depends $H(X_{\mathcal{A}})$. In fact, [10] which introduced quasi-uniform codes focused on their information theoretic properties, rather than on their coding properties.

The goal of this study is to address the construction and understanding of such quasi-uniform codes from a constructive point of view. We will use the group theoretic approach proposed in [6, 12] for constructing quasi-uniform random variables from finite groups.

More precisely, in Section 7.1, we recall the construction of quasi-uniform codes from groups, and compute the size of the corresponding code as a function of the group G we started with. We then consider abelian groups in Section 7.1.1, and compute the alphabet as well as the minimum distance as a function of G . We next move to nonabelian groups in Section 7.1.2. The structure of group, even though it is a nonabelian one, allows in some cases to mimic a definition for the minimum distance of the code. Potential applications to the design of almost affine codes is mentioned in Section 8.5.

Note that codes coming from groups have been studied in the literature (see for example [16],[43],[25],[17],[1] or [32]). Looking at constructions of codes from groups is here motivated by the need to design non-linear codes for network coding (see [13] and also [7] for applications of quasi-uniform codes to network coding), apart from designing almost affine codes as mentioned above.

7.1 Quasi-Uniform Codes from Groups

Let G be a finite group of order $|G|$ with n subgroups $G_1, \dots, G_n$, and $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$.

Recall from Section 4.1.4 that the number of (left) cosets of G_i in G is called the index of G_i in G , denoted by $[G : G_i]$ and Lagrange's Theorem (Theorem 28) states that $[G : G_i] = |G|/|G_i|$.

If G_i is normal, the sets of cosets $G/G_i := \{gG_i : g \in G\}$ are themselves groups, called quotient groups (Section 4.1.5).

	G_1	$\dots$	G_n
g_1	$g_1 G_1$		$g_1 G_n$
g_2	$g_2 G_1$		$g_2 G_n$
$\vdots$	$\vdots$		$\vdots$
$g_{ G }$	$g_{ G } G_1$	$\dots$	$g_{ G } G_n$

It is clear from Theorem 41 that we can obtain quasi-uniform random variables $X_1, \dots, X_n$ corresponding to $G_1, \dots, G_n$ such that for all non-empty subsets $\mathcal{A}$ of $\mathcal{N}$,

$$P(X_{\mathcal{A}} = x_{\mathcal{A}}) = |G_{\mathcal{A}}|/|G|.$$

Recall that the random variable $X_i = XG_i$ has for support the $[G : G_i]$ cosets of G_i in G where X is a random variable uniformly distributed over G with probability $P(X = g) = 1/|G|$ for any $g \in G$.

This shows that quasi-uniform random variables may be obtained from finite groups.

Quasi-uniform codes are obtained from these quasi-uniform distributions by taking the support $\lambda(X_{\mathcal{N}})$, as explained before. Codewords (of length n) can then be described explicitly by letting the random variable X take every possible values in the group G , and by computing the corresponding cosets as follows: Each row corresponds to one codeword of length n . The cardinality $|C|$ of the code obtained seems to be $|G|$, but in fact, it depends on the subgroups $G_1, \dots, G_n$. Indeed, it could be that the above table yields several copies of the same code.

Lemma 78. Let C be a quasi-uniform code obtained from a group G and its subgroups $G_1, \dots, G_n$. Then $|C| = |G|/|G_{\mathcal{N}}|$. In particular, if $|G_{\mathcal{N}}| = 1$, then $|C| = |G|$.

Proof. Let $G_{\mathcal{N}} = \{h_1, \dots, h_m\}$ be the intersection of all the subgroups $G_1, \dots, G_n$. Then there are $|G|/|G_{\mathcal{N}}|$ cosets of $G_{\mathcal{N}}$ in G .

Let us compute a first coset, say $g_1 G_{\mathcal{N}} = \{g_1, g_1 h_2, \dots, g_1 h_m\}$, by assuming wlog that h_1 is the identity element of $G_{\mathcal{N}}$. In words, we observe that every element in $g_1 G_{\mathcal{N}}$ is a multiple of a non-trivial element of $G_{\mathcal{N}}$.

Thus, when computing the above table, we have (by reordering the elements of G so as to list first the elements in $g_1 G_{\mathcal{N}}$): where $g_1 h_i G_j = g_1 G_j$, for all $i, j = 1, \dots, n$ because by definition of $G_{\mathcal{N}}$, $h_i \in G_j$.

Since the cosets of $G_{\mathcal{N}}$ partition G , each element repeats $|G|/|G_{\mathcal{N}}|$ times and hence we get $|G|/|G_{\mathcal{N}}|$ copies of a code C . $\square$

	G_1	$\dots$	G_n
g_1	$g_1 G_1$		$g_1 G_n$
$g_1 h_2$	$g_1 h_2 G_1 = g_1 G_1$		$g_1 h_2 G_n = g_1 G_n$
$\vdots$	$\vdots$		$\vdots$
$g_1 h_m$	$g_1 h_m G_1 = g_1 G_1$	$\dots$	$g_1 h_m G_n = g_1 G_n$
$\vdots$	$\vdots$		$\vdots$

One of the motivations to consider quasi-uniform codes is that they allow to go beyond abelian structures. Nevertheless, we will start by considering the case of abelian groups, which is relatively unexplored when different alphabets are involved.

7.1.1 Quasi-Uniform Codes from Abelian Groups

Suppose that G is an abelian group, with subgroups $G_1, \dots, G_n$. The procedure from Section 7.1 explains how to obtain a quasi-uniform distribution of n random variables, and thus a quasi-uniform code of length n from G . To avoid getting several copies of the same code, as exposed in Lemma 78, notice that since G is abelian, all the subgroups $G_1, \dots, G_n$ are normal, and thus so is G_N . If $|G_N| > 1$, we consider instead of G the quotient group G/G_N , and we can thus assume wlog that $|G_N| = 1$.

Lemma 79. Let C be a quasi-uniform code obtained from a group G and its subgroups $G_1, \dots, G_n$. Then the alphabet size of C is $\sum_{i=1}^n [G : G_i]$.

Proof. It is enough to show that all the cosets that appear in the table are distinct. By definition, every column contains $[G : G_i]$ distinct cosets. If $|G_i| \neq |G_j|$, the respective cosets will have different sizes, so let us assume that $|G_i| = |G_j|$. If $gG_i = g'G_j$, then $g^{-1}gG_i = G_i = g^{-1}g'G_j$ and it must be that $g^{-1}g'G_j$ is a subgroup. This implies that $g^{-1}g' \in G_j$, and thus that $g^{-1}g'G_j = G_j$. $\square$

The size of the alphabet can often be reduced, as explained next.

Let π_i denote the canonical projection $\pi_i : G \rightarrow G/G_i$.

Since G/G_i is itself an abelian group, let us denote this group isomorphism explicitly by $\psi_i : G/G_i \rightarrow H_i$, for some abelian group H_i .

Then $\pi_i(g) = gG_i \mapsto h \in H_i$ via $\psi_i(gG_i) = h, i = 1, \dots, n$.

Proposition 80. Let G be an abelian group with subgroups $G_1, \dots, G_n$. Then its corresponding quasi-uniform code is defined over $H_1 \times \dots \times H_n$.

Proof. Let X be again this random variable defined over G by $Pr(X = g) = 1/|G|$.

Define a new random variable Z_i by $Z_i = \psi_i(\pi_i(X))$ which takes values directly in H_i .

Then

$$\begin{aligned} \Pr(Z_i = h) &= \Pr(\psi_i(\pi_i(X)) = h) \\ &= \Pr(\pi_i(X) = gG_i) \\ &= \frac{1}{|G : G_i|} \\ &= |G_i|/|G|. \end{aligned}$$

Similarly, if $\Pr(Z_i = h_i, i \in \mathcal{A}) > 0$, then

$$\begin{aligned} \Pr(Z_i = h_i : i \in \mathcal{A}) &= \Pr(\psi_i(\pi_i(X)) = h_i : i \in \mathcal{A}) \\ &= \Pr(\pi_i(X) = gG_i : i \in \mathcal{A}) \\ &= |G_{\mathcal{A}}|/|G|. \end{aligned}$$

This implies that the random variables $Z_1, \dots, Z_n$ are quasi-uniform with the same joint probabilities as $X_1, \dots, X_n$. $\square$

In other words, we get a labeling of the cosets which respects the group structures componentwise. The next result then follows naturally.

Corollary 81. A quasi-uniform code C obtained from an abelian group is itself an abelian group.

Proof. First notice that the zero codeword is in C , since the codeword corresponding to the identity element in G is $(\psi_1(\pi_1(G_1)), \dots, \psi_n(\pi_n(G_n))) = (0, \dots, 0)$ where each 0 corresponds to the identity element in each abelian group H_i .

Let $(\psi_1(\pi_1(gG_1)), \dots, \psi_n(\pi_n(gG_n)))$ and $(\psi_1(\pi_1(g'G_1)), \dots, \psi_n(\pi_n(g'G_n)))$ be two codewords in C .

Then note that the codeword in C corresponding to the element $g + g' \in G$ is

$(\psi_1(\pi_1((g+g')G_1)), \dots, \psi_n(\pi_n((g+g')G_n)))$ where $\psi_i(\pi_i(g+g')) = \psi_i(\pi_i(g)) + \psi_i(\pi_i(g'))$, $i = 1, \dots, n$.

Every codeword has an additive inverse for the same reason. It forms an abelian group because every group law componentwise is commutative. $\square$

Because every H_i is an abelian group, we can freely use 0 since it corresponds to the identity element of H_i , as well as the operation $+$ and $-$, since $+$ is the group law for H_i , and $-$ is the additive inverse.

However, the alphabet H_i is possibly any abelian group, in particular, different abelian groups might be used for different components of the codewords.

The classification of abelian groups tells us that each H_i can be expressed as the direct sum of cyclic subgroups of order a prime power.

In the particular case where we have only one cyclic group, then

- (1) the group C_{p^r} is isomorphic to the integers mod p^r , and
- (2) the group C_p is isomorphic to the integers mod p , which in fact has a field structure, and we deal with the usual finite field $\mathbb{F}_p$.

If all the subgroups $G_1, \dots, G_n$ have index p , then we get an $[n, k]$ linear code over $\mathbb{F}_p$ (see Example 7.1).

Minimum Distance. The minimum distance of an abelian quasi-uniform code is encoded in its weight enumerator, but is however not easily read from the expression given in Theorem 77. We can easily express it in terms of the subgroups $G_1, \dots, G_n$.

Lemma 82. The minimum distance $\min_{c \in C} wt(c)$ of a quasi-uniform code C generated by an abelian group G and its subgroups $G_i, i = 1, \dots, n$ is $n - \max_{\mathcal{A} \in \mathcal{N}, G_{\mathcal{A}} \neq \{0\}} |\mathcal{A}|$.

Proof. The minimum distance $\min_{c \neq c' \in C} |\{c_i \neq c'_i\}|$ can be written as $\min_{c \neq c' \in C} wt(c - c')$ since $c - c'$ makes sense. Furthermore, since $c - c' \in C$, it reduces, as for linear codes over finite fields, to $\min_{c \in C} wt(c)$, and we are left to find the weight of the codeword having maximum number of zeros. This corresponds to finding the maximum number of subgroups whose intersection contains a non-trivial element of G . $\square$

As an illustration, here is a simple family of abelian groups that generate $[n, k]$ linear codes over $\mathbb{F}_p$ of length n and dimension k .

Lemma 83. The elementary abelian group $C_p \times C_p$ generates a $[p + 1, 2]$ linear code over $\mathbb{F}_p$ with minimum distance p .

Proof. The group $G = C_p \times C_p$ contains $p + 1$ non-trivial subgroups, of the form $\langle (1, i) \rangle$ where $i = 0, 1, \dots, p - 1$ and $\langle (0, 1) \rangle$. They all have index p and trivial pairwise intersection. We thus get a code of length $n = p + 1$, containing p^2 codewords, which is linear over $\mathbb{F}_p$ (by using that C_p is isomorphic to the integers mod p).

Since the pairwise intersection of subgroups is trivial, the minimum distance is p . $\square$

Example 7.1. Consider the elementary abelian group $G = C_3 \times C_3 \simeq \{0, 1, 2\} \times \{0, 1, 2\}$ and the four subgroups:

	$\langle(10)\rangle$	$\langle(01)\rangle$	$\langle(11)\rangle$	$\langle(12)\rangle$
(00)	$\langle(10)\rangle$	$\langle(01)\rangle$	$\langle(11)\rangle$	$\langle(12)\rangle$
(01)	(01)(11)(21)	$\langle(01)\rangle$	(01)(12)(20)	(01)(10)(22)
(02)	(02)(12)(22)	$\langle(01)\rangle$	(10)(21)(02)	(02)(11)(20)
(10)	$\langle(10)\rangle$	(10)(11)(12)	(10)(21)(02)	(01)(10)(22)
(11)	(01)(11)(21)	(10)(11)(12)	$\langle(11)\rangle$	(02)(11)(20)
(12)	(02)(12)(22)	(10)(11)(12)	(01)(12)(20)	$\langle(12)\rangle$
(20)	$\langle(10)\rangle$	(20)(21)(22)	(01)(12)(20)	(02)(11)(20)
(21)	(01)(11)(21)	(20)(21)(22)	(10)(21)(02)	$\langle(12)\rangle$
(22)	(02)(12)(22)	(20)(21)(22)	$\langle(11)\rangle$	(01)(10)(22)

TABLE 7.1: Quasi-uniform code constructed from $C_3 \times C_3 \simeq \{0, 1, 2\} \times \{0, 1, 2\}$.

