Applications of the Discrete Fourier Transform in Information Theory and Cryptology

Matthew Warren Dodd

2003

Thesis submitted to the University of London for the degree of
Doctor of Philosophy

Royal Holloway and Bedford New College
University of London

To my wife and parents.

Abstract

This work explores two problems, one in information theory and one in cryptology, and shows that in both cases the Discrete Fourier Transform (DFT) can be usefully applied.

The first problem is the following:

> Suppose that an originator generates an n-bit vector V according to the uniform probability distribution on such vectors, and sends V over a binary symmetric channel with error probability $p < \frac{1}{2}$ to a receiver, who receives the n bits as a vector W. Is it possible for the originator and receiver to agree on choices for balanced n-bit to 1-bit functions f and g prior to the generation of V in such a way that $f(V)$ and $g(W)$ agree with probability greater than $1 - p$?

We show that the answer is "no" if we can prove a generalisation couched in terms of information-theoretic measures of Rényi order α. This we do for $\alpha = 2$, using the DFT, and extensively explore the generalisation for the case of Shannon information — $\alpha = 1$ — making use of the DFT and related ideas. Finally, we prove that the generalisation does not hold for all $\alpha \geq 1$, n and f.

The second problem is that of known plaintext cryptanalysis of certain types of stream cipher constructed from regularly clocking binary shift reg-

isters. It is shown that various types of maximum likelihood attack on the ciphers may be approximated by reformulations in terms of DFT coefficients, and implemented using the DFT. Moreover, the effectiveness of the attacks, in terms of their unicity distance, can be given by expressions in the DFT coefficients. We generalise the idea of a correlation attack, introduced by Siegenthaler, to that of a simultaneous correlation attack, and study a number of variants; we also show that fast correlations attacks can take advantage of simultaneous correlation.

Contents

Abstract **3**

Contents **5**

Acknowledgements **9**

1 Background Material **11**

1.1 Overview of the Thesis 11

1.2 Concerning This Chapter 12

1.3 The Discrete Fourier Transform 12

1.3.1 Introduction 12

1.3.2 The DFT on an Abelian Group 15

1.3.2.1 Construction of a Transform 15

1.3.2.2 Properties of the DFT on an Abelian Group . 18

1.3.3 The Walsh-Hadamard Transform 20

1.3.3.1 Construction of the Walsh-Hadamard Transform 21

1.3.3.2 Properties of the Walsh-Hadamard Transform 22

1.3.3.3 Computation of the Walsh-Hadamard Transform 23

1.3.3.4 The "Pile-up" Lemma 24

1.3.4 The DFT on a Finite Group 24
1.3.4.1 Survey of Principal Results 24
1.4 Information Theoretic Preliminaries 30
1.4.1 Directed Divergence 30
1.4.2 Information and Mutual Information 34
1.4.3 Information in Bernoulli Distributions 37
1.5 Some Theory Of Ciphers 38
1.5.1 Stream Ciphers 39
1.5.2 Implementing Stream Ciphers 40
1.5.3 Known Plaintext Attacks on Stream Ciphers 40
1.5.4 Gallager's Decoding Algorithm 41

2 A Problem in Information Theory 47
2.1 Introduction 47
2.2 The Problem 48
2.3 Information-Theoretic Reformulation 48
2.4 Proof of the Conjecture for $\alpha = 2$ 51
2.5 Towards a Proof for $\alpha = 1$ 55
2.6 The Generalised Conjecture 73
2.6.1 Proof of Generalised Conjecture for $\alpha = 1$, $n = 1$ . . . 74
2.6.2 Extrema of the Generalised Conjecture 77
2.7 Conjecture and Generalised Conjecture for Large α 80

3 A Problem In Cryptology 83
3.1 Introduction 83
3.1.1 Concerning Notation used in this Chapter 84
3.2 Maximum Likelihood Attacks on Stream Ciphers 84
3.2.1 The Keystream Generator 84

3.2.2 Maximum Likelihood Attack 85
3.2.3 Uniqueness of Maximum Likelihood Solutions 87
3.2.4 How Large N Should Be 88
3.2.4.1 A Theorem of Brynielsson [2] 88
3.2.4.2 A Corollary 89
3.2.4.3 N for Reduced Output Function 91
3.2.4.4 N for Reduced Key Space 91
3.2.5 An Observation Concerning "Correlation Immunity" . 92
3.3 Correlation Attacks on Linearly Clocking Stream Ciphers . . . 92
3.3.1 Linear Correlations to f 93
3.3.2 Linear Correlation Attack 94
3.3.3 The Walsh-Hadamard Transform 95
3.3.4 Reed-Muller Decoding Algorithm 95
3.3.5 Complexity of the Reed-Muller Attack 96
3.3.6 Two Significant Observations 96
3.4 Characterising a Function by Correlations to Linear Functionals 97
3.5 The Unicity Distance N in Terms of Correlations 98
3.5.1 Information in Terms of Correlations 98
3.5.2 Unicity Distance when Reducing the Output Function 100
3.5.3 Unicity Distance when Reducing the Key Space 101
3.6 Maximum Likelihood Attacks in Terms of Correlations 103
3.6.1 Reformulation of a Maximum Likelihood Condition . . 103
3.6.2 Reformulation for Reduced Output Function 103
3.6.3 Cryptanalytic Applications 104
3.6.3.1 Siegenthaler's Method 104
3.6.3.2 Generalised "Reed-Muller Method" 104
3.6.4 Reformulation for Reduced Key Space 105

3.6.5 More Cryptanalytic Applications 105
3.6.6 Why Less Keystream May Be Required 107
3.6.7 Example 107
3.6.8 A Brief Observation Concerning Probabilistic f 108
3.7 An Hybrid Attack and Its Evaluation 108
3.7.1 Unicity Distance of Equations (3.27) 109
3.7.2 A Method for Solving Equations (3.27) 111
3.7.3 The Particular Case Outlined in Section 3.6.6 112
3.7.4 The Case $U_2 = \langle v \rangle$ 113
3.8 Simultaneous Correlation and "Fast Correlation Attacks" . . . 116
3.8.1 Finding Suitable Relations 120
3.8.2 Concerning the Effectiveness of this Algorithm 121

4 Some Concluding Observations **122**
4.1 The DFT and Probability Distributions 122
4.1.1 The "Pile-up" Lemma Revisited 124
4.2 Induced Distributions and the DFT 125
4.3 The DFT and Order 2 Information 126

References **129**

Acknowledgements

As is doubtless usual for a project of such longevity, there are many people whose contributions to this work are a pleasure to acknowledge.

I am indebted to Professor Fred Piper, my supervisor, who encouraged me to register as a Ph.D. student some considerable number of years ago when first employed by the now sadly defunct Racal Comsec Limited, and has been greatly supportive of my work since then. More recently, Dr. Peter Wild has directed my studies with very useful words of practical guidance, and has, along with Dr. Sean Murphy, put aside time to discuss my work.

I would like to thank the former directors of Racal Comsec for allowing me time to pursue the first years of my doctoral research, and for funding these studies. All my immediate colleagues there did much to support my work, but I owe a particular debt of gratitude to Dr. Steve Babbage. His insights have illuminated many discussions over the years, and, specifically, he is responsible as much as I for formulating the problem defined in section 2.2, and proposed the information-theoretic formulation (2.3). In addition, he proofread the paper [5] on which chapter 3 is based, and has provided support and encouragement over many years.

I have been much inspired by the work of Lennart Brynielsson, whose theorem 3.8 is important in chapter 3. Dr. Simon Blackburn has made an important, but unwitting contribution, by lending me his copy of [4] for rather

longer than I think he realises. Dr. David Callaghan has been particularly supportive of my efforts finally to complete this thesis.

Last, but of course not least, I particularly thank all my family, without whose loving and patient support this work would never have been completed.

Chapter 1

Background Material

1.1 Overview of the Thesis

The material in this work has been divided into four chapters. Each describes its own content in an introductory section, as section 1.2 does for this chapter. However, it is hoped that it may assist the reader if we present here an outline of what is covered by the whole work.

This first chapter presents standard material on which subsequent chapters of the thesis will rely, in an attempt to make the thesis as self-contained as is reasonably possible. Chapter 2 tackles a particular problem concerning communication over a binary symmetric channel, which it formulates and generalises a number of times, and presents a number of results concerning information measures and the Discrete Fourier Transform (DFT) on the elementary abelian group of order 2^n, the Walsh-Hadamard transform. In Chapter 3, we concern ourselves with another specific problem: that of cryptanalysis of a linearly-clocking stream ciphers with memoryless output function. We show that here again the Walsh-Hadamard transform can assume an important role, from both theoretical and practical standpoints.

Finally, in chapter 4, we present some generalisations derived from the preceding material.

1.2 Concerning This Chapter

The purpose of the remainder of this introductory chapter is to equip a reader already blessed with a general mathematical education with specific concepts and results concerning the principal topics of this thesis: the Discrete Fourier Transform on a finite group, information theory and cryptology. These are dealt with in turn in sections 1.3, 1.4 and 1.5. In doing this, we will also establish some notation which will assist subsequent exposition. The material presented in this chapter is all well established in the published literature of the particular field. This is in contrast to that of subsequent chapters, which, unless stated otherwise, is all original work of the author.

1.3 The Discrete Fourier Transform

1.3.1 Introduction

Ever since Fourier asserted in 1807 that an "arbitrary" function can be expressed as a linear combination of sines and cosines, the theory of such decompositions and their application to the understanding of a variety of physical phenomena have developed together. Important in this context is the concept of a Hilbert space, which is an *inner product space* (see definition 1.1) with respect to whose *norm* (definition 1.2) it is complete i.e. every Cauchy sequence converges. A body of theory applies to the expansion of elements of the Hilbert space in terms of an *orthogonal subset* (definition 1.3). More generally, the notion of a complete normed vector space — *a Banach space*

— is of fundamental importance in modern functional analysis.

Finite-dimensional inner product spaces — which are necessarily Hilbert spaces — are of particular interest to us. In fact, elements of the space $\mathbb{C}^n$ (for any fixed integer $n \geq 1$) have a Fourier decomposition as linear combinations of elements of the orthogonal set $\{(\omega^{ij})_{i=0}^{n-1} : 0 \leq j < n\}$, where $\omega \in \mathbb{C}$ is a primitive n^{th} root of unity (see section 1.3.2). The transformation mapping an element of $\mathbb{C}^n$ to a corresponding linear combination — determined by the vector's components — of elements of the orthogonal set is a *Discrete Fourier Transform* (DFT), and there is a close relationship between the DFT and its inverse, which allows the coefficients in a Fourier decomposition of a given element to be computed. In 1965, Cooley and Tookey presented an algorithm — the Fast Fourier Transform (FFT) — to compute this DFT, or inverse DFT, in time $\approx n \log_2 n$, a dramatic improvement on the complexity n^2 of the most obvious calculations, and of significant importance in applications. In fact for any abelian group G, we have an algebra isomorphism, or DFT, $D : \mathbb{C}G \to \mathbb{C}^{|G|}$ which preserves inner products, as we discuss in section 1.3.2. Of immediate importance to us in chapters 2 and 3 will be the case where G is the elementary abelian 2-group, when the Walsh-Hadamard transform is a corresponding DFT: in view of its importance, we explicitly describe this map and its properties in section 1.3.3. In fact, for any finite group G, there is an isomorphism $\text{DFT} : \mathbb{C}G \to \bigoplus_{j=1}^{h} \mathbb{C}^{d_j \times d_j}$ (theorem 1.24); if G is *supersolvable* then such a DFT can be computed in time approximately $|G| \log_2 |G|$ (theorem 1.41).

We now give the definitions 1.1–1.3 cited above, which will also be of immediate use in the following sections.

Definition 1.1. An *inner product space* is a vector space V over a field $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ with an inner product $\langle , \rangle : V \times V \to \mathbb{F}$ such that

1. $\langle v, v\rangle \in \mathbb{R}$ for all $v \in V$, and is ≥ 0 with equality precisely when $v = 0$;

2. $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$, $\langle u+v, w\rangle = \langle u, v\rangle + \langle u, w\rangle$ for all $u, v, w \in V$;

3. $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$, $\langle u, \lambda v\rangle = \bar{\lambda}\langle u, v\rangle$ for all $u, v \in V$, $\lambda \in \mathbb{F}$.

Definition 1.2. For any $v \in V$, the *norm* $\|v\|$ of v is the quantity $\langle v, v\rangle^{\frac{1}{2}}$.

Definition 1.3. Two elements in an inner product space V are *orthogonal* precisely when their inner product is 0. A set of elements in V is orthogonal when its elements are non-zero and pairwise orthogonal.

Lemma 1.4. *An orthogonal set S in an inner product space V is a linearly independent set.*

Proof. Suppose S is orthogonal, and $\sum_{s\in S} \lambda_s s = 0$. Then for each $t \in S$

$$0 = \langle \sum_{s\in S} \lambda_s s, t\rangle = \sum_{s\in S} \lambda_s \langle s, t\rangle = \lambda_t \langle t, t\rangle,$$

which, since $t \neq 0$ implies that $\lambda_t = 0$. □

Finally in this section, we define two different multiplication operations on the elements of V. Their interaction with the DFT will be of great interest in due course.

Definition 1.5. For group G (written multiplicatively), field $\mathbb{F}$, and functions $f, g : G \to \mathbb{F}$,

1. the *convolution* $f \otimes g$ of f and g is defined by

$$(f \otimes g)(a) = \sum_{\substack{b,c\in G:\\ a=bc}} f(b)g(c) \qquad \text{for all } a \in G;$$

2. the *pointwise product* fg of f and g is defined by

$$(fg)(a) = f(a)g(a) \qquad \text{for all } a \in G.$$

1.3.2 The DFT on an Abelian Group

1.3.2.1 Construction of a Transform

Let G denote any finite abelian group, and let $V = \mathbb{C}^G$, the vector space of maps $G \to \mathbb{C}$. V becomes a complex inner product space if we define

$$\langle f_1, f_2 \rangle = \sum_{g \in G} f_1(g)\overline{f_2(g)} \quad \text{for any } f_1, f_2 \in V.$$

As we point out in remark 1.32 in section 1.3.4, V has an orthogonal subset $S = \{f_g : g \in G\}$ of $|G|$ elements for which

- f_{gh} is the (pointwise) product $f_g f_h$ for all $g, h \in G$; and
- f_g is a group homomorphism $G \to \mathbb{C}$ for each $g \in G$.

In fact we have the following:

Lemma 1.6. *Suppose that $S = \{f_g : g \in G\}$ is a subset of $|G|$ elements of V for which*

- *f_{gh} is the (pointwise) product $f_g f_h$ for all $g, h \in G$; and*
- *f_g is a group homomorphism $G \to \mathbb{C}$ for each $g \in G$.*

Then

1. *$f_1(a) = 1$ for all $a \in G$.*
2. *For any $g \in G \setminus \{1\}$, $\sum_{a \in G} f_g(a) = 0$.*
3. *$|f_g(a)| = 1$ for each $f_g \in S$ and any $a \in G$.*
4. *S is orthogonal.*
5. *$\|f_g\|^2 = |G|$ for each $g \in G$.*

Proof. Part 1 follows from the identity $f_1(a) = f_1(a)f_1(a)$, and the observation that we cannot have $f_1(a) = 0$ for any $a \in G$, since then $f_1(b) = f_1(a)f_1(a^{-1}b) = 0$ for all $b \in G$, and $f_g = f_1 f_g = 0$ for all $g \in G$, contradicting $|S| = |G|$.

For part 2, if $g \neq 1$ then $f_g \neq f_1$, so there exists $b \in G$ such that $f_g(b) \neq 1$. Then

$$\begin{aligned} \sum_{a \in G} f_g(a) &= \sum_{a \in G} f_g(ab) \\ &= \left(\sum_{a \in G} f_g(a) \right) f_g(b), \end{aligned}$$

from which it follows that $\sum_{a \in G} f_g(a) = 0$.

Part 3 follows from the fact that $f_g(a)^{|G|} = f_{g^{|G|}}(a) = f_1(a) = 1$ (by part 1).

To prove parts 4 and 5, suppose that $g, h \in G$. Then

$$\begin{aligned} \langle f_g, f_h \rangle &= \sum_{a \in G} f_g(a)\overline{f_h(a)} \\ &= \sum_{a \in G} f_g(a) f_h(a)^{-1} \qquad \text{since } |f_h(a)| = 1 \text{ by part 3} \\ &= \sum_{a \in G} f_g(a) f_{h^{-1}}(a) \qquad \text{since } g \mapsto f_g \text{ is a homomorphism} \\ &= \sum_{a \in G} f_{gh^{-1}}(a) \\ &= \begin{cases} 0 & \text{if } g \neq h, \text{ by part 2} \\ |G| & \text{if } g = h, \text{ by part 1,} \end{cases} \end{aligned}$$

as required. □

Note that the elements of S are linearly independent, by lemma 1.4, therefore form a basis of the $|G|$-dimensional vector space V.

Example 1.7. *Let $G = C_n$, the cyclic group of order n, which we write additively and identify with the integers $\{0, 1, \ldots, n-1\}$ under addition* $\bmod n$. *Fix any primitive n^{th} root of unity $\omega \in \mathbb{C}$. Then the set of functions $\{f_i : 0 \leq i < n\}$, where $f_i : j \mapsto \omega^{ij}$ $(0 \leq j < n)$, form such a set S.*

We are now in a position to define a Discrete Fourier Transform of an abelian group G.

Definition 1.8. We define a *Discrete Fourier Transform* $D : V \to V$ by the rule

$$D(f)(g) = \sum_{h \in G} f(h) f_h(g) \qquad \text{for all g} \in G,\ f \in V.$$

We can express this in an alternative — and perhaps clearer — way after introducing some new notation.

Notation 1.9. For each $g \in G$ we denote simply by g the element of V which maps $g \mapsto 1$ and $h \mapsto 0$ for $h \in G \backslash \{g\}$.

Now we can write any function $f \in V$ as

$$f = \sum_{g \in G} f(g) g$$

and

$$D(\sum_{g \in G} f(g) g) = \sum_{g \in G} f(g) f_g$$

Also of interest is the inverse DFT. We have the following result.

Lemma 1.10. *For any $f \in V$,*

$$D^{-1}(f)(g) = \frac{1}{|G|} \langle f, f_g \rangle.$$

Proof. Since S is a basis of V, we can write

$$f = \sum_{g \in G} \lambda_g f_g \tag{1.11}$$

for some $\lambda_g \in \mathbb{C}$. Then, by definition of D,

$$D^{-1}(f)(g) = \lambda_g$$

Taking inner products of both sides of equation (1.11) with f_h, for any $h \in G$,

$$\begin{aligned} \langle f, f_h \rangle &= \langle \sum_{g \in G} \lambda_g f_g, f_h \rangle \\ &= \lambda_h \langle f_h, f_h \rangle \qquad \text{by linearity and orthogonality} \\ &= |G| \, \lambda_h \qquad \text{by lemma 1.6, part 5.} \end{aligned}$$

Thus

$$D^{-1}(f)(g) = \lambda_g = \frac{1}{|G|} \langle f, f_g \rangle$$

as required. □

Note that, by the definition of inner product on V, we can equivalently write this

$$D^{-1}(f)(g) = \frac{1}{|G|} \sum_{h \in G} f(h) \overline{f_g(h)}.$$

1.3.2.2 Properties of the DFT on an Abelian Group

Some properties of the DFT on a general abelian group G are given by the following theorem.

Theorem 1.12. *For any elements $f_1, f_2 \in V$,*

1. $\langle D(f_1), D(f_2) \rangle = |G| \, \langle f_1, f_2 \rangle$
2. $D(f_1 \otimes f_2) = D(f_1) D(f_2)$

3. $|G|\, D(f_1 f_2) = D(f_1) \otimes D(f_2)$

Proof. For the first part we compute

$$\begin{aligned}
\langle D(f_1), D(f_2)\rangle &= \langle \sum_{g\in G} f_1(g) f_g, \sum_{h\in G} f_2(h) f_h\rangle \\
&= \sum_{g\in G}\sum_{h\in G} f_1(g)\overline{f_2(h)}\langle f_g, f_h\rangle \\
&= \sum_{g\in G} f_1(g)\overline{f_2(g)}\,\|f_g\|^2 \qquad \text{since S is orthogonal} \\
&= |G|\,\langle f_1, f_2\rangle \qquad \text{by lemma 1.6, part 5.}
\end{aligned}$$

The second is an immediate consequence of the definition of D and the multiplicative property of S:

$$\begin{aligned}
D(f_1 \otimes f_2)(g) &= \sum_{h\in G} (f_1 \otimes f_2)(h) f_h(g) \\
&= \sum_{h\in G} \sum_{\substack{a,b\in G:\\ ab=h}} f_1(a) f_2(b) f_{ab}(g) \\
&= \sum_{a,b\in G} f_1(a) f_2(b) f_a(g) f_b(g) \\
&= D(f_1)(g) D(f_2)(g)
\end{aligned}$$

For the third part, note first that for any $F_1, F_2 \in V$

$$\begin{aligned}
&D^{-1}(F_1 \otimes F_2)(g) \\
&= \frac{1}{|G|}\langle F_1 \otimes F_2, f_g\rangle \\
&= \frac{1}{|G|}\sum_{a\in G} (F_1 \otimes F_2)(a)\overline{f_g(a)} \\
&= \frac{1}{|G|}\sum_{\substack{a,b,c\in G\\ a=bc}} F_1(b) F_2(c)\overline{f_g(a)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{|G|}\sum_{b,c\in G} F_1(b)\overline{f_g(b)}F_2(c)\overline{f_g(c)} \qquad \text{because } f_g(a) = f_g(b)f_g(c)\\
&= |G|\,\frac{1}{|G|}\langle F_1, f_g\rangle\frac{1}{|G|}\langle F_2, f_g\rangle\\
&= |G|\,(D^{-1}(F_1)D^{-1}(F_2))(g).
\end{aligned}
$$

so that $D^{-1}(F_1 \otimes F_2) = |G|\, D^{-1}(F_1)D^{-1}(F_2)$. Now with the choice $F_i = D(f_i)$ for $i = 1, 2$, this becomes

$$D^{-1}(F_1 \otimes F_2) = |G|\, f_1 f_2$$

from which

$$D(f_1) \otimes D(f_2) = |G|\, D(f_1 f_2)$$

as required. □

Expressed informally, theorem 1.12 states that

1. D preserves inner product (up to a factor of $|G|$);
2. convolution in the "time domain" corresponds to pointwise product in the "frequency domain";
3. pointwise product in the "time domain" corresponds to convolution in the "frequency domain" (up to a factor of $|G|$).

1.3.3 The Walsh-Hadamard Transform

In view of its particular importance in chapters 2 and 3, we devote this section to the DFT on the elementary abelian 2-group, the Walsh-Hadamard transform.

1.3.3.1 Construction of the Walsh-Hadamard Transform

First we make the following (probably superfluous) definition:

Definition 1.13. A *bit* (short for binary digit) is an element of the set $\{0, 1\}$.

Now the elementary abelian group $G = C_2^n$ of order 2^n may be identified with the set $\{0, 1\}^n$ of n-bit vectors, with group operation $+$, written additively, defined to be componentwise addition modulo 2. As in section 1.3.2 above, the set $\mathbb{C}^G$ of complex-valued functions on G is a complex inner product space V of dimension 2^n.

Now we define a set S of 2^n functions in V by the rule

$$f_v(u) = (-1)^{u \cdot v} \qquad (u \in G)$$

for each $v \in G$, where for any vectors $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n)$ $\in G$ we define

$$u \cdot v = \sum_{i=1}^{n} u_i v_i \pmod 2.$$

Lemma 1.14. *For all $u, v, g, h \in G$,*

1. $f_{u+v}(g) = f_u(g) f_v(g)$; *and*
2. $f_v(g + h) = f_v(g) f_v(h)$.

Proof. Both of these results are immediate consequences of the definitions:

$$f_{u+v}(g) = (-1)^{(u+v)\cdot g} = (-1)^{u \cdot g}(-1)^{v \cdot g} = f_u(g) f_v(g)$$

and

$$f_v(g + h) = (-1)^{v \cdot (g+h)} = (-1)^{v \cdot g}(-1)^{v \cdot h} = f_v(g) f_v(h)$$

as required. □

Thus the conditions of lemma 1.6 are satisfied. In this context, definition 1.8 becomes:

Definition 1.15. The *Walsh-Hadamard transform* $D : V \to V$ is given by

$$D(f)(g) = \sum_{h \in G} f(h)(-1)^{g \cdot h} \quad \text{for all } g \in G,\ f \in V.$$

By lemma 1.10,

$$D^{-1}(f)(g) = 2^{-n} \sum_{h \in G} f(h)(-1)^{g \cdot h} \quad \text{for all } g \in G,\ f \in V.$$

Notice that if f is real-valued, then so is $D(f)$, and *vice versa*.

1.3.3.2 Properties of the Walsh-Hadamard Transform

The following result summarises some important properties of the Walsh-Hadamard transform.

Theorem 1.16. *For elements $f, g \in V$,*

1. $\langle f, g \rangle = 2^{-n} \langle D(f), D(g) \rangle$; *in particular,* $\|f\|^2 = 2^{-n} \|D(f)\|^2$
2. $D(f \otimes g) = D(f) D(g)$
3. $D(fg) = 2^{-n} D(f) \otimes D(g)$
4. $D(D(f)) = 2^n f$

Proof. Parts 1, 2 and 3 are immediate consequences of theorem 1.12. Part 4 follows from the equations for the Walsh-Hadamard transform and its inverse. □

1.3.3.3 Computation of the Walsh-Hadamard Transform

For $n \geq 1$, we have a natural isomorphism $G \simeq G' \times C_2$, where G' denotes the elementary abelian group of order 2^{n-1}, so that any element $u \in G$ can be written $u = (u', u_n)$ for an $(n-1)$-bit vector $u' \in G'$ and a bit $u_n \in C_2$. This observation allows us to compute the Walsh-Hadamard transform of a function $f \in V$ inductively: for $u = (u', u_n) \in G$,

$$\begin{aligned} D(f)(u) &= \sum_{v \in G} (-1)^{u \cdot v} f(v) \\ &= \sum_{(v', v_n) \in G} (-1)^{u' \cdot v' + u_n v_n} f(v', v_n) \\ &= \sum_{v' \in G'} (-1)^{u' \cdot v'} f(v', 0) + \sum_{v' \in G'} (-1)^{u' \cdot v' + u_n} f(v', 1) \\ &= D'(f_0)(u') + (-1)^{u_n} D'(f_1)(u') \end{aligned} \tag{1.17}$$

where D' denotes the DFT on G' defined for any $f' : G' \to \mathbb{C}$ by

$$D'(f')(u') := \sum_{v' \in G'} (-1)^{u' \cdot v'} f(v')$$

and $f_i : G' \to \mathbb{C}$ is defined by

$$f_i(u') := f(u', i) \qquad \text{for all } u' \in G'.$$

Thus we can compute the Walsh-Hadamard transform of a function on the elementary abelian 2-group of order 2^n in terms of the Walsh-Hadamard transform of two functions on the elementary abelian 2-group of order 2^{n-1}.