- $G_1 = \langle(1, 0)\rangle = \{(0, 0), (1, 0), (2, 0)\}$,
- $G_2 = \langle(0, 1)\rangle = \{(0, 0), (0, 1), (0, 2)\}$,
- $G_3 = \langle(1, 1)\rangle = \{(0, 0), (1, 1), (2, 2)\}$, and
- $G_4 = \langle(1, 2)\rangle = \{(0, 0), (1, 2), (2, 1)\}$.

Using the method of Section 7.1, we obtain the following codewords (we write ij instead of (i, j) for brevity): Now let $H_1 = G/\langle(10)\rangle$, $H_2 = G/\langle(01)\rangle$, $H_3 = G/\langle(11)\rangle$ and $H_4 = G/\langle(12)\rangle$. Note that $H_i \simeq C_3 = \{0, 1, 2\}$ for all i . If we replace the subgroups by their quotients in the above table, we get the following code from Proposition 80:

	$\langle(10)\rangle$	$\langle(01)\rangle$	$\langle(11)\rangle$	$\langle(12)\rangle$
(00)	0	0	0	0
(01)	1	0	1	1
(02)	2	0	2	2
(10)	0	1	2	1
(11)	1	1	0	2
(12)	2	1	1	0
(20)	0	2	1	2
(21)	1	2	2	0
(22)	2	2	0	1

It is a ternary linear code of length $p + 1 = 4$ and minimum distance $p = 3$ with generator matrix $\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 \end{pmatrix}$.

Since $H(X_{\mathcal{N}}) = \log_q 9$, $H(X_{\mathcal{A}}) = \log_q 3$ when $|\mathcal{A}| = 1$ and $H(X_{\mathcal{A}}) = \log_q 9$ if $|\mathcal{A}| \geq 2$, we have that $q^{H(X_{\mathcal{N}})-H(X_{\mathcal{A}})} = 9$ when $\mathcal{A}$ is empty, $q^{H(X_{\mathcal{N}})-H(X_{\mathcal{A}})} = q^{\log_q(9/3)} = 3$ when $|\mathcal{A}| = 1$ and 1 otherwise, so that the weight enumerator of C using Theorem 77 is

$$\begin{aligned} W_C(x, y) &= \sum_{\mathcal{A} \subseteq \mathcal{N}} q^{H(X_{\mathcal{N}})-H(X_{\mathcal{A}})} (x-y)^{|\mathcal{A}|} y^{n-|\mathcal{A}|} \\ &= 9y^4 + \binom{4}{1} 3(x-y)y^3 + \binom{4}{2} (x-y)^2 y^2 + \binom{4}{3} (x-y)^3 y + (x-y)^4 \\ &= x^4 + 8xy^3, \end{aligned}$$

which can be verified using the usual methods.

7.1.2 Quasi-Uniform Codes from Nonabelian Groups

Suppose now that G is a nonabelian group. The resulting quasi-uniform codes might end up being very different depending on the nature of the subgroups considered. We next treat the different possible cases that can occur.

7.1.2.1 The Case of Quotient Groups

If G is a nonabelian group, but $G_1, \dots, G_n$ are normal subgroups, then the intersection $G_{\mathcal{N}}$ of all subgroups $G_1, \dots, G_n$ is a normal subgroup. Following Lemma 78, we can further consider the quotient $G/G_{\mathcal{N}}$. As a result, some nonabelian groups generate quasi-uniform codes which can be reduced to codes obtained from some abelian groups. This is the case of some dihedral groups (see Lemma 88). Let

$$D_{2m} = \langle r, s \mid r^m = s^2 = 1, rs = sr^{-1} \rangle$$

be the dihedral group of order $2m$.

We look at the normal subgroups of D_{2m} (which is not explained in Section 5.2.1) in order to have a deep understanding of the subgroup structure.

7.1.2.2 Normal Subgroups of D_{2m}

All subgroups of D_{2m} are of the form $\langle r^{d'} \rangle : d' | m$ and $\langle r^{d'}, r^i s \rangle : d' | m, 0 \leq i \leq d' - 1$ but not all of them are normal.

Proposition 84. Subgroups of the form $\langle r^{d'} \rangle : d' | m$ are normal in D_{2m} .

Proposition 85. When m is even, $\langle r^2, rs \rangle$ and $\langle r^2, s \rangle$ are the only normal subgroups of the form $\langle r^{d'}, r^i s \rangle$ and when m is odd, there is no such normal subgroups for D_{2m} .

From the above two propositions, we have

Proposition 86. The normal subgroups of D_{2m} are:

- $\{\langle r^{d'}; d'|m \rangle\}$ when m is odd
- $\{\langle r^{d'} \rangle; d'|m, \langle r^2, s \rangle, \langle r^2, rs \rangle\}$ when m is even.

The proofs of the above three propositions are given in Appendix A.

The total number of non-trivial normal subgroups is therefore the number of divisors of $m-1$ when m is odd (excluding $\langle r^m \rangle = \{1\}$) and the number of divisors $-1 + 2 = \text{Number of divisors} + 1$ when m is even.

From number theory we have the total number of divisors of $m = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$ is

$$\tau(m) = (r_1 + 1)(r_2 + 1) \dots (r_k + 1).$$

Hence the following result:

Proposition 87. # Normal subgroups = $(r_1 + 1) \dots (r_k + 1) - 1$ when m is odd and
Normal subgroups = $(r_1 + 1) \dots (r_k + 1) + 1$ when m is even.

Lemma 88. Quasi-uniform codes obtained from dihedral groups of order a power of 2 and some (possibly all) of their normal subgroups are obtained from the abelian group D_4 .

Proof. Normal subgroups H of D_{2m} are known: H is either a subgroup of $\langle r \rangle$, or $2|m$ and H is one of the two maximal subgroups of index 2 $\langle r^2, s \rangle, \langle r^2, rs \rangle$. Since m is a power of 2, the intersection G_N of (any choice of) these normal subgroups is necessarily a subgroup of order some power of 2, and we can take the quotient D_{2m}/G_N , which is a dihedral group of smaller order also a power of 2 and the code generated by D_{2m} and D_{2m}/G_N will be the same (see Lemma 78). Since the normal subgroups of D_{2m}/G_N too are of the same form, the process can be iterated until we reach D_4 which is abelian. $\square$

The case when G is nonabelian and all the subgroups $G_1, \dots, G_n$ are normal but some (possibly all) of the quotient groups G/G_i are nonabelian is very interesting.

Indeed, in that case, Proposition 80 still holds (with H_i nonabelian), and in fact, the corresponding quasi-uniform code still has a group structure, but that of a nonabelian group.

This gives the opportunity to go beyond abelian structures, and yet to keep a group structure.

As above, we assume wlog that $|G_{\mathcal{N}}| = 1$.

Lemma 89. A quasi-uniform code C obtained from a nonabelian group G with normal subgroups $G_1, \dots, G_n$ where at least one quotient group G/G_i is nonabelian forms a non-abelian group.

The proof is identical to that of Corollary 81, but one has to be cautious that the group law is not commutative. The group law $*$ is defined componentwise, and the identity element is the codeword $(1_{H_1}, \dots, 1_{H_n})$.

It is then possible to mimic the definition of minimum distance. The weight of a codeword is then the number of components which are not an identity element. The identity codeword plays the role of the whole zero codeword.

Minimum Distance. The minimum distance $\min_{c \neq c' \in C} |\{c_i \neq c'_i\}|$ is then

$$d = \min_{c \neq c' \in C} wt(c * (c')^{-1}).$$

Indeed, $(c')^{-1} = ((c'_1)^{-1}, \dots, (c'_n)^{-1})$, and noncommutativity is not an issue, since every inverse componentwise is both a left and a right inverse. This gives a counterpart to Lemma 106.

Corollary 90. The *minimum distance* $\min_{c \in C} wt(c)$ of a quasi-uniform code C generated by a nonabelian group G and its normal subgroups $G_i, i = 1, \dots, n$ is

$$d = n - \max_{\mathcal{A} \in \mathcal{N}, G_{\mathcal{A}} \neq \{0\}} |\mathcal{A}|,$$

where the weight $wt(c)$ is understood as the number of components which are not an identity element in some H_i .

Example 7.2. The dihedral group D_{12} has two normal subgroups $G_1 = \langle r^3 \rangle$ and $G_2 = \langle r^2, s \rangle$ with trivial intersection (see Figure 7.1). We can create quasi-uniform codes of length n by choosing $n - 2$ other normal subgroups. Since $D_{12}/G_1 \simeq D_6$, this gives a nonabelian quotient H_1 .

This example illustrates the difference between an information theoretic view of these codes, where one focuses on the joint entropy of the corresponding quasi-uniform codes, and a coding perspective, where the actual code, its alphabet, and its structure are of interest.

FIGURE 7.1: On the right, the dihedral group D_{12} , and on the left, the abelian group $C_3 \times C_2 \times C_2$, both with some of their subgroups.

We observe on Figure 7.1 that if we care about having two (normal) subgroups G_1, G_2 with respective order 2 and 6, in a group of order 12, we could have done that with the abelian group $C_3 \times C_2 \times C_2$. The difference is in the alphabet: D_{12}/G_1 is a nonabelian group, while $C_3 \times C_2 \times C_2/1 \times 1 \times C_2$ is abelian.

This question of distinguishing nonabelian groups whose entropic vectors can or cannot be obtained from abelian groups was addressed more generally in Chapter 5.

Recall the definition of group representability: for a given group G and subgroups $G_1, \dots, G_n$, $(A, A_1, \dots, A_n)$ is said to *represent* $(G, G_1, \dots, G_n)$ if for every non-empty $\mathcal{A} \subseteq \mathcal{N}$, $[G : G_{\mathcal{A}}] = [A : A_{\mathcal{A}}]$ where A is an abelian group with subgroups $A_1, \dots, A_n$.

It follows immediately that if $(A, A_1, \dots, A_n)$ represents $(G, G_1, \dots, G_n)$, then the quasi-uniform codes generated by both groups have the same joint entropy, and in turn the same weight enumerator, by Theorem 77.

From Chapter 5 this is the case for dihedral and quasi-dihedral 2-groups as well as dicyclic 2-groups, which are abelian group representable for any number of subgroups, and all nilpotent groups for $n = 2$ [37].

If $(G, G_1, \dots, G_n)$ cannot be abelian represented, then we can build a quasi-uniform code using exactly the same subgroups $G_1, \dots, G_n$ and be sure that its weight enumerator cannot be obtained from an abelian group. It was also shown in [37] that dihedral groups whose order is a power of 2 are abelian group representable, however Lemma 88 is stronger, in that it shows that the code alphabet will be the same.

Proposition 91. There exists a quasi-uniform code obtained from a non-nilpotent group which cannot be obtained by any abelian group.

Proof. Let G be a non-nilpotent group. Then G is not abelian group representable for all n [37]. That is, there exists some subgroups G_i such that $[G : G_{\mathcal{A}}] \neq [A : A_{\mathcal{A}}]$ for some

abelian group and subgroups. The proof uses the fact that in a group which is not nilpotent, there exists a Sylow subgroup which is not normal. $\square$

The above proposition guarantees that non-nilpotent groups can be used to construct quasi-uniform codes having different entropies and weight enumerators than abelian groups (refer Theorem 77).

We conclude this section by looking at the maximal length code obtained from a dihedral group D_{2m} and its normal subgroups.

7.1.2.3 Quasi-Uniform Codes from D_{2m} of Maximum Length

Suppose D_{2m} is a dihedral group of order $2m$. From Proposition 87, the total number of non-trivial normal subgroups is given by

- $(r_1 + 1) \dots (r_k + 1) - 1$ when m is odd
- $(r_1 + 1) \dots (r_k + 1) + 1$ when m is even.

Minimum distance of a quasi-uniform code $d = n - \max_{G_{\mathcal{A}} \neq \{1\}} |\mathcal{A}|$ and $\max_{G_{\mathcal{A}} \neq \{1\}} |\mathcal{A}|$ is the maximum number of subgroups with non-trivial mutual intersection.

Consider the set of all non-trivial subgroups of D_{2m} which do not contain any other non-trivial subgroups.

From the subgroup structure we can see that the following subgroups fit into that category.

$$\langle r^{r_1-1} p_1^{r_1-1} p_2^{r_2} \dots p_k^{r_k} \rangle, \langle r^{r_1-1} p_1^{r_1-1} p_2^{r_2-1} \dots p_k^{r_k} \rangle, \dots, \langle r^{r_1-1} p_1^{r_1-1} p_2^{r_2} \dots p_k^{r_k-1} \rangle.$$

Let $r_i = \max \{r_1, \dots, r_k\}$. Then the subgroup $H = \langle r^{r_1-1} p_1^{r_1-1} \dots p_i^{r_i-1} \dots p_k^{r_k} \rangle$ is the non-trivial intersection of maximum number of subgroups of D_{2m} .

The total number of divisors of $p_1^{r_1} \dots p_i^{r_i-1} \dots p_k^{r_k}$ is

$$h = (r_1 + 1) \dots (r_{i-1} + 1) r_i (r_{i+1} + 1) \dots (r_k + 1)$$

and hence there are h subgroups with mutual intersection H so that

$$\begin{aligned} \max_{G_{\mathcal{A}} \neq \{1\}} |\mathcal{A}| &= (r_1 + 1) \dots r_i \dots (r_k + 1) \quad \text{for odd } m \\ &= (r_1 + 1) \dots r_i \dots (r_k + 1) + 2 \quad \text{for even } m. \end{aligned}$$

	$\langle(12)\rangle$	$\langle(13)\rangle$	$\langle(23)\rangle$
$()$	$\langle(12)\rangle$	$\langle(13)\rangle$	$\langle(23)\rangle$
(12)	$\langle(12)\rangle$	$(12), (132)$	$(12), (123)$
(13)	$(13), (123)$	$\langle(13)\rangle$	$(13), (132)$
(23)	$(23), (132)$	$(23), (123)$	$\langle(23)\rangle$
(123)	$(13), (123)$	$(23), (123)$	$(12), (123)$
(132)	$(23), (132)$	$(12), (132)$	$(13), (132)$

TABLE 7.2: Quasi-uniform code constructed from S_3 and some nonnormal subgroups

Now $d = n - h$ and $n - (h + 2)$ for odd and even m respectively. After expanding n and h , it can be seen that

$$d = (r_1 + 1) \dots (r_{i-1} + 1)(r_{i+1} + 1) \dots (r_k + 1) - 1 \text{ for odd and even } m.$$

7.1.2.4 The Case of Nonnormal Subgroups

Let G be a nonabelian group, with subgroups $G_1, \dots, G_n$, where some (possibly all) of the subgroups are not normal. In that case, we lose the group structure on the set of cosets. To start with, Lemma 78 will still hold, since it does not depend on the normality of the subgroups, however in general, we cannot take the quotient by G_N anymore, and the copy of C we are left with does not have a group structure. This is an example of situations mentioned in the introduction, where the minimum distance is then not of interest.

Example 7.3. Consider the group $S_3 = \{(), (12), (13), (23), (123), (132)\}$ of permutations on three elements described in cycle notation. The corresponding quasi-uniform code is: Note that we could label the cosets using integers, however then one should keep in mind that these integers do not have any algebraic meaning.

Losing the group structure of the code however has the advantage that we have more flexibility in choosing the subgroups we deal with, and thus have more choices in terms of possible intersections that we are getting. Some applications of this kind of codes are given in Chapter 8.

7.2 Some Classical Bounds for Quasi-Uniform Codes

Finding bounds is important to understand the performance a code. In this section we go through some of the known classical bounds and wish to rewrite them in terms of groups or codes over different alphabets.

7.2.1 Singleton Bound for Quasi-Uniform Codes

When we think about classical bounds, Singleton bound is the first one which comes into our minds. It states that given a length n , q -ary code C of size $|C|$ having minimum distance d ,

$$|C| \leq q^{n-d+1}.$$

For linear codes in particular, it simplifies to

$$k \leq n - d + 1$$

where k is the dimension and $|C| = q^k$.

This bound is no longer valid for quasi-uniform codes since the alphabet size of each component varies. Therefore, we propose a new Singleton bound for quasi-uniform codes which serves as the usual singleton bounds for linear codes or q -ary codes. The result uses similar techniques as those from [31].

Proposition 92. The size of the quasi-uniform code C generated by a group G and subgroups $G_i : i = 1, \dots, n$ of length n and minimum distance d is bounded by

$$|C| \leq \sum_{l=0}^{n'} C(n', l) \quad (7.2)$$

where $C(n', l)$ is the number of elements of G which are in all but l subgroups and $n' = n - d + 1$.

Proof. Let G be a group and $G_1, \dots, G_n$ are n subgroups. Let C be the corresponding quasi-uniform code of length n and minimum distance d where $d = \min_{c \in C} wt(c)$. Without loss of generality assume that no words in C are repeated.

Let α be a subset of $\mathcal{N} = \{1, \dots, n\}$ and C_α denotes the projection of C onto the coordinate space X . If $|\alpha| = n - 1$, C_α has length $n - 1$ and the minimum distance of C_α , $d_{min}(C_\alpha) \geq d - 1$. (If we remove one entire column, the weight can be reduced at most by 1.)

Similarly if $|\alpha| = n - \beta$, $d_{min}(C_\alpha) \geq d - \beta$. As long as $d - \beta > 0$, $|C| = |C_\alpha|$. So for $|C| = |C_\alpha|$, β can be at most $d - 1$. That is if $\beta = d - 1$, the minimum distance $d_{min}(C_\alpha) \geq 1$ and $|C| \leq |C_\alpha|$.

Then the length of C_α is $n' = n - (d - 1) = n - d + 1$ and

$$|C| \leq \sum_{l=0}^{n'} \text{Number of words of weight } l \text{ in } C_\alpha.$$

(Note that if all G_i s are normal, then $d = n - \max_{\mathcal{A}, G_{\mathcal{A}} \neq 0} |\mathcal{A}|$ and $n' = \max |\mathcal{A}| + 1$.)

Observe that C_α is defined over $G/G_1 \times G/G_2 \times \dots \times G/G_{|\mathcal{A}|+1}$ and a codeword of weight 1 has all symbols except one are zeros. That is, the corresponding element in the group G is in all subgroups except one.

Similarly all words of weight l are words corresponding to group elements which are in all but l subgroups.

Denote by $C(n', l)$ **the number of elements of group G which are not in exactly l subgroups**. Therefore

$$|C| \leq \sum_{l=0}^{n'} C(n', l).$$

□

Computation of $C(n', l)$

We have the weight enumerator of C_X is given by

$$W_{C_\alpha}(x, y) = \sum_{\mathcal{A} \subseteq \mathcal{N}' = \{1, \dots, n'\}} q^{H(X_{\mathcal{N}'}) - H(X_{\mathcal{A}})} (x - y)^{|\mathcal{A}|} y^{n' - |\mathcal{A}|},$$

where $H(X_{\mathcal{A}}) = \log_q(|\lambda(X_{\mathcal{A}})|)$ (Theorem 77).

That is,

$$W_{C_\alpha}(x, y) = \sum_{\mathcal{A} \subseteq \mathcal{N}'} \frac{|\lambda(X_{\mathcal{N}'})|}{|\lambda(X_{\mathcal{A}})|} (x - y)^{|\mathcal{A}|} y^{n' - |\mathcal{A}|}.$$

Then $C(n', l)$, which is the number of codewords of weight l is the numerical coefficient in the term containing y^l for $l = 0, 1, \dots, n'$.

Lemma 93. If C is a q -ary code or a linear code of dimension k , the bound $|C| \leq \sum_{l=0}^{n'} C(n', l)$ reduces to the usual Singleton bound.

Proof. Suppose the length n quasi-uniform code C is linear over $\mathbb{F}_q$ having dimension k and minimum distance d . Then $|C| = q^k$.

Now $n' = n - d + 1$ and $\sum_{l=0}^{n'} C(n', l) = |C_\alpha|$ where C_α is the projection of C of length n' , which is also linear over $\mathbb{F}_q$ [10]. Then we have

$$|C| \leq \sum_{k=0}^{n'} C(n', k) = |C_\alpha| \leq q^{n'} = q^{n-d+1} \quad (7.3)$$

and hence $q^k \leq q^{n-d+1}$.

In particular, for a general q -ary code, it follows from (7.3) that

$$|C| \leq q^{n-d+1}.$$

□

7.2.1.1 Examples of Quasi-Uniform Codes Satisfying the Above Bound

In this section we describe some families of quasi-uniform codes which satisfies the proposed Singleton bound above. The following lemma is useful in this discussion.

Lemma 94. An elementary abelian group $G = C_p \times \dots \times C_p = (C_p)^m$ of order p^m has $\frac{p^m-1}{p-1}$ subgroups of order p ; where p , prime and $C_p = \{0, 1, \dots, p-1\}$.

Proof. Let G_i be a subgroup of order p . Since p is a prime, G_i is cyclic, spanned by an element in G .

Now each element in G has m components and each component takes p values and therefore the total number of subgroups spanned by elements of the form $\langle(0, \dots, i, 0, \dots, 0)\rangle$ is m .

Similarly, the total number of subgroups spanned by elements of the form

$\langle(0, \dots, 1, 0, \dots, i, 0, \dots, 0)\rangle$ is $\binom{m}{2}(p-1)$.

(Note that $\langle(0, \dots, i, 0, \dots, 1, 0, \dots, 0)\rangle = \langle(0, \dots, 1, 0, \dots, j, 0, \dots, 0)\rangle$ for some indices i and j .)

Using similar type of arguments the total number of subgroups of order p is

$$n = m + \binom{m}{2}(p-1) + \binom{m}{3}(p-1)^2 + \dots + \binom{m}{m-1}(p-1)^{m-2} + \binom{m}{m}(p-1)^{m-1}.$$

After simplification,

$$n = \frac{1}{p-1} \left[\sum_{i=0}^m \binom{m}{i} (p-1)^i - 1 \right] = \frac{1}{p-1} [(1 + (p-1))^m - 1] = \frac{p^m - 1}{p-1}.$$

□

Remark 10. The union of all subgroups G_i of order p is G with trivial mutual intersection.

Proof. Since G_i 's are of order p , they have trivial pairwise intersection and there is a total of $\frac{p^m-1}{p-1}$ subgroups of order p .

Using a counting argument, the total non-trivial elements is $\frac{p^m-1}{p-1}(p-1) = p^m - 1$ and hence the result follows. $\square$

Proposition 95. The quasi-uniform code generated by elementary abelian group $G = C_p \times \dots \times C_p$, ($C_p = \{0, 1, \dots, p-1\}$, $|G| = p^m$) and subgroups of order p satisfies the Singleton bound (7.2).

Proof. From Lemma 94 we have $\frac{p^m-1}{p-1}$ subgroups of order p so that the maximum length of the corresponding quasi-uniform code C is $n = \frac{p^m-1}{p-1}$ and since the mutual intersection of all subgroups is trivial, $|C| = |G|/|G_{\mathcal{N}}| = p^m$.

Now the minimum distance $d = n - \max_{G_{\mathcal{A}} \neq 0} |\mathcal{A}|$ since all subgroups are normal (G is abelian), which is equal to $d = n - 1$ since the pairwise intersection is trivial. That is,

$$d = \frac{p^m - 1}{p - 1} - 1 = \frac{p^m - p}{p - 1}.$$

Since $d = n - 1$, the length n' of C_X is $n' = n - d + 1 = 2$ so that the Singleton bound is $\sum_{l=0}^2 C(n', l)$ which is obtained from the weight enumerator as follows:

The weight enumerator of a quasi-uniform code is given by

$$W_C(x, y) = \sum_{\mathcal{A} \subseteq \mathcal{N}} q^{\log_q[G:G_{\mathcal{N}}] - \log_q[G:G_{\mathcal{A}}]} (x - y)^{|\mathcal{A}|} y^{n - |\mathcal{A}|} = \sum_{\mathcal{A} \subseteq \mathcal{N}} \frac{|G_{\mathcal{A}}|}{|G_{\mathcal{N}}|} (x - y)^{|\mathcal{A}|} y^{n - |\mathcal{A}|}.$$

Here $\mathcal{N} = \{1, 2\}$ and $\mathcal{A} = \{\phi, \{1\}, \{2\}, \{1, 2\}\}$. Therefore

$$W_C(x, y) = p^m y^2 + 2p(x - y)y + (x - y)^2 = (p^m - 2p + 1)y^2 + (2p - 2)xy + x^2.$$

Now $C(n', l)$ is the numerical coefficient of y^l so that

$$\sum_{l=0}^2 C(n', l) = p^m = |C|,$$

the Singleton bound is achieved. $\square$

We generalize the above result as given in the next proposition.

Proposition 96. The quasi-uniform code C generated by a group G and non-trivial normal subgroups $G_1, \dots, G_n$ having trivial pairwise intersection attains the Singleton bound (7.2).

Proof. Since the pairwise intersection of all G_i s is trivial, $|C| = |G|/|G_{\mathcal{N}}| = |G|$ and $\max_{G_{\mathcal{A}} \neq 1} |\mathcal{A}| = 1$ so that $d = n - 1$.

That is, C is a $(n, |G|, n-1)$ code defined over $H_1 \times \dots \times H_n$ where $H_i \cong G/G_i$.