In fact, the number of additions/subtractions required for this computation is $n2^n$. To prove this by induction, it suffices to note that it is clearly true for $n = 1$, and if the number for the elementary abelian 2-group of order 2^{n-1} is $(n-1)2^{n-1}$, then the number for the group of order 2^n is seen from equation (1.17) to be $2(n-1)2^{n-1} + 2^n = n2^n$.

1.3.3.4 The "Pile-up" Lemma

Here we present a result often referred to as the "Pile-up" Lemma, which we will use later in this chapter. In fact, it can be seen to be a consequence of the DFT, but that discussion is deferred to section 4.1.1. An elegant elementary proof can be found in [6].

Definition 1.18. For a distribution X on $\{0,1\}$, define

$$c(X) := \Pr(X = 0) - \Pr(X = 1).$$

Lemma 1.19 (Pile-up Lemma). *If $X_1, \ldots X_m$ are m independent distributions on $\{0,1\}$, then*

$$c(X_1 \oplus \cdots \oplus X_m) = c(X_1) \ldots c(X_m)$$

1.3.4 The DFT on a Finite Group

In this section we present some principal results from classical group representation theory. These results are almost all taken from [4], to which a reader seeking proofs or other details is referred.

1.3.4.1 Survey of Principal Results

Definition 1.20. An *algebra* A over a field $\mathbb{F}$ is a set which forms both a ring and a vector space over $\mathbb{F}$ such that the additive group structure in each case is the same, and ring multiplication commutes with scalar multiplication, i.e.

$$\lambda(ab) = (\lambda a)b = a(\lambda b) \qquad \text{for all } a, b \in A, \lambda \in \mathbb{F}.$$

An *algebra homomorphism* $\phi : A \to B$ is a vector space homomorphism for which

$$\begin{aligned} \phi(ab) &= \phi(a)\phi(b) \qquad \text{for all } a, b \in A, \text{ and} \\ \phi(1_A) &= 1_B. \end{aligned}$$

Lemma 1.21. *The set of maps $G \to \mathbb{C}$ forms an algebra $\mathbb{C}G$ over $\mathbb{C}$ with multiplication given by convolution:*

$$(ab)(g) = \sum_{g_1, g_2 : g_1 g_2 = g} a(g_1) a(g_2).$$

Definition 1.22. $\mathbb{C}G$ is the *group algebra* of G over $\mathbb{C}$.

Definition 1.23. A *representation* of $\mathbb{C}G$ of dimension d is an algebra homomorphism $\mathbb{C}G \to \mathrm{End}(\mathbb{C}^d)$, the algebra of linear transformations $\mathbb{C}^d \to \mathbb{C}^d$.

Theorem 1.24 (Wedderburn). *The group algebra $\mathbb{C}G$ of a finite group G is isomorphic to an algebra of block diagonal matrices:*

$$\mathbb{C}G \simeq \bigoplus_{j=1}^{h} \mathbb{C}^{d_j \times d_j}, \tag{1.25}$$

where h is the number of conjugacy classes in G and the d_j are determined up to permutation by G.

Definition 1.26. An isomorphism $D : \mathbb{C}G \to \bigoplus_{j=1}^{h} \mathbb{C}^{d_j \times d_j}$ is a *discrete Fourier transform (DFT)* for G.

Notation 1.27. Given such an isomorphism, we write D_j for the induced $\mathbb{C}$ algebra surjection $\mathbb{C}G \to \mathbb{C}^{d_j \times d_j}$.

Theorem 1.28. *The group homomorphisms $G \to \mathbb{C}$ form a group isomorphic to G/G', where G' denotes the commutator subgroup $\langle g^{-1}h^{-1}gh : g, h \in G \rangle$.*

Definition 1.29. A *character* $\chi : \mathbb{C}G \to \mathbb{C}$ is a map derived from a representation D of $\mathbb{C}G$ by setting $\chi(g) := \mathrm{Tr}(D(g))$ for $g \in G$, and extending its definition to $\mathbb{C}G$ linearly; here Tr denotes the trace of a square matrix, that is, the sum of its diagonal entries. We write χ_i for the character corresponding to the representation D_i.

Definition 1.30. The inner product $\langle \chi, \chi' \rangle$ of two characters χ and χ' is

$$\sum_{g \in G} \chi(g)\overline{\chi'(g)}.$$

Theorem 1.31. *The characters χ_i are pairwise orthogonal; that is, for $i \neq j$, $\langle \chi_i, \chi_j \rangle = 0$.*

Remark 1.32. We have now gathered enough group-theoretic machinery to derive the result cited at the beginning of section 1.3.2.1. Suppose that G is abelian, so that $h = |G|$. Since the vector space dimensions of both sides of equation (1.25) are equal, we have the equality

$$|G| = \sum_{j=1}^{|G|} d_j^2$$

from which it follows that $d_j = 1$ for $1 \leq j \leq |G|$. Thus each D_i is an algebra homomorphism $\mathbb{C}G \to \mathbb{C}$, so equals its corresponding character χ_i. The characters χ_i are pairwise orthogonal, and therefore, considered as functions $G \to \mathbb{C}$, are distinct, and group homomorphisms, so by theorem 1.28 form a group isomorphic to $G/G' \simeq G$. Thus we can label them χ_g for $g \in G$ in such a way that

$$\chi_{gh} = \chi_g \chi_h \qquad \text{for all } g, h \in G.$$

The set $\{\chi_g : G \to \mathbb{C}\}_{g \in G}$ is thus seen to be an orthogonal set with the properties cited in section 1.3.2.1.

A DFT $D : \mathbb{C}G \to \bigoplus_{j=1}^{h} \mathbb{C}^{d_j \times d_j}$ determines a $|G| \times |G|$ matrix $\boldsymbol{D}$ according to the rule

$$\boldsymbol{D}_{(j,k,l),g} := D_j(g)_{k,l} \qquad \text{for } g \in G,\ 1 \leq j \leq h,\ 1 \leq k,l \leq d_j.$$

Up to permutations of the rows and columns, $\boldsymbol{D}$ is uniquely determined by D.

Proposition 1.33. *The cyclic group of order n has a DFT matrix DFT_n with entries $(DFT_n)_{ij} = \omega^{(i-1)(j-1)}$ $(1 \leq i, j \leq n)$, for any fixed primitive n^{th} root ω of unity.*

In fact, we have already seen this from example 1.7 and the work of section 1.3.2.

Definition 1.34. The *Kronecker product* $A \otimes B$ of square matrices A and B is given by

$$A \otimes B := \begin{pmatrix} A_{11}B & A_{12}B & \cdots \\ A_{21}B & A_{22}B & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix},$$

where A_{ij} denotes the entry of A in the i^{th} row and j^{th} column.

Theorem 1.35. *If A is a DFT matrix for a group G and B is a DFT matrix for a group H, then $A \otimes B$ is a DFT matrix for the direct product $G \times H$.*

Corollary 1.36. *The direct product V_n of n copies of the cyclic group of order 2 has a $2^n \times 2^n$ DFT matrix with $(i+1, j+1)^{th}$ entry $(-1)^{i \cdot j}$, where for integers $i = \sum_{k=0}^{n-1} 2^k i_k$ and $j = \sum_{k=0}^{n-1} 2^k j_k$ we denote by $i \cdot j$ the quantity*

$$i \cdot j := \sum_{k=0}^{n-1} i_k j_k \pmod 2.$$

Proof. Let $T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, a DFT matrix for the cyclic group of order 2 by proposition 1.33. By theorem 1.35, $T^n = T \otimes \cdots \otimes T$ (n terms) is a DFT matrix for V_n. We prove by induction on n that the $(i+1, j+1)^{\text{th}}$ entry $T_{ij}^{(n)}$ of T^n is $(-1)^{i \cdot j}$.

The result certainly holds for $n = 1$, so suppose $n > 1$ and the result is true for $n - 1$. By definition of the Kronecker product, for indices $i =$

$2^{n-1}i_{n-1} + i'$ $(0 \leq i' < 2^{n-1})$ and $j = 2^{n-1}j_{n-1} + j'$ $(0 \leq j' < 2^{n-1})$, the $(i+1, j+1)^{\text{th}}$ entry of T^n is the $(i+1, j+1)^{\text{th}}$ entry of T^{n-1} iff[1] either i_{n-1} or j_{n-1} is 0, otherwise the $(i+1, j+1)^{\text{th}}$ entry of T^n is the (-1) times the $(i+1, j+1)^{\text{th}}$ entry of T^{n-1}. Thus

$$\begin{aligned} T^{(n)}_{ij} &= (-1)^{i_{n-1}j_{n-1}} T^{(n-1)}_{i'j'} \\ &= (-1)^{i_{n-1}j_{n-1}} (-1)^{i' \cdot j'} \qquad \text{by the inductive hypothesis} \\ &= (-1)^{i \cdot j}. \end{aligned}$$

Thus the result is true for n.

Hence, by induction, the result is true for all $n \geq 1$. □

Theorem 1.37. *With the same notation D_i and d_i as above, we have the following:*

1. *(Fourier inversion formula) For $a \in \mathbb{C}G$,*

$$a(g) = \frac{1}{|G|} \sum_{i=1}^{h} d_i \operatorname{Tr}(D_i(g^{-1}) D_i(a)).$$

2. *(Plancherel formula) For two elements $a, b \in \mathbb{C}G$,*

$$\sum_{g \in G} a(g) b(g) = \frac{1}{|G|} \sum_{i=1}^{h} d_i \operatorname{Tr}(D_i(\tilde{a}) D_i(b)),$$

where $\tilde{a} \in \mathbb{C}G$ is defined by the rule $\tilde{a}(g) = a(g^{-1})$ $(g \in G)$.

Finally, we present a result from [4] about computing a DFT for finite groups.

Definition 1.38. A *chief sequence* for a finite group G is a sequence of normal subgroups $G_1, \ldots, G_r$ such that

$$G = G_r \rhd G_{r-1} \rhd \cdots \rhd G_1 = \langle 1 \rangle$$

[1]Here, and subsequently in this thesis, we use the abbreviation 'iff' for 'if and only if'.

and there is no normal subgroup N of G such that $G_i \rhd N \rhd G_{i-1}$ for any i $(2 \leq i \leq r)$. A finite group G is *supersolvable* iff it has a chief sequence $G_1, \ldots, G_r$ for which each G_i/G_{i-1} $(2 \leq i \leq r)$ is cyclic of prime order.

Definition 1.39. Let $x_1, \ldots, x_n$ be indeterminates over $\mathbb{C}$. Then a sequence $(g_1, \ldots g_{n+r})$ of linear forms $g_i \in \mathbb{C}x_1 + \cdots + \mathbb{C}x_n$ $(1 \leq i \leq n+r)$ is a *linear computation sequence* of length r iff $g_1 = x_1, \ldots, g_n = x_n$ and whenever $n < k \leq n + r$ either

1. $g_k = z_k g_i$ for some $z_k \in \mathbb{C}$ and index i satisfying $1 \leq i < k$; or
2. $g_k = \epsilon_k g_i + \delta_k g_j$ for $\epsilon_k, \delta_k \in \{1, -1\}$ and indices i, j satisfying $1 \leq i, j < k$.

Definition 1.40. For an $m \times n$ matrix $A = (A_{ij})$ with entries in $\mathbb{C}$, $L_\infty(A)$ is the minimal integer r for which there is a linear computation sequence of length r computing all the forms $\sum_{j=1}^{n} A_{ij}x_j$ $(i = 1, \ldots m)$.

Theorem 1.41. *If G is supersolvable and D is any DFT matrix for G, $L_\infty(D) \leq 8.5\,|G| \log_2 |G|$.*

Moreover, a construction is given in [4] to construct a DFT from a power-commutator presentation of G and chief sequence for G in at most $14|G| \log_2 |G| + 5\,|G|$ "basic operations".

Finally, on this topic, we have the following result relating the complexity of computing a DFT matrix and its inverse.

Theorem 1.42. *For any DFT matrix D on a finite group G,*

$$\left|L_\infty(D) - L_\infty(D^{-1})\right| \leq |G|\,.$$

1.4 Information Theoretic Preliminaries

In this section we present some basic ideas in information theory, which will be used in subsequent chapters. In section 1.4.1 we follow ideas of Rényi to define directed divergence of general order $\alpha \geq 1$, and explore some of its properties. Then in section 1.4.2 we define Rényi information and mutual information in terms of directed divergence, and derive some further results. Finally, in section 1.4.3, we consider the particular case of Bernouilli distributions, which will be of particular importance in chapter 2.

1.4.1 Directed Divergence

Definition 1.43. The *directed divergence* of order α, for any real $\alpha \geq 1$, between two probability distributions X and Y on a set S is

$$d_\alpha(X;Y) := \begin{cases} \frac{1}{\alpha-1}\left(\left(\sum_{s\in S} \Pr_Y(s)\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)^\alpha\right)-1\right) & \text{if } \alpha > 1 \\ \sum_{s\in S} \Pr_X(s) \ln \frac{\Pr_X(s)}{\Pr_Y(s)} & \text{if } \alpha = 1. \end{cases}$$

To interpret these equations correctly, we need to give a couple of points of guidance: first, the sums $\sum_{s\in S}$ are over elements $s \in S$ for which $\Pr_Y(s) > 0$; second, whenever $\Pr_X(s) = 0$ then the term $\Pr_X(s) \ln \frac{\Pr_X(s)}{\Pr_Y(s)}$ is taken to be 0.

Although it is not immediately apparent, if we fix any X and Y, $d_\alpha(X;Y)$ is a continuous function of α on $[1,\infty)$. To establish this it is sufficient to show that whenever $\Pr_X(s)$ and $\Pr_Y(s)$ are both non-zero

$$\frac{1}{\alpha-1}\left(\left(\sum_{s\in S} \Pr_Y(s)\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)^\alpha\right)-1\right) \to \sum_{s\in S} \Pr_X(s) \ln \frac{\Pr_X(s)}{\Pr_Y(s)} \text{ as } \alpha \downarrow 1.$$

Since the L.H.S. of this can be re-expressed

$$\begin{aligned} &\frac{1}{\alpha-1}\left(\left(\sum_{s\in S} \Pr_Y(s)\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)^\alpha\right)-1\right) \\ &= \sum_{s\in S}\left(\Pr_Y(s)\frac{1}{\alpha-1}\left(\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)^\alpha - \frac{\Pr_X(s)}{\Pr_Y(s)}\right)\right), \end{aligned}$$

we see that the result is an immediate corollary of the following lemma.

Lemma 1.44. *For any real $a > 0$,*

$$\lim_{\alpha \to 1} \frac{1}{\alpha - 1} (a^\alpha - a) = a \ln a.$$

Proof. Write $f(\alpha) = a^\alpha$. f is differentiable on $(-\infty, \infty)$ with derivative $f'(\alpha) = a^\alpha \ln a$. In particular, f is differentiable at $\alpha = 1$, so

$$\lim_{\alpha \to 1} \frac{1}{\alpha - 1} (f(\alpha) - f(1)) = f'(1)$$

$$\text{i.e. } \lim_{\alpha \to 1} \frac{1}{\alpha - 1} (a^\alpha - a) = a \ln a \qquad \square$$

An important property of directed divergence is given in the following proposition.

Proposition 1.45. *For any $\alpha \geq 1$, $d_\alpha(X;Y) \geq 0$ with equality if and only if $\Pr_X(s) = \Pr_Y(s)$ for each $s \in S$ such that $\Pr_Y(s) > 0$.*

Proof. First of all, we show that $f(x) := x \ln x$ and $g(x) := x^\alpha$ (for $\alpha > 1$) are convex on $(0, \infty)$ by computing their second derivatives and observing that they are strictly positive on that interval:

$$\begin{aligned} f'(x) &= 1 + \ln x \\ f''(x) &= 1/x > 0 \qquad \text{for } x > 0 \end{aligned}$$

and, for $\alpha > 1$,

$$g''(x) = \alpha(\alpha - 1)x^{\alpha - 2} > 0 \qquad \text{for } x > 0.$$

Now we use the fact that if $h(x)$ is any convex function on $(0, \infty)$, x_i $(i = 1, \ldots, n)$ are reals in the interval $(0, \infty)$ and p_i $(i = 1, \ldots, n)$ are reals in the range $(0, 1]$ whose sum is 1, then

$$\sum_{i=1}^{n} p_i h(x_i) \geq h(\sum_{i=1}^{n} p_i x_i)$$

with equality iff all x_i are equal.

Applying this first to f,

$$\begin{aligned} d_1(X;Y) &= \sum_{s\in S} \Pr_X(s) \ln \frac{\Pr_X(s)}{\Pr_Y(s)} \\ &= \sum_{s\in S} \Pr_Y(s) f\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right) \\ &\geq f(1) \\ &= 0, \end{aligned}$$

with equality iff $\Pr_X(s) = \Pr_Y(s)$ whenever $\Pr_Y(s) > 0$.

Similarly, when $\alpha > 1$ we can apply it to g:

$$\begin{aligned} d_\alpha(X;Y) &= \frac{1}{\alpha-1}\left(\left(\sum_{s\in S} \Pr_Y(s) \left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)^\alpha\right) - 1\right) \\ &= \frac{1}{\alpha-1}\left(\left(\sum_{s\in S} \Pr_Y(s) g\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)\right) - 1\right) \\ &\geq \frac{1}{\alpha-1}\left(g(1) - 1\right) \\ &= 0, \end{aligned}$$

again with equality iff $\Pr_X(s) = \Pr_Y(s)$ whenever $\Pr_Y(s) > 0$. □

Finally in this section we prove a result about directed divergence which will be of use to us in the next section.

Proposition 1.46. *With notation as above,*

$$d_\alpha(X;Y) \geq d_\alpha(h(X);h(Y)) \qquad \text{for any } \alpha \geq 1,$$

where $h : S \to S'$ (for any set S') is any function, and the corresponding induced distribution $h(X)$ on S' is given by

$$\Pr_{h(X)}(s') := \sum_{\substack{s\in S:\\ h(s)=s'}} \Pr_X(s) \qquad \text{for each } s' \in S',$$

and similarly for $h(Y)$.

Proof. As in the proof of proposition 1.45, define $f(x) := x \ln x$ and $g(x) := x^\alpha$ (for $\alpha > 1$) for $x \in (0, \infty)$. For $\alpha = 1$ we have

$$\begin{aligned}
d_1(X;Y) &= \sum_{s\in S} \Pr_X(s) \ln \frac{\Pr_X(s)}{\Pr_Y(s)} \\
&= \sum_{s\in S} \Pr_Y(s) f\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right) \\
&= \sum_{s'\in S'} \Pr_{h(Y)}(s') \sum_{\substack{s\in S:\\ h(s)=s'}} \frac{\Pr_Y(s)}{\Pr_{h(Y)}(s')} f\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right) \\
&\geq \sum_{s'\in S'} \Pr_{h(Y)}(s') f\left(\sum_{\substack{s\in S:\\ h(s)=s'}} \frac{\Pr_Y(s)}{\Pr_{h(Y)}(s')} \frac{\Pr_X(s)}{\Pr_Y(s)}\right) \qquad \text{by convexity} \\
&= \sum_{s'\in S'} \Pr_{h(Y)}(s') f\left(\frac{\Pr_{h(X)}(s')}{\Pr_{h(Y)}(s')}\right) \\
&= d_1(h(X);h(Y)).
\end{aligned}$$

As usual the sums here are over $s \in S$ for which $\Pr_Y(s) > 0$ and $s' \in S'$ for which $\Pr_{h(Y)}(s') > 0$; note that if $\Pr_Y(s) > 0$ then $\Pr_{h(Y)}(s') > 0$ for $s' = h(s)$.

For $\alpha > 1$, a similar argument applies:

$$\begin{aligned}
&d_\alpha(X;Y) \\
&= \frac{1}{\alpha - 1}\left(\left(\sum_{s\in S} \Pr_Y(s) \left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)^\alpha\right) - 1\right) \\
&= \frac{1}{\alpha - 1}\left(\left(\sum_{s\in S} \Pr_Y(s) g\left(\frac{\Pr_X(s)}{\Pr_Y(s)}\right)\right) - 1\right)
\end{aligned}$$

$$= \frac{1}{\alpha - 1} \left(\left(\sum_{s' \in S'} \Pr_{h(Y)}(s') \sum_{\substack{s \in S: \\ h(s) = s'}} \frac{\Pr_Y(s)}{\Pr_{h(Y)}(s')} g\left(\frac{\Pr_X(s)}{\Pr_Y(s)} \right) \right) - 1 \right)$$

$$\geq \frac{1}{\alpha - 1} \left(\left(\sum_{s' \in S'} \Pr_{h(Y)}(s') g\left(\sum_{\substack{s \in S: \\ h(s) = s'}} \frac{\Pr_Y(s)}{\Pr_{h(Y)}(s')} \frac{\Pr_X(s)}{\Pr_Y(s)} \right) \right) - 1 \right)$$

by convexity

$$= \frac{1}{\alpha - 1} \left(\left(\sum_{s' \in S'} \Pr_{h(Y)}(s') g\left(\frac{\Pr_{h(X)}(s)}{\Pr_{h(Y)}(s')} \right) \right) - 1 \right)$$

$$= d_\alpha(h(X); h(Y)). \qquad \square$$

Note that a proof for the case $\alpha = 1$ could alternatively be derived from a proof for the case $\alpha > 1$ by a continuity argument.

1.4.2 Information and Mutual Information

Definition 1.47. For any real $\alpha \geq 1$, the *information* $I_\alpha(X)$ of order α of (or in) a distribution X on a set S is $d_\alpha(X; U)$, where U denotes the uniform distribution on S. When $\alpha = 1$, we also refer to $I_\alpha(X)$ as the *Shannon information* of (or in) X.

An immediate consequence of proposition 1.45 is:

Lemma 1.48. *For any real $\alpha \geq 1$, $I_\alpha(X) \geq 0$ with equality iff $X = U$.*

Definition 1.49. Let (X, Y) denote a probability distribution on the Cartesian product $S \times T$ of two sets S and T. From (X, Y) we can form new

distributions X on S, Y on T and X, Y on $S \times T$ by defining

$$\begin{aligned} \Pr_X(s) &= \sum_{t \in T} \Pr_{(X,Y)}(s,t), \\ \Pr_Y(t) &= \sum_{s \in S} \Pr_{(X,Y)}(s,t) \qquad \text{and} \\ \Pr_{X,Y}(s,t) &= \Pr_X(s) \Pr_Y(t). \end{aligned}$$

The *mutual information* $I_\alpha(X;Y)$ *of order* α between X and Y is defined to be $d_\alpha((X,Y);X,Y)$, for any real $\alpha \geq 1$. If $\alpha = 1$ we also refer to $I_\alpha(X;Y)$ as the *Shannon mutual information* between X and Y.

Having made these definitions, the following result follows readily:

Lemma 1.50. *For any real* $\alpha \geq 1$, $I_\alpha(X;Y) \geq 0$ *with equality if and only if* X *and* Y *are independent.*

Proof. From lemma 1.48, $I_\alpha(X;Y) \geq 0$ with equality if and only if

$$\Pr_{(X,Y)}(s,t) = \Pr_X(s) \Pr_Y(t)$$

whenever $\Pr_X(s) \Pr_Y(t) > 0$. Note though that if $\Pr_X(s) \Pr_Y(t) = 0$, either $\Pr_X(s) = 0$ or $\Pr_Y(t) = 0$, so that, either way, $\Pr_{(X,Y)}(s,t) = 0$. Hence $I_\alpha(X;Y) = 0$ if and only if

$$\Pr_{(X,Y)}(s,t) = \Pr_X(s) \Pr_Y(t)$$

for all $s \in S$, $t \in T$; this is, if and only if X and Y are independent. □

Note also that it follows immediately from the definitions that, for all distributions X and Y, $I_\alpha(X;Y) = I_\alpha(Y;X)$.

Definition 1.51. Retaining the same notation, the *conditional information* $I_\alpha(X|Y)$ *of order* α between X and Y is defined by the expression

$$I_\alpha(X|Y) := \sum_{t \in T} \Pr_Y(t) I_\alpha(X|Y = t)$$

where $I_\alpha(X|Y=t)$ is the information in the distribution $X|(Y=t)$ on S defined by

$$\Pr_{X|Y=t}(s) := \Pr(X=s|Y=t) = \Pr_{(X,Y)}(s,t)/\Pr_Y(t).$$

For Shannon information, if we interpret $I_1(X;Y)$ as the amount of information knowledge of Y reveals about X, then it seems apt that $I_1(X;Y)$ should equal the information in X when Y is known minus the information in X before Y is known. This is borne out by the following lemma.

Lemma 1.52. *$I_1(X;Y) = I_1(X|Y) - I_1(X)$.*

Proof. Let U_S denote the uniform distribution on S. Now we calculate

$$\begin{aligned}
&I_1(X;Y)\\
&= \sum_{\substack{s\in S\\ t\in T}} \Pr_{(X,Y)}(s,t) \ln \frac{\Pr_{(X,Y)}(s,t)}{\Pr_X(s)\Pr_Y(t)}\\
&= \sum_{\substack{s\in S\\ t\in T}} \Pr_{(X,Y)}(s,t) \ln \frac{\Pr_{(X,Y)}(s,t)}{\Pr_{U_S}(s)\Pr_Y(t)} - \sum_{\substack{s\in S\\ t\in T}} \Pr_{(X,Y)}(s,t) \ln \frac{\Pr_X(s)}{\Pr_{U_S}(s)}\\
&= \sum_{t\in T} \Pr_Y(t) \sum_{s\in S} \frac{\Pr_{(X,Y)}(s,t)}{\Pr_Y(t)} \ln \frac{\Pr_{(X,Y)}(s,t)}{\Pr_{U_S}(s)\Pr_Y(t)} - \sum_{s\in S} \Pr_X(s) \ln \frac{\Pr_X(s)}{\Pr_{U_S}(s)}\\
&= \sum_{t\in T} \Pr_Y(t) \sum_{s\in S} \Pr(s|t) \ln \frac{\Pr(s|t)}{\Pr_{U_S}(s)} - I_1(X)\\
&= I_1(X|Y) - I_1(X),
\end{aligned}$$

as required. □

It is also intuitive that the mutual information $I_\alpha(X;Y)$ which X reveals about Y should be at least as great as the information $I_\alpha(f(X);Y)$, where $f(X)$ denotes the distribution induced on S' by a function $f : S \to S'$:

$$\Pr_{f(X)}(s') = \sum_{\substack{s\in S:\\ f(s)=s'}} \Pr_X(s) \qquad \text{for each } s' \in S'.$$

Proposition 1.53. *With notation as above,*

$$I_\alpha(f(X);Y) \leq I_\alpha(X;Y).$$

Proof. By definition

$$I_\alpha(X;Y) = d_\alpha((X,Y);X,Y),$$

and

$$I_\alpha(f(X);Y) = d_\alpha((\tilde{f}((X,Y));\tilde{f}(X,Y)),$$

where we define the map $\tilde{f} : S \times T \to S \times T$ by the rule

$$\tilde{f}(s,t) = (f(s),t).$$

The result now follows immediately from proposition 1.46. □

1.4.3 Information in Bernoulli Distributions

In chapter 2 it will prove convenient to have made the following definition.