To check whether C attains the Singleton bound, we compute $C(n', l)$ for l from 0 to $n' = n - d + 1 = 2$. Now

$$\begin{aligned}
 C(n', 0) &= \text{Number of elements in } C_X \text{ with weight zero} = 1. \\
 C(n', 2) &= \text{Number of elements of weight two} \\
 &= \text{Number of elements not in any subgroup components of } C_X \\
 &= |(G_i \cup G_j)^c| = |G \setminus (G_i \cup G_j)| \\
 &= |G| - (|G_i| + |G_j| - |G_{ij}|) = |G| - (|G_i| + |G_j| - 1) \\
 C(n', 1) &= \text{Number of elements not in exactly 1 subgroup} \\
 &= |G_i \cap G_j^c| + |G_i^c \cap G_j| \\
 &= |G_i \setminus G_j| + |G_j \setminus G_i| \\
 &= |G_i| + |G_j| - 2.
 \end{aligned}$$

Therefore

$$\sum_{k=0}^2 C(n', k) = 1 + |G| - [|G_i| + |G_j|] + 1 + |G_i| + |G_j| - 2 = |G| = |C|.$$

□

Example 17. From Section 7.1.2.3, the quasi-uniform code of length n , $d = n - 1$ and cardinality $2m$ generated by the dihedral group

$$D_{2m} = \langle r, s : r^m = s^2 = 1, rs = sr^{-1} \rangle; m = p_1^{r_1} p_2^{r_2} \dots p_n^{r_n}, p_i \text{ prime}$$

and subgroups

$$\langle r^{p_1^{r_1-1} p_2^{r_2} \dots p_n^{r_n}} \rangle, \langle r^{p_1^{r_1} p_2^{r_2-1} \dots p_n^{r_n}} \rangle, \dots, \langle r^{p_1^{r_1} p_2^{r_2} \dots p_n^{r_n-1}} \rangle$$

satisfies the Singleton bound using Proposition 96.

Corollary 97. Let G be a group and $G_1, \dots, G_n$ are normal subgroups whose mutual intersection is another normal subgroup say G_k . Then the quasi-uniform code generated by G/G_k and subgroups G_i/G_k satisfies the Singleton bound.

Proof. Let $G = G/G_k$ and $G_i = G_i/G_k$ where $G_i/G_k \trianglelefteq G/G_k$. Then the result follows from Lemma 78 and Proposition 96. □

7.2.2 Gilbert-Varshamov Bound

We define a ball of radius r centered at a codeword c .

Definition 46. A ball of radius r centered at c is defined as

$$B(c, r) = \{v \in H_1 \times \dots \times H_n \mid d(c, v) \leq r\}$$

where $d(c, v) = |\{i : c_i \neq v_i, i = 1, \dots, n\}|$ is the Hamming distance between c and v .

Suppose that $|H_i| = q_i$ for $i = 1, \dots, n$. Then the cardinality of the ball $B(c, r)$,

$$\begin{aligned} |B(c, r)| &= 1 + \sum_{i=1}^n q_i - n + \sum_{i_1 \leq i_2} (q_{i_1} - 1)(q_{i_2} - 1) \\ &\quad + \sum_{i_1 \leq i_2 \leq i_3} (q_{i_1} - 1)(q_{i_2} - 1)(q_{i_3} - 1) + \dots \\ &\quad + \sum_{i_1 \leq i_2 \leq \dots \leq i_r} (q_{i_1} - 1) \dots (q_{i_r} - 1). \end{aligned} \quad (7.4)$$

where $\sum_{i=1}^n q_i - n$ is the number of terms at distance 1 from c , $\sum_{i_1 \leq i_2} (q_{i_1} - 1)(q_{i_2} - 1)$ is the number of terms at distance 2 from c and so on.

When $q_i = q$ for all i , then

$$|B(c, r)| = \sum_{j=0}^r (q-1)^j,$$

the number of elements in a ball of radius r defined on a q -ary code.

Theorem 98. [33] Let M denote the maximum possible size of a q -ary code C with length n and minimum distance d . Then the Gilbert-Varshamov bound is given by

$$M \geq \frac{q^n}{\sum_{j=0}^{d-1} (q-1)^j}$$

We generalize this bound to the quasi-uniform case.

Proposition 99. If M denotes the maximum possible size of a quasi-uniform code C with length n and minimum distance d defined over $H_1 \times \dots \times H_n$ with $|H_i| = q_i$. Then the Gilbert-Varshamov bound is given by

$$M \geq \frac{q_1 \dots q_n}{|B(c, d-1)|}$$

where $B(c, d-1)$ is a ball of radius $d-1$ centered at $c \in C$.

Proof. Consider an (n, M, d) quasi-uniform code C (not necessarily linear), where the size $|C| = M$ is maximum for a given length n and minimum distance d defined over $H_1 \times H_2 \times \dots \times H_n$ where $H_i = q_i$.

For all $x \in H_1 \times H_2 \times \dots \times H_n$ not in C , there exists at least one code word $c_x \in C$ such that $d(x, c_x) \leq d - 1$.

Otherwise, we can add x to the code C while maintaining the minimum distance d , a contradiction to the maximality of $|C|$.

Hence $H_1 \times H_2 \times \dots \times H_n \subseteq \cup_{c \in C} B(c, d - 1)$ where $B(c, d - 1)$ is a ball of radius $d - 1$ with center $c \in C$.

That is,

$$H_1 \times H_2 \times \dots \times H_n \leq \sum_{c \in C} |B(c, d - 1)| = |C| |B(c, d - 1)|.$$

That is, $H_1 \times H_2 \times \dots \times H_n \leq M \cdot |B(c, d - 1)|$ and therefore

$$M \geq \frac{q_1 \cdot \dots \cdot q_n}{|B(c, d - 1)|}$$

where $|B(c, d - 1)|$ is obtained from 7.4. □

When all components are defined on the same alphabet, $q_i = q$ for all i and the above bound becomes,

$$M \geq \frac{q^n}{|B(c, d - 1)|} = \frac{q^n}{\sum_{j=0}^{d-1} \binom{n}{j} (q - 1)^j},$$

the usual Gilbert-Varshamov bound.

7.2.3 Hamming Bound

Theorem 100. [33] If C is a linear code over $\mathbb{F}_q$ of length n and minimum distance d having size $M = q^k$ where k is the dimension of C , then

$$M \leq \frac{q^n}{\sum_{i=0}^t \binom{n}{i} (q - 1)^i}$$

where $t = \lfloor \frac{d-1}{2} \rfloor$, called the *Hamming bound*. Note that $\sum_{i=0}^t \binom{n}{i} (q - 1)^i = |B(c, t)|$, the cardinality of the ball of radius t .

We can rewrite the Hamming bound in terms of a length n quasi-uniform code C of size M (not necessarily maximum) and minimum distance d defined over $H_1 \times H_2 \times \dots \times H_n$

where $|H_i| = q_i$. From 7.4, we get the cardinality $|B(c, t)|$. Therefore

$$M \leq \frac{q_1 \cdots q_n}{|B(c, t)|}.$$

The result came from the following facts: All spheres of radius t are disjoint and the total number of elements in a sphere of radius $t = |B(c, t)|$.

Since there are M codewords in C , the total number of elements in spheres around each word is

$$M \cdot |B(c, t)| \leq |H_1 \times \cdots \times H_n|.$$

7.2.4 Plotkin Bound

Theorem 101. [33] For any (n, M, d) binary code C with $2d > n$, where M is the size of the code (not the dimension), we have

$$M \leq \frac{2d}{2d - n}.$$

This bound is known as *Plotkin bound*. We generalize this bound to the quasi-uniform case.

Proposition 102. For any (n, M, d) code C defined over $q_1 \times \cdots \times q_n$ of length n , size M , minimum distance d having the size of the i th codeword symbol q_i , we have

$$M \leq \left\lfloor \frac{d}{\left[\frac{1}{q_1} + \cdots + \frac{1}{q_n}\right] - (n - d)} \right\rfloor$$

where $d + \left(\frac{1}{q_1} + \cdots + \frac{1}{q_n}\right) > n$.

Proof. We prove this result using similar tricks used in the binary case [33].

Consider the sum $S = \sum_{u \in C} \sum_{v \in C} d(u, v)$ where $d(u, v)$ is the hamming distance between u and v .

Since $d(u, v) \geq d$ for all $u \neq v \in C$, we have $S \geq M(M - 1)d$. (u and v counted twice.)

Now consider C as an $M \times n$ matrix with the i th column contains $q_i = |\{0, 1, \dots, q_i - 1\}|$ elements.

Let $x_{(i,j)} = \#j$'s in the i th column where $j = 0, 1, \dots, q_i - 1$ such that $M = \sum_{j=0}^{q_i-1} x_{(i,j)}$.

Then the i th column contributes $\sum_{j=0}^{q_i-1} x_{(i,j)}(M - x_{(i,j)})$ towards S .

That is

$$\begin{aligned}
 \sum_{j=0}^{q_i-1} x_{(i,j)}(M - x_{(i,j)}) &= x_{(i,0)}(M - x_{(i,0)}) + \dots + x_{(i,q_i-1)}(M - x_{(i,q_i-1)}) \\
 &= M \left(\sum_{j=0}^{q_i-1} x_{(i,j)} \right) - \sum_{j=0}^{q_i-1} x_{(i,j)}^2 \\
 &= M^2 - \sum_{j=0}^{q_i-1} x_{(i,j)}^2.
 \end{aligned}$$

The Cauchy-Schwartz inequality is given by

$$\left(\sum_{i=1}^n x_i y_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right)$$

where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$.

Now let $x = (x_{(i,0)}, \dots, x_{(i,q_i-1)})$ and $y = (1, \dots, 1) \in \mathbb{R}^{q_i}$. Then

$$\left(\sum_{j=0}^{q_i-1} x_{(i,j)} \right)^2 \leq \left(\sum_{j=0}^{q_i-1} x_{(i,j)}^2 \right) (q_i) \Rightarrow \frac{1}{q_i} \left(\sum_{j=0}^{q_i-1} x_{(i,j)} \right)^2 \leq \sum_{j=0}^{q_i-1} x_{(i,j)}^2.$$

Therefore

$$\sum_{j=0}^{q_i-1} x_{(i,j)}(M - x_{(i,j)}) \leq M^2 - \frac{1}{q_i} \left(\sum_{j=0}^{q_i-1} x_{(i,j)} \right)^2 = M^2 \left(1 - \frac{1}{q_i} \right).$$

Now the total sum $S \leq M^2(n - [\frac{1}{q_1} + \dots + \frac{1}{q_n}])$.

Writing the upper and lower bounds together, we have

$$M(M-1)d \leq S \leq M^2(n - [\frac{1}{q_1} + \dots + \frac{1}{q_n}]). \quad (7.5)$$

Therefore

$$\begin{aligned}
 1 - \frac{1}{M} &\leq \frac{(n - [\frac{1}{q_1} + \dots + \frac{1}{q_n}])}{d} \\
 \frac{d - n + [\frac{1}{q_1} + \dots + \frac{1}{q_n}]}{d} &\leq \frac{1}{M} \\
 M &\leq \frac{d}{(\frac{1}{q_1} + \dots + \frac{1}{q_n}) - (n-d)}
 \end{aligned}$$

where $d + (\frac{1}{q_1} + \dots + \frac{1}{q_n}) > n$.

Since $M \in \mathbb{N}$, we can rewrite the bound as

$$M \leq \left\lfloor \frac{d}{\left(\frac{1}{q_1} + \dots + \frac{1}{q_n}\right) - (n - d)} \right\rfloor$$

provided $d + \left(\frac{1}{q_1} + \dots + \frac{1}{q_n}\right) > n$.

□

In the case of quasi-uniform codes, each element in a column repeats the same number of times to preserve the quasi-uniformity.

That is, $x_{i,j} = M/q_i$, $\forall j = 0, \dots, q_i - 1$ and therefore

$$S = M^2 \left(n - \left[\frac{1}{q_1} + \dots + \frac{1}{q_n} \right] \right).$$

Then the above proof follows from Equation 7.5.

Note that each number in the set $\{0, 1, \dots, q_i - 1\}$ is associated to a coset of G_i in G where 0 represents the coset G_i .

In terms of group theory, we have

$$d = n - \max_{G_A \neq 1} |\mathcal{A}| \text{ and } q_i = [G : G_i] = \frac{|G|}{|G_i|}$$

if all G_i s are normal in G .

Therefore the bound can be rewritten as

$$M \leq \left\lfloor \frac{|G|d}{(|G_1| + \dots + |G_n|) - |G|(n - d)} \right\rfloor$$

provided the denominator is positive.

For example, consider the quasi-uniform code in Example 7.1, where $|G| = 9$, $|G_i| = 3$ for $i = 1, 2, 3, 4$ and $d = 3$. Then

$$\frac{|G|d}{(|G_1| + \dots + |G_4|) - |G|(n - d)} = 9 \times 3 / (12 - 9) = 9$$

which is exactly the size M of the code and the bound achieved.

7.2.4.1 q -ary Plotkin bound

When $q_i = q$ for all i , the above bound becomes the usual q -Plotkin bound.