Definition 1.54. For the specific case $S = \{0,1\}$, we define, for each $\alpha \geq 1$, a function on $[-1,1]$ by the rule

$$i_\alpha(q-p) := I_\alpha(X_p),$$

where X_p is the distribution on S taking the values 0 and 1 with respective probabilities p and $q := 1-p$.

Hence, more explicitly,

$$i_\alpha(x) = \begin{cases} \frac{1}{\alpha-1}(\frac{1}{2}((1+x)^\alpha + (1-x)^\alpha) - 1) & \text{for } \alpha > 1 \\ \frac{1+x}{2}\ln(1+x) + \frac{1-x}{2}\ln(1-x) & \text{for } \alpha = 1. \end{cases}$$

As above, we interpret the term $\frac{1+x}{2}\ln(1+x)$ as having value 0 when $x = -1$.

Lemma 1.55. *For any real $\alpha \geq 1$, $i_\alpha(x)$ is convex and $i_\alpha(x) = i_\alpha(-x)$ on $[-1, 1]$.*

Proof. Certainly $i_\alpha(x) = i_\alpha(-x)$ for all $x \in [-1, 1]$.

To show that $i_\alpha(x)$ is convex it is necessary and sufficient to show that the second derivative $i''_\alpha(x) > 0$ on $(-1, 1)$. We compute the first and second derivatives of $i_\alpha(x)$ with respect to x:

$$i'_\alpha(x) = \begin{cases} \frac{\alpha}{2(\alpha-1)} \left((1+x)^{\alpha-1} - (1-x)^{\alpha-1}\right) & \text{for } \alpha > 1 \\ \frac{1}{2}\ln(1+x) - \frac{1}{2}\ln(1-x) & \text{for } \alpha = 1 \end{cases}$$

and

$$i''_\alpha(x) = \frac{\alpha}{2}\left((1+x)^{\alpha-2} + (1-x)^{\alpha-2}\right) \qquad \text{for all } \alpha \geq 1;$$

the result follows immediately from this expression. □

1.5 Some Theory Of Ciphers

It is generally acknowledged that the birth of the formal mathematical study of cipher systems is marked by the seminal paper of Shannon [11]. There he makes the following definition:

Definition 1.56 (Shannon). A *secrecy system* is a family of uniquely reversible transformations T_i of a set of possible messages into a set of cryptograms, the transformation T_i having an associated probability p_i.

Each index i is also called a *key*. It is understood that the number of indices i, or keys, is finite. Likewise, the set of possible messages is finite. It is assumed that an *enemy*, i.e. anyone attempting to derive information about message from observations of cryptograms, knows the set of transformations $\{T_i\}$ and the probability p_i associated with each key i.

We also use the word *cipher* to mean a secrecy system defined in this way.

An originating party uses the system by choosing a key i according to the distribution defined by the p_i, and communicating it over a secure channel to his intended recipient. Subsequently when he wishes to send a message over a channel on which an enemy may be eavesdropping, he can apply T_i to his message, often called the *plaintext*, and transmit the resulting *ciphertext* over the insecure channel.

1.5.1 Stream Ciphers

In this section we define a particular type of cipher known as a *stream cipher*.

Definition 1.57. A *stream cipher* is a cipher system in which, for some integer $N \geq 1$, the messages are N-bit sequences and each encrypting transformation T_i is of the form

$$m_1 m_2 \dots m_N \mapsto c_1 c_2 \dots c_N$$

where

$$b_1 b_2 \dots b_N$$

is an N-bit *keystream sequence* determined by the key i, and each ciphertext bit is obtained from corresponding plaintext and keystream bits according to the rule

$$c_j = m_j \oplus b_j \qquad (j = 1, \dots, N),$$

where $\oplus$ denotes modulo 2 addition, or *exclusive-or*.

Stream ciphers find favour in certain practical situations because they are not error-propagating, in the sense that each message bit m_j a recipient recovers using T_i^{-1} from $c_1 c_2 \dots c_N$ is in error due to channel corruptions iff the corresponding c_j is likewise in error.

1.5.2 Implementing Stream Ciphers

In section 1.5.1 we saw that an essential component of a stream cipher is a rule to derive from a key value i an associated keystream sequence

$$b_1 b_2 \dots b_N$$

A construction used frequently in practice to achieve this is as follows.

Definition 1.58. Let V be the set of n-bit vectors, for some fixed n. A *keystream generator* is a finite state machine with state space V, a state transition function $T : V \to V$, and an output function $f : V \to \{0, 1\}$. The initial state x of the keystream generator is either the key i or some fixed function of i. The keystream sequence

$$b_1 b_2 \dots b_N$$

produced from i is the corresponding output sequence from the finite state machine:

$$b_i = f(T^i x) \qquad (i = 1, \dots, N).$$

For any good stream cipher constructed from a keystream generator the output function f will be, approximately, balanced — that is, it takes the values 0 and 1 (approximately) equally often.

1.5.3 Known Plaintext Attacks on Stream Ciphers

An enemy attacking a stream cipher will usually have some knowledge of the plaintext for which he has observed corresponding ciphertext, typically in the form of a probability distribution on the possible values of the plaintext. As a worst case assumption, from the point of view of the communicating parties, we presume that the enemy has complete knowledge of plaintext as

well as ciphertext: thus we say that his attack is a *known plaintext* attack. In consequence, he knows the keystream sequence

$$b_1 b_2 \dots b_N.$$

This knowledge may, or may not, depending on the nature of the stream cipher, allow the corresponding key to be calculated (of course, this will not be possible for a good stream cipher).

In a practical situation, knowledge of the key for a particular message may of much greater use to an enemy than knowledge of a particular keystream sequence $b_1 b_2 \dots b_N$. For example, a system may encrypt each individual message by initialising a keystream generator with a simple function of a key and message-specific value. In this case a known plaintext attack on any one message may allow an enemy to determine the initial state of the keystream generator when encrypting that message, and hence to deduce the initial keystream generator state — and hence keystream — used to encrypt other messages. Notice also, that if an enemy can determine the keystream generator state corresponding to the start of a segment of keystream, he can recompute the entire keystream sequence.

1.5.4 Gallager's Decoding Algorithm

Finally in this section, we present an algorithm due to Gallager [6] which is effective in decoding certain binary block codes. Our interest in it stems from applications in the cryptanalysis of certain stream ciphers, such as those attacks described in [3]. We will consider an attack based on this algorithm in section 3.8, but here we concentrate on presenting the theory of the decoding algorithm, as presented in [6] and [3].

Suppose that a source generates a sequence $x_1, \dots, x_N$ of bits uniformly,

subject to m relations $\bigoplus_{j \in J_k} x_j = 0$ $(1 \leq k \leq m)$, where each $J_k \subseteq \{1, \ldots, N\}$ and $|J_k| \geq 2$. The sequence $(x_i)_{i=1}^N$ is transmitted over a binary symmetric channel — that is, each bit is corrupted with probability p independently of each other bit — and is received as a sequence $y_1, \ldots, y_N$.

In this section we denote by $\Pr_{\text{source}}$ a probability taken over sequences $(x_i)_{i=1}^N$, each with associated probability that with which it is generated by the source, and denote simply by Pr a probability taken over sequences $(x_i)_{i=1}^N$ with the uniform distribution (i.e. each sequence with probability 2^{-N}). Let p_i denote the probability $\Pr(x_i = 0 | (y_j)_{j=1}^N)$ $(1 \leq i \leq N)$, so that

$$p_i = \begin{cases} 1-p & \text{if } y_i = 0 \\ p & \text{if } y_i = 1. \end{cases}$$

For any index d $(1 \leq d \leq N)$, we can compute

$$\begin{aligned} &\Pr_{\text{source}}(x_d = 0 | (y_i)_{i=1}^N) \\ &= \Pr(x_d = 0 | (y_i)_{i=1}^N \text{ and relations } J_k \ (1 \leq k \leq m) \text{ hold}) \\ &= \frac{\Pr(\text{relations } J_k \ (1 \leq k \leq m) \text{ hold} | x_d = 0, (y_i)_{i=1}^N) \Pr(x_d = 0 | (y_i)_{i=1}^N)}{\Pr(\text{relations } J_k \ (1 \leq k \leq m) \text{ hold} | (y_i)_{i=1}^N)} \end{aligned} \tag{1.59}$$

by the definition of conditional probabilities. Now

$$\begin{aligned} &\Pr(\text{relations } J_k \ (1 \leq k \leq m) \text{ hold} | x_d = 0, (y_i)_{i=1}^N) \\ &= \prod_{k=1}^m \Pr(\text{relation } J_k \text{ holds} | x_d = 0, (y_i)_{i=1}^N) \qquad \text{if the } J_k \backslash \{d\} \text{ are disjoint} \\ &= \prod_{\substack{k=1 \\ d \in J_k}}^m \left(\frac{1}{2} \left(1 + \prod_{j \in J_k \backslash \{d\}} (2p_j - 1) \right) \right) \prod_{\substack{k=1 \\ d \notin J_k}}^m \left(\frac{1}{2} \left(1 + \prod_{j \in J_k} (2p_j - 1) \right) \right) \end{aligned} \tag{1.60}$$

if the probabilities $\Pr(x_i = 0|(y_j)_{j=1}^N)$ for $1 \le i \le N$ are independent, so that we can use lemma 1.19. In fact this is generally not true, but Gallager argues that equation (1.60) is nonetheless a good approximation.

Thus, with these assumptions, we have, by (1.59),

$$\begin{aligned}&\Pr_{\text{source}}(x_d = 0|(y_i)_{i=1}^N) \\ &= \frac{p_d \prod_{\substack{k=1 \\ d\in J_k}}^m \left(\frac{1}{2}\left(1 + \prod_{j\in J_k\setminus\{d\}}(2p_j - 1)\right)\right) \prod_{\substack{k=1 \\ d\notin J_k}}^m \left(\frac{1}{2}\left(1 + \prod_{j\in J_k}(2p_j - 1)\right)\right)}{\Pr(\text{relations } J_k \ (1 \le k \le m) \text{ hold}|(y_i)_{i=1}^N)}\end{aligned} \tag{1.61}$$

Similarly

$$\begin{aligned}&\Pr_{\text{source}}(x_d = 1|(y_i)_{i=1}^N) \\ &= \frac{(1-p_d) \prod_{\substack{k=1 \\ d\in J_k}}^m \left(\frac{1}{2}\left(1 - \prod_{j\in J_k\setminus\{d\}}(2p_j - 1)\right)\right) \prod_{\substack{k=1 \\ d\notin J_k}}^m \left(\frac{1}{2}\left(1 + \prod_{j\in J_k}(2p_j - 1)\right)\right)}{\Pr(\text{relations } J_k \ (1 \le k \le m) \text{ hold}|(y_i)_{i=1}^N)}\end{aligned} \tag{1.62}$$

Hence from (1.61) and (1.62)

$$\frac{\Pr_{\text{source}}(x_d = 0|(y_i)_{i=1}^N)}{\Pr_{\text{source}}(x_d = 1|(y_i)_{i=1}^N)} = \frac{p_d}{1-p_d} \prod_{\substack{k=1 \\ d\in J_k}}^m \frac{1 + \prod_{j\in J_k\setminus\{d\}}(2p_j - 1)}{1 - \prod_{j\in J_k\setminus\{d\}}(2p_j - 1)}$$

These calculations motivate the decoding algorithm due to Gallager [6]:

1. For $1 \le i \le N$, set
$$p_i := \begin{cases} 1-p & \text{if } y_i = 0 \\ p & \text{if } y_i = 1 \end{cases}$$
2. Given the sequence of probabilities $(p_i)_{i=1}^N$, generate a new sequence of probabilities $(p'_i)_{i=1}^N$ according to the equations
$$\frac{p'_i}{1-p'_i} = \frac{p_i}{1-p_i} \prod_{\substack{k=1 \\ i\in J_k}}^m \frac{1 + \prod_{j\in J_k\setminus\{i\}}(2p_j - 1)}{1 - \prod_{j\in J_k\setminus\{i\}}(2p_j - 1)} \qquad (1 \le i \le N) \tag{1.63}$$

3. Set $p_i := p'_i$ $(1 \leq i \leq N)$.

4. If the sequence $(p_i)_{i=1}^N$ has not converged, goto 2.

5. Recover a codeword $x'_1, \ldots, x'_N$ by setting

$$x'_i := \begin{cases} 0 & \text{if } p_i \geq \frac{1}{2} \\ 1 & \text{if } p_i < \frac{1}{2} \end{cases}$$

In fact, Gallager suggests that the contribution to p'_i due to a relation J_k should be computed using probability estimates p_i which did not make use of the relation J_k when they themselves were last re-estimated. If we denote by $p_{i,k}$ an estimate for $\Pr(x_i = 0|(y_i)_{i=1}^N)$ which didn't make use of relation J_k in the previous round, we obtain the following version of the algorithm:

1. For $1 \leq i, l \leq N$, set

$$p_{i,l} := \begin{cases} 1-p & \text{if } y_i = 0 \\ p & \text{if } y_i = 1 \end{cases}$$

2. Given the sequences of probabilities $(p_{i,l})_{i=1}^N$ $(1 \leq l \leq N)$, generate new sequences of probabilities $(p'_{i,l})_{i=1}^N$ $(1 \leq l \leq N)$ according to the equations

$$\frac{p'_{i,l}}{1-p'_{i,l}} = \frac{p_{i,l}}{1-p_{i,l}} \prod_{\substack{k=1 \\ i \in J_k, k \neq l}}^{m} \frac{1 + \prod_{j \in J_k \setminus \{i\}} (2p_{j,k} - 1)}{1 - \prod_{j \in J_k \setminus \{i\}} (2p_{j,k} - 1)} \qquad (1 \leq i \leq N)$$

3. Set $p_{i,l} := p'_{i,l}$ for $1 \leq i, l \leq N$.

4. If not all the sequences $(p_{i,l})_{i=1}^N$ have converged, goto 2.

5. Recover a codeword $x'_1, \ldots, x'_N$ by setting

$$x'_i := \begin{cases} 0 & \text{if } p_{i,1} \geq \frac{1}{2} \\ 1 & \text{if } p_{i,1} < \frac{1}{2} \end{cases}$$

Gallager also notes that the calculations (1.63) can be simplified by using log-likelihood ratios; if we write

$$L(x) := \ln \frac{x}{1-x} \qquad (-1 < x < 1)$$

and

$$s(x) := \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

then, after some calculations, we can re-express (1.63) as

$$L(p_i') = L(p_i) + \sum_{\substack{k=1 \\ i \in J_k}}^{m} \left\{ \left(\prod_{j \in J_k \setminus \{i\}} s(L(p_j)) \right) f \left[\sum_{j \in J_k \setminus \{i\}} f(|L(p_j)|) \right] \right\},$$

where

$$f(\beta) := \ln \frac{e^\beta + 1}{e^\beta - 1}$$

Canteaut and Trabbia [3] cite an approximation from [7], which in this context becomes

$$L(p_i') \approx L(p_i) + \sum_{\substack{k=1 \\ i \in J_k}}^{m} \left\{ \left(\prod_{j \in J_k \setminus \{i\}} s(L(p_j)) \right) \min_{j \in J_k \setminus \{i\}} |L(p_j)| \right\}$$

This leads to the following reformulation of the basic Gallager algorithm:

1. For $1 \leq i \leq N$, set

$$L_i := \begin{cases} L(p) & \text{if } y_i = 0 \\ -L(p) & \text{if } y_i = 1 \end{cases}$$

2. For $1 \leq i \leq N$, set

$$L_i' := L_i + \sum_{\substack{k=1 \\ i \in J_k}}^{m} \left(\prod_{j \in J_k \setminus \{i\}} s(L_j) \right) \min_{j \in J_k \setminus \{i\}} |L_j|$$

3. Set $L_i := L_i'$ $(1 \leq i \leq N)$.

4. If the sequence $(L_i)_{i=1}^N$ has not converged, goto 2.

5. Recover a codeword $x'_1, \ldots, x'_N$ by setting

$$x'_i := \begin{cases} 0 & \text{if } L_i \geq 0 \\ 1 & \text{if } L_i < 0 \end{cases}$$

Canteaut and Trabbia consider the particular case where

- each J_k has cardinality s;
- the relations J_k satisfy the shift property

$$J_k \text{ is a relation, } l \in \mathbb{Z} \text{ and } J_k + l \subseteq \{1, \ldots N\} \Rightarrow J_k + l \text{ is a relation.}$$

In these circumstances, the authors claim, on the basis of experimental evidence, that the condition for the reformulated algorithm to converge to the correct solution is approximately

$$m(s) \geq K_s / i_1((1-2p)^{s-2}),$$

where

$$m(s) = |\{k : 1 \leq k \leq N \text{ and } 1 \in J_k\}|$$

and

$$K_3 \approx 2, K_s \approx 1 \text{ for } s \geq 4.$$

Chapter 2

A Problem in Information Theory

2.1 Introduction

In this chapter we take as our starting point a problem in communication theory, which we present in section 2.2. This, as we show in section 2.3, can be reformulated in terms of the information function I_α introduced in section 1.4.2 to produce a result we conjecture to be true, at least for some values of α. We show how the DFT provides a simple and natural proof of this conjecture for $\alpha = 2$, from which we are able to solve our original problem. We then work towards a proof of the conjecture, or a generalised form of it, for Shannon information ($\alpha = 1$), deriving, on the way, a number of results of interest in their own right concerning the DFT and information theoretic functions. Unfortunately, despite strenuous efforts, the truth of these conjectures for $\alpha = 1$ remains undetermined. Finally, we prove that neither the conjecture nor the generalised conjecture holds for all $\alpha \geq 1$.

2.2 The Problem

Suppose that an originator generates an n-bit vector V[1] according to the uniform probability distribution on such vectors, and sends the n bits of V over a channel to a receiver, who receives the n bits as a vector W. We suppose that the channel is *binary symmetric*, so that each bit of W is the corresponding bit of V corrupted with error probability $p \leq \frac{1}{2}$ independently of any corruptions in other positions. We ask the following question: is it possible for the originator and receiver to agree on choices for balanced n-bit to 1-bit functions f and g prior to the generation of V in such a way that $f(V)$ and $g(W)$ agree with probability greater than $1 - p$? Note that by a *balanced* function we mean one that assumes the values 0 and 1 for equally many arguments.

2.3 Information-Theoretic Reformulation

For any real $\alpha \geq 1$, we can re-express the problem of section 2.2 by asking whether or not

$$I_\alpha(f(V) \oplus g(W)) \leq i_\alpha(q - p)$$

for all balanced f and g, where we have written $q := 1 - p$. We can prove the following result:

Proposition 2.1. *Fix any real $\alpha \geq 1$. Then with the notation of section 2.2, $f(V)$ and $g(W)$ agree with probability $\leq 1 - p$ for all choices of balanced functions f and g if and only if*

$$I_\alpha(f(V) \oplus g(W)) \leq i_\alpha(q - p) \tag{2.2}$$

[1]Throughout this chapter V will represent an n-bit vector as described here, and not a map $G \to \mathbb{C}$ as in the previous chapter.

for all balanced f and g.

Proof. Write $r := \Pr(f(V) = g(W))$, and $s := 1 - r$. Then

$$I_\alpha(f(V) \oplus g(W)) = i_\alpha(r - s)$$

and

$$\begin{aligned} I_\alpha(f(V) \oplus g(W)) &\leq i_\alpha(q - p) \\ \Leftrightarrow \quad i_\alpha(r - s) &\leq i_\alpha(q - p) \\ \Leftrightarrow \quad |r - s| &\leq q - p, \end{aligned}$$

since, by lemma 1.55, $i_\alpha(x)$ is convex and $i_\alpha(x) = i_\alpha(-x)$ for all $x \in [-1, 1]$. But

$$\begin{aligned} |r - s| &\leq q - p \\ \Leftrightarrow \quad 2r - 1 &\leq 2q - 1 \text{ and } 1 - 2r \leq 1 - 2p \\ \Leftrightarrow \quad p &\leq r \leq 1 - p. \end{aligned}$$

So if (2.2) holds, $f(V)$ and $g(W)$ agree with probability $\leq 1-p$. And if $f(V)$ and $g(W)$ agree with probability $\leq 1-p$ for all balanced f and g, then, noting that f is balanced if and only if its pointwise complement $1 - f$ is balanced, so also $f(V)$ and $g(W)$ agree with probability $\geq p$ for all balanced f and g, so equation (2.2) also holds for all balanced f and g. □

However, we also have the following result.

Lemma 2.3. *If X and Y are balanced Bernouilli random variables, then, for any real $\alpha \geq 1$, $I_\alpha(X \oplus Y) = I_\alpha(X; Y)$.*

Proof. Let a denote the probability $\Pr(X = 0, Y = 0)$. Since $\Pr(X = 0) = \Pr(Y = 0) = \frac{1}{2}$, we must have $\Pr(X = 0, Y = 0) = \Pr(X = 1, Y = 1) = a$,

and $\Pr(X=0,Y=1) = \Pr(X=1,Y=0) = \frac{1}{2} - a$. For $\alpha > 1$, it follows from definitions 1.47 and 1.49 of section 1.4 that

$$\begin{aligned} I_\alpha(X \oplus Y) &= \frac{1}{\alpha - 1}\left(\frac{1}{2}\left(2a/\frac{1}{2}\right)^\alpha + \frac{1}{2}\left((1-2a)/\frac{1}{2}\right)^\alpha\right) - 1 \quad \text{and} \\ I_\alpha(X;Y) &= \frac{1}{\alpha - 1}\left(2.\frac{1}{4}\left(a/\frac{1}{4}\right)^\alpha + 2.\frac{1}{4}\left((\frac{1}{2}-a)/\frac{1}{4}\right)^\alpha\right) - 1. \end{aligned}$$

Hence $I_\alpha(X \oplus Y) = I_\alpha(X;Y)$ for all $\alpha > 1$. The definitions and the continuity result of section 1.4.1 imply that equality holds for $\alpha = 1$ too. □

Hence equation (2.2) can be written

$$I_\alpha(f(V); g(W)) \leq i_\alpha(q-p),$$

or

$$I_\alpha(f(V); g(V \oplus E)) \leq i_\alpha(q-p)$$

if we write E for the error vector associated with transmission over the binary symmetric channel.

In fact, we conjecture that a rather stronger result holds, at least for some values $\alpha \geq 1$. As in section 1.3.3, we let G denote — here and for the remainder of this chapter — the elementary abelian 2-group of n-bit vectors under componentwise modulo 2 addition. Then for any function $g : G \to \{0,1\}$ (not necessarily balanced)

$$I_\alpha(V; g(V \oplus E)) \leq i_\alpha(q-p).$$

Note that this really is a stronger result by virtue of proposition 1.53. We strengthen this reformulation further by letting g map each element $v \in G$ to a distribution X_v on $\{0,1\}$, determined by a real $g(v) \in [0,1]$ for which

$$\Pr_{X_v}(1) = g(v).$$

If $g(v) = 0$, X_v always assumes the value 0, and if $g(v) = 1$, X_v always assumes the value 1, so this new definition is genuinely an extension of the previous one.

We state this new conjecture formally, in terms of a parameter $\alpha \geq 1$.

Conjecture 2.4. *Suppose that V and E are two independent random variables whose values range over all n-bit vectors, assumed with uniform probability in the case of V, and with probability given by $\Pr(e) = p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e}$ in the case of E (where $\mathrm{wt}(e)$ is the weight of e, i.e. the number of 1s in e). Let f be any map $G \to [0, 1]$, which we identify with a map $G \to \{$distributions on $\{0, 1\}\}$, in the manner described above. Then the following inequality holds for all real $p \in [0, 1]$:*

$$I_\alpha(V; f(V \oplus E)) \leq i_\alpha(q - p). \tag{2.5}$$

Corollary 2.6. *If conjecture 2.4 holds for at least one value $\alpha \geq 1$, then the question posed in section 2.2 can be answered in the negative.*

Proof. If conjecture 2.4 holds for any $\alpha \geq 1$, then for any balanced f and g

$$\begin{aligned} I_\alpha(f(V) \oplus g(W)) &= I_\alpha(f(V); g(V \oplus E)) \qquad \text{by lemma 2.3} \\ &\leq I_\alpha(V; g(V \oplus E)) \qquad \text{by proposition 1.53} \\ &\leq i_\alpha(q - p) \qquad \text{by conjecture 2.4,} \end{aligned}$$

and then the result follows from proposition 2.1. □

2.4 Proof of the Conjecture for $\alpha = 2$

In this section we make use of the Walsh-Hadamard transform to prove conjecture 2.4 for order 2 information. As in section 1.3.3, let G denote the

elementary abelian 2 group of order 2^n, and for each $f \in \mathbb{C}^G$, let $D(f)$ denote the DFT of f given by

$$D(f)(v) = \sum_{u \in G} (-1)^{u \cdot v} f(u).$$

Lemma 2.7. *Define $b \in \mathbb{C}^G$ by $b(v) := p^{\mathrm{wt}\, v} q^{n - \mathrm{wt}\, v}$ for each $v \in G$. Then $D(b)$ is given by $D(b)(v) = (q-p)^{\mathrm{wt}\, v}$.*

Proof. By induction. The result is certainly true for $n = 1$, when $b(0) = q$ and $b(1) = p$ so $D(b)(0) = q + p = 1$ and $D(b)(1) = q - p$. For any $n > 1$, assume that it is true for $n - 1$. Fix any $v \in G$, and, as in section 1.3.3.3, write $v = (v', v_n)$ for an $(n-1)$-bit vector v' and bit v_n. Then

$$\begin{aligned}
&D(b)(v) \\
&= \sum_{u \in G} (-1)^{u \cdot v} b(u) \\
&= \sum_{u=(u',0) \in G} (-1)^{u \cdot v} p^{\mathrm{wt}\, u} q^{n - \mathrm{wt}\, u} + \sum_{u=(u',1) \in G} (-1)^{u \cdot v} p^{\mathrm{wt}\, u} q^{n - \mathrm{wt}\, u} \\
&\qquad\qquad \text{where } u' \text{ is an } (n-1)\text{-bit vector} \\
&= q \sum_{(u',0) \in G} (-1)^{u' \cdot v'} p^{\mathrm{wt}\, u'} q^{n-1-\mathrm{wt}\, u'} \\
&\qquad + (-1)^{v_n} p \sum_{(u',1) \in G} (-1)^{u' \cdot v'} p^{\mathrm{wt}\, u'} q^{n-1-\mathrm{wt}\, u'} \\
&= (q + (-1)^{v_n} p)(q-p)^{\mathrm{wt}\, v'} \qquad \text{by the inductive hypothesis} \\
&= (q-p)^{\mathrm{wt}\, v}.
\end{aligned}$$

Thus the result holds for n. Hence by induction it holds for all n. □

Now we can use the properties of the Walsh-Hadamard transform to obtain our proof.

Proposition 2.8. *Conjecture 2.4 is true for the particular value $\alpha = 2$.*

Proof. First of all, let g be any map in $\mathbb{R}^G$. For convenience write a for the map in $\mathbb{R}^G$ defined by $a(v) := \sum_{e \in G} p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e} g(v \oplus e)$, and b for the map $b(v) := p^{\mathrm{wt}\, v} q^{n - \mathrm{wt}\, v}$ $(v \in G)$. Note that $a = b \otimes g$. Hence

$$\begin{aligned}
&\sum_{v \in G} \left(\sum_{e \in G} p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e} g(v \oplus e) \right)^2 \\
&= \sum_{v \in G} a(v)^2 \\
&= \|a\|^2 \\
&= 2^{-n} \|D(a)\|^2 \qquad \text{by theorem 1.16, part 1} \\
&= 2^{-n} \|D(b) D(g)\|^2 \qquad \text{by theorem 1.16, part 2} \\
&= 2^{-n} \sum_{v \in G} (q - p)^{2\, \mathrm{wt}\, v} \left(D(g)(v) \right)^2 \qquad \text{by lemma 2.7.}
\end{aligned}$$

Now suppose that f is any map $G \to [0, 1]$. Observe first that

$$2^{-n} D(f)(0) = 2^{-n} \sum_{u \in G} f(u) = \Pr(f(u) = 1 | u \text{ is uniformly distributed})$$

and similarly

$$2^{-n} D(1 - f)(0) = \Pr(f(u) = 0 | u \text{ is uniformly distributed}).$$

We will also use the fact that

$$\begin{aligned}
2^{-n} \sum_{v \in G} \left(D(f)(v) \right)^2 &= \sum_{v \in G} \left(f(v) \right)^2 \qquad \text{by theorem 1.16, part 1} \\
&\leq \sum_{v \in G} f(v)
\end{aligned}$$

and similarly

$$2^{-n} \sum_{v \in G} \left(D(1 - f)(v) \right)^2 \leq \sum_{v \in G} (1 - f)(v).$$

Now we compute

$$
\begin{aligned}
& I_2(V; f(V \oplus E)) \\
&= \left(\sum_{\substack{v \in G \\ i=0,1}} \Pr(f(V \oplus E) = i) \Pr(V = v) \left(\frac{\Pr(f(V \oplus E) = i, V = v)}{\Pr(f(V \oplus E) = i) \Pr(V = v)} \right)^2 \right) \\
&\qquad - 1 && \text{by definitions 1.49 and 1.43} \\
&= \left(2^{-n} \sum_{\substack{v \in G \\ i=0,1}} \frac{\Pr(f(V \oplus E) = i | V = v)^2}{\Pr(f(V \oplus E) = i)} \right) - 1 \\
&= 2^{-n} \sum_{v \in G} \frac{\left(\sum_{e \in G} p^{\operatorname{wt} e} q^{n - \operatorname{wt} e} f(v \oplus e) \right)^2}{2^{-n} D(f)(0)} \\
&\quad + 2^{-n} \sum_{v \in G} \frac{\left(\sum_{e \in G} p^{\operatorname{wt} e} q^{n - \operatorname{wt} e} (1 - f(v \oplus e)) \right)^2}{2^{-n} D(1-f)(0)} - 1 \\
&= \frac{2^{-n}}{D(f)(0)} \left(\sum_{v \in G} (q-p)^{2 \operatorname{wt} v} \left(D(f)(v) \right)^2 \right) \\
&\quad + \frac{2^{-n}}{D(1-f)(0)} \left(\sum_{v \in G} (q-p)^{2 \operatorname{wt} v} \left(D(1-f)(v) \right)^2 \right) - 1 \\
&\qquad && \text{by the considerations above} \\
&\leq \frac{2^{-n}}{D(f)(0)} \left(\left(D(f)(0) \right)^2 + \sum_{v \in G \setminus \{0\}} (q-p)^2 \left(D(f)(v) \right)^2 \right) \\
&\quad + \frac{2^{-n}}{D(1-f)(0)} \left(\left(D(1-f)(0) \right)^2 + \sum_{v \in G \setminus \{0\}} (q-p)^2 \left(D(1-f)(v) \right)^2 \right) \\
&\quad - 1 && \text{separating } \textstyle\sum_v \text{ into } \sum_{v=0} + \sum_{v \neq 0}
\end{aligned}
$$

$$= \frac{2^{-n}}{D(f)(0)} \sum_{v \in G \setminus \{0\}} (q-p)^2 \left(D(f)(v)\right)^2$$

$$+ \frac{2^{-n}}{D(1-f)(0)} \sum_{v \in G \setminus \{0\}} (q-p)^2 \left(D(1-f)(v)\right)^2$$

since $D(f)(0) + D(1-f)(0) = 2^n$

$$\leq \frac{2^{-n}}{D(f)(0)} (q-p)^2 \left(2^n D(f)(0) - D(f)(0)^2\right)$$

$$+ \frac{2^{-n}}{D(1-f)(0)} (q-p)^2 \left(2^n D(1-f)(0) - D(1-f)(0)^2\right)$$

by the results above

$$= (q-p)^2 \left(1 - 2^{-n} D(f)(0) + 1 - 2^{-n} D(1-f)(0)\right)$$

$$= (q-p)^2$$

$$= i_2(q-p).$$

Thus we have proved the reformulation (2.5) of conjecture 2.4 for the case $\alpha = 2$. □

As an immediate consequence of corollary 2.6, we have the following:

Corollary 2.9. *The question posed in section 2.2 can be answered in the negative.*

2.5 Towards a Proof for $\alpha = 1$

We saw in the previous section that the DFT gave a natural proof of conjecture 2.4 for order 2 information. In this section we work toward a proof the result for Shannon information, i.e. the case $\alpha = 1$. Although we don't reach our goal, we do succeed in deriving a number of results of interest in their own right.

We start with a definition.

Definition 2.10. For any function $f : G \to [0,1]$ we denote by $\tilde{f}$ the function $G \to [-1,1]$ defined by the rule

$$\tilde{f}(v) := 1 - 2f(v) \qquad (v \in G).$$

Now we can rewrite equation (2.5) of conjecture 2.4 in a new form which will prove to be convenient in due course. We do this by expanding its L.H.S.:

$$\begin{aligned}
&I_1(V; f(V \oplus E)) \\
&= I_1(f(V \oplus E)|V) - I_1(f(V \oplus E)) \qquad \text{by lemma 1.52} \\
&= 2^{-n} \sum_{v \in G} i_1 \left(\Pr(f(V \oplus E) = 0 | V = v) - \Pr(f(V \oplus E) = 1 | V = v)\right) \\
&\quad - i_1 \left(\Pr(f(V \oplus E) = 0) - \Pr(f(V \oplus E) = 1)\right) \\
&\qquad\qquad \text{by definitions 1.51 and 1.54.}
\end{aligned}$$

Now

$$\begin{aligned}
&\Pr(f(V \oplus E) = 0 | V = v) - \Pr(f(V \oplus E) = 1 | V = v) \\
&= \sum_{e \in G} \Pr(E = e) \left(\Pr(f(V \oplus E) = 0 | V = v, E = e)\right. \\
&\qquad \left. - \Pr(f(V \oplus E) = 1 | V = v, E = e)\right) \\
&= \sum_{e \in G} p^{\operatorname{wt} e} q^{n - \operatorname{wt} e} \left(1 - f(v \oplus e) - f(v \oplus e)\right) \\
&= \sum_{e \in G} p^{\operatorname{wt} e} q^{n - \operatorname{wt} e} \tilde{f}(v \oplus e).
\end{aligned}$$

Also

$$\begin{aligned}
&\Pr(f(V \oplus E) = 0) - \Pr(f(V \oplus E) = 1) \\
&= 2^{-n} \sum_{e \in G} (1 - f(v \oplus e)) - 2^{-n} \sum_{e \in G} f(v \oplus e) \\
&= 2^{-n} \sum_{v \in G} \tilde{f}(v).
\end{aligned}$$

Thus we have proved:

Lemma 2.11. *With our usual notation,*

$$\begin{aligned}
&I_1(V; f(V \oplus E)) \\
&= 2^{-n} \sum_{v \in G} i_1 \left(\sum_{e \in G} p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e} \tilde{f}(v \oplus e) \right) - i_1 \left(2^{-n} \sum_{v \in G} \tilde{f}(v) \right),
\end{aligned}$$

and for $\alpha = 1$ equation (2.5) becomes

$$2^{-n} \sum_{v \in G} i_1 \left(\sum_{e \in G} p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e} \tilde{f}(v \oplus e) \right) \leq i_1 \left(2^{-n} \sum_{v \in G} \tilde{f}(v) \right) + i_1(q - p).$$

Although we did not highlight the fact, it was intrinsic to the proof of conjecture 2.4 for the case $\alpha = 2$ that

$$\frac{1}{2} \left(i_2(a) + i_2(b) \right) = i_2 \left(\frac{a+b}{2} \right) + i_2 \left(\frac{a-b}{2} \right)$$

for all a and b in $[-1, 1]$. This suggests that we should consider a similar result for $\alpha = 1$.

Lemma 2.12. *For $a, b \in [-1, 1]$,*

$$\frac{1}{2} (i_1(a) + i_1(b)) \geq i_1 \left(\frac{a+b}{2} \right) + i_1 \left(\frac{a-b}{2} \right). \tag{2.13}$$

We will shortly offer two proofs of this, but first we establish some preliminary results.

Lemma 2.14. *The Taylor expansion of i_1 about 0 is given by*

$$i_1(x) = \sum_{i=1}^{\infty} \frac{x^{2i}}{2i(2i-1)}$$

which converges for all $x \in (-1, 1)$.

Proof. Since, for $x \in (-1, 1)$,

$$i_1(x) = \frac{1}{2}(1+x)\ln(1+x) + \frac{1}{2}(1-x)\ln(1-x),$$

the first and second derivatives of i_1 on $(-1, 1)$ are given by

$$\begin{aligned} i_1'(x) &= \frac{1}{2}\ln(1+x) - \frac{1}{2}\ln(1-x) \quad \text{and} \\ i_1''(x) &= \frac{1}{1-x^2}. \end{aligned}$$

Hence $i_1''(x)$ is given by the series

$$i_1''(x) = \sum_{i=1}^{\infty} x^{2i},$$

convergent on $(-1, 1)$, from which, given that $i_1(x) = 1$ and $i_1'(x) = 0$, the required result follows by integrating the R.H.S. twice, term by term; this is justified by, for example, theorem 9.23 on page 236 of [1]. □

Lemma 2.15. *For $a, b \in [-1, 1]$, write $x = \frac{a+b}{2}$ and $y = \frac{a-b}{2}$. Then, provided $x \neq 1$ or -1,*

$$\frac{1}{2}(i_1(x+y) + i_1(x-y)) = i_1(x) + \frac{1}{2}(1-x)i_1\left(\frac{y}{1-x}\right) + \frac{1}{2}(1+x)i_1\left(\frac{y}{1+x}\right).$$

Proof. We expand the R.H.S.:

$$\begin{aligned}
&i_1(x) + \frac{1}{2}(1-x)i_1\left(\frac{y}{1-x}\right) + \frac{1}{2}(1+x)i_1\left(\frac{y}{1+x}\right) \\
&= i_1(x) \\
&\quad + \frac{1}{2}(1-x)\left(\frac{1-x+y}{2(1-x)}\ln\left(\frac{1-x+y}{1-x}\right) + \frac{1-x-y}{2(1-x)}\ln\left(\frac{1-x-y}{1-x}\right)\right) \\
&\quad + \frac{1}{2}(1+x)\left(\frac{1+x+y}{2(1+x)}\ln\left(\frac{1+x+y}{1+x}\right) + \frac{1+x-y}{2(1+x)}\ln\left(\frac{1+x-y}{1+x}\right)\right) \\
&= i_1(x) - \left(\frac{1-x}{2}\ln(1-x) + \frac{1+x}{2}\ln(1+x)\right) \\
&\quad + \left(\frac{1+(x+y)}{4}\right)\ln\left(1+(x+y)\right) + \left(\frac{1-(x+y)}{4}\right)\ln\left(1-(x+y)\right) \\
&\quad + \left(\frac{1+(x-y)}{4}\right)\ln\left(1+(x-y)\right) + \left(\frac{1-(x-y)}{4}\right)\ln\left(1-(x-y)\right) \\
&= \frac{1}{2}(i_1(x+y) + i_1(x-y)),
\end{aligned}$$

which is the L.H.S., as required. □

Now we are in a position to give the two proofs we promised above.

Proof using Taylor expansion of i_1. Suppose a and b are both in the range $(-1, 1)$. Write

$$\begin{aligned}
A &:= \frac{a+b}{2} \qquad \text{and} \\
B &:= \frac{a-b}{2},
\end{aligned}$$

so that A, B, $A+B$ and $A-B$ are all in the range $(-1, 1)$. Now, for any integer $r \geq 0$,

$$\begin{aligned}
(A+B)^{2r} + (A-B)^{2r} &= \sum_{i=0}^{r}\binom{2r}{2i}A^{2i}B^{2r-2i} \\
&\geq 2\left(A^{2r} + B^{2r}\right).
\end{aligned}$$

Hence, comparing corresponding terms in the Taylor expansions,

$$i_1(A+B) + i_1(A-B) \geq 2\left(i_1(A) + i_1(B)\right),$$

i.e.

$$\frac{1}{2}(i_1(a) + i_1(b)) \geq i_1\left(\frac{a+b}{2}\right) + i_1\left(\frac{a-b}{2}\right),$$

as required. This result for a, b in the range $(-1, 1)$ extends to $[-1, 1]$ by continuity of $i_1(x)$ on the closed interval. □

Our second proof is as follows:

Proof using convexity of i_1. Suppose a and b are in the range $(-1, 1)$, and write, as above,

$$\begin{aligned} A &:= \frac{a+b}{2} \qquad \text{and} \\ B &:= \frac{a-b}{2}, \end{aligned}$$

so that A, B, $A+B$ and $A-B$ are all in the range $(-1, 1)$. Then

$$\begin{aligned} &\frac{1}{2}(i_1(a) + i_1(b)) \\ &= \frac{1}{2}(i_1(A+B) + i_1(A-B)) \\ &= i_1(A) + \frac{1}{2}(1-A) i_1\left(\frac{B}{1-A}\right) + \frac{1}{2}(1+A) i_1\left(\frac{B}{1+A}\right) \quad \text{by lemma 2.15} \\ &\geq i_1(A) \\ &\qquad + i_1\left(\frac{1}{2}(1-A)\left(\frac{B}{1-A}\right) + \frac{1}{2}(1+A)\left(\frac{B}{1+A}\right)\right) \quad \text{by convexity of } i_1 \\ &= i_1(A) + i_1(B) \\ &= i_1\left(\frac{a+b}{2}\right) + i_1\left(\frac{a-b}{2}\right), \end{aligned}$$

as required. As before, a continuity argument extends the result to $a, b \in [-1, 1]$. □

Having established the truth of lemma 2.12, we see now that it implies the following interesting result concerning the DFT and i_1.

Proposition 2.16. *Suppose that f is a function defined on the elementary abelian group G of order 2^n which takes real values in the range $[-1, 1]$. Then*

$$2^{-n} \sum_{v \in G} i_1(f(v)) \geq \sum_{v \in G} i_1(2^{-n} D(f)(v)). \tag{2.17}$$

Proof. Recall that

$$D(f)(v) = \sum_{u \in G} (-1)^{u \cdot v} f(u).$$

The proof proceeds by induction on n.

For $n = 1$, we need to show that

$$\frac{1}{2} \left(i_1(a) + i_1(b) \right) \geq i_1 \left(\frac{a+b}{2} \right) + i_1 \left(\frac{a-b}{2} \right)$$

for a, b in $[-1, 1]$, which is precisely the assertion of lemma 2.12.

For any $n > 1$ suppose that the result holds for $n - 1$. As observed in 1.3.3.3, we have a natural isomorphism $G \simeq G' \times C_2$ so that an element $v \in G$ can be written $v = (v', v_n)$ for an $(n-1)$-bit vector $v' \in G'$ and bit $v_n \in C_2$. Now

$$\begin{aligned}
& 2^{-n} \sum_{v \in G} i_1(f(v)) \\
& = 2^{-n} \sum_{v=(v',0) \in G} i_1(f(v)) + 2^{-n} \sum_{v=(v',1) \in G} i_1(f(v)) \\
& = 2^{-(n-1)} \sum_{v' \in G'} \frac{1}{2} \left(i_1(f((v',0))) + i_1(f((v',1))) \right) \\
& \geq 2^{-(n-1)} \sum_{v' \in G'} i_1 \left(\frac{f((v',0)) + f((v',1))}{2} \right) + \\
& \quad 2^{-(n-1)} \sum_{v' \in G'} i_1 \left(\frac{f((v',0)) - f((v',1))}{2} \right) \quad \text{by lemma 2.12}
\end{aligned}$$

$$\geq \sum_{v' \in G'} i_1 \left(2^{-(n-1)} \sum_{u' \in G'} (-1)^{u' \cdot v'} \left(\frac{f((u',0)) + f((u',1))}{2} \right) \right) +$$

$$\sum_{v' \in G'} i_1 \left(2^{-(n-1)} \sum_{u' \in G'} (-1)^{u' \cdot v'} \left(\frac{f((u',0)) - f((u',1))}{2} \right) \right)$$

by the inductive hypothesis

$$= \sum_{v' \in G'} i_1 \left(2^{-n} \sum_{\substack{u' \in G' \\ b=0,1}} (-1)^{(u',b)\cdot(v',0)} f((u',b)) \right) +$$

$$\sum_{v' \in G'} i_1 \left(2^{-n} \sum_{\substack{u' \in G' \\ b=0,1}} (-1)^{(u',b)\cdot(v',1)} f((u',b)) \right)$$

$$= \sum_{v \in G} i_1 \left(2^{-n} \sum_{u \in G} (-1)^{u \cdot v} f(u) \right).$$

Thus the result also holds for n.

The result follows for all $n \geq 1$ by induction. □

Next we apply this result in the context of the conjecture to obtain the following:

Corollary 2.18. *With our usual notation,*

$$I_1(V; f(V \oplus E)) \geq \sum_{v \in G \setminus \{0\}} i_1 \left(2^{-n} D(\tilde{f})(v)(q-p)^{\operatorname{wt} v} \right).$$

Proof. As before, write a for the map in $\mathbb{C}^G$ defined by

$$a(v) = \sum_{e \in G} p^{\operatorname{wt} e} q^{n - \operatorname{wt} e} \tilde{f}(v \oplus e),$$

and b for the map

$$b(v) = p^{\operatorname{wt} v} q^{n - \operatorname{wt} v}.$$

Then

L.H.S.

$$= 2^{-n} \sum_{v \in G} i_1 \left(\sum_{e \in G} p^{\operatorname{wt} e} q^{n - \operatorname{wt} e} \tilde{f}(v \oplus e) \right) - i_1 \left(2^{-n} \sum_{v \in G} \tilde{f}(v) \right)$$

by lemma 2.11

$$= 2^{-n} \sum_{v \in G} i_1(a(v)) - i_1 \left(2^{-n} D(\tilde{f})(0) \right)$$

$$\geq \sum_{v \in G} i_1 \left(2^{-n} D(a)(v) \right) - i_1 \left(2^{-n} D(\tilde{f})(0) \right) \qquad \text{by the preceding result}$$

$$= \sum_{v \in G} i_1 \left(2^{-n} D(b \otimes \tilde{f})(v) \right) - i_1 \left(2^{-n} D(\tilde{f})(0) \right)$$

$$= \sum_{v \in G} i_1 \left(2^{-n} D(b)(v) D(\tilde{f})(v) \right) - i_1 \left(2^{-n} D(\tilde{f})(0) \right)$$

by theorem 1.16, part 2

$$= \sum_{v \in G} i_1 \left(2^{-n} (q-p)^{\operatorname{wt} v} D(\tilde{f})(v) \right) - i_1 \left(2^{-n} D(\tilde{f})(0) \right) \qquad \text{by lemma 2.7}$$

$$= \sum_{v \in G \setminus \{0\}} i_1 \left(2^{-n} (q-p)^{\operatorname{wt} v} D(\tilde{f})(v) \right)$$

$$= \text{ R.H.S.} \qquad \square$$

This result is interesting because it gives us an inequality on the quantity $I_1(V; f(V \oplus E))$ of the opposite kind to that of conjecture 2.4. The functions

$$f(x_1, x_2, \ldots, x_n) = x_r$$

for $r = 1, \ldots, n$ and their mod 2 complements are precisely those choices for which $D(\tilde{f})(v)$ vanishes except on vectors v of weight $\operatorname{wt} v = 1$, and are precisely those functions for which the bounds of conjecture 2.4 and corollary 2.18 are both attained by $I(V; f(V \oplus E))$. We can establish this claim with the help of the following result:

Proposition 2.19. *Equality is achieved in equation (2.17) of proposition 2.16 iff f is a map $v \mapsto a(-1)^{u \cdot v}$ for some $a \in [-1, 1]$ and $u \in G$.*

Proof. First we consider lemma 2.12. From the proof we gave using the Taylor expansion of i_1, it is clear that we have equality in equation (2.13) iff $A = 0$ or $B = 0$, where $A = \frac{a+b}{2}$ and $B = \frac{a-b}{2}$: i.e. iff $a = b$ or $a = -b$.

We can now prove proposition 2.19 by induction on n, where $|G| = 2^n$.

For $n = 1$, proposition 2.16 follows directly from lemma 2.12, and we have equality iff

$$f(0) = a, f(1) = a \text{ or } f(0) = a, f(1) = -a \quad \text{for some } a \in [-1, 1]$$

i.e. iff

$$f : v \mapsto a(-1)^{u \cdot v} \quad \text{for some } a \in [-1, 1], u \in G.$$

Suppose now that $n > 1$ and this result holds for $n - 1$. As previously, we identify G with $G' \times C_2$, where G' is elementary abelian of order 2^{n-1} and C_2 is cyclic of order 2, and write any $v \in G$ as $v = (v', v_n)$ for an $(n-1)$-bit vector $v' \in G'$ and bit $v_n \in C_2$. By the proof of proposition 2.16, we see that we have equality in equation (2.17) iff

$$f(v', 0) = f(v', 1) \text{ or } f(v', 0) = -f(v', 1) \text{ for each } v' \in G' \tag{2.20}$$

and the maps

$$v' \mapsto \frac{f(v', 0) + f(v', 1)}{2}$$

and

$$v' \mapsto \frac{f(v', 0) - f(v', 1)}{2}$$

both give equality in proposition 2.16. Suppose these conditions are satisfied. By the inductive hypothesis,

$$\begin{aligned} \frac{f(v', 0) + f(v', 1)}{2} &= a_0(-1)^{u_0' \cdot v'} \text{ for all } v' \in G, \text{ and} \\ \frac{f(v', 0) - f(v', 1)}{2} &= a_1(-1)^{u_1' \cdot v'} \text{ for all } v' \in G \end{aligned}$$

for some values $a_0, a_1 \in [-1, 1]$ and $u'_0, u'_1 \in G'$. Considering $v' = 0$ in equation (2.20), either $f(v', 0) = f(v', 1)$ or $f(v', 0) = -f(v', 1)$. In the first case, $a_1 = 0$ and, for all $v' \in G', v_n \in C_2$,

$$\begin{aligned} f(v', v_n) &= \left(\frac{f(v', 0) + f(v', 1)}{2}\right) + (-1)^{v_n}\left(\frac{f(v', 0) - f(v', 1)}{2}\right) \\ &= a_0(-1)^{u'_0 \cdot v'} \\ &= a_0(-1)^{(u'_0, 0)\cdot(v', v_n)} \end{aligned}$$

In the second case, $a_0 = 0$ and, for all $v' \in G', v_n \in C_2$,

$$\begin{aligned} f(v', v_n) &= \left(\frac{f(v', 0) + f(v', 1)}{2}\right) + (-1)^{v_n}\left(\frac{f(v', 0) - f(v', 1)}{2}\right) \\ &= (-1)^{v_n} a_1(-1)^{u'_0 \cdot v'} \\ &= a_1(-1)^{(u'_0, 1)\cdot(v', v_n)} \end{aligned}$$

Thus, either way, f is of the required form $v \mapsto a(-1)^{u \cdot v}$. Conversely, it is not hard to see that if f is of this form then the two conditions above are satisfied, and we have equality in equation (2.17).

The result follows for all $n \geq 1$ by induction. □

Corollary 2.21. *If $q - p \neq 0$, the functions*

$$f(x_1, x_2, \ldots, x_n) = x_r$$

and

$$f(x_1, x_2, \ldots, x_n) = x_r \oplus 1$$

for $r = 1, \ldots, n$ are precisely those functions for which the bounds of conjecture 2.4 and corollary 2.18 are both attained by $I(V; f(V \oplus E))$.