That is

$$M \leq \left\lfloor \frac{d}{n/q - (n-d)} \right\rfloor = \left\lfloor \frac{qd}{qd - (q-1)n} \right\rfloor$$

where $qd > (q-1)n$.

7.2.5 Shortening

When a quasi-uniform code is shortened by at least one position preserving the minimum distance, we have the following observations about the size.

Suppose C is an (n, M, d) quasi-uniform code without any repetitions. Denote the size of the shortened code C_{n-1} by M_{n-1} .

Proposition 103. $M = q_i M_{n-1}$ for all i .

Proof. Suppose we shorten the given (n, M, d) code C at the i th symbol of alphabet size q_i keeping the minimum distance d .

That is, we take the all codewords with i th position zero and delete the entire column. The new code C_{n-1} has length $n-1$ and minimum distance d having size equal to the number of zeros in the i th position of C .

Since C is quasi-uniform, all the elements in the i th coordinate repeats M/q_i times and hence $M_{n-1} = M/q_i$. $\square$

Rewrite M and q_i in group terms, $M = |G|/|G_{\mathcal{N}}| = |G|$ since $G_{\mathcal{N}}$ assumed to be trivial and $q_i = |G|/|G_i|$.

Then we have $|G| = M_{n-1}|G|/|G_i| \Rightarrow M_{n-1} = |G_i|$ for all i .

In general, if the code is shortened by k positions then $M_{n-k} = |G_{i_1 \dots i_k}|$.

7.2.6 Litsyn-Laihonen Bound

Theorem 104. [2] Let C be a q -ary code of length n and minimum distance d of maximum size, say $M_q(n, d)$. Suppose that $0 \leq d \leq n$, $d - 2r \leq n - t$, $0 \leq r \leq t$ and $0 \leq r \leq d/2$. Then the *Litsyn-Laihonen* bound is given by

$$M_q(n, d) \leq \frac{q^t}{B_t(x, r)} M_q(n - t, d - 2r)$$

where $B_t(x, r)$ is the cardinality of the ball of radius r centered at x of length t .

We rewrite the above bound for a maximum size code C defined over the alphabet $q_1 \times \dots \times q_n$. Denote the size of the code C by $M_{q_1 \dots q_n}(n, d) = M$ and of the shortened code by M' . Then

Proposition 105.

$$M \leq \frac{q_{i_1} \dots q_{i_t}}{B_t(x, r)} M'(n - t, d - 2r)$$

Proof. We can follow the same proof of Theorem 104 except changing the alphabet.

Suppose C be the maximal quasi-uniform code of length n and minimum distance d defined over $q_1 \times \dots \times q_n$ having size M .

Choose t components from n and shorten the code in such a way that, taking the codewords formed by the t components which are in a ball of radius r centered at some $x \in q_{i_1} \times \dots \times q_{i_t}$ and delete those t components.

Then the new code C' is of length $n - t$ and minimum distance $d - 2r$ since the removed t components from the set of codewords are at most $2r$ distance apart.

Let $y_1, \dots, y_M \in q_{i_1} \times \dots \times q_{i_t}$.

Define an indicator function

$$\begin{aligned} \chi_x(y) &= 1 \text{ if } y \in B(x, r) \\ &= 0 \text{ otherwise.} \end{aligned}$$

Now the total number of y_i s in all spheres of radius r

$$\begin{aligned} \sum_{x \in q_{i_1} \times \dots \times q_{i_t}} \sum_{i=1}^M \chi_x(y_i) &= \sum_x \sum_{i=1}^M 1(y_i \in B(x, r)) \\ &= \sum_{i=1}^M \sum_x 1(y_i \in B(x, r)) \\ &= \sum_{i=1}^M \sum_x 1(x \in B(y_i, r)) \\ &= M \cdot B_t(y_i, r) \end{aligned}$$

The above argument holds since the quasi-uniform code is distance invariant.

Normalize the above sum by dividing with $q_{i_1} \dots q_{i_t}$. That is,

$$\frac{1}{q_{i_1} \dots q_{i_t}} \sum_x \sum_{i=1}^M \chi_x(y_i) = \frac{M}{q_{i_1} \dots q_{i_t}} B_t(y_i, r).$$

The above equation says that the average number of elements y_i in each ball of radius r is $\frac{M}{q_{i_1} \dots q_{i_t}} B_t(y_i, r)$.

This implies that the ball with maximum number of elements

$$M'(n - t, d - 2r) \geq \frac{M}{q_{i_1} \dots q_{i_t}} B_t(y_i, r)$$

where $B_t(y_i, r) = B_t(x, r)$ and the t components are $q_{i_1}, \dots, q_{i_t}$. □

The bound is not tight since we are unaware of the t components to choose from n as d increases.

Quasi-uniform codes were known to be constructed from groups. In this chapter, we were interested in relating the properties of the obtained code as a function of the corresponding group. We determined the size of the code, its alphabet, and its minimum distance, both for abelian groups, but also for some nonabelian groups where the group structure allows to mimic the definition of minimum distance. Also some of the usual bounds were translated into the language of quasi-uniform bounds.

Next chapter discusses some possible applications of quasi-uniform codes to distributed storage.

Applications of Quasi-Uniform Codes

We have seen the construction of quasi-uniform codes from groups and their basic properties in Chapter 7. In this chapter we examine some of the possible applications of quasi-uniform codes to storage and network coding. We will see the construction of almost affine codes via quasi-uniform codes towards the end.

Recall the definition of a code in a very general setting (Definition 43):

A *code* C of length n is an arbitrary nonempty subset of $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ where $\mathcal{X}_i$ is the alphabet for the i th codeword symbol, and each $\mathcal{X}_i$ might be different.

We are interested in quasi-uniform codes (Definition 45), whose codeword coefficients indeed live in possibly different alphabets. Quasi-uniform codes are defined with respect to an underlying probability distribution which is quasi-uniform [5], and quasi-uniform distributions (and in turn quasi-uniform codes [9]) can be constructed from finite groups and their subgroups as explained in Section 7.1[12]. When quasi-uniform codes come from finite groups, the underlying group structure can be exploited to derive code properties, such as the minimum distance [39] (Sections 7.1.1 and 7.1.2).

In Section 8.1, we present a family of quasi-uniform codes built from the dihedral 2-groups. In this case, codeword coefficients are either binary, or living in an alphabet of size a power of 2. Unlike the cases we have in Chapter 7, we consider nonnormal subgroups also for code constructions. We compute the code parameters, in particular its minimum distance (it is not exactly the same because of the inclusion of nonnormal subgroups), and study how the behavior of the minimum distance with respect to the coefficient alphabets.

We give concrete code instances in Section 8.1.1, where we discuss how such codes can actually be encoded. Properties of quasi-uniform codes coming from dihedral 2-groups include: (1) binary, or power of 2 alphabets, (2) good minimum distance which is easily

characterized by the subgroup structure of the group used, (3) control of how many coefficients are binary. These properties motivate us to consider their applications to storage, which is discussed in Section 8.2, with an explanation of the storage model that we assume.

A bound on the minimum distance is given for a class of p -groups, which takes into account the number of p -ary coefficients of the codeword is given in Section 8.3.

Applications of quasi-uniform codes in network coding and in the construction of almost affine codes are given in Sections 8.4 and 8.5 respectively.

Recall the code construction from Section 7.1 using finite groups and subgroups. In Section 7.1.2.4, we have introduced the code construction using subgroups of a finite group which are non-normal. We find examples and properties of such codes constructed and their applications here.

It is clear from Section 7.1.2.4 that if some subgroups used for code construction are non-normal, the code loses the group structure and the concept of minimum distance does not follow simply as in the case of normal subgroups. We investigate this scenario to see whether the concept of abelian group representability has any role to resolve this issue. Fortunately, it turns out that if we use abelian representable groups, we will get a corresponding code with same length, weight enumerator and more importantly having an abelian group structure.

Recall the following lemma on the minimum distance of a quasi-uniform code in terms of the subgroups $G_1, \dots, G_n$.

Lemma 106. [39] The minimum distance $\min_{c \in C} wt(c)$ of a quasi-uniform code C generated by a (possibly) nonabelian group G and its normal subgroups $G_i, i = 1, \dots, n$ is

$$n - \max_{\mathcal{A} \in \mathcal{N}, G_{\mathcal{A}} \neq \{0\}} |\mathcal{A}|,$$

where the weight $wt(c)$ is understood as the number of components which are not an identity element in the respective group where each coefficient takes values.

8.1 Quasi-Uniform Codes from Dihedral 2-Groups

Consider the dihedral group $D_{2^{k+1}} = \langle r, s : r^{2^k} = s^2 = 1, rs = sr^{-1} \rangle$ of order 2^{k+1} .

A quasi-uniform code C of length n built from $D_{2^{k+1}}$ consists of codewords of length n , where the i th coefficient takes values in the set of cosets of the subgroup $G_i, i = 1, \dots, n$.

Since not every subgroup of $D_{2^{k+1}}$ is normal, not every coefficient alphabet will end up having a group structure. However, since we are interested in the subgroup structure, and not in the non-abelian binary operation of this dihedral group, we may alternatively consider its abelian representation [37], that is an abelian group A with subgroups $A_1, \dots, A_n$ such that $|G|/|G_{\mathcal{A}}| = |A|/|A_{\mathcal{A}}|$ for every choice of $\mathcal{A}$ in $\mathcal{N}$.

The abelian group representation of $D_{2^{k+1}}$ [37] (refer Section 5.2.1.1) is explicitly given by

$$\begin{aligned} \psi : D_{2^{k+1}} &\rightarrow A = (\mathbb{Z}_2)^{k+1} \\ r^a s^j &\mapsto \sum_{i=0}^{k-1} a_i e_i + j e_k \end{aligned} \quad (8.1)$$

where a_i 's are obtained from the binary representation of $a = \sum_{i=0}^{k-1} a_i 2^i$, $j = 0, 1$ and $(e_0, \dots, e_k)$ is the standard basis of $(\mathbb{Z}_2)^{k+1}$.

The code $\psi(C)$ thus has an abelian group structure, but the length n and minimum distance (where the weight is understood as the number of coefficients not equal to G_i) of both codes are the same, since ψ preserves the group structure.

Code Length. To determine the largest value of n , we only need to count the number of non-trivial subgroups of $D_{2^{k+1}}$ (or $\psi(D_{2^{k+1}})$, whichever one finds most convenient).

Subgroups of $D_{2^{k+1}}$ are of the form [37]:

1. $\langle r^{2^i}, r^c s \rangle$ where $1 \leq i \leq k$, $0 \leq c < 2^i$ and
2. $\langle r^{2^{i-1}} \rangle$; $1 \leq i \leq k$.

Note that $|\langle r^{2^i}, r^c s \rangle| = |\langle r^{2^{i-1}} \rangle| = 2^{k-i+1}$, hence

$$[D_{2^{k+1}} : \langle r^{2^i}, r^c s \rangle] = [D_{2^{k+1}} : \langle r^{2^{i-1}} \rangle] = 2^i$$

which implies that there are $2^i + 1$ subgroups of index 2^i .

The total number of proper subgroups is then

$$\sum_{i=1}^k 2^i + k = 2[2^k - 1] + k = 2^{k+1} + k - 2.$$

The center of $D_{2^{k+1}}$ is $\langle r^{2^{k-1}} \rangle$, which is the intersection of all subgroups except subgroups of order two (or of index 2^k), and thus at least one subgroup of order 2 should be taken to build a code, to ensure that $G_{\mathcal{N}} = 1$.

FIGURE 8.1: The dihedral group D_8 and its lattice of subgroups.

Minimum Distance. Consider the quasi-uniform code C of maximum length formed by all non-trivial subgroups of $G = D_{2^{k+1}}$. Its minimum distance is by Lemma 106 and abelian group representability,

$$d = n - \max_{G_{\mathcal{A} \neq 1}} |\mathcal{A}|.$$

Since there are 2^k subgroups of order 2, and that they are the only ones which do not intersect the center, the number of subgroups intersecting the center is $(2^{k+1} + k - 2) - 2^k = 2^k + k - 2$. The minimum distance d is then

$$d = (2^{k+1} + k - 2) - (2^k + k - 2) = 2^k.$$

To summarize, the code parameters of C are

$$(n, |C|, d) = (2^{k+1} + k - 2, 2^{k+1}, 2^k), \quad (8.2)$$

with alphabet $(\mathbb{Z}_2)^k$ for $2^k + 1$ codeword coefficients.

Remark 11. Quasi-dihedral 2-groups and generalized quaternion groups of order 2^{k+1} also generate abelian representable codes coming from $A = (\mathbb{Z}_2)^{k+1}$ and its subgroups [37]. These two families of 2-groups thus do not provide new code constructions.

8.1.1 Quasi-Uniform Codes from D_8

Consider the dihedral group $D_8 = \langle r, s : r^4 = s^2 = 1, rs = sr^{-1} \rangle$, whose subgroup lattice diagram is shown in Figure 8.1.

Now the map $\psi : D_8 \rightarrow \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$ from (8.1) explicitly gives the abelian representation of D_8 (to simplify the notation, we write abc for $(a, b, c) \in \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$):

$$1 \mapsto 000, r \mapsto 100, r^2 \mapsto 010, r^3 \mapsto 110,$$

$$s \mapsto 001, rs \mapsto 101, r^2s \mapsto 011, r^3s \mapsto 111,$$

and the subgroups are mapped accordingly:

$$\begin{aligned} \langle r \rangle &\mapsto \langle 100, 010 \rangle = G_1, & \langle r^2, s \rangle &\mapsto \langle 010, 001 \rangle = G_2, \\ \langle r^2, rs \rangle &\mapsto \langle 010, 101 \rangle = G_3, & \langle r^3s \rangle &\mapsto \langle 111 \rangle = G_4, \\ \langle s \rangle &\mapsto \langle 001 \rangle = G_5, & \langle r^2 \rangle &\mapsto \langle 010 \rangle = G_6, \\ \langle r^2s \rangle &\mapsto \langle 011 \rangle = G_7, & \langle rs \rangle &\mapsto \langle 101 \rangle = G_8, \end{aligned}$$

resulting in a lattice of subgroups of $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$ which is identical to that of D_8 . We compute next the corresponding code.

Code Construction. Since G_1, G_2, G_3 have order 4, their corresponding quotient groups have order 2 and are isomorphic to $\mathbb{Z}_2$. The other proper subgroups have order 2, and thus the corresponding quotient groups are $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

Pick for example $g = 100 \in \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$. Then

$$g + G_1 = 100 + \langle 100, 010 \rangle = 100 + \{100, 010, 000, 110\} = G_1.$$

This is thus the identity element of the quotient group.

Next $g + G_6 = 100 + \langle 010 \rangle = 100 + \{010, 000\}$, which we map to 10 (the Klein group $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ has 3 elements of order 2, thus any choice will do).

What matters for the minimum distance is the identity element, for which there is no choice.

The code is shown in Table 8.1, where we have used that $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ contains an isomorphic copy of $\mathbb{Z}_2$, namely $\mathbb{Z}_2 \oplus \{0\}$, to emphasize binary coefficients.

From (8.2), which can be seen directly as well, its parameters are a length of $n = 8$, a cardinality $|C| = 8$ and a minimum distance of $d = 4$.

Encoding. The term “generator matrix” is probably risky in this context, since we do not have a vector space, however, we do have a $\mathbb{Z}_2$ -basis for the code, that is, a set of linearly independent vectors such that any binary linear combination generates the code. By defining

$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$	G_6	G_5	G_7	G_8	G_4	G_1	G_2	G_3
000	0	0	0	0	0	0	0	0
100	1	1	1	1	1	0	1	1
010	0	01	01	01	01	0	0	0
110	1	11	11	11	11	0	1	1
001	01	0	01	1	11	1	0	1
101	11	1	11	0	01	1	1	0
011	01	01	0	11	1	1	0	1
111	11	11	1	01	0	1	1	0

TABLE 8.1: A $(8, |C|, 3)$ code constructed from D_8 , $|C| = 8$. Pairs are elements in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

$a(b, c) = (ab, ac)$ for $a, b, c \in \mathbb{Z}_2$, we may encode the above codewords by

$$[u_1, u_2, u_3] \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 01 & 01 & 01 & 01 & 0 & 0 & 0 \\ 01 & 0 & 01 & 1 & 11 & 1 & 0 & 1 \end{bmatrix} \quad (8.3)$$

for $u_1, u_2, u_3 \in \mathbb{Z}_2$.

For example, to encode the vector $111 = (1, 1, 1) \in \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$, consider

$$\begin{aligned}
& [1, 1, 1] \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 01 & 01 & 01 & 01 & 0 & 0 & 0 \\ 01 & 0 & 01 & 1 & 11 & 1 & 0 & 1 \end{bmatrix} \\
&= (10 + 00 + 01, 10 + 01 + 00, 10 + 01 + 01, \\
&\quad 10 + 01 + 10, 10 + 01 + 11, 0 + 0 + 1, 1 + 0 + 0, 1 + 0 + 1) \\
&= (11, 11, 10, 01, 00, 1, 1, 0) \\
&= (11, 11, 1, 01, 0, 1, 1, 0),
\end{aligned}$$

which is the case from Table 8.1.

8.2 Storage Applications

Suppose we have n nodes across which data is stored using an erasure code C of length n . Then every coefficient of a codeword of C is stored at every node. In case of node failure(s), we should be able to retrieve the stored data. Whether this is possible depends on the number of failures and on the minimum distance of the code. We thus start by comparing the minimum distance of the proposed codes to known codes.

$(n, k + 1)$	d over $\mathbb{F}_2$	d over $\mathbb{F}_4$	New d
$(8, 3)$	4	5	4
$(7, 3)$	4	4	4
$(6, 3)$	4	4	4
$(5, 3)$	4	3	4
$(17, 4)$	8	12	8
$(16, 4)$	8	11	8
$(15, 4)$	8	10	8
$(14, 4)$	7	9	8
$(13, 4)$	6	8	8
$(12, 4)$	6	7	8
$(11, 4)$	5	6	8
$(10, 4)$	4	6	8
$(9, 4)$	4	5	8

TABLE 8.2: Minimum distance comparison with known codes [29].

8.2.1 Code Comparisons

We compare in Table 8.2 the codes obtained from dihedral groups with binary codes [29], more precisely, if a dihedral code has length n and cardinality 2^{k+1} , we compare it with a binary $(n, k + 1, d)$ code, where $k + 1$ is the dimension of the code.

- When $k = 2$, the code from D_8 has the same minimum distance as that of a known binary code of length 8 and dimension 3. When this code is punctured at the components corresponding to G_1, G_2, G_3 respectively, the minimum distance stays $d = 4$, but so is the minimum distance of the above binary code when punctured.
- When $k = 3$, the code from D_{16} also has the same minimum distance compared to a binary code of length 16 and dimension 4. However the behavior improves after puncturing at least three of the components corresponding to subgroups of order 8.

Since the finite field $\mathbb{F}_4$ can be seen as a vector space over $\mathbb{F}_2$ by fixing a $\mathbb{F}_2$ -basis (we write $\mathbb{F}_2$ to emphasize the field structure, and $\mathbb{Z}_2$ to emphasize the additive group one), we could see the coefficients of the code in $\mathbb{F}_4$. We also put parameters of codes over $\mathbb{F}_4$ in Table 8.2, however a code over $\mathbb{F}_4$ contains 4^{k+1} , so for a fair comparison, we should really compute the minimum distance of the proposed codes as codes over $\mathbb{F}_4$ (which we skip here, since we are interested in the minimum distance as predicted by the subgroup structure).

8.2.2 A Storage Example

We illustrate how the codeword (8.3) can be used over $n = 8$ nodes. Suppose a data object (u_1, u_2, u_3) needs to be stored. Consider the following storage allocation [40]:

$$\begin{array}{ll} \text{node 1: } (u_1, u_3) & \text{node 5: } (u_1 + u_3, u_2 + u_3) \\ \text{node 2: } (u_1, u_2) & \text{node 6: } (u_3, 0) \\ \text{node 3: } (u_1, u_2 + u_3) & \text{node 7: } (u_1, 0) \\ \text{node 4: } (u_1 + u_3, u_2) & \text{node 8: } (u_1 + u_3, 0) \end{array}$$

We easily notice that three failures at most can be tolerated (corresponding to a minimum distance of 4 indeed): u_2 is present only in nodes 2,3,4 and 5, so if these four nodes fail, u_2 is lost.

In case of one node failure, this node is repaired easily: indeed, this codeword is created from $\mathbb{Z}_2$ -linear combinations. For example:

- node 1 is repaired by downloading u_1 (from node 2 or node 7) and u_3 (from node 6),
- node 2 is repaired by downloading u_2 from node 4 and u_1 (from node 1 or 7).

Apart from repairs obtained using $\mathbb{Z}_2$ -linear combination, a distinctive feature of this code is the fact that we control how many coefficients live in $\mathbb{Z}_2$, and how many live in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

8.3 Bounds on the Minimum Distance

It is known that the minimum distance of a linear code over $\mathbb{F}_q$ depends on the choice of q . The same phenomenon appears here: the minimum distance depends on the different alphabets in which live the codeword coefficients.

In particular, the more coefficients we want small, the more subgroups of high index we need, which reduces the minimum distance.

In this section, we will give a bound on the minimum distance of a quasi-uniform code, when it is obtained from a p -group as described above from a dihedral 2-group (that is using its abelian group representation A).

Definition 47. The Frattini subgroup $\phi(G)$ of a group G is the intersection of all maximal subgroups of G .

In particular, $\phi(G) = G$, if there are no maximal subgroups.

It is also known that $\phi(G)$ is normal in G . [15]

To prove the main result of this section, we need the following lemmas.

Lemma 107. [22] Every maximal subgroup of a finite p -group is normal with index p .

Lemma 108. Let $\phi(G)$ be the Frattini subgroup of a finite p -group G . Then $G/\phi(G)$ is elementary abelian.

Proof. If M is maximal in G , then G/M is of order p by Lemma 107, and G/M is abelian. The commutator subgroup G' is then a subgroup of M for all maximal subgroups $M \leq G$ and then $G' \leq \phi(G)$ implies that $G/\phi(G)$ is abelian [15].

Since $|G/M| = p$, $(gM)^p = M$ for all $g \in G$, that is, $g^p M = M$ for all $g \in G$ and for all maximal subgroups M and hence $g^p \in G/\phi(G)$ and $g^p \phi(G) = \phi(G)$.

Therefore, if $g\phi(G) \in G/\phi(G)$, its order $|g\phi(G)| = p$ for all $g \in G$, and $G/\phi(G)$ is an abelian group where each non-trivial element has order p . $\square$

Now we restate Lemma 94 which counts the subgroups of order p of an elementary abelian p -group where an elementary abelian group is of the form $(\mathbb{Z}_p)^m$, by the classification of finitely generated abelian groups [15].

Lemma 109. Let G be an elementary abelian p -group of order p^m . Then there are $\frac{p^m-1}{p-1}$ subgroups of order p .

From Lemma 108 and Lemma 109, it is clear that there are $\frac{p^m-1}{p-1}$ subgroups of $G/\phi(G)$ of order p .

Proposition 110. The number of maximal subgroups of a p -group having Frattini subgroup of index p^m is $\frac{p^m-1}{p-1}$.

Proof. Let G be a group of order p^k and suppose $[G : \phi(G)] = p^m$.

By Lemma 109, the number of subgroups of order p of the quotient group $G/\phi(G)$ is $\frac{p^m-1}{p-1}$.

Now by the lattice isomorphism theorem 34, the subgroups of G containing $\phi(G)$ and subgroups of $G/\phi(G)$ are in 1-1 correspondence.

Let M be a maximal subgroup of G of index p (Lemma 107). Then $|M| = p^{k-1}$.

Consider the subgroup, $M/\phi(G) \leq G/\phi(G)$. Since $|G/\phi(G)| = p^m$ and $|M/\phi(G)| = p^{k-1-(k-m)} = p^{m-1}$ and therefore $[G/\phi(G) : M/\phi(G)] = p$.

From [15] and [30], the number of index p -subgroups and the number of order p -subgroups of an elementary abelian p -group are the same, $\frac{p^m-1}{p-1}$ in total. $\square$

Let G be a p -group which is abelian group representable. Recall that it means that we can find an abelian group A and subgroups A_i such that $[G : G_{\mathcal{A}}] = [A : A_{\mathcal{A}}]$ for all $\mathcal{A} \subseteq \mathcal{N}$. Note that A may have more subgroups than G .

Suppose that the Frattini subgroup $\phi(G)$ of G is non-trivial and of index p^m .

Because A may contain more subgroups than G , $|\phi(G)|$ and $|\phi(A)|$ may not be equal. For example, $|\phi(D_8)| = 2 \neq 1 = |\phi(\mathbb{Z}_2^3)|$.

By Proposition 110, there are $\frac{p^m-1}{p-1}$ maximal subgroups of G of index p .

Suppose $\psi : G \rightarrow A$ is a mapping which preserves the subgroup structure of G , used to prove that A is an abelian group representation of G . Under this mapping, the subgroup lattice of G is embedded into that of A and $[G : \phi(G)] = [A : \psi(\phi(G))] = p^m$.

Suppose we construct a quasi-uniform code C of length n from G using its abelian representation A and assume that all subgroups corresponding to the maximal subgroups of G are included. Then the minimum distance of C is bounded as follows:

Proposition 111. Let G be a p -group which is abelian representable and suppose G has a non-trivial Frattini subgroup, $\phi(G)$ of index p^m . Then the quasi-uniform code obtained from the abelian representation of G has a minimum distance d such that:

$$1 \leq d \leq n - \frac{p^m - 1}{p - 1},$$

where n is the length of the code.