Proof. Assume that $q - p \neq 0$. By the proof of corollary 2.18 and proposition 2.19, it follows that

$$I_1(V; f(V \oplus E)) = \sum_{v \in G \setminus \{0\}} i_1\left(2^{-n} D(\tilde{f})(v)(q - p)^{\mathrm{wt}\, v}\right)$$

iff $\tilde{f}(v) = a(-1)^{u \cdot v}$ for all $v \in V$, for some $a \in [-1, 1]$ and some $u \in G$, or, equivalently, iff

$$2^{-n} D(\tilde{f})(v) = \begin{cases} a & \text{if } v = u; \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the equation

$$I_1(V; f(V \oplus E)) = i_1(q - p)$$

holds also iff $u \neq 0$ and

$$i_1(a(q-p)^{\operatorname{wt} u}) = i_1(q - p)$$

i.e. iff $a = 1$ or -1 and $\operatorname{wt} u = 1$. These parameters correspond precisely to the functions specified in the statement of the result. □

Consider now the expression

$$2^{-n} \sum_{v \in G} i_\alpha (\sum_{e \in G} p^{\operatorname{wt} e} q^{n - \operatorname{wt} e} \tilde{f}(v \oplus e)).$$

We have seen that this is

$$2^{-n} \sum_{v \in G} i_\alpha ((b \otimes \tilde{f})(v))$$

(with b defined as in the proof of corollary 2.18); this in turn can be expressed

$$2^{-n} \sum_{v \in G} i_\alpha (2^{-n} D(D(b \otimes \tilde{f}))(v))$$

i.e.

$$2^{-n} \sum_{v \in G} i_\alpha \left(\sum_{u \in G} (-1)^{u \cdot v} 2^{-n} D(\tilde{f})(u) (q-p)^{\operatorname{wt} u} \right)$$

We record this fact for future reference:

Lemma 2.22. *With the usual notation,*

$$\begin{aligned} & 2^{-n} \sum_{v \in G} i_\alpha \left(\sum_{e \in G} p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e} \tilde{f}(v \oplus e) \right) \\ = \;& 2^{-n} \sum_{v \in G} i_\alpha \left(\sum_{u \in G} (-1)^{u \cdot v} 2^{-n} D(\tilde{f})(u) (q-p)^{\mathrm{wt}\, u} \right) \end{aligned}$$

We note also that $2^{-n} D(\tilde{f})(0) = 2^{-n} \sum_{v \in G} \tilde{f}(v)$ is the average value of $\tilde{f}$. These considerations motivated the following conjecture:

Conjecture 2.23. *Suppose we have m real values $x_r \in [-1, 1]$ for $1 \leq r \leq m$. Write $\bar{x}$ for the average of the x_r, $\bar{x} = \frac{1}{m} \sum_{r=1}^{m} x_r$. Then for all real $\lambda \in [0, 1]$,*

$$\frac{1}{m} \sum_{r=1}^{m} i_1(\bar{x} + \lambda(x_r - \bar{x})) \leq i_1(\bar{x}) + i_1(\lambda).$$

However, even if it can be proved, conjecture 2.23 does not provide the immediate proof of conjecture 2.4 for $\alpha = 1$ that we are seeking. To see this, set $n = 3$, and define

$$\tilde{f}(v) = \begin{cases} 1 & \text{if } \mathrm{wt}\, v \leq 1 \\ -1 & \text{if } \mathrm{wt}\, v \geq 2 \end{cases}$$

Thus $\tilde{f}$ is balanced. But consider the term for $v = 0$ in

$$\sum_{v \in G} i_1 (\sum_{e \in G} p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e} \tilde{f}(v \oplus e)).$$

This equals

$$i_1(q^3 + 3q^2 p - 3qp^2 - p^3),$$

and

$$\begin{aligned} & q^3 + 3q^2p - 3qp^2 - p^3 \\ = \;& (q-p)(q^2 + pq + p^2 + 3pq) \\ = \;& (q-p)((p+q)^2 + 2pq) \\ = \;& (q-p)(1 + 2pq) \\ > \;& q - p \qquad \text{if } p, q \neq 0. \end{aligned}$$

To continue our search for a proof of conjecture 2.4 for $\alpha = 1$, we return to proposition 2.16 and corollary 2.18. For convenience, write, for each $v \in G$,

$$c_v := 2^{-n} D(\tilde{f})(v).$$

By corollary 2.18,

$$I_1(V; f(V \oplus E)) \geq \sum_{v \in G \setminus \{0\}} i_1(c_v (q-p)^{\operatorname{wt} v}).$$

We can derive an upper bound for the R.H.S. of this equation using proposition 2.16 in two different ways. First, observe that for the function

$$\tilde{g} : v \mapsto c_0 + (q-p)(\tilde{f}(v) - c_0)$$

we have

$$2^{-n} D(\tilde{g})(v) = \begin{cases} c_0 & \text{if } v = 0 \\ (q-p)c_v & \text{if } v \neq 0 \end{cases}$$

Therefore

$$\begin{aligned} & \sum_{v \in G \setminus \{0\}} i_1(c_v (q-p)^{\operatorname{wt} v}) \\ \leq \;& \sum_{v \in G \setminus \{0\}} i_1(c_v (q-p)) \\ \leq \;& -i_1(c_0) + 2^{-n} \sum_{v \in G} i_1(c_0 + (q-p)(\tilde{f}(v) - c_0)) \qquad \text{by proposition 2.16} \end{aligned}$$

Alternatively,

$$\begin{aligned}
&\sum_{v\in G\setminus\{0\}} i_1(c_v(q-p)^{\mathrm{wt}\,v}) \\
&\leq \sum_{v\in G\setminus\{0\}} i_1(c_v(q-p)) \\
&= -i_1(c_0(q-p)) + \sum_{v\in G} i_1(c_v(q-p)) \\
&\leq -i_1(c_0(q-p)) + 2^{-n}\sum_{v\in G} i_1((q-p)\tilde{f}(v)) \qquad \text{by proposition 2.16.}
\end{aligned}$$

We can now hope, optimistically, that the effect of the $\geq$ inequality is more than compensated for by the $\leq$ one, and that one of the inequalities

$$I_1(V; f(V\oplus E)) \leq -i_1(c_0) + 2^{-n}\sum_{v\in G} i_1(c_0 + (q-p)(\tilde{f}(v) - c_0)) \qquad (2.24)$$

and

$$I_1(V; f(V\oplus E)) \leq -i_1(c_0(q-p)) + 2^{-n}\sum_{v\in G} i_1((q-p)\tilde{f}(v)) \qquad (2.25)$$

holds for all $f : G \to [0,1]$. Recall that, by lemma 2.11,

$$I_1(V; f(V\oplus E)) = 2^{-n}\sum_{v\in G} i_1(\sum_{e\in G} p^{\mathrm{wt}\,e} q^{n-\mathrm{wt}\,e}\tilde{f}(v\oplus e)) - i_1(c_0)\,.$$

The first of these, (2.24), is a promising looking inequality, because, if true, conjecture 2.4 for $\alpha = 1$ might follow by applying conjecture 2.23 to its R.H.S.. Moreover, when $q-p=0$, both sides equal 0; when $q-p=1$, both sides equal $2^{-n}\sum_{v\in G} i_1(\tilde{f}(v)) - i_1(c_0)$; and both sides are equal when $n=1$. However a computer search reveals that it does not always hold, even for $n=2$ (and hence for all $n\geq 2$). For example, with the parameters

$$\tilde{f}(00) = -0.4,\ \tilde{f}(01) = 0.5,\ \tilde{f}(10) = 0.5,\ \tilde{f}(11) = 1,\ q-p = 0.9,$$

the L.H.S. $> 1.004 \times$ R.H.S..

The second of these, inequality (2.25), is not true even for $n = 1$, when in fact precisely the opposite inequality holds:

Proposition 2.26. *For any $A, B \in [-1, 1]$ and $\lambda \in [0, 1]$*

$$\begin{aligned} &\frac{1}{2}\left(i_1(A + \lambda B) + i_1(A - \lambda B)\right) - i_1(A) \\ &\quad \geq \quad \frac{1}{2}\left(i_1(\lambda(A + B)) + i_1(\lambda(A - B))\right) - i_1(\lambda A) \end{aligned}$$

with equality iff $A = 0$ or $B = 0$ or $\lambda = 1$.

Proof. By continuity, it suffices to establish the result for $A, B \in (-1, 1)$. For any $r \geq 1$,

$$\begin{aligned} &\frac{1}{2}\left((A + \lambda B)^{2r} + (A - \lambda B)^{2r}\right) - A^{2r} \\ &\quad = \quad \sum_{i=0}^{r} \binom{2r}{2i} A^{2i}(\lambda B)^{2r-2i} - A^{2r} \\ &\quad = \quad \sum_{i=1}^{r} \binom{2r}{2i} A^{2i}(\lambda B)^{2r-2i} \\ &\quad \geq \quad \sum_{i=1}^{r} \binom{2r}{2i} (\lambda A)^{2i}(\lambda B)^{2r-2i} \\ &\quad = \quad \frac{1}{2}\left((\lambda(A + B))^{2r} + (\lambda(A - B))^{2r}\right) - (\lambda A)^{2r}, \end{aligned}$$

with equality iff $A = 0$ or $B = 0$ or $\lambda = 1$. The result follows after dividing both sides by $2r(2r - 1)$ and summing for $r = 1, 2, \ldots$, on account of the Taylor expansion

$$i_1(x) = \sum_{r=1}^{\infty} \frac{x^{2r}}{2r(2r - 1)}$$

established in lemma 2.14. □

To finish off this rather lengthy section, we present (in proposition 2.28) another inequality in $I_1(V; f(V \oplus E))$. First though, we need a preliminary lemma.

Lemma 2.27. *For reals $a, b \in [-1, 1]$ and real $\lambda \in [0, 1]$,*

$$i_1\left(\frac{a+b}{2} + \lambda\frac{a-b}{2}\right) + i_1\left(\frac{a+b}{2} - \lambda\frac{a-b}{2}\right) \leq i_1(a) + i_1(b).$$

Proof. Denote the L.H.S. by $g(\lambda)$. g is twice differentiable on $(0, 1)$ with

$$\begin{aligned} g'(0) &= 0; \\ g''(\lambda) &= \left(\frac{a-b}{2}\right)^2 i_1''\left(\frac{a+b}{2} + \lambda\frac{a-b}{2}\right) + \\ &\quad \left(\frac{a-b}{2}\right)^2 i_1''\left(\frac{a+b}{2} - \lambda\frac{a-b}{2}\right) \\ &\geq 0 \qquad \text{by convexity of } i_1. \end{aligned}$$

Hence $g(x)$ is non-decreasing on $[0, 1]$, so is bounded above on that interval by $g(1) = i_1(a) + i_1(b)$. □

Proposition 2.28. *With notation as usual,*

$$2^{-n}\sum_{v\in G} i_1\Big(\sum_{e\in G} p^{\mathrm{wt}\, e} q^{n-\mathrm{wt}\, e}\tilde{f}(v \oplus e)\Big) \leq 2^{-n}\sum_{v\in G} i_1(\tilde{f}(v)).$$

Proof. We proceed by induction. For $n = 1$ the L.H.S. equals

$$\begin{aligned} &\frac{1}{2}\Big(i_1(q\tilde{f}(0) + p\tilde{f}(1)) + i_1(p\tilde{f}(0) + q\tilde{f}(1))\Big) \\ &= \frac{1}{2}\left(i_1\left(\frac{\tilde{f}(0) + \tilde{f}(1)}{2} + (q-p)\frac{\tilde{f}(0) - \tilde{f}(1)}{2}\right) + \right. \\ &\quad \left. + i_1\left(\frac{\tilde{f}(0) + \tilde{f}(1)}{2} - (q-p)\frac{\tilde{f}(0) - \tilde{f}(1)}{2}\right)\right) \\ &\leq \frac{1}{2}\Big(i_1(\tilde{f}(0)) + i_1(\tilde{f}(1))\Big) \qquad \text{by lemma 2.27,} \end{aligned}$$

as required.

Suppose now that $n > 1$, and the result holds for $n - 1$. As in section 1.3.3.3, we identify the group G of n-bit vectors with the group $G' \times C_2$

of pairs (u', u_n), where $u' \in G'$ is an $(n-1)$-bit vector and $u_n \in C_2$ is a bit. For convenience we denote by $\tilde{f}_i$, for $i = 1, 2$, the map $G' \to [-1, 1]$ given by $v' \mapsto \tilde{f}(v', i)$. Now we calculate

$$
\begin{aligned}
& 2^{-n} \sum_{v \in G} i_1 \left(\sum_{e \in G} p^{\mathrm{wt}\, e} q^{n - \mathrm{wt}\, e} \tilde{f}(v \oplus e) \right) \\
&= 2^{-n} \left[\sum_{v=(v',0) \in G} i_1 \left(\sum_{e=(e',0) \in G} p^{\mathrm{wt}\, e} q^{n-\mathrm{wt}\, e} \tilde{f}(v \oplus e) \right.\right. \\
&\quad \left. + \sum_{e=(e',1) \in G} p^{\mathrm{wt}\, e} q^{n-\mathrm{wt}\, e} \tilde{f}(v \oplus e) \right) \\
&\quad + \sum_{v=(v',1) \in G} i_1 \left(\sum_{e=(e',0) \in G} p^{\mathrm{wt}\, e} q^{n-\mathrm{wt}\, e} \tilde{f}(v \oplus e) \right. \\
&\quad \left.\left. + \sum_{e=(e',1) \in G} p^{\mathrm{wt}\, e} q^{n-\mathrm{wt}\, e} \tilde{f}(v \oplus e)) \right) \right] \\
&= 2^{-n} \left[\sum_{(v',0) \in G} i_1 \left(q \sum_{(e',0) \in G} p^{\mathrm{wt}\, e'} q^{n-1-\mathrm{wt}\, e'} \tilde{f}_0(v' \oplus e') \right.\right. \\
&\quad \left. + p \sum_{(e',1) \in G} p^{\mathrm{wt}\, e'} q^{n-1-\mathrm{wt}\, e'} \tilde{f}_1(v' \oplus e') \right) \\
&\quad + \sum_{(v',1) \in G} i_1 \left(q \sum_{(e',0) \in G} p^{\mathrm{wt}\, e'} q^{n-1-\mathrm{wt}\, e'} \tilde{f}_1(v' \oplus e') \right. \\
&\quad \left.\left. + p \sum_{(e',1) \in G} p^{\mathrm{wt}\, e'} q^{n-1-\mathrm{wt}\, e'} \tilde{f}_0(v' \oplus e')) \right) \right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq 2^{-n} \sum_{v' \in G'} \left[i_1 \left(\sum_{e' \in G'} p^{\mathrm{wt}\, e'} q^{n-1-\mathrm{wt}\, e'} \tilde{f}_0(v' \oplus e') \right) \right. \\
&\quad \left. + i_1 \left(\sum_{e' \in G'} p^{\mathrm{wt}\, e'} q^{n-1-\mathrm{wt}\, e'} \tilde{f}_1(v' \oplus e') \right) \right] \qquad \text{by lemma 2.27} \\
&\leq 2^{-n} \sum_{v' \in G'} \left[i_1 \left(\tilde{f}_0(v') \right) + i_1 \left(\tilde{f}_1(v') \right) \right] \qquad \text{by the inductive hypothesis} \\
&= 2^{-n} \sum_{v \in G} i_1 \left(\tilde{f}(v) \right),
\end{aligned}
$$

so the result holds for n.

The result is therefore proven by induction. □

2.6 The Generalised Conjecture

In the previous section we considered two inequalities, equations (2.24) and (2.25), neither of which lead to a proof of conjecture 2.4. In this section we consider the following variant of equation (2.25)

$$I_\alpha(V; f(V \oplus E)) \leq 2^{-n} \sum_{v \in G} i_\alpha((q-p)\tilde{f}(v)),$$

to which we give the epithet "generalised conjecture". If it holds for a function f then so does conjecture 2.4 for $\alpha = 1$, since

$$i_\alpha((q-p)a) \leq i_\alpha(q-p)$$

for any $a \in [-1, 1]$. Indeed, if f is $\{0,1\}$-valued, then it is precisely the inequality of conjecture 2.4.

Using lemma 2.11, we can equivalently express the generalised conjecture for $\alpha = 1$ as

$$2^{-n} \sum_{v \in G} i_1 \left(\sum_{e \in G} p^{\mathrm{wt}\, e} q^{n-\mathrm{wt}\, e} \tilde{f}(v \oplus e) \right) \leq i_1 \left(\overline{\tilde{f}} \right) + 2^{-n} \sum_{v \in G} i_1 \left((q-p)\tilde{f}(v) \right) \tag{2.29}$$

where $\overline{\tilde{f}} = 2^{-n}\sum_{v\in G}\tilde{f}(v)$, or, using lemma 2.22, as

$$2^{-n}\sum_{v\in G} i_1\left(\sum_{u\in G}(-1)^{u\cdot v}2^{-n}D(\tilde{f})(u)(q-p)^{\mathrm{wt}\,u}\right)$$
$$\leq i_1\left(2^{-n}D(\tilde{f})(0)\right) + 2^{-n}\sum_{v\in G} i_1\left(2^{-n}\sum_{u\in G}(-1)^{u\cdot v}D(\tilde{f})(u)(q-p)\right) \quad (2.30)$$

2.6.1 Proof of Generalised Conjecture for $\alpha = 1$, $n = 1$

In this section we prove the generalised conjecture for the case $\alpha = 1$ and $n = 1$. For convenience, write

$$\begin{aligned} A &:= \frac{1}{2}D(\tilde{f})(0), \\ B &:= \frac{1}{2}D(\tilde{f})(1), \text{ and} \\ \lambda &:= q-p, \end{aligned}$$

so that (2.30) becomes

$$\frac{1}{2}\left(i_1(A+\lambda B) + i_1(A-\lambda B)\right) \leq i_1\left(A\right) + \frac{1}{2}\left(i_1(\lambda\left(A+B\right)) + i_1(\lambda\left(A-B\right))\right)$$

i.e.

$$\begin{aligned} i_1\left(A\right) + \frac{1}{2}\left(i_1(\lambda\left(A+B\right)) + i_1(\lambda\left(A-B\right))\right) \\ - \frac{1}{2}\left(i_1(A+\lambda B) + i_1(A-\lambda B)\right) \geq 0 \end{aligned}$$

The value of the L.H.S. is 0 at $\lambda = 0$, so it suffices to show that its derivative with respect to λ is ≥ 0 on $[0,1]$, i.e.

$$\begin{aligned} \frac{1}{2}\left(\left(A+B\right)i_1'(\lambda\left(A+B\right)) + \left(A-B\right)i_1'(\lambda\left(A-B\right))\right) \\ - \frac{1}{2}\left(Bi_1'(A+\lambda B) - Bi_1'(A-\lambda B)\right) \geq 0. \end{aligned}$$

The value of the L.H.S. of this new inequality is also 0 at $\lambda = 0$, so, applying the same argument once again, it suffices to show that *its* derivative with respect to λ is ≥ 0 on $[0, 1]$, i.e.

$$\frac{1}{2}\left((A+B)^2 i_1''(\lambda(A+B)) + (A-B)^2 i_1''(\lambda(A-B))\right) - \frac{1}{2}B^2\left(i_1''(A+\lambda B) + i_1''(A-\lambda B)\right) \geq 0.$$

Hence it suffices to show that

$$\frac{(A+B)^2}{1-\lambda^2(A+B)^2} + \frac{(A-B)^2}{1-\lambda^2(A-B)^2} - \frac{B^2}{1-(A+\lambda B)^2} - \frac{B^2}{1-(A-\lambda B)^2} \geq 0.$$

To do this we first consider the differences

$$\begin{aligned} &\frac{(A+B)^2}{1-\lambda^2(A+B)^2} + \frac{(A-B)^2}{1-\lambda^2(A-B)^2} - \left(\frac{B^2}{1-\lambda^2B^2} + \frac{B^2}{1-\lambda^2B^2}\right) \\ &= 2A^2\frac{1+3\lambda^2B^2-\lambda^2A^2}{\left(1-\lambda^2(A+B)^2\right)\left(1-\lambda^2(A-B)^2\right)\left(1-\lambda^2B^2\right)} \end{aligned}$$

and

$$\begin{aligned} &\left(\frac{B^2}{1-(A+\lambda B)^2} + \frac{B^2}{1-(A-\lambda B)^2}\right) - \left(\frac{B^2}{1-\lambda^2B^2} + \frac{B^2}{1-\lambda^2B^2}\right) \\ &= 2B^2A^2\frac{1+3\lambda^2B^2-A^2}{\left(1-(A+\lambda B)^2\right)\left(1-(A-\lambda B)^2\right)\left(1-\lambda^2B^2\right)} \end{aligned}$$

We readily see that it suffices to show that

$$\frac{1}{\left(1-\lambda^2(A+B)^2\right)\left(1-\lambda^2(A-B)^2\right)} - \frac{B^2}{\left(1-(A+\lambda B)^2\right)\left(1-(A-\lambda B)^2\right)} \geq 0,$$

or

$$\begin{aligned} &\left(1-(A+\lambda B)^2\right)\left(1-(A-\lambda B)^2\right) \\ &\quad - B^2\left(1-\lambda^2(A+B)^2\right)\left(1-\lambda^2(A-B)^2\right) \geq 0. \end{aligned} \tag{2.31}$$

The L.H.S. equals

$$\begin{aligned}
&\left(1-A^2-\lambda^2B^2-2\lambda AB\right)\left(1-A^2-\lambda^2B^2+2\lambda AB\right)\\
&\quad -B^2\left(1-\lambda^2A^2-\lambda^2B^2-2\lambda^2AB\right)\left(1-\lambda^2A^2-\lambda^2B^2+2\lambda^2AB\right)\\
&=\left(1-A^2-\lambda^2B^2\right)^2-4\lambda^2A^2B^2\\
&\quad -B^2\left(\left(1-\lambda^2B^2-\lambda^2A^2\right)^2-4\lambda^4A^2B^2\right)\\
&=\left(1-A^2-\lambda^2B^2\right)^2-B^2\left(1-\lambda^2B^2-\lambda^2A^2\right)^2\\
&\quad -4\lambda^2A^2B^2(1-\lambda^2B^2).
\end{aligned}$$

However,

$$\begin{aligned}
&\left(1-\lambda^2B^2+\lambda^2A^2\right)^2-\left(1-\lambda^2B^2-\lambda^2A^2\right)^2\\
&\quad = \quad 4\lambda^2A^2\left(1-\lambda^2B^2\right),
\end{aligned}$$

so we can rewrite inequality (2.31) as

$$\left(1-A^2-\lambda^2B^2\right)^2-B^2\left(1-\lambda^2B^2+\lambda^2A^2\right)^2\geq 0.$$

Since our original expression was symmetric in B and $-B$ we can assume without any loss of generality that $B\geq 0$. Hence it is sufficient to show that

$$1-A^2-\lambda^2B^2-B\left(1-\lambda^2B^2+\lambda^2A^2\right)\geq 0.$$

However

$$\begin{aligned}
&1-A^2-\lambda^2B^2(1-B)-B-\lambda^2A^2B\\
&\quad \geq \quad 1-A^2-B^2(1-B)-B-A^2B\\
&\quad = \quad (1+B)\left(1-(A+B)\right)\left(1+(A-B)\right)\\
&\quad \geq \quad 0,
\end{aligned}$$

as required.

2.6.2 Extrema of the Generalised Conjecture

In this section we consider a new approach towards a proof of the generalised conjecture for $\alpha = 1$. Once again we use techniques from calculus, although the strategy differs from that used in the previous section. The principle here is to extremise the difference

$$2^{-n} \sum_{v \in G} i_1 \left(\sum_{u \in G} (-1)^{u \cdot v} c_u (q-p)^{\mathrm{wt}\, u} \right) - \left(i_1(c_0) + 2^{-n} \sum_{v \in G} i_1 \left(\sum_{u \in G} (-1)^{u \cdot v} c_u (q-p) \right) \right) \tag{2.32}$$

over reals c_v $(v \in G)$ in the region $T \subseteq \mathbb{R}^{|G|}$ defined by

$$(c_v)_{v \in G} \in T \Leftrightarrow \left| \sum_{u \in G} c_u (-1)^{u \cdot v} \right| < 1 \quad \text{for all } v \in G.$$

For convenience, write λ for $q-p$, as before. Setting the partial derivative of equation (2.32) with respect to each c_w $(w \in G)$ equal to 0, we obtain the following set of equations satisfied at an extremum $(c_v)_{v \in G}$ of (2.32):

$$\sum_{v \in G} i_1' \left(\sum_{u \in G} (-1)^{u \cdot v} c_u \lambda^{\mathrm{wt}\, u} \right) (-1)^{w \cdot v} \lambda^{\mathrm{wt}\, w}$$

$$= \begin{cases} \sum_{v \in G} i_1' \left(\sum_{u \in G} (-1)^{u \cdot v} c_u \lambda \right) (-1)^{w \cdot v} \lambda & \text{for all } w \in G \setminus \{0\}; \\ 2^n i_1'(c_0) + \sum_{v \in G} i_1' \left(\sum_{u \in G} (-1)^{u \cdot v} c_u \lambda \right) \lambda & \text{at } w = 0. \end{cases}$$

Note that $i_1'(-x) = -i_1'(x)$ for all $x \in (-1, 1)$, so these equations can be rewritten

$$\sum_{v \in G} i_1' \left(\sum_{u \in G} (-1)^{(u+w) \cdot v} c_u \lambda^{\mathrm{wt}\, u} \right) \lambda^{\mathrm{wt}\, w}$$

$$= \begin{cases} \sum_{v \in G} i_1' \left(\sum_{u \in G} (-1)^{(u+w) \cdot v} c_u \lambda \right) \lambda & (w \in G \setminus \{0\}); \\ 2^n i_1'(c_0) + \sum_{v \in G} i_1' \left(\sum_{u \in G} (-1)^{u \cdot v} c_u \lambda \right) \lambda & (w = 0). \end{cases}$$

Exponentiating each side of this equation, we obtain the following equations, which determine an extremum:

$$\left(\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\operatorname{wt}u}}{1-\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\operatorname{wt}u}}\right)^{\lambda^{(\operatorname{wt}w)-1}}=\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda}{1-\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda} \tag{2.33a}$$

for all $w\in G\setminus\{0\}$, and, corresponding to $w=0$,

$$\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{u\cdot v}c_u\lambda^{\operatorname{wt}u}}{1-\sum_{u\in G}(-1)^{u\cdot v}c_u\lambda^{\operatorname{wt}u}}=\left(\frac{1+c_0}{1-c_0}\right)^{2^n}\left(\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{u\cdot v}c_u\lambda}{1-\sum_{u\in G}(-1)^{u\cdot v}c_u\lambda}\right)^{\lambda} \tag{2.33b}$$

We have the following rather limited result concerning extrema.

Theorem 2.34. *If the function*

$$\tilde{f}:v\mapsto d_{u_1}(-1)^{v\cdot u_1}+d_{u_2}(-1)^{v\cdot u_2}$$

for distinct elements $u_1,u_2\in G\setminus\{0\}$ defines a map $G\to[-1,1]$, $\tilde{f}$ extremises the generalised conjecture for any $\lambda=q-p\in(0,1)$ iff $\operatorname{wt}u_1=1$ if $d_{u_1}\neq 0$ and $\operatorname{wt}u_2=1$ if $d_{u_2}\neq 0$.

Proof. For this function $\tilde{f}$, the values $c_v:=2^{-n}D(\tilde{f})(v)$ $(v\in G)$ satisfy

$$c_v=\begin{cases} d_v & \text{if } v=u_1 \text{ or } v=u_2;\\ 0 & \text{otherwise.}\end{cases}$$

First, suppose that the latter condition is satisfied: we will show then that $\tilde{f}$ extremises the generalised conjecture, i.e. equations (2.33) hold for all $w\in G$. They certainly hold whenever $\operatorname{wt}w=1$, so fix any $w\in G$ for which $\operatorname{wt}w\neq 1$. Choose $v'\in G$ so that $(u_i+w)\cdot v'=1$ for each $i\in\{0,1\}$ for which $d_{u_i}\neq 0$.

Then

$$\begin{aligned}
&\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}{1-\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}\\
&\quad=\quad\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{(u+w)\cdot (v+v')}c_u\lambda^{\mathrm{wt}\,u}}{1-\sum_{u\in G}(-1)^{(u+w)\cdot (v+v')}c_u\lambda^{\mathrm{wt}\,u}}\\
&\quad=\quad\prod_{v\in G}\frac{1+(-1)^{(u+w)\cdot v'}\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}{1-(-1)^{(u+w)\cdot v'}\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}\\
&\quad=\quad 1\Big/\left(\prod_{v\in G}\frac{1-\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}{1+\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}\right)
\end{aligned}$$

from which we conclude that

$$\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}{1-\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda^{\mathrm{wt}\,u}}=1.$$

By the same argument

$$\prod_{v\in G}\frac{1+\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda}{1-\sum_{u\in G}(-1)^{(u+w)\cdot v}c_u\lambda}=1,$$

and, noting, for the case $w=0$, that $c_0=0$, it follows that equation (2.33) holds. Hence $\tilde{f}$ extremises the generalised conjecture.

Suppose, conversely, that $\tilde{f}$ extremises the generalised conjecture. By interchanging the labels u_1 and u_2 if necessary, we may assume that $\mathrm{wt}\,u_1\leq\mathrm{wt}\,u_2$. Suppose for a contradiction that $1\leq\mathrm{wt}\,u_1<\mathrm{wt}\,u_2$ and $d_{u_2}\neq 0$. Since $\tilde{f}$ is an extremum iff $-\tilde{f}$ is, we may assume also that $d_{u_2}>0$. Choose any $v'\in G$ for which $v'\cdot(u_1+u_2)=1$. Now, with the intention of demonstrating that the extremum equation for $w=u_2$ does not hold, we compute

$$\begin{aligned}
&\left(\prod_{v\in G}\frac{1+(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1-(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}\right)^2\\
&\quad=\quad\prod_{v\in G}\left[\left(\frac{1+(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1-(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}\right)\right.\\
&\qquad\left.\left(\frac{1+(-1)^{(u_1+u_2)\cdot (v+v')}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1-(-1)^{(u_1+u_2)\cdot (v+v')}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \prod_{v\in G}\left[\left(\frac{1+(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1-(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}\right)\right. \\
&\qquad \left.\left(\frac{1-(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1+(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}\right)\right] \\
&= \prod_{v\in G}\left(\frac{\left(1+d_{u_2}\lambda^{\mathrm{wt}\,u_2}\right)^2-d_{u_1}^2\lambda^{2\,\mathrm{wt}\,u_1}}{\left(1+d_{u_2}\lambda^{\mathrm{wt}\,u_2}\right)^2-d_{u_1}^2\lambda^{2\,\mathrm{wt}\,u_1}}\right) \qquad \text{(note this is } \geq 1) \\
&< \prod_{v\in G}\left(\frac{\left(1+d_{u_2}\lambda\right)^2-d_{u_1}^2\lambda^2}{\left(1-d_{u_2}\lambda\right)^2-d_{u_1}^2\lambda^2}\right) \\
&= \left(\prod_{v\in G}\frac{1+(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1-(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}\right)^2
\end{aligned}
$$

where the last line is obtained from the previous by an argument similar to that connecting all the lines above those two. Therefore

$$
\begin{aligned}
&\left(\prod_{v\in G}\frac{1+(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1-(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}\right)^{\lambda^{\mathrm{wt}\,u_2-1}} \\
&\qquad < \prod_{v\in G}\frac{1+(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}+d_{u_2}\lambda^{\mathrm{wt}\,u_2}}{1-(-1)^{(u_1+u_2)\cdot v}d_{u_1}\lambda^{\mathrm{wt}\,u_1}-d_{u_2}\lambda^{\mathrm{wt}\,u_2}}
\end{aligned}
$$

contradicting the extremum equation corresponding to $w = u_2$. The result follows. □

2.7 Conjecture and Generalised Conjecture for Large α

In this concluding section of this chapter we show that neither the conjecture (conjecture 2.4) nor the generalised conjecture (section 2.6) holds for all f and all $\alpha \geq 1$.

Consider the quantity $I_\alpha(V; f(V \oplus E))$ for balanced f and $\alpha > 1$. Let (X, Y) denote the joint distribution $(V, f(V \oplus E))$, so that for any $v \in V$

and $i \in \{0,1\}$,

$$\begin{aligned}\Pr_{(X,Y)}(v,i) &= 2^{-n}\Pr(f(V\oplus E)=i|V=v)\\ &= \begin{cases} 2^{-n}\sum_{e\in G} p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}(1-f(v\oplus e)) & \text{for } i=0;\\ 2^{-n}\sum_{e\in G} p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}f(v\oplus e) & \text{for } i=1.\end{cases}\end{aligned}$$

Hence

$$\begin{aligned}&I_\alpha(V; f(V\oplus E))\\ &= d_\alpha((X,Y);X,Y)\\ &= \frac{1}{\alpha-1}\left[\sum_{v\in G}2^{-n}\frac{1}{2}\left(\frac{\sum_{e\in G} f(v\oplus e)p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}}{1/2}\right)^\alpha\right.\\ &\quad\left.+\sum_{v\in G}2^{-n}\frac{1}{2}\left(\frac{\sum_{e\in G}(1-f(v\oplus e))p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}}{1/2}\right)^\alpha-1\right]\end{aligned}$$

Since

$$i_\alpha(q-p)=\frac{1}{\alpha-1}\left(\frac{1}{2}\left((1+(q-p))^\alpha+(1-(q-p))^\alpha\right)-1\right)$$

it follows that

$$\begin{aligned}&I_\alpha(V; f(V\oplus E)) \le i_\alpha(q-p)\\ &\Leftrightarrow \sum_{v\in G}2^{-n}\frac{1}{2}\left(\frac{\sum_{e\in G} f(v\oplus e)p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}}{1/2}\right)^\alpha\\ &\quad+\sum_{v\in G}2^{-n}\frac{1}{2}\left(\frac{\sum_{e\in G}(1-f(v\oplus e))p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}}{1/2}\right)^\alpha\\ &\quad\le\frac{1}{2}\left((1+(q-p))^\alpha+(1-(q-p))^\alpha\right)\\ &\Leftrightarrow \left[\sum_{v\in G}2^{-n}\left(\frac{\sum_{e\in G} f(v\oplus e)p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}}{1/2}\right)^\alpha\right.\\ &\quad\left.+\sum_{v\in G}2^{-n}\left(\frac{\sum_{e\in G}(1-f(v\oplus e))p^{\mathrm{wt}\, e}q^{n-\mathrm{wt}\, e}}{1/2}\right)^\alpha\right]^{1/\alpha}\\ &\quad\le[(2q)^\alpha+(2p)^\alpha]^{1/\alpha}\end{aligned}$$

If n and f are fixed, then for sufficiently large α this condition becomes

$$\begin{aligned}\max_{v\in G}\left\{\frac{\sum_{e\in G} f(v\oplus e)p^{\mathrm{wt}\,e}q^{n-\mathrm{wt}\,e}}{1/2}, \frac{\sum_{e\in G}(1-f(v\oplus e))p^{\mathrm{wt}\,e}q^{n-\mathrm{wt}\,e}}{1/2}\right\}\\ \leq \quad 2q\end{aligned}$$

i.e.

$$\max_{v\in G}\left\{\left|\sum_{e\in G} f(v\oplus e)p^{\mathrm{wt}\,e}q^{n-\mathrm{wt}\,e} - \sum_{e\in G}(1-f(v\oplus e))p^{\mathrm{wt}\,e}q^{n-\mathrm{wt}\,e}\right|\right\} \leq q-p$$

However, we have already seen a counterexample to this — the function f defined, for $n=3$, by

$$\tilde{f}(v) = \begin{cases} 1 & \text{if } \mathrm{wt}\,v \leq 1; \\ -1 & \text{if } \mathrm{wt}\,v \geq 2. \end{cases}$$

Hence neither conjecture nor generalised conjecture holds for all f, n and α.

Chapter 3

A Problem In Cryptology

3.1 Introduction

In section 1.5 of chapter 1, we introduced the definition of a stream cipher, and of a known plaintext attack on a stream cipher. In this chapter we concern ourselves with the problem of known plaintext cryptanalysis of stream ciphers arising from keystream generators whose state space is an elementary abelian 2-group of order 2^n, which we identify with the vector space of n-bit vectors over the field GF(2) with 2 elements. In this chapter we let V denote this group (rather than the space of maps from the group to $\mathbb{C}$, as we had in chapter 1).

In section 3.2 we consider the implications of replacing the deterministic equations for the keystream bits in terms of the initial state by approximations, which will, generally, be probabilistic. We assess how many such derived equations are needed to obtain a result about the initial state. From section 3.3 onwards, we focus our attention on the case when state transition is vector space isomorphism of V — the stream cipher is "linearly clocking". It turns out that the DFT is useful, both as a tool for theoretical analy-

sis, and also as one yielding practical solutions to our original problem. In section 3.8 we show how the cryptanalytic ideas arising from simultaneous correlation lead to improved attacks using a method derived from Gallager's techniques (described in section 1.5.4).

The material in sections 3.2 to 3.6 of this chapter is essentially that of [5]. Section 3.7 presents the additional material cited in section 3.2 of the paper [5], while section 3.8 contains more recent material.

3.1.1 Concerning Notation used in this Chapter

The information function I used in this chapter denotes Shannon information measured in bits, that is, with logarithms to base 2, which we denote by log; we retain the notation ln for natural logarithms. Hence $I = (\log e)I_1$.

3.2 Maximum Likelihood Attacks on Stream Ciphers

3.2.1 The Keystream Generator

Throughout this chapter we shall be considering stream ciphers derived from keystream generators as defined in section 1.5.2. Thus, with the notation introduced there, the keystream bits $b_1, \ldots, b_N$ produced by such a keystream generator satisfy

$$b_i = f(T^i x) \qquad (i = 1, \ldots, N) \tag{3.1}$$

when x assumes the value of the key, or key-dependent value which, in this chapter, we will not distinguish from the key itself. We identify V with the set of possible n-bit keystream generator states, and assume that T is bijective on V.

3.2.2 Maximum Likelihood Attack

When analysing a stream cipher, an attacker can sometimes benefit by modelling some pseudo-random bits within the system as genuinely random ones. The correlation attacks described by Siegenthaler [13] are a well-known example of this. In the case of a function g with input consisting of m pseudo-random bits, we can derive a new — in general, non-deterministic — function $\bar{g}$ by modelling certain arguments as truly random inputs. These thoughts motivate a definition we make shortly, after a preliminary one.

Definition 3.2. The *orthogonal complement* (with respect to the standard basis of V) of a subspace S of V is denoted $S^{\perp}$ and defined to be the set

$$\{u \in V : u \cdot v = 0 \text{ for all } v \in S\}$$

(Recall that we defined inner product on V just before lemma 1.14.)

It is not hard to see that $S^{\perp}$ is itself a subspace of V. We are now able to make the definition mentioned above:

Definition 3.3. For any function $g : V \to \{0, 1\}$ and subspace $S \leq V$, we define a *reduced* version, $\overline{g}_S$, of g, by taking, for each $x \in V$, $\overline{g}_S(x)$ to be the random variable which assumes each of the values 0 and 1 with the probability that g assumes that value on a uniformly selected element of the coset $x + S^{\perp}$.[1]

Returning to our keystream generator, we observe that for any subspace $U \leq V$, the equations (3.1) can be reduced, in this technical sense, to

$$b_i = \overline{f}_U(T^i x) \qquad (i = 1, \ldots, N), \tag{3.4}$$

[1]Note that we use cosets of $S^{\perp}$ in this definition in order that section 3.4 can derive a result concerning a DFT which equals 0 for arguments outside S. We also show in the same section that $(S^{\perp})^{\perp} = S$, so we could instead have made our definition here in terms of cosets of S, and derived a result in section 3.4 about a DFT vanishing outside $S^{\perp}$.

corresponding to a reduction of the output function, or to

$$b_i = \overline{f \circ T^i}_U(x) \qquad (i = 1, \ldots, N), \tag{3.5}$$

corresponding to a reduction of the key space; we will also use a common symbolism

$$b_i = \overline{g}_i(x) \qquad (i = 1, \ldots, N). \tag{3.6}$$

For either of these reductions we may attempt to determine the most likely $x \in V$ given the probabilistic equations hold: that is, maximise

$$\begin{aligned} &\Pr(x|\overline{g}_i(x) = b_i \ \forall i \in \{1, \ldots, N\}) \qquad (3.7)\\ = \ &\Pr(\overline{g}_i(x) = b_i \, \forall i|x) \Pr(x)/\Pr(b_1, \ldots, b_N). \end{aligned}$$

Thus for equiprobable initial states and given keystream, x equivalently maximises

$$\Pr(\overline{g}_i(x) = b_i \, \forall i) = \prod_{i=1}^{N} \Pr(\overline{g}_i(x) = b_i),$$

by (pairwise) independence of the $\overline{g}_i(x)$ $(1 \leq i \leq N)$.

In section 3.6 we shall use a reformulation of the maximum likelihood condition, which makes use of the limit $\ln(1+x) = x + O(x^2)$ as $x \to 0$ in the case when all $\Pr(\overline{g}_i(x) = 0) \approx \frac{1}{2}$: in this case, maximising (3.7) is equivalent to maximising

$$\begin{aligned} &\ln\left((\frac{1}{2})^{-N} \prod_{i=1}^{N} \Pr(\overline{g}_i(x) = b_i)\right)\\ = \ &\sum_{i=1}^{N} \ln(\Pr(\overline{g}_i(x) = b_i)/\frac{1}{2})\\ \approx \ &\sum_{i=1}^{N} (2\Pr(\overline{g}_i(x) = b_i) - 1) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} \left[\Pr(\overline{g}_i(x) \oplus b_i = 0) - \Pr(\overline{g}_i(x) \oplus b_i = 1)\right] \\
&= \sum_{i=1}^{N} E\left((-1)^{\overline{g}_i(x) \oplus b_i}\right),
\end{aligned}
$$

where E denotes the expected value of its argument.

3.2.3 Uniqueness of Maximum Likelihood Solutions

In this section we observe that the reduced equations (3.4) or (3.5) may not yield a unique most likely solution x. This is apparent in the case of (3.5), where x can only possibly be determined up to a coset $x + U^{\perp}$.

To see that ambiguity is possible also for (3.4), observe that we can have a subspace W of V containing U such that T maps every coset $x + W^{\perp}$ to some other coset of $W^{\perp}$. (Such a W can arise when we have a decomposition of the keystream generator states into those of two sub-generators: that is, up to a reordering of positions in a vector of V, we have a cartesian product $V = W \times W^{\perp}$ (identifying W with the subspace $\{(w, 0) : w \in W\}$, and $W^{\perp}$ with $\{(0, w) : w \in W^{\perp}\}$) where W (and $W^{\perp}$) is closed under T.) For such a subspace W, given any $x \in V$, $T(x + W^{\perp}) = y + W^{\perp}$ for some $y \in V$. If this condition holds, $T(x) \in T(x + W^{\perp}) = y + W^{\perp}$, so $T(x) = y + w$ for some $w \in W^{\perp}$ and $T(x + W^{\perp}) = T(x) - w + W^{\perp} = T(x) + W^{\perp}$. On the other hand, $\overline{f}_U$ is the same distribution on all elements in a coset $x + U^{\perp} \supseteq x + W^{\perp}$, and hence is well-defined on cosets $x + W^{\perp}$. Thus the properties of T and $\overline{f}_U$ together imply that equation (3.4) holds for x if and only if it holds for any other element x' of $x + W^{\perp}$.

3.2.4 How Large N Should Be

We now address the question: if N is large will the most likely coset of solutions to equations (3.6) be the correct one, i.e. the coset of the initial state of the generator, and, if so, how large does N need to be? We define the least such N to be the *unicity distance* of the equations. We answer this by way of a corollary to the following theorem.

3.2.4.1 A Theorem of Brynielsson [2]

Theorem 3.8. *Let $\underline{a}$ and $\underline{b}$ be two distributions on K objects, taking values with probabilities a_j and b_j $(j = 1, \ldots, K)$ respectively; X be a uniform random variable on $\{1, \ldots, M\}$; Y_i $(i = 1, \ldots, M)$ be independent random variables having the multinomial distribution $M(N, \underline{a})$ for $i = X$, but the multinomial distribution $M(N, \underline{b})$ for $i \neq X$; and, lastly, y_i be an observation of Y_i $(i = 1, \ldots, M)$. Following definition 1.43, denote by $d_1(\underline{a}, \underline{b})$ the order 1 directed divergence $\sum_{j=1}^{K} a_j \log(a_j/b_j)$, and denote by p_i $(i = 1, \ldots, M)$ the probability*

$$\Pr(X = i | Y_1 = y_1, \ldots, Y_M = y_M).$$

Then the ordering on the i induced by the p_i is the same as that induced by the likelihood ratio

$$\frac{\Pr(y_i \text{ is an observation of } M(N, \underline{a}))}{\Pr(y_i \text{ is an observation of } M(N, \underline{b}))},$$

and the probability that X is amongst the k greatest values of i under this ordering (for any k such that $K \ll \log(M/k)/d_1(\underline{a}, \underline{b})$) is approximately

$$\begin{cases} 0 & \text{if } N < \log(M/k)/d_1(\underline{a}, \underline{b}) \\ 1 & \text{if } N > \log(M/k)/d_1(\underline{a}, \underline{b}) \end{cases}$$

3.2.4.2 A Corollary

Corollary 3.9. *Suppose that we have an array* $\overline{g}_i(x)$ *(*$i = 1, \ldots, N; x \in V$*) of Bernoulli random variables, taking value 0 with probabilities* $p_{i,x}$*, for which (for any* i*)* $\overline{g}_i(x)$ *is the same distribution as* $\overline{g}_i(y)$ *when* x *and* y *are in the same coset, but otherwise are independent distributions; the* $\overline{g}_i(x)$ *(for fixed* x*) are independent; the parameters* $p_{i,x}$ *are themselves (for pairs of* i *and pairs of* x *in different cosets) known independent realisations of some* a priori *distribution* X *with mean* $\frac{1}{2}$*; and* $(b_i)_{i=1}^N$ *is an observation of* $(\overline{g}_i(x))_{i=1}^N$ *for (any)* x *in some particular coset — that of* x_0*, say. Then, for large number* 2^s *of cosets, the maximum likelihood method determines the coset of* x_0 *after* $N \approx s/\overline{I}(\overline{g}_i(x); x)$ *observations of* $\overline{g}_i(x_0)$*, where* $\overline{I}(\overline{g}_i(x); x)$ *denotes the average order* 1 *mutual information (averaged over* X*) between* $\overline{g}_i(x)$ *and* x*.*

Proof. If necessary, replace X by an approximation taking

$$K \ll s/\overline{I}(\overline{g}_i(x); x)$$

values, and make appropriate approximations in what follows.

Fix any x. For each i, there are $2K$ possibilities:

$$b_i = b \text{ and } \Pr(\overline{g}_i(x) = 0) = p,$$

for $b = 0, 1$ and each probability p associated with X, and these possibilities themselves have probabilities pairwise independent for distinct indices i. The sequence $y_x = (N_{x;b,p})_{b=0,1;p\in X}$, where

$$N_{x;b,p} := \text{ number of times } b_i = b \text{ and } p_{i,x} = p,$$

is an observation of the multinomial distribution on N observations of independent events with these $2K$ probabilities.

For $x \in$ the coset of x_0, these $2K$ probabilities are

$$\begin{aligned}\Pr(b_i = 0, p_{i,x} = p) &= p \Pr(p_{i,x} = p) \\ \Pr(b_i = 1, p_{i,x} = p) &= (1-p) \Pr(p_{i,x} = p),\end{aligned}$$

defining a distribution $\underline{a}$, while for other x

$$\begin{aligned}&\Pr(b_i = b, p_{i,x} = p) \\ &= \Pr(b_i = b) \Pr(p_{i,x} = p) \qquad \text{by independence,} \\ &= \frac{1}{2} \Pr(p_{i,x} = p) \qquad \text{by balance,}\end{aligned}$$

defining a distribution $\underline{b}$.

Writing, for convenience, $\Pr(p)$ for $\Pr(p_{i,x} = p)$, the directed divergence $d_1(\underline{a}, \underline{b})$ is

$$\begin{aligned}&\left(\sum_p p \Pr(p) \log \frac{p \Pr(p)}{\frac{1}{2} \Pr(p)} \right) + \left(\sum_p (1-p) \Pr(p) \log \frac{(1-p) \Pr(p)}{\frac{1}{2} \Pr(p)} \right) \\ &= \sum_p \Pr(p) (p \log(p/\frac{1}{2}) + (1-p) \log((1-p)/\frac{1}{2})) \\ &= \overline{I}(\overline{g}_i(x); x).\end{aligned}$$

Thus by theorem 3.8,

$$\frac{\Pr(y_x \text{ is an observation of } M(N, \underline{a}))}{\Pr(y_x \text{ is an observation of } M(N, \underline{b}))}$$

will be maximised for x in the correct coset (i.e. that of x_0) for

$$N \approx (\log 2^s)/\overline{I}(\overline{g}_i(x); x) = s/\overline{I}(\overline{g}_i(x); x).$$

But

$$\begin{aligned}
&\frac{\Pr(y_x \text{ is an observation of } M(N,\underline{a}))}{\Pr(y_x \text{ is an observation of } M(N,\underline{b}))} \\
&= \frac{\binom{N}{(N_{x;b,p})_{b,p}} \prod_p (p\Pr(p)))^{N_{x;0,p}} \prod_p ((1-p)\Pr(p)))^{N_{x;1,p}}}{\binom{N}{(N_{x;b,p})_{b,p}} \prod_p (\frac{1}{2}\Pr(p)))^{N_{x;0,p}} \prod_p (\frac{1}{2}\Pr(p)))^{N_{x;1,p}}} \\
&= 2^N \prod_p p^{N_{x;0,p}} (1-p)^{N_{x;1,p}} \\
&= 2^N \Pr((b_i)_{i=1}^N \text{ comes from } (\overline{g}_i(x))_{i=1}^N),
\end{aligned}$$

whose maximisation is equivalent to the maximum likelihood method described in section 3.2.2. □

3.2.4.3 N for Reduced Output Function

In this case, we are considering the equations (3.4). We assume here, and subsequently when considering unicity distance for reduced output function, that the assumptions of the corollary apply for cosets of $W^\perp$, where W is an m-dimensional subspace of V containing U.

$$I(\overline{g}_i(x); x) = I(\overline{f}_U(T^i x); x) = I(\overline{f}_U(x); x),$$

since T is bijective, so the unicity distance is

$$N \approx m / I(\overline{f}_U(x); x). \tag{3.10}$$

3.2.4.4 N for Reduced Key Space

Similarly with equations (3.5), if we make analogous assumptions about the cosets of $U^\perp$, we can apply the corollary to learn that the unicity distance is

$$N \approx m' / \overline{I}(\overline{f \circ T^i}_U(x); x),$$

where m' denotes the dimension of U.

We shall develop this result further in section 3.5.

3.2.5 An Observation Concerning "Correlation Immunity"

Following Siegenthaler [12], we define a function $f(x_1, \ldots, x_n)$ to be *correlation-immune* to a set $\{i_1, \ldots, i_m\}$ of input positions iff

$$I(f(x_1, \ldots, x_n); x_{i_1}, \ldots, x_{i_m}) = 0.$$

This mutual information is equal to

$$I(\overline{f}_U(x); x)$$

where U is the subspace of V generated by the standard basis vectors with non-zero component in respective positions $i_1, \ldots, i_m$. Thus, by equation (3.10), if f is correlation-immune, the maximum likelihood method will fail no matter how much keystream is considered.

3.3 Correlation Attacks on Linearly Clocking Stream Ciphers

Henceforth we focus our attention on the case where state transition is an invertible linear transformation on V; we will represent it as A, rather than T, and consider A to be an $n \times n$ matrix. (As a point of interest, we note that A is similar to a matrix in rational canonical form, so that, by a suitable choice of basis for V, the generator can be taken to comprise one or more separate Galois-clocking LFSRs. Moreover, a change of basis preserves the dimension of the smallest subspace of V on which f depends. The importance of this subspace will become clear later in this chapter.)

This section will review some "traditional" theory of correlation attacks;

in section 3.6 we will reformulate these ideas in the language of section 3.2, which will cast new light on existing attacks and produce some new ones.

3.3.1 Linear Correlations to f

Linear correlation attacks exploit correlation between f and linear functions on V i.e. functionals $\in V^*$. For any function $g : V \to \{0,1\}$ we define the *correlation* $c_{g,v}$ of g to the functional "$\cdot v$" by

$$c_{g,v} := \Pr(g(x) = x \cdot v) - \Pr(g(x) \neq x \cdot v).$$

For convenience, we will write c_v for a correlation $c_{f,v}$ of our keystream generator's output function f.

As we saw in section 1.3.3, the set of all ± 1-valued functionals $\{x \mapsto (-1)^{x.v}\}_{v \in V}$ is a complete orthogonal subset of the complex vector space of complex-valued functions $V \to \mathbb{C}$ with inner product

$$\langle g_1, g_2 \rangle = \sum_{v \in V} g_1(v)\overline{g_2(v)}.$$

In this context, we can re-express $c_{g,v}$:

$$\begin{aligned} c_{g,v} &= \Pr(g(x) = x \cdot v) - \Pr(g(x) \neq x \cdot v) \\ &= 2^{-n} \sum_{x \in V} (-1)^{g(x) \oplus x \cdot v} \\ &= 2^{-n} \langle (-1)^g, (-1)^{\cdot v} \rangle \\ &= 2^{-n} D((-1)^g)(v). \end{aligned}$$

where, as in section 1.3.3, D is the DFT corresponding to the above orthogonal subset. We can apply theorem 1.16, part 1, to $(-1)^g$ to learn that

$$2^{-n} \left\| (-1)^g \right\|^2 = \left\| D((-1)^g) \right\|^2,$$

whence

$$\sum_{v \in V} c_{g,v}^2 = 1.$$

In particular, $c_{g,v}$ is non-zero for at least one $v \in V$.

3.3.2 Linear Correlation Attack

Fix any $v \in V$ for which $c_v \neq 0$. The sequences of bits $(f(A^i x))_i$ generated by the KG can be written $(((A^i x) \cdot v) \oplus e_{i,x})_i$, where the $e_{i,x}$ are modelled as independent Bernoulli random variables which take the value 0 with probability $(1 + c_v)/2 \neq \frac{1}{2}$. Thus the KG generates "noisy LFSR sequences". Considerable research effort has been directed towards the problem of the efficient implementation of minimum distance decoding of such sequences (when the error probability $< \frac{1}{2}$) to the underlying linear sequence. Siegenthaler [13] presents the straightforward method for solving this problem: to examine each underlying linear sequence in turn and maximise (or minimise) the number of agreements ("correlation") between each of these linear sequences and the keystream. This method has[2] time complexity $N2^m$. Later, in section 3.6, we shall characterise Siegenthaler's attack in terms of the language of section 3.2.