Proof. Since no codeword is repeated, $G_{\mathcal{N}} = 1$ implies $n > \max_{G_{\mathcal{A}} \neq 1} |\mathcal{A}|$ so that

$$d = n - \max_{G_{\mathcal{A}} \neq 1} |\mathcal{A}| \geq 1. \quad (8.4)$$

From Proposition 110, $\max_{G_{\mathcal{A}} \neq 1} |\mathcal{A}| \geq \frac{p^m-1}{p-1}$ since all maximal subgroups are included and hence

$$d \leq n - \frac{p^m - 1}{p - 1}. \quad (8.5)$$

From (8.4) and (8.5),

$$1 \leq d \leq n - \frac{p^m - 1}{p - 1}.$$

$\square$

The above bound is tight. To see this, consider the $(7, |C|, 4)$ code C , obtained by puncturing the first column of the $(8, |C|, 4)$ code proposed in Section 8.1.1. That is, the puncturing is done by removing a subgroup which is not maximal. Its minimum distance is $d = 4$.

Since $[D_8 : \phi(D_8)] = 4$ and $n - (2^2 - 1) = n - 3 = 4 = d$, equality holds.

Corollary 112. Consider the quasi-uniform code construction of the above proposition, but where we do not include l maximal subgroups out of $\frac{p^m-1}{p-1}$ in the code construction. Then $\max_{G_{\mathcal{A}} \neq 1} |\mathcal{A}| \geq \frac{p^m-1}{p-1} - l$ and the bound on the minimum distance becomes:

$$1 \leq d \leq n + l - \frac{p^m - 1}{p - 1}.$$

It is worth repeating that the number of maximal subgroups also decide the number of components of the codewords that will live in $\mathbb{Z}_p$, thus this bound gives a trade-off between coefficients in $\mathbb{Z}_p$ and minimum distance.

Corollary 113. Let C be a quasi-uniform code of length n formed from the dihedral 2-group $D_{2^{k+1}}$. Suppose t subgroups of order 2, distinct from the center, are included. Then the minimum distance d of the code is $d = t \leq 2^k$, that is, it only depends on the number of subgroups of order 2.

Proof. We know that the center $\langle r^{2^{k-1}} \rangle$ is of order 2 and is contained in all but subgroups of order 2.

Then $\max_{G_{\mathcal{A}} \neq 1} |\mathcal{A}| = n - t$ implies $d = n - (n - t) = t$.

Since there are 2^k subgroups of order two except the center, $t \leq 2^k$. □

8.4 Quasi-Uniform Codes in Network Coding

Quasi-uniform codes have applications in network coding. Dougherty et al in [13] showed that linear codes do not achieve the capacity in general network coding problems. Therefore it is necessary to have non-linear codes or non-linear network coding strategies to increase the maximum throughput. In [8], it is shown that quasi-uniform codes are sufficient to achieve the maximal throughput.

The importance of quasi-uniform codes is also shown in [10], where the network considered is an incremental multi-cast network. We rewrite that particular problem in terms of quasi-uniform codes obtained from groups.

Consider a network $\mathcal{G}$ where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an acyclic directed graph where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ is the set of directed edges.

We assume that $\mathcal{G}$ has exactly one source node denoted by s which does not have any incoming edges. Suppose s aims to transmit two messages M_1 and M_2 to a group of sink nodes. Since we consider the incremental multi-cast problem, there are two classes of sink nodes D_1 and D_2 such that sink nodes in the set D_1 request only message M_1 where as D_2 requires both M_1 and M_2 . A network coding problem is hence specified by the triplet $(\mathcal{G}, D_1, D_2)$ [10].

An application of incremental multi-cast is for distribution of multimedia contents, which are usually encoded into multiple frames or layers. The quality (or the level of distortion) of the reconstructed signal depends on the subset of frames a receiver receives. Moreover, a received frame improves the signal quality only if all other frames in the lower layers are also received [10].

Define $in(e)$ as the set of incoming edges to $e \in \mathcal{E}$, where $in(e) = s$ if e is starting from the source and $in(u)$ as the set of incoming edges to the node $u \in \mathcal{V}$.

Definition 48. A quasi-uniform network code [8, 10] C (with respect to the network coding problem $(\mathcal{G}, D_1, D_2)$) is a quasi-uniform vector $(X_1, \dots, X_n)$ which satisfies the following constraints:

- 1) For any $e \in \mathcal{E}$, $H(X_e|X_i; i \in in(e)) = 0$.
- 2) $H(X_1|X_e, e \in in(u)) = 0$ if $u \in D_1$ and $H(X_1, X_2|X_e, e \in in(u)) = 0$ if $u \in D_2$, where the random variables X_1, X_2 corresponding to messages M_1 and M_2 are $X_3, \dots, X_n$ correspond to all edges in $\mathcal{E}$.

A fundamental question in network coding is thus to determine what the maximum achievable throughput and what the optimal network codes are. Specifically, we are interested in the following optimization problem in the above situation [9]:

Maximize $H(X_1, X_2)$ subject to

$$\begin{aligned}
 (X_1, \dots, X_n) & \quad \text{is a quasi-uniform code} & (8.6) \\
 H(X_e|X_i; i \in in(e)) & = 0 \quad \forall e \in \mathcal{E} \\
 H(X_1|X_e, e \in in(u)) & = 0 \quad \text{for } u \in D_1 \\
 H(X_1, X_2|X_e, e \in in(u)) & = 0 \quad \text{for } u \in D_2 \\
 H(X_2|X_1) & \geq \alpha.
 \end{aligned}$$

It is also mentioned in [10] that it is extremely hard to solve this problem. However, this problem might be easier if assume that the quasi-uniform code $(X_1, \dots, X_n)$ are generated

by a finite group G and subgroups $G_1, \dots, G_n$. Then $H(X_1, X_2) = \log \frac{|G|}{|G_{12}|}$ where $G_{12} = G_1 \cap G_2$.

Now

$$H(X_e|X_i) = H(X_e, X_i) - H(X_i) = \log \frac{|G|}{|G_{ie}|} - \log \frac{|G|}{|G_i|} = \log \frac{|G_i|}{|G_{ie}|}.$$

Therefore $H(X_e|X_i; i \in \text{in}(e)) = 0 \forall e \in \mathcal{E}$ becomes

$$\log \frac{|G_i|}{|G_{ie}|} = 0 \iff |G_i| = |G_{ie}| \iff G_i \leq G_e; i \in \text{in}(e) \forall e \in \mathcal{E}.$$

Similarly, $H(X_1|X_e, e \in \text{in}(u)) = 0$ for $u \in D_1$ becomes

$$\log \frac{|G_e|}{|G_{1e}|} = 0 \iff |G_e| = |G_{1e}| \iff G_e \leq G_1; e \in \text{in}(u), \text{ for } u \in D_1,$$

and $H(X_1, X_2|X_e, e \in \text{in}(u)) = 0$ for $u \in D_2$ becomes

$$\log \frac{|G_e|}{|G_{12e}|} = 0 \iff |G_e| = |G_{12e}| \iff G_e \leq G_1, G_2; e \in \text{in}(u), \text{ for } u \in D_2.$$

Finally, $H(X_2|X_1) \geq \alpha \iff \log \frac{|G_1|}{|G_{12}|} \geq \alpha$.

That is, we can rewrite (8.6) as:

Maximize $\frac{|G|}{|G_{12}|}$ subject to

$$\begin{aligned} (X_1, \dots, X_n) \quad & \text{is a quasi-uniform code defined over} & (8.7) \\ & \text{the cosets of } G_i : i = 1, \dots, n \\ G_i \leq & G_e; i \in \text{in}(e) \forall e \in \mathcal{E} \\ G_e \leq & G_1; e \in \text{in}(u), \text{ for } u \in D_1 \\ G_e \leq & G_1, G_2; e \in \text{in}(u), \text{ for } u \in D_2 \\ \log \frac{|G_1|}{|G_{12}|} \geq & \alpha. \end{aligned}$$

Now the problem reduces to finding group G with maximal cardinality which satisfies the above subgroup conditions.

8.5 Almost Affine Codes from Groups

Consider a non-empty subset $\mathcal{A}$ of $\mathcal{N}$ and let $C_{\mathcal{A}}$ be the projection of the code C into the coordinate space $\mathcal{A}$, that is, all the words of C are restricted to the positions in $\mathcal{A}$.

Definition 49. [35] A q -ary code C of length n is said to be *almost affine* if it satisfies the condition

$$\log_q(|C_{\mathcal{A}}|) \in \mathbb{N}, \text{ for all } \mathcal{A} \subseteq \mathcal{N}.$$

Almost affine codes were introduced in [35] as a generalization of affine codes, which are themselves generalizations of linear codes over finite fields. It was shown in [9] that almost affine codes are quasi-uniform. It is thus natural to look for such codes among codes built from groups. Because of the definition of almost affine codes, p -groups are the first candidates that come to mind.

Lemma 114. Let G be a p -group, and let $G_1, \dots, G_n$ be subgroups of index p . The corresponding quasi-uniform code is almost affine.

Proof. First note that $G_1, \dots, G_n$ are normal subgroups of G , thus so is their intersection, and wlog we may assume that $|G_{\mathcal{N}}| = 1$. Then Proposition 80 holds (though the H_i might be nonabelian), and we obtain a p -ary quasi-uniform code. Since any intersection $G_{\mathcal{A}}$ will have order a power of p , the code obtained is almost affine. $\square$

There are other ways to get almost affine codes from p -groups, and p -groups are not the only finite groups that can provide almost affine codes.

Future Works

This thesis opens questions in many directions. They can be summarized as follows:

- It is proven that nilpotent groups are abelian group representable if and only if p -groups are. Also non-nilpotent groups are not abelian group representable. Hence in order to complete the problem of abelian group representability, we have to check whether all nilpotent groups or which of the different classes of p -groups are abelian group representable.
- We obtained the smallest violations of only three of the 24 DFZ inequalities which characterize the linear rank inequalities in dimension 5. Finding more violations requires deep understanding of the connections between the inequalities and the subgroup structure of the group, which is an open problem. Also it may be possible to extend the smallest violators to a class of violators.
- The next problem is to analyze the impact of these violators and the entropic vectors that they generate in determining the region of entropic vectors in the corresponding dimension.
- Construction of quasi-uniform codes from groups and their algebraic properties are studied. Also we obtained some classical bounds for these codes. The next step is to compare these bounds with other codes to understand their performance.
- The components of quasi-uniform codes belong to different alphabets and most of the bounds we obtained depend on the ball of radius r of these codes. So learning

more properties of these balls may lead to a refinement of the bounds and to understand the encoding and decoding procedures of quasi-uniform codes.

- Some applications of quasi-uniform codes to storage are given. It may be possible to use quasi-uniform codes to construct non-linear storage codes or linear storage codes over different alphabets.
- The initial motivation of this work was to construct non-linear codes that attain capacity bounds in networks where linear codes are insufficient. Algebraic quasi-uniform codes are potential candidates for this which is yet to be studied.

Normal Subgroups of Dihedral Groups

A dihedral group of order $2m$ is defined as

$$D_{2m} = \langle r, s \rangle : r^m = s^2 = 1, rs = sr^{-1} \}.$$

To compute the normal subgroups of D_{2m} we have to identify the conjugacy classes of it, where a conjugacy class is defined as follows:

Definition 50. For an element $g \in G$, the conjugacy class of x is defined as

$$Cl(x) = \{gxg^{-1} : g \in G\}.$$

That is, $Cl(g)$ contains all elements of G which are conjugate to g .

Now consider the set $gHg^{-1} = \{ghg^{-1} : h \in H\}$ where $H \leq G$.

If $gHg^{-1} = H \forall g \in G$, we say that H is normal in G . That is possible if and only if the conjugacy classes of all elements in H is H itself.

Also note that different conjugacy classes are disjoint and the size of a conjugacy class divides $|G|$ [15].

Therefore, we identify different conjugacy classes of D_{2m} in order to compute its normal subgroups.

A.1 Conjugacy Classes of D_{2m}

Case 1: m is odd

All elements of D_{2m} are either of the form $r^j, r^j s$ for $1 \leq j \leq m$.

Consider the conjugacy classes of elements of the form r^j : we have $r^i(r^j)r^{-i} = r^j \forall i$ and $r^i s(r^j)(r^i s)^{-1} = r^i s(r^j)r^i s = (r^i s)(r^i s)r^{-j} = r^{-j}, \forall i$ where $(r^i s)^{-1} = r^i s$. Therefore,

$$Cl(r^j) = \{r^j, r^{-j}\}$$

and

$$Cl(r^m) = Cl(1) = \{1\}.$$

Since m is odd, there are $\frac{m-1}{2}$ different conjugacy classes of the form $\{r^j, r^{-j}\}$.

Consider the conjugacy class of s : $r^i s r^{-i} = r^{2i} s$ and $r^i s(s)(r^i s)^{-1} = r^{2i} s \forall i$.

Since m is odd, all integers modulo m are multiples of 2 and hence

$$\{r^{2i} s : 1 \leq i \leq m\} = \{r^i s : 1 \leq i \leq m\}.$$

Therefore

$$Cl(s) = \{r^i s : 1 \leq i \leq m\} = \{\text{All reflections}\}$$

and there are m elements in this set.