As previously observed by Mund *et al.* [10], a more efficient method, borrowed from the theory of decoding first-order Reed-Muller codes (see, e.g., MacWilliams and Sloane [8]), makes use of the Walsh-Hadamard transform. Before describing this method, we introduce some notation concerning the Walsh-Hadamard transform and review the results from section 1.3.3.

[2]This situation can be described as "output function reduction", and we can take m to be the dimension of any subspace W of V containing v for which $W^\perp$ is A-invariant. See also sections 3.2.3 and 3.5.2.

3.3.3 The Walsh-Hadamard Transform

For any subspace $S \leq V$ and real-valued function g on S, the Walsh-Hadamard transform $D_S(g)$ of g on S is also a real-valued function on S, defined by

$$D_S(g)(v) = \sum_{s \in S} (-1)^{s \cdot v} g(s) \qquad (v \in S).$$

For simplicity we write D for D_V.

Denoting the dimension of S by d, we have the following results as consequences of our definitions and of section 1.3.3:

1. The array $D_S(g)(v)$ $(v \in S)$ for a function $g : S \to \mathbb{R}$ can be computed in time $d2^d$ and space 2^d real storage locations.

2. $D_S(D_S(g)) = 2^d g$ for any $g : S \to \mathbb{R}$.

3. If g is a $\{0,1\}$-valued function on V, $D((-1)^g)(v) = 2^n c_{g,v}$ (we defined $c_{g,x}$ in section 3.3.1).

4. $\sum_{s \in S} (D_S(g)(s))^2 = 2^d \sum_{s \in S} (g(s))^2$ $(g : S \to \mathbb{R})$.

3.3.4 Reed-Muller Decoding Algorithm

We now return to the situation introduced in section 3.3.2, but assume also that $W^\perp$ is A-invariant. (In section 3.5.3 we shall prove that this implies W is invariant under the transpose A^* of A.[3])

[3]We use the notation A^* rather than A^T to emphasise that this map is dual to A.

First of all, observe that for any $x, v \in W$

$$\begin{aligned}
(A^i x) \cdot v &= (A^i x)^T v \\
&= x^T (A^*)^i v \qquad \text{where } A^* \text{ denotes the transpose of } A \\
&= x \cdot ((A^*)^i v) \\
&= x \cdot v_i \qquad \text{where } v_i = (A^*)^i v \in W.
\end{aligned}$$

Now we can compute

$$\begin{aligned}
&|\{i\ (1 \le i \le N) : x \cdot v_i = b_i\}| - |\{i\ (1 \le i \le N) : x \cdot v_i \ne b_i\}| \\
&= \sum_{i=1}^{N} (-1)^{(x \cdot v_i) \oplus b_i} \\
&= \sum_{w \in W} (-1)^{x \cdot w} h(w) \qquad \text{putting } h(w) := \sum_{i : v_i = w} (-1)^{b_i} \\
&= D_W(h)(x),
\end{aligned}$$

the Walsh-Hadamard transform of h on W.

3.3.5 Complexity of the Reed-Muller Attack

Using the results cited in section 3.3.3, we can summarise the complexity of the Reed-Muller attack (the corresponding parameters for Siegenthaler's attack are shown in brackets):

- time complexity = $m2^m + N$ ($N2^m$);
- space complexity = 2^m (0).

3.3.6 Two Significant Observations

These two attacks, relying on correlation to the single functional "$\cdot v$", appear, heuristically, to waste information as compared to the maximum likelihood

attack. Generally, many c_v will be non-zero, corresponding to *simultaneous correlation* of the keystream to many linear sequences. Moreover we can see from the results of section 3.3.3 that any function $f : V \mapsto \{0,1\}$ is characterised by $D((-1)^f)$ and so also by its correlations $(c_v)_{v \in V}$ to linear functionals.

These observations will be explored subsequently in this chapter.

3.4 Characterising a Function by Correlations to Linear Functionals

Given a real-valued function g on V, what function is characterised by the correlations of g to the functionals in a subspace S^* of V^* i.e. by the correlations $(c_{g,s})_{s \in S}$ for a subspace S of V? We now demonstrate that these correlations in fact characterise $\overline{g}_S$.

Lemma 3.11. *For any $S \leq V$,*

$$D(E((-1)^{\overline{g}_S}))(v) = \begin{cases} D((-1)^g)(v) & \text{if } v \in S \\ 0 & \text{otherwise,} \end{cases}$$

where E denotes expected value.

Proof.

$$\begin{aligned} D(E((-1)^{\overline{g}_S}))(v) &= \sum_{v' \in V} (-1)^{v' \cdot v} E((-1)^{\overline{g}_S(v')}) \\ &= \sum_{v' \in V} (-1)^{v' \cdot v} \frac{1}{|S^\perp|} \sum_{s \in S^\perp} (-1)^{g(v'+s)} \\ &= \frac{1}{|S^\perp|} \sum_{v'' \in V, s \in S^\perp} (-1)^{(v''+s) \cdot v} (-1)^{g(v'')} \qquad (v'' = v' + s) \\ &= \frac{1}{|S^\perp|} \sum_{s \in S^\perp} (-1)^{s \cdot v} D((-1)^g)(v) \end{aligned}$$

Now $S = (S^\perp)^\perp$ ($S \subseteq (S^\perp)^\perp$ and they have the same dimension[4]), so that if $v \notin S$, $\exists s' \in S^\perp$ such that $v \cdot s' = 1$; then

$$\begin{aligned}\sum_{s\in S^\perp}(-1)^{s\cdot v} &= \sum_{s\in S^\perp}(-1)^{(s+s')\cdot v} \qquad \text{since } s' + S^\perp = S^\perp \\ &= -\sum_{s\in S^\perp}(-1)^{s\cdot v},\end{aligned}$$

so this sum is 0. If $v \in S$, $\sum_{s\in S^\perp}(-1)^{s\cdot v} = \left|S^\perp\right|$. Hence the result. □

Corollary 3.12. *$\overline{g}_S$ is characterised by the $(c_{g,s})_{s\in S}$.*

3.5 The Unicity Distance N in Terms of Correlations

In this section we demonstrate that the expressions for the unicity distances obtained in section 3.2.4 can be couched in terms of the correlations c_v of the output function f.

3.5.1 Information in Terms of Correlations

Proposition 3.13. *If $S \leq V$ is any subspace and $g : V \to \mathbb{R}$ any balanced function for which all $\Pr(\overline{g}_S(x) = 0) \approx \frac{1}{2}$, then we can approximate*

$$I(x; \overline{g}_S(x)) \approx \frac{1}{2\ln 2}\sum_{s\in S} c_s^2.$$

[4]Let X be any subspace of V, and M a matrix whose columns are a basis of X. $\operatorname{im} M \simeq V/\ker M$, so $\dim M = \dim V - \dim \ker M$; but $\operatorname{im} M = X$ and $\ker M = X^\perp$, so $\dim X = \dim V - \dim X^\perp$. Applying this twice, $\dim(S^\perp)^\perp = \dim V - \dim S^\perp = \dim V - (\dim V - \dim S) = \dim S$.

Proof. From the Taylor expansion which we established in lemma 2.14:

$$i_1(x) = \sum_{i=1}^{\infty} \frac{x^{2i}}{2i(2i-1)}$$

we deduce the approximation

$$i_1(x) \approx \frac{1}{2}x^2 \qquad \text{for } x \approx 0.$$

Now we compute

$$\begin{aligned}
I(x; \overline{g}_S(x)) &= I(\overline{g}_S(x)|x) - I(\overline{g}_S(x)) \qquad \text{by lemma 1.52} \\
&= \frac{1}{\ln 2} 2^{-n} \sum_{x \in V} i_1(\Pr(\overline{g}_S(x) = 0) - \Pr(\overline{g}_S(x) = 1)) \\
&\qquad\qquad (I(\overline{g}_S(x)) = 0 \text{ because } g \text{ is balanced}) \\
&\approx \frac{1}{2 \ln 2} 2^{-n} \sum_{x \in V} (\Pr(\overline{g}_S(x) = 0) - \Pr(\overline{g}_S(x) = 1))^2 \\
&\qquad\qquad \text{since all } \Pr(\overline{g}_S(x) = 0) \approx \tfrac{1}{2} \\
&= \frac{1}{2 \ln 2} 2^{-n} \sum_{x \in V} E((-1)^{\overline{g}_S(x)})^2 \\
&= \frac{1}{2 \ln 2} 2^{-n} 2^{-n} \sum_{v \in V} \left(D(E((-1)^{\overline{g}_S(x)}))(v) \right)^2 \qquad \text{by 3.3.3, point 4} \\
&\qquad\qquad (3.14) \\
&= \frac{1}{2 \ln 2} 2^{-2n} \sum_{s \in S} (D((-1)^g)(s))^2 \qquad \text{by section 3.4} \\
&= \frac{1}{2 \ln 2} \sum_{s \in S} c_{g,s}^2 \qquad \text{by 3.3.3, point 3.}
\end{aligned}$$

□

(Notice that, with our assumption that g is balanced, $c_{g,0} = 0$.)

3.5.2 Unicity Distance when Reducing the Output Function

Now that state transition is linear, A is a well-defined map $V/S \to V/S$ on the cosets of any subspace S of V if and only if that subspace is A-invariant: the "if" direction is immediate, and for "only if", observe that $A(S)$ is both a coset of S and a subspace of V, therefore must equal S. Moreover, we can prove the following lemma:

Lemma 3.15. *For any subspace U of V, there is a smallest subspace W of V containing U for which $W^\perp$ is A-invariant.*

Proof. Let $\mathcal{S} = \{\text{subspaces } S \text{ of } V : S \text{ contains } U \text{ and } S^\perp \text{ is } A\text{-invariant}\}$. Certainly $V \in \mathcal{S}$; and using corollary 3.19, which we prove later,

$$\begin{aligned}
S_1, S_2 \in \mathcal{S} \quad &\Rightarrow \quad U \subseteq S_1,\ U \subseteq S_2,\ S_1^\perp \text{ is } A\text{-invariant},\ S_2^\perp \text{ is } A\text{-invariant} \\
&\Rightarrow \quad U \subseteq S_1 \cap S_2,\ S_1 \text{ is } A^*\text{-invariant},\ S_2 \text{ is } A^*\text{-invariant} \\
&\Rightarrow \quad U \subseteq S_1 \cap S_2,\ S_1 \cap S_2 \text{ is } A^*\text{-invariant} \\
&\Rightarrow \quad U \subseteq S_1 \cap S_2,\ (S_1 \cap S_2)^\perp \text{ is } A\text{-invariant} \\
&\Rightarrow \quad S_1 \cap S_2 \in \mathcal{S}
\end{aligned}$$

Thus $W = \bigcap_{S \in \mathcal{S}} S$ is the (unique) required subspace. □

As before, let $m = \dim W$. Then by equation (3.10), the unicity distance when reducing the output function f to $\overline{f}_U$ is

$$N \approx m / I(x; \overline{f}_U(x)).$$

Applying the result of proposition 3.13, we see that

$$N \approx \frac{2m \ln 2}{\sum_{u \in U} c_u^2},$$

or

$$N \approx \frac{m}{\sum_{u \in U} c_u^2}.$$

3.5.3 Unicity Distance when Reducing the Key Space

Similarly in this case, section 3.2.4 gives the unicity distance in terms of $\overline{I}(x; \overline{(f \circ A^i)}_U)$. To compute this information using proposition 3.13, we first compute

$$\begin{aligned}
D((-1)^{f \circ A^i})(v) &= \sum_{v' \in V} (-1)^{v \cdot v'} (-1)^{f(A^i v')} \\
&= \sum_{v' \in V} (-1)^{v^T v'} (-1)^{f(A^i v')} \\
&= \sum_{v'' \in V} (-1)^{v^T A^{-i} v''} (-1)^{f(v'')} \\
&= \sum_{v'' \in V} (-1)^{((A^*)^{-i} v)^T v''} (-1)^{f(v'')} \\
&= D((-1)^f)((A^*)^{-i} v) \\
&= 2^n c_{((A^*)^{-i} v)} \quad \text{by section 3.3.3, point 3} \qquad (3.16)
\end{aligned}$$

Therefore, by proposition 3.13,

$$I(x; \overline{f \circ A^i}_U(x)) \approx \frac{1}{2 \ln 2} \sum_{v \in U} c^2_{((A^*)^{-i} v)}. \qquad (3.17)$$

Now we wish to average this over values i, but in order to do this we need some preliminary results.

Lemma 3.18. *For any subspace $S \leq V$, $A^{-1} S^\perp = (A^* S)^\perp$.*

Proof. For any $v, v' \in V$,

$$(Av) \cdot v' = (Av)^T v' = v^T (A^* v') = v \cdot (A^* v');$$

Therefore

$$\begin{aligned} v \in (A^*S)^\perp &\Leftrightarrow v \cdot (A^*s) = 0 \qquad \text{for all } s \in S \\ &\Leftrightarrow Av \cdot s = 0 \qquad \text{for all } s \in S \\ &\Leftrightarrow Av \in S^\perp \\ &\Leftrightarrow v \in A^{-1}S^\perp. \end{aligned}$$

□

Corollary 3.19. *A subspace $S \leq V$ is A^*-invariant $\Leftrightarrow$ $S^\perp$ is A-invariant.*

Proof.

$$\begin{aligned} S = A^*S &\Leftrightarrow S^\perp = (A^*S)^\perp \\ &\Leftrightarrow S^\perp = (A^{-1})S^\perp \\ &\Leftrightarrow AS^\perp = S^\perp \end{aligned}$$

□

Now we see that W is also the smallest A^*-invariant subspace of V containing U, and that the subspaces $(A^*)^{-i}U$ $(i = 1, \ldots, N)$ will include the non-zero elements of W with approximately equal probabilities $2^{m'-m}$ (recall that $m' = \dim U$). And since $c_0 = 0$, the average value of the right hand side of equation (3.17) is

$$\frac{1}{2\ln 2} 2^{m'-m} \sum_{w \in W} c_w^2.$$

Hence

$$N \approx 2^{m-m'} \frac{m'}{\sum_{w \in W} c_w^2}. \tag{3.20}$$

In particular, if $W = V$, or if each $c_v^2 \approx 2^{-n}$,

$$N \approx m' 2^{n-m'}. \tag{3.21}$$

3.6 Maximum Likelihood Attacks in Terms of Correlations

3.6.1 Reformulation of a Maximum Likelihood Condition

In section 3.2, we saw that elements x of the most likely coset of initial keystream generator states given observed keystream $(b_i)_{i=1}^N$ maximise

$$\sum_{i=1}^{N} E((-1)^{\overline{g}_i(x)\oplus b_i}) \tag{3.22}$$

in the case where all $\Pr(\overline{g}_i(x) = 0) \approx \frac{1}{2}$.

In the following sections, we shall reformulate this condition — in the case of reduced output function or reduced key space — in terms of the correlations c_v of f, and explore cryptanalytic methods they suggest.

3.6.2 Reformulation for Reduced Output Function

Considering (3.22) in the case where $\overline{g}_i(x) = \overline{f}_U(A^i x)$, we first note that

$$\begin{aligned} D(E((-1)^{\overline{f}_U}))(v) &= \begin{cases} 2^n c_v & \text{if } v \in U \\ 0 & \text{otherwise} \end{cases} \\ &= D\left(\sum_{u\in U} c_u (-1)^{x\cdot u}\right)(v). \end{aligned}$$

Therefore

$$E((-1)^{\overline{f}_U(x)}) = \sum_{u\in U} c_u (-1)^{x\cdot u},$$

and we can rewrite (3.22) as follows:

$$\begin{aligned}\sum_{i=1}^{N} E((-1)^{\overline{g}_i(x)\oplus b_i}) &= \sum_{i=1}^{N}(-1)^{b_i} E((-1)^{\overline{f}_U(A^i x)}) \\ &= \sum_{i=1}^{N}\sum_{u\in U} c_u (-1)^{(A^i x)\cdot u \oplus b_i}. \end{aligned} \tag{3.23}$$

3.6.3 Cryptanalytic Applications

In this section we present two observations concerning the use of expression (3.23) in maximum likelihood attacks. Throughout the section we suppose $V = W \oplus W^{\perp}$, so that each coset of $W^{\perp}$ has a (unique) representative in W, and consequently we need only evaluate (3.23) for $x \in W$.

3.6.3.1 Siegenthaler's Method

If we put $U = \langle v \rangle$, the vector space generated by v, maximising (3.23) over x amounts to maximising

$$c_v \sum_{i=1}^{N} (-1)^{(A^i x)\cdot v \oplus b_i},$$

which is just Siegenthaler's "closest fit" method. A corollary of section 3.5 is that this method succeeds with about m/c_v^2 bits of keystream.

3.6.3.2 Generalised "Reed-Muller Method"

A speedup akin to that of section 3.3.4 can be obtained for the task of finding maximum likelihood solutions for any output function reduction, by writing (3.23) as

$$\begin{aligned}\sum_{i=1}^{N}\sum_{u\in U} c_u (-1)^{(A^i x)\cdot u \oplus b_i} &= \sum_{i=1}^{N}\sum_{v\in U} c_v (-1)^{x\cdot (A^{*i} v)\oplus b_i} \\ &= D_W(h)(x),\end{aligned}$$

where $h(x) = \sum_{v,i:A^{*i}v=x} c_v(-1)^{b_i}$ $(x \in W)$.

The vector of all values of h can be computed in time $2^{m'}N$, so (3.23) can be computed for all cosets of $W^\perp$, i.e. for each $x \in W$, in time $m2^m + N2^{m'}$ and space 2^m, rather than time $N2^m$ and negligible space: quite a remarkable result!

3.6.4 Reformulation for Reduced Key Space

In this case, $\overline{g}_i(x) = \overline{f \circ A^i}_U(x)$, and expression (3.22) is

$$\begin{aligned}
&\sum_{i=1}^{N} E((-1)^{\overline{f\circ A^i}_U(x)\oplus b_i}) \\
&= 2^{-n} D\left(D\left(\sum_{i=1}^{N} E((-1)^{\overline{f\circ A^i}_U \oplus b_i})\right)\right)(x) \qquad \text{by section 3.3.3, point 2} \\
&= 2^{-n} D\left(\sum_{i=1}^{N} D(E((-1)^{\overline{f\circ A^i}_U \oplus b_i}))\right)(x) \qquad \text{by the linearity of } D \\
&= \sum_{u\in U}(-1)^{x\cdot u}\Big(\sum_{i=1}^{N}(-1)^{b_i} c_{(A^*)^{-i}u}\Big) \quad \text{using lemma 3.11 and (3.16).} \qquad (3.24)
\end{aligned}$$

3.6.5 More Cryptanalytic Applications

In this section, we show that (3.24) can provide a practical vehicle for cryptanalytic attack if the correlations c_v of f vanish outside some subspace X of dimension r which is not too large. (This will be the case if f depends only on a small number r of state bits.)

We assume $V = U \oplus U^\perp$, so that each coset of $U^\perp$ has a unique representative in U, and we can perform a maximum likelihood attack by maximising (3.24) over $x \in U$. Thus the outer sum in (3.24) is a Walsh-Hadamard transform on U, which we can compute in time $m'2^{m'}$ and space $2^{m'}$.

Terms of the inner sum in (3.24) contribute only when $c_{(A^*)^{-i}u} \neq 0$, so we need only compute $c_{(A^*)^{-i}u}$ for those $u \in U$ for which $(A^*)^{-i}u \in X$ i.e. for $u \in U \cap A^{*i}X$. To see how many such u there are for each i, we apply the following lemma.

Lemma 3.25. *For a random r-dimensional subspace S of V,* $\dim(S \cap U) \approx r + m' - n$.

Proof. The expected size of $S \cap U \backslash \{0\}$ is $(2^n - 1)\times$ the probability that a random element of $V \backslash \{0\}$ is in both $S \backslash \{0\}$ and $U \backslash \{0\}$, i.e.

$$(2^n - 1) \times (2^r - 1)/(2^n - 1) \times (2^{m'} - 1)/(2^n - 1) \approx 2^{r+m'-n}. \qquad \square$$

Thus if we model each $A^{*i}X$ as a random r-dimensional subspace of V, $\dim\,(U \cap A^{*i}X) \approx r + m' - n$, and its elements can be efficiently computed[5] in time $\max\{1, 2^{r+m'-n}\}$.

Combining all this, we see that we can compute (3.24) for all cosets of x with

- time complexity $\approx m'2^{m'} + N \max\{1, 2^{r+m'-n}\}$;
- space complexity $= 2^{m'}$;
- number N of required keystream bits given by (3.20).

For $r \leq n/2$, and $W = V$, the value for $m' = \dim U$ which minimises the time complexity is $m' = n/2$, when

- time complexity $\approx n2^{n/2}$;

[5]In the case when U is a subspace whose elements are precisely those with 0 entries in certain coordinate positions, straightforward Gaussian elimination can be used.

- space complexity $= 2^{n/2}$;
- number N of required keystream bits $\approx (n/2)2^{n/2}$, by equation (3.21), even with the "worst case" assumption that $c_v^2 \approx 2^{-r}$ for $v \in X \setminus \{0\}$.

3.6.6 Why Less Keystream May Be Required

Suppose $r < n/2$ and let s be maximal subject to $rs \leq n/2$. Given N' bits of keystream, we can construct $\binom{N'}{s}$ "reduced" equations for the initial key:

$$\overline{g_{(i_1,\ldots,i_s)U}}(x) = b_{i_1} \oplus \cdots \oplus b_{i_s}, \qquad (1 \leq i_j \leq N', j = 1, \ldots, s) \tag{3.26}$$

where we have defined

$$g_{(i_1,\ldots,i_s)}(x) := \bigoplus_{j=1}^{s} f(A^{i_j}x).$$

The transform of $(-1)^{g_{(i_1,\ldots,i_s)}}$ is the convolution of the transforms of the $(-1)^{f\circ A^{i_j}}$; consequently it vanishes outside a subspace of dimension rs. Now essentially the method of the previous section applies, with required number N' of required known keystream bits satisfying

$$\binom{N'}{s} \approx N,$$

i.e., for $s \ll N$,

$$N' \approx (N.s!)^{1/s}.$$

3.6.7 Example

To perform a known keystream attack on the sequence generator illustrated in figure 3.1, we can choose U to be the set of states whose first 32 bits are 0. Then $U^\perp$ is the set of states whose last 32 bits are 0, $W = V$, $n = m = 64$, $m' = 32$, $N = 32.2^{64-32}$, $r = 8$, $s = 4$, and the attack will determine the last 32 bits of the initial state with

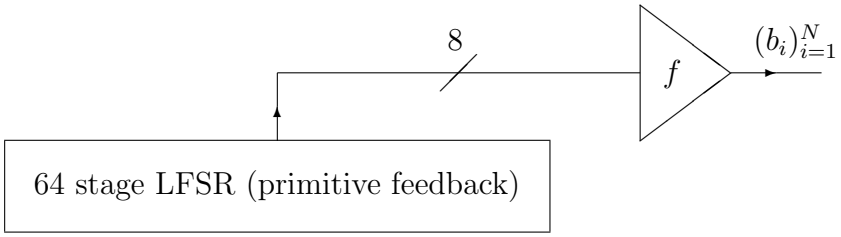

Figure 3.1: Example keystream generator

- time complexity $\approx 64.2^{32} = 2^{38}$;
- space complexity $= 2^{32}$;
- number N' of required keystream bits $\approx (32.2^{32}.4!)^{1/4} \approx 2^{10 \cdot 4}$.

3.6.8 A Brief Observation Concerning Probabilistic f

The techniques of this section may be applicable even when some inputs to f are not known linear functions of the initial state, but instead can be modelled as independent random bits: their effect may then be absorbed by a suitable choice of U.

3.7 An Hybrid Attack and Its Evaluation

This section is the first of two in this chapter which contain material extending the work in [5]. The results of this section are in fact those cited in section 3.2 of the paper [5], and generalise and expand the idea on which section 3.6.6 is based.

Suppose that we have observed N' bits of keystream. Fix arbitrary subspaces U_1 and U_2 of V, and an arbitrary integer $s \geq 1$. Following the example of section 3.6.6, construct $\binom{N'}{s}$ equations in the initial key x:

$$\overline{g_{(i_1,\ldots,i_s)}}_{U_1}(x) = b_{i_1} \oplus \cdots \oplus b_{i_s} \qquad (1 \leq i_j \leq N', j = 1, \ldots, s) \tag{3.27}$$

where now

$$g_{(i_1,\ldots,i_s)}(x) := \bigoplus_{j=1}^{s} \overline{f}_{U_2}(A^{i_j}x).$$

3.7.1 Unicity Distance of Equations (3.27)

These equations can only determine the correct value x up to a coset $x + U_1^{\perp}$; suppose that the requirements of corollary 3.9 are satisfied for U_1. Write $m' = \dim U_1$.