Then the different conjugacy classes of D_{2m}, m is odd are

1. 1 conjugacy class of order m : $\{r^i s : 1 \leq i \leq m\}$.
2. $\frac{m-1}{2}$ conjugacy classes of order 2 of the form $\{r^i, r^{-i}\}$.
3. $Cl(1) = \{1\}$.

Note that the total number of elements in all conjugacy classes is $2m$.

Case 2: m is even

The conjugacy class of 1, $Cl(1) = \{1\}$.

$Cl(r^{m/2})$ is computed as : $r^i r^{m/2} r^{-i} = r^{m/2}$ and $r^i s r^{m/2} (r^i s)^{-1} = r^{-m/2} = r^{m/2}$ implies $Cl(r^{m/2}) = \{r^{m/2}\}$.

Similarly as in Case 1, $Cl(r^j) = \{r^j, r^{-j}\}$ for $1 \leq j \leq \frac{m}{2} - 1$ since m is even. Also note that there are $m/2 - 1$ conjugacy classes of such form.

The conjugacy class of s is given similarly as before:

$$Cl(s) = \{r^{2i}s : 1 \leq i \leq m\}$$

and its order is $m/2$ since m is even.

The conjugacy class of rs :

$$Cl(rs) = \{r^{2i+1}s : 1 \leq i \leq m\}$$

which is of order $m/2$.

Therefore the different conjugacy classes of D_{2m} , m is even are

1. 2 conjugacy class of order $m/2$: $\{r^{2i}s : 1 \leq i \leq m\}$ and $\{r^{2i+1}s : 1 \leq i \leq m\}$.
2. $\frac{m}{2} - 1$ conjugacy classes of order 2 of the form $\{r^i, r^{-i}\}$.
3. $Cl(1) = \{1\}$ and $Cl(r^{m/2}) = \{r^{m/2}\}$.

So the total number of elements is $2 + 2(\frac{m}{2} - 1) + (\frac{m}{2})2 = 2m$.

A.2 Normal Subgroups

All subgroups of D_{2m} are either of the form:

- 1) $\langle r^d \rangle : d|m$
- 2) $\langle r^d, r^i s \rangle : d|m, 0 \leq i \leq d - 1$ (refer Section 5.2.1).

Now we check which of the above subgroups are normal in D_{2m} .

Proposition 115. Subgroups of the form $\langle r^d \rangle$, $d|m$ are normal in D_{2m} .

Proof. All elements of $\langle r^d \rangle$ are of the form r^{kd} for some integer k .

Now $Cl(r^{kd}) = \{r^{kd}, r^{-kd}\}$ for all k , from the previous section.

Since $\langle r^d \rangle$ is a subgroup $r^{-kd} \in \langle r^d \rangle$ for element $r^{kd} \in \langle r^d \rangle$ and therefore

$$Cl(\langle r^d \rangle) = \langle r^d \rangle$$

and hence it is normal. $\square$

Proposition 116. When m is odd, no subgroup of the form $\langle r^d, r^i s \rangle$, $d|m, 0 \leq i \leq d-1$ is proper normal in D_{2m} .

Proof. If $d = 1$, then $\langle r^d, r^i s \rangle = \langle r, s \rangle = D_{2m}$, not proper.

If $d = m$, then $\langle r^d, r^i s \rangle = \langle r^i s \rangle = \{r^i s, 1\}$, which is not normal since $Cl(r^i s)$ is the set of all reflections.

Suppose $d \neq 1$ and $d \neq m$. Consider $\langle r^d, r^i s \rangle$. The conjugacy class $Cl(r^i s)$ is the set of reflections and $Cl(r^i s) \subset \langle r^d, r^i s \rangle$ if and only if $d = 1$, implies that $Cl(\langle r^d, r^i s \rangle) \neq \langle r^d, r^i s \rangle$, it is not normal. $\square$

Proposition 117. When m is even, $\langle r^2, rs \rangle$ and $\langle r^2, s \rangle$ are the only normal subgroups of the form $\langle r^d, r^i s \rangle$.

Proof. Consider the subgroup $\langle r^2, s \rangle$. The conjugacy class of s , $Cl(s) = \{r^{2i} s : 1 \leq i \leq m\}$, the set of all reflections with even power of r and the conjugacy class of r^j , $Cl(r^j) = \{r^{\pm j}\}$.

Now the elements of $\langle r^2, s \rangle$ are of the form $r^{2i} s, r^{2i}, 1 \leq i \leq m$. Therefore

$$Cl(\langle r^2, s \rangle) = \langle r^2, s \rangle,$$

implies it is normal.

Consider the subgroup $\langle r^2, rs \rangle$. Its elements are of the form $r^{2i}, r^{2i+1} s$ for all i . Since the conjugacy class, $Cl(r^j) = \{r^{\pm j}\}$ and $Cl(rs) = \{r^{2i+1} s : 1 \leq i \leq m\}$ from the previous section, we have

$$Cl(\langle r^2, rs \rangle) = \langle r^2, rs \rangle,$$

proves the normality of $\langle r^2, rs \rangle$.

Now suppose that $d \neq 2$, $d|m$ and $d \geq 4$ (since m is even). All elements of the subgroup $\langle r^d, r^i s \rangle$, $d|m, 0 \leq i \leq d-1$ are of the form r^{kd} and r^{kd+i} for some integer k .

Consider the conjugacy class of r^{kd+i} :

$$Cl(r^{kd+i}) = r^{2i} s, \text{ when } kd + i \text{ is even}$$

and

$$Cl(r^{kd+i}) = r^{2i+1} s, \text{ when } kd + i \text{ is odd}$$

Since $d \geq 4$, the subgroup $\langle r^d, r^i s \rangle$ does not contain all the elements of either of the above two conjugacy classes and hence it is not normal. $\square$

To summarize, the normal subgroups of D_{2m} are:

- $\langle r^d \rangle$ for all $d|m$; when m is odd.
- $\langle r^2, s \rangle$, $\langle r^2, rs \rangle$ and $\langle r^d \rangle$ for all $d|m$; when m is even.

Bibliography

- [1] J. Bali and B.S. Rajan. Block-coded psk modulation using two-level group codes over dihedral groups. *IEEE Transactions on Information Theory*, 44(4), 1998.
- [2] E. Bellini, E. Gerrini, and M. Sala. On capacity regions of non-multicast networks. *IEEE Transactions on Information Theory*, 60(3):1475--1480, February 2014.
- [3] N. Boston and T.T. Nan. Large violations of theingleton inequality. *Allerton '12 Proceedings of the 47th annual Allerton conference on communication, control and computing*, 2012.
- [4] N. Boston and T.T. Nan. A refinement of the four-atom conjecture. *International Symposium on Network Coding(NetCod)*, 2013.
- [5] T. H. Chan. A combinatorial approach to information inequalities. *Communications in Information and Systems*, 1(3):1--14, September 2001.
- [6] T. H. Chan. Group characterizable entropy functions. *IEEE International Symposium on Information Theory (ISIT), Nice, France*, 24-29 June 2007.
- [7] T. H. Chan and A. Grant. On capacity regions of non-multicast networks. *IEEE International Symposium on Information Theory (ISIT)*, 2010.
- [8] T. H. Chan and A. Grant. Network coding capacity regions via entropy functions. *IEEE Transactions on Information Theory*, 2012.
- [9] T. H. Chan, A. Grant, and T. Britz. Properties of quasi-uniform codes. *IEEE International Symposium on Information Theory (ISIT)*, June 2010.
- [10] T. H. Chan, A. Grant, and T. Britz. Quasi-uniform codes and their applications. *IEEE Transactions on Information Theory*, 59(12), 2013.
- [11] T. H. Chan and R. W. Yeung. On a relation between information inequalities and group theory. *IEEE Transactions on Information Theory*, 48:1992--1995, July 2002.

- [12] T.H. Chan. Aspects of information inequalities and its applications. *M.Phil Thesis, Dept. of Information Engineering, The Chinese University of Hong Kong*, September 1998.
- [13] R. Dougherty, C. Freiling, and K. Zeger. Insufficiency of linear network coding in network information flow. *IEEE Transactions on Information Theory*, pages 2745--2759, 2005.
- [14] R. Dougherty, C. Freiling, and K. Zeger. Linear rank inequalities on five or more variables. <http://arxiv.org/pdf/0910.0284v3.pdf>, 2010.
- [15] D. S. Dummit and R. M. Foote. Abstract algebra (3rd edition). *John Wiley and Sons*, 2004.
- [16] G. D. Forney. Geometrically uniform codes. *IEEE Transactions on Information Theory*, 35(5), 1991.
- [17] R. Garelo and S. Benedetto. Multilevel construction of block and trellis group codes. *IEEE Transactions on Information Theory*, 41(5), 1995.
- [18] L. Guillé, T.H. Chan, and A. Grant. The minimal set of ingleton inequalities. *IEEE Transactions on Information Theory*, 57(4):1849--1864, 2011.
- [19] D. Hammer, A. Romashchenko, A. Shen, and N. Vereshchagin. Inequalities for shannon entropy and kolmogorov complexity. *Journal of Computer and System Sciences*, 60, June 2000.
- [20] B. Hassibi and S. Shadbakht. Normalized entropy vectors, network information theory and convex optimization. *Information Theory Workshop (ITW)*, July 2007.
- [21] B. Hassibi and S. Shadbakht. Normalized entropy vectors, network information theory and convex optimization. *Proceedings of the 2007 IEEE Information Theory Workshop*, July 2007.
- [22] K. Igusa. p-groups. *Lecture Notes*, <http://people.brandeis.edu/~igusa/Math101b/pGroups.pdf>.
- [23] A. Ingleton. Representation of matroids. *Combinatorial Mathematics and its Applications*, 149--167:18, 1971.
- [24] H. Li and E. K. P. Chong. On connections between group homomorphisms and the ingleton inequality. *International Symposium on Information Theory (ISIT)*, June 2007.
- [25] H. A. Loeliger. Signal sets matched to groups. *IEEE Transactions on Information Theory*, 37(6), 1991.

- [26] W. Mao and B. Hassibi. Violating theingleton inequality with finite groups. *Allerton'09 Proceedings of the 47th annual Allerton conference on communication,control and computing*, pages 1053--1060, 2009.
- [27] N. Markin, E. Thomas, and F. Oggier. Groups and information inequalities in 5 variables. *Fifty-first Annual Allerton Conference*, October 2013.
- [28] F. Matúš. Conditional independences among four random variables i. *Combinatorics, Probability and Computing*, 4:269--278, 1995.
- [29] M.Grassl. Bounds on the minimum distance of linear codes and quantum codes. www.codetables.de, 24 April 2014.
- [30] Properties of the image and kernel of the p -power map on a finite abelian group. <http://crazyproject.wordpress.com/2010/06/16/properties-of-the-image-and-kernel-of-the-p-power-map-on-a-finite-abelian-group/>. 2010.
- [31] S.B. Pai and B.S. Rajan. A lattice singleton bound. *International Symposium on Information Theory (ISIT)*, 2013.
- [32] E.J. Rossin, N.T. Sindhushayana, and C.D. Heegard. Trellis group codes for the gaussian channel. *IEEE Transactions on Information Theory*, 41(5), 1995.
- [33] L. San and X. Chaoping. Coding theory: A first course. *Cambridge University Press*, 2004.
- [34] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379--423, 623--656, July, October 1948.
- [35] J. Simonis and A. Ashikhmin. Almost affine codes. *Designs, Codes and Cryptography*, 14(2), 1998.
- [36] R. Stancu and F. Oggier. Finite nilpotent and metacyclic groups never violate theingleton inequality. *International Symposium on Network Coding(NetCod)*, June 2012.
- [37] E. Thomas, N. Markin, and F. Oggier. On abelian group representability of finite groups. *Advances in Mathematics of Communications*, 8(2):139--152, May 2014.
- [38] E. Thomas and F. Oggier. A note on quasi-uniform distributions and abelian group representability. *International Conference on Signal Processing and Communications (SPCOM)*, 21-25 July 2012.
- [39] E. Thomas and F. Oggier. Explicit constructions of quasi-uniform codes from groups. *International Symposium on Information Theory (ISIT)*, July 2013.

- [40] E. Thomas and F. Oggier. Applications of quasi-uniform codes to storage. *International Conference on Signal Processing and Communications (SPCOM)*, July 2014.
- [41] X. Yan, R. Yeung, and Z. Zhang. The capacity region for multi-source multi-sink network coding. *Proceedings of the 2007 IEEE International Symposium on Information Theory*, June 2007.
- [42] R. W. Yeung. Information theory and network coding. *Springer*, November 2007.
- [43] A.A. Zain and B.S. Rajan. Algebraic characterization of mds group codes over cyclic groups. *Designs, Codes and Cryptography*, 41(6), 1995.
- [44] Z. Zhang and R. W. Yeung. A non-shannon-type conditional inequality of information quantities. *IEEE Transactions on Information Theory*, 43(6):1982--1986, November 1997.