First of all we compute the transform of $E((-1)^{\overline{g_{(i_1,\ldots,i_s)}}_{U_1}})$:

$$D\left(E((-1)^{\overline{g_{(i_1,\ldots,i_s)}}_{U_1}})\right)(v) = \begin{cases} D\left(E((-1)^{g_{(i_1,\ldots,i_s)}})\right)(v) & \text{if } v \in U_1 \\ 0 & \text{otherwise} \end{cases}$$

by lemma 3.11, and

$$\begin{aligned} & D\left(E((-1)^{g_{(i_1,\ldots,i_s)}})\right)(v) \\ &= D\left(E((-1)^{\bigoplus_{j=1}^{s} \overline{f}_{U_2}(A^{i_j}x)})\right)(v) \\ &= D\left(\prod_{j=1}^{s} E((-1)^{\overline{f}_{U_2}(A^{i_j}x)})\right)(v) \\ &= 2^{-n(s-1)} \sum_{\substack{v_1,\ldots,v_s \in V \\ \sum_{j=1}^{s} v_j = v}} \left[\prod_{j=1}^{s} D\left(E((-1)^{\overline{f}_{U_2}(A^{i_j}x)})\right)(v_j)\right] \end{aligned}$$

by theorem 1.16, part 3. Also

$$\begin{aligned} D\left(E((-1)^{\overline{f}_{U_2}(A^{i_j}x)})\right)(v) &= \sum_{x\in V}(-1)^{v\cdot x}E((-1)^{\overline{f}_{U_2}(A^{i_j}x)}) \\ &= \sum_{x\in V}(-1)^{v^TA^{-i_j}x}E((-1)^{\overline{f}_{U_2}(x)}) \\ &= D\left((-1)^{\overline{f}_{U_2}}\right)((A^*)^{-i_j}v) \\ &= \begin{cases} 2^n c_{(A^*)^{-i_j}v} & \text{if } (A^*)^{-i_j}v \in U_2 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Assembling all these components, we obtain the transform of $\overline{g_{(i_1,\ldots,i_s)}}_{U_1}$:

$$D\left(E((-1)^{\overline{g_{(i_1,\ldots,i_s)}}_{U_1}})\right)(v) = 2^n \sum_{\substack{u_1,\ldots,u_s\in U_2 : \\ v=\sum_{j=1}^s (A^*)^{i_j}u_j \in U_1}} \prod_{j=1}^s c_{u_j} \qquad \text{for all } v \in V.$$

By equation (3.14) in the proof of proposition 3.13,

$$\begin{aligned} I(x;\overline{g_{(i_1,\ldots,i_s)}}_{U_1}(x)) &\approx \frac{1}{2\ln 2}2^{-2n}\sum_{v\in V}\left(D\left(E((-1)^{\overline{g_{(i_1,\ldots,i_s)}}_{U_1}(x)})\right)(v)\right)^2 \\ &= \frac{1}{2\ln 2}\sum_{v\in U_1}\left(\sum_{\substack{u_1,\ldots,u_s\in U_2 : \\ v=\sum_{j=1}^s (A^*)^{i_j}u_j}} \prod_{j=1}^s c_{u_j}\right)^2 \end{aligned}$$

so

$$\begin{aligned} &\overline{I}(x;\overline{g_{(i_1,\ldots,i_s)}}_{U_1}(x)) \\ &\approx \sum_{1\le i_1,\ldots,i_s\le N'} \frac{1}{2\ln 2}\sum_{v\in U_1}\left(\sum_{\substack{u_1,\ldots,u_s\in U_2 : \\ v=\sum_{j=1}^s (A^*)^{i_j}u_j}} \prod_{j=1}^s c_{u_j}\right)^2 \Big/ \binom{N'}{s}, \end{aligned}$$

and by corollary 3.9, the least number N' of keystream bits required to determine the correct coset of x using the maximum likelihood method satisfies

$$\binom{N'}{s} \approx \frac{m'}{\overline{I}(x;\overline{g_{(i_1,\ldots,i_s)}}_{U_1}(x))}$$

from which we obtain

$$m' \approx \sum_{\substack{1 \le i_1,\ldots,i_s \le N' \\ v \in U_1}} \left(\sum_{\substack{u_1,\ldots,u_s \in U_2 : \\ v = \sum_{j=1}^{s} (A^*)^{i_j} u_j}} \prod_{j=1}^{s} c_{u_j} \right)^2 \tag{3.28}$$

3.7.2 A Method for Solving Equations (3.27)

In this section we present a method for deriving the maximum likelihood solution of equations (3.27).

As in section 3.6, the elements x of the most likely coset of initial keystream generator states given the observed keystream $b_1, \ldots, b_{N'}$ maximise

$$\begin{aligned}
&\sum_{1 \le i_1,\ldots,i_s \le N'} E((-1)^{\overline{g_{(i_1,\ldots,i_s)}}_{U_1}(x) \oplus b_{i_1} \oplus b_{i_2} \oplus \cdots \oplus b_{i_s}}) \\
&= 2^{-n} \sum_{1 \le i_1,\ldots,i_s \le N'} (-1)^{b_{i_1} \oplus b_{i_2} \oplus \cdots \oplus b_{i_s}} D\left(D\left(E((-1)^{\overline{g_{(i_1,\ldots,i_s)}}_{U_1}})\right)\right)(x) \\
&= 2^{-n} \sum_{1 \le i_1,\ldots,i_s \le N'} (-1)^{b_{i_1} \oplus b_{i_2} \oplus \cdots \oplus b_{i_s}} \sum_{v \in V} (-1)^{x \cdot v} D\left(E((-1)^{\overline{g_{(i_1,\ldots,i_s)}}_{U_1}})\right)(v) \\
&= \sum_{1 \le i_1,\ldots,i_s \le N'} (-1)^{b_{i_1} \oplus b_{i_2} \oplus \cdots \oplus b_{i_s}} \sum_{v \in U_1} (-1)^{x \cdot v} \sum_{\substack{u_1,\ldots,u_s \in U_2 : \\ v = \sum_{j=1}^{s} (A^*)^{i_j} u_j}} \prod_{j=1}^{s} c_{u_j} \\
&= \sum_{v \in U_1} (-1)^{x \cdot v} \left(\sum_{1 \le i_1,\ldots,i_s \le N'} (-1)^{b_{i_1} \oplus b_{i_2} \oplus \cdots \oplus b_{i_s}} \sum_{\substack{u_1,\ldots,u_s \in U_2 : \\ v = \sum_{j=1}^{s} (A^*)^{i_j} u_j}} \prod_{j=1}^{s} c_{u_j} \right)
\end{aligned}$$

Thus the method of attack is to construct the array of values

$$\sum_{1 \le i_1,\ldots,i_s \le N'} (-1)^{b_{i_1} \oplus b_{i_2} \oplus \cdots \oplus b_{i_s}} \sum_{\substack{u_1,\ldots,u_s \in U_2 : \\ v = \sum_{j=1}^{s} (A^*)^{i_j} u_j}} \prod_{j=1}^{s} c_{u_j} \tag{3.29}$$

for indices $v \in U_1$ and compute the Walsh-Hadamard transform D_{U_1} of this array; then the index v corresponding to the greatest entry in the transformed

array is an element of the most likely coset, provided N' is at least as large as the value given by equation (3.28).

3.7.3 The Particular Case Outlined in Section 3.6.6

In section 3.6.6, we considered the case when the correlations c_v of f vanish outside a subspace X of dimension $r < n/2$, and took s to be the greatest integer such that $rs \leq n/2$, and U to be any subspace for which $V = U \oplus U^\perp$.

If we set $U_1 = U$ and $U_2 = X$, then, since $\overline{f}_{U_2} = f$, equations (3.27) become precisely the equations (3.26) of section 3.6.6.

As in sections 3.6.5 and 3.6.6, the method of attack consists of two stages: firstly, constructing the array (3.29), and, secondly, computing its transform in order to obtain the coset $x_0 + U^\perp$.

For the first task, choose a coordinate system for V so that $U^\perp$ is spanned by the first $n-m'$ standard basis vectors and U by the remaining m' standard basis vectors (recall that $m' = \dim U$). Now, given an s-tuple $(i_1, \ldots, i_s)$, form a matrix of rs vectors whose first r rows are a basis of $(A^*)^{i_1}U_2$, whose next r rows are a basis of $(A^*)^{i_2}U_2$, and so on. Performing Gaussian elimination on this array allows one to compute all vectors $v \in \sum_{j=1}^{s}(A^*)^{i_j}U_2$ whose first $n - m'$ components are 0 — (at least) $2^{rs-(n-m')}$ vectors in all. Once this has been done for one s-tuple $(i_1, \ldots, i_s)$, the process is repeated for the next such s-tuple; however, if the s-tuples are considered in increasing order of $\sum_{j=0}^{s-1}(N')^{js}i_{s-j}$, then generally only the last r rows of the matrix of rs vectors need be changed and subjected to Gaussian elimination, and some unnecessary repetition in the calculation is avoided.

Since the second stage consists of a Walsh-Hadamard transform of an m' dimensional array, the method described here runs with

- time complexity $= m'2^{m'} + \binom{N'}{s} \max\{1, 2^{rs+m'-n}\}$; and

- space complexity $= 2^{m'}$,

just as in section 3.6.5.

Finally, we compute the required number N' of keystream bits from (3.28) for the case $c_v^2 \approx 2^{-r}$ for $v \in X \backslash \{0\}$. We suppose that $rs < n - \dim U_1$ so that for each $v \in U_1$ and each s-tuple $(i_1, \ldots i_s)$ $(1 \leq i_j \leq N')$ there is generally at most one s-tuple $(u_1, \ldots, u_s)$ $(u_i \in U_2)$ such that $v = \sum_{j=1}^{s} (A^*)^{i_j} u_j$. Then (3.28) gives

$$\begin{aligned} m' &\approx \binom{N'}{s} (2^{rs}/2^{n-m'})(2^{-r})^s \\ &= \binom{N'}{s} 2^{m'-n}. \end{aligned}$$

The time complexity is thus approximately

$$m' 2^{m'} + m' 2^{n-m'},$$

minimised at $m' = n/2$ when it takes the value $n2^{n/2}$. The corresponding number of required keystream bits is

$$\begin{aligned} N' &\approx (s! m' 2^{n-m'})^{1/s} \\ &= (s!(n/2)2^{n/2})^{1/s} \end{aligned}$$

These conclusions agree with those of sections 3.6.5 and 3.6.6.

3.7.4 The Case $U_2 = \langle v \rangle$

In this section we revisit the problem introduced in section 3.3.2, and seek to solve it using the techniques of this section. We take $U_2 = \langle v \rangle$ and, having fixed any coordinate system for V, U_1 to be the subspace of V consisting of vectors with components equal to 0 outside the first m' positions.

The method now involves computing the array specified by (3.29), with indices $u \in U_1$ and corresponding values

$$\sum_{\substack{1 \le i_1, \ldots, i_s \le N' \\ u = \sum_{j=1}^{s} (A^*)^{i_j} u_j \\ u_j \in \{0, v\}}} (-1)^{b_{i_1} \oplus b_{i_2} \oplus \cdots \oplus b_{i_s}} c_v^s.$$

(Recall that $c_0 = 0$ given the assumption that f is balanced.) This can be done using a technique that takes advantage of the Birthday Paradox. This method is well established in the published literature, and an early reference can be found in section 5, page 174, of [9]. First, construct a table whose entries are the vectors $(i_1, \ldots, i_{\lfloor s/2 \rfloor})$ for $1 \le i_1, \ldots, i_{\lfloor s/2 \rfloor} \le N'$, indexed by the last $n - r$ components of $\sum_{j=1}^{\lfloor s/2 \rfloor} (A^*)^{i_j} v$ (or, in practice, by a suitable hash of this value). Then for each $(i_{\lfloor s/2 \rfloor + 1}, \ldots, i_s)$ $(1 \le i_{\lfloor s/2 \rfloor + 1}, \ldots, i_s \le N')$ compute the last $n - r$ components of $\sum_{j=\lfloor s/2 \rfloor + 1}^{s} (A^*)^{i_j} v$ and use it (or its hash) as an index to the table; in addition use the last $n - r$ components of $\sum_{j=\lfloor s/2 \rfloor + 2}^{s} (A^*)^{i_j} v$ (or hash) as an index: if, in either case, the table has a corresponding entry, we obtain a vector $u = \sum_{i \in I} (A^*)^i v \in U_1$ for a subset I of $\{1, \ldots, N'\}$ of cardinality $|I| \le s$ for each such corresponding entry, and all vectors $u = \sum_{j=1}^{s} (A^*)^{i_j} u_j \in U$ $(1 \le i_1, \ldots, i_s \le N', u_j \in \{0, v\})$ can be found in this way. This technique computes the array in time $\approx \binom{N'}{\lfloor s/2 \rfloor} + \binom{N'}{\lceil s/2 \rceil} \approx \binom{N'}{\lceil s/2 \rceil}$ and space $\binom{N'}{\lfloor s/2 \rfloor}$. The DFT of this array can then be computed in time $m' 2^{m'}$ and space $2^{m'}$.

In this case, the unicity distance equation (3.28) is

$$m' \approx \sum_{\substack{1 \le i_1, \ldots, i_s \le N' \\ u \in U_1}} \left(\sum_{\substack{u = \sum_{j=1}^{s} (A^*)^{i_j} u_j \\ u_j \in \{0, v\}}} \prod_{j=1}^{s} c_{u_j} \right)^2$$

$$\approx \sum_{\substack{1\le i_1,\ldots,i_s\le N' \\ u=\sum_{j=1}^{s}(A^*)^{i_j}u_j \\ u_j\in\{0,v\}}} \prod_{j=1}^{s} c_{u_j}^2$$

if $s < n - m'$ when each $u \in U_1$ will generally have at most one expansion

$$u = \sum_{j=1}^{s} (A^*)^{i_j} u_j \qquad (1 \le i_j \le N'; u_j \in \{0, v\}).$$

If we assume that approximately 1 in $2^{n-m'}$ of the $\binom{N'}{s}$ sums $\sum_{j=1}^{s}(A^*)^{i_j}v$ has 0s in the last $n - m'$ coordinate positions, this becomes

$$\begin{aligned} m' &\approx \sum_{t=1}^{s} \binom{N'}{t} 2^{m'-n} c_v^{2t} \\ &\approx 2^{m'-n} \sum_{t=1}^{s} \frac{(N'c_v^2)^t}{t!} \\ &\approx 2^{m'-n} \frac{(N'c_v^2)^s}{s!} \qquad \text{for practical values of } m' \text{ and } s \end{aligned}$$

from which

$$N' \approx \left(s!m'2^{n-m'}\right)^{1/s} c_v^{-2}. \tag{3.30}$$

In order to gain some appreciation of the usefulness of this method, we make the simplifying assumption that s is even, and compute the time complexity:

$$\begin{aligned} \binom{N'}{\lceil s/2 \rceil} + m'2^{m'} &\approx (N')^{s/2}/(s/2)! + m'2^{m'} \\ &\approx \left(\left(s!m'2^{n-m'}\right)^{1/s} c_v^{-2}\right)^{s/2} /(s/2)! + m'2^{m'} \\ &\approx \sqrt{s!m'}2^{(n-m')/2 - s\log c_v}/(s/2)! + m'2^{m'}. \end{aligned}$$

In any case where $\log\left(\sqrt{s!m'}/(s/2)!\right)$ and $\log m'$ differ by a quantity $\ll m'$, we can choose

$$m' \approx (n - m')/2 - s \log c_v$$

i.e.

$$m' \approx \frac{n - s \log c_v}{3}$$

to achieve a running time of about $2m'2^{m'}$. From equation (3.30), we can see that increasing s can dramatically reduce the required number N' of keystream bits.

3.8 Simultaneous Correlation and "Fast Correlation Attacks"

Our original problem (3.1) was to solve the simultaneous equations

$$b_i = f(T^i x) \qquad (i = 1, \ldots, N).$$

We consider the situation where T is a linear transformation A on V, and, as in section 3.6.5, the correlations c_v of f vanish outside a subspace X of dimension r, so that f depends only on a small number r of state bits. Choose a basis of V so that f is a function of the first r coordinates of $A^i x$. We denote the j^{th} coordinate of a vector $v \in V$ as $(v)_j$.

Using Gallager's algorithm (see section 1.5.4) as a model, we attempt an iterative reconstruction of the sequence of vectors $(x_i)_{i=1}^N$ ($x_i \in X$), where the first r components of x_i are of the first r components of $A^i x$ for the correct value x (and the other components = 0). To do this we store N distributions on r-bit vectors — one corresponding to each x_i — and make use of linear relations amongst the rN bits $(A^i x)_j$ ($1 \le i \le N$, $1 \le j \le r$).

Given an s-tuple $(i_1, \ldots, i_s)$ of integers i_j ($1 \le i_j \le N$) and an s-tuple of

vectors $(v_1, \ldots, v_s)$ $(v_j \in X)$,

$$\begin{aligned}
\bigoplus_{j=1}^{s} v_j \cdot x_{i_j} &= \bigoplus_{j=1}^{s} v_j \cdot (A^{i_j} x) \\
&= \bigoplus_{j=1}^{s} v_j^T A^{i_j} x \\
&= \bigoplus_{j=1}^{s} ((A^*) v_j)^T x \\
&= \left(\sum_{j=1}^{s} (A^*)^{i_j} v_j \right) \cdot x \\
&\equiv 0 \text{ for all } x \in V \Leftrightarrow \sum_{j=1}^{s} (A^*)^{i_j} v_j = 0.
\end{aligned}$$

Let J_k denote a relation $\bigoplus_{j=1}^{s} v_{k,j} \cdot x_{i_{k,j}} = 0$ arising from an s-tuple of distinct integers $(i_{k,1}, \ldots, i_{k,s})$ $(1 \leq i_{k,j} \leq N)$ and corresponding s-tuple $(v_{k,1}, \ldots, v_{k,s})$ of vectors $\in X$ which satisfy the condition $\sum_{j=1}^{s} (A^*)^{i_{k,j}} v_{k,j} = 0$. For convenience, we write $i \in J_k$ iff $i \in \{i_{k,1}, \ldots, i_{k,s}\}$. Suppose we have found m such relations, for $k = 1, \ldots, m$.

Denote by $p_{i,y}$ the probability $\Pr(x_i = y | f(A^i x) = b_i)$ for $1 \leq i \leq N$ and $y \in X$, and by $(c_{i,y})_{y \in X}$ the correlations corresponding to the Walsh-Hadamard transform of the array $(p_{i,y})_{y \in X}$ for each fixed i $(1 \leq i \leq N)$:

$$c_{i,y} = 2^{-r} \sum_{v \in X} (-1)^{v \cdot y} p_{i,v} \qquad (y \in X)$$

so that

$$c_{i,y} = \Pr_{v \in V}(v \cdot y = 0 | f(A^i x) = b_i) - \Pr_{v \in V}(v \cdot y = 1 | f(A^i x) = b_i)$$

for any $y \in X$. Then, as in section 1.5.4, we can compute, for any $y \in X$,

$$\begin{aligned}
&\Pr(x_d = y | (b_i)_{i=1}^N \text{ and relations } J_k \ (1 \le k \le m) \text{ hold}) \\
&= \frac{\Pr(\text{relations } J_k \ (1 \le k \le m) \text{ hold} | x_d = y, (b_i)_{i=1}^N) \Pr(x_d = y | (b_i)_{i=1}^N)}{\Pr(\text{relations } J_k \ (1 \le k \le m) \text{ hold} | (b_i)_{i=1}^N)}
\end{aligned} \tag{3.31}$$

and

$$\begin{aligned}
&\Pr(\text{relations } J_k \ (1 \le k \le m) \text{ hold} | x_d = y, (b_i)_{i=1}^N) \\
&= \prod_{k=1}^m \Pr(\text{relation } J_k \text{ holds} | x_d = y, (b_i)_{i=1}^N) \qquad \text{if the } J_k \backslash \{d\} \text{ are disjoint} \\
&= \prod_{\substack{k=1 \\ d \in J_k}}^m \left(\frac{1}{2} \left(1 + (-1)^{y \cdot v_{k,j_{k,d}}} \prod_{\substack{j=1 \\ i_{k,j} \neq d}}^s c_{i_{k,j}, v_{k,j}} \right) \right) \prod_{\substack{k=1 \\ d \notin J_k}}^m \left(\frac{1}{2} \left(1 + \prod_{j=1}^s c_{i_{k,j}, v_{k,j}} \right) \right),
\end{aligned}$$

where for each k $(1 \le k \le m)$ and $d \in J_k$ we define $j_{k,d}$ so that $i_{k,j_{k,d}} = d$. The argument justifying the last line is analogous to that following equation (1.60) in chapter 1. Assuming the validity of that expression,

$$\begin{aligned}
&\Pr(x_d = y | (b_i)_{i=1}^N \text{ and relations } J_k \ (1 \le k \le m) \text{ hold}) \\
&= \frac{p_{d,y} \prod_{\substack{k=1 \\ d \in J_k}}^m \left(1 + (-1)^{y \cdot v_{k,j_{k,d}}} \prod_{\substack{j=1 \\ i_{k,j} \neq d}}^s c_{i_{k,j}, v_{k,j}} \right)}{\sum_{y \in X} \left[p_{d,y} \prod_{\substack{k=1 \\ d \in J_k}}^m \left(1 + (-1)^{y \cdot v_{k,j_{k,d}}} \prod_{\substack{j=1 \\ i_{k,j} \neq d}}^s c_{i_{k,j}, v_{k,j}} \right) \right]}
\end{aligned}$$

These equations suggest the following version of Gallager's algorithm:

1. For $1 \le i \le N$ and $y \in X$, set

$$p_{i,y} := \begin{cases} 1/k_{b_i} & \text{if } f(y) = b_i \\ 0 & \text{otherwise} \end{cases}$$

where $k_b := |\{y \in X : f(y) = b\}|$ for $b = 0, 1$.

2. Given the sequence of probabilities distributions $((p_{i,y})_{y\in X})_{i=1}^{N}$, compute the corresponding sequence of correlations $((c_{i,y})_{y\in X})_{i=1}^{N}$ given by $c_{i,y} = 2^{-r}\sum_{v\in X}(-1)^{v\cdot y}p_{i,v}$ $(1 \leq i \leq N,\, y \in X)$.

3. Generate a new sequence of probability distributions $(p'_{i,y})_{i=1}^{N}$ according to the equations

$$p'_{i,y} = \alpha p_{i,y}\prod_{\substack{k=1\\ i\in J_k}}^{m}\left(1 + (-1)^{y\cdot v_{k,j_{k,i}}}\prod_{\substack{j=1\\ i_{k,j}\neq i}}^{s} c_{i_{k,j},v_{k,j}}\right),$$

where α is chosen so that $\sum_{y\in X} p'_{i,y} = 1$ i.e.

$$\alpha = \left\{\sum_{y\in X}\left[p_{i,y}\prod_{\substack{k=1\\ i\in J_k}}^{m}\left(1 + (-1)^{y\cdot v_{k,j_{k,i}}}\prod_{\substack{j=1\\ i_{k,j}\neq i}}^{s} c_{i_{k,j},v_{k,j}}\right)\right]\right\}^{-1}$$

.

4. Set $p_{i,y} := p'_{i,y}$ $(1 \leq i \leq N, y \in X)$.

5. If the sequence $((p_{i,y})_{y\in Y})_{i=1}^{N}$ has not converged, goto 2.

6. Recover a codeword $x'_1, \ldots, x'_N$ by choosing each $x'_i \in X$ so that $p_{i,x'_i} = \max_{y\in X} p_{i,y}$.

As in section 1.5.4, it is clear that we can readily modify this algorithm in order to accommodate Gallager's suggestion that the contribution to $(p'_{i,y})_{y\in X}$ due to a relation J_k should be computed using probability estimates $(p_{i,y})_{y\in X}$ which did not make use of the relation J_k when they themselves were last re-estimated.

3.8.1 Finding Suitable Relations

Note first of all that it is sufficient to find relations J for which $1 \in J$, since for indices $1 \le i_1 < \cdots < i_s \le N$ and vectors $v_1, \ldots, v_s \in X$

$$\sum_{j=1}^{s} v_j \cdot (A^{i_j} x) = 0 \qquad \text{for all } x \in V$$

$$\Leftrightarrow \quad \sum_{j=1}^{s} v_j \cdot (A^{i_j - i_1 + 1} x) = 0 \qquad \text{for all } x \in V$$

by the invertibility of A on V.

The number $m(s)$ of relations $1 = i_1 < \cdots < i_s \le N$ and vectors $v_1, \ldots, v_s \in X \setminus \{0\}$ involving s sequence positions is

$$m(s) \approx \binom{N-1}{s-1} (2^r - 1)^s 2^{-n} \approx \frac{N^{s-1}}{(s-1)!} 2^{rs-n}$$

These relations can be found using a Gaussian elimination technique like that in section 3.7.3. For each sequence of indices $1 = i_1 < \cdots < i_s \le N$, form a matrix of rs vectors whose first r rows are a basis of $(A^*)^{i_1} X$, whose next r rows are a basis of $(A^*)^{i_2} X$, and so on. Perform Gaussian elimination to compute all linear combinations $\in \sum_{j=1}^{s} (A^*)^{i_s} X$ which equal 0. As before, once this has been done for one s-tuple $(i_1, \ldots, i_s)$, the process is repeated for the next such s-tuple; however, if the s-tuples are considered in increasing order of $\sum_{j=0}^{s-1} N^{js} i_{s-j}$, then generally only the last r rows of the matrix of rs vectors need be changed and subjected to Gaussian elimination, and some unnecessary repetition in the calculation is avoided. The time complexity of this process is approximately $r\binom{N-1}{s-1}$.

Alternatively, the relations can be found using a technique taking advantage of the Birthday Paradox. The idea, in this context, is first to construct a table indexed by vectors $v \in V$ (or, in practice, by a suitable hash of such vectors) whose entries are a list of all sequences $(i_2, \ldots, i_{\lfloor s/2 \rfloor}, v_1, \ldots, v_{\lfloor s/2 \rfloor})$

$(1 < i_2 < \cdots < i_{\lfloor s/2 \rfloor} \leq N; v_1, \ldots, v_{\lfloor s/2 \rfloor} \in X)$ for which $v = \sum_{j=1}^{\lfloor s/2 \rfloor} (A^*)^{i_j} v_j$. Note that any list with more than one entry yields a relation $\sum_{j=1}^{\lfloor s/2 \rfloor} (A^*)^{i_j} v_j + \sum_{j=1}^{\lfloor s/2 \rfloor} (A^*)^{i'_j} v'_j = 0$ for any two distinct entries $(i_2, \ldots, i_{\lfloor s/2 \rfloor}, v_1, \ldots, v_{\lfloor s/2 \rfloor})$ and $(i'_2, \ldots, i'_{\lfloor s/2 \rfloor}, v'_1, \ldots, v'_{\lfloor s/2 \rfloor})$. Once the table is constructed, we compute, for each $(i_{\lfloor s/2 \rfloor + 1}, \ldots, i_s, v_{\lfloor s/2 \rfloor + 1}, \ldots, v_s)$ $(1 < i_{\lfloor s/2 \rfloor + 1} < \cdots < i_s \leq N; v_{\lfloor s/2 \rfloor + 1}, \ldots, v_s \in X)$ the vector $v = \sum_{j=\lfloor s/2 \rfloor + 1}^{s} (A^*)^{i_j} v_j$ and use the value v as an index into the table. If there is any entry in the table corresponding to v, then for each such entry $(i_2, \ldots, i_{\lfloor s/2 \rfloor}, v_1, \ldots, v_{\lfloor s/2 \rfloor})$ for which $|\{i_2, \ldots, i_s\}| = s - 1$ we have a relation $\sum_{j=1}^{s} (A^*)^{i_j} v_j = 0$. Moreover, it is clear that every such relation can be found using this technique. The table can be constructed in time and space $\approx \binom{N-1}{\lfloor s/2 \rfloor - 1} 2^{r \lfloor s/2 \rfloor}$ and construction of relations from it in additional time $\approx \binom{N-1}{\lceil s/2 \rceil} 2^{r \lceil s/2 \rceil}$.

3.8.2 Concerning the Effectiveness of this Algorithm

An evaluation of attacks based on the algorithm described in the previous section requires a criterion for its convergence, analogous to that provided by Canteaut and Trabbia [3] for the convergence of the original Gallager algorithm (see section 1.5.4). The determination of such a criterion remains a topic for further study.

Chapter 4

Some Concluding Observations

In this final chapter, we make some general observations — and present some general results — concerning the Discrete Fourier Transform and its applications which arose naturally during the specific considerations of the previous two chapters.

4.1 The DFT and Probability Distributions

The DFT, defined in section 1.3.4, takes arguments which are complex-valued functions on a finite group G. Of particular interest to us is the case where these complex-valued functions in fact correspond to probability distributions on G:

$$g \mapsto \Pr(g)$$

If A_1 and A_2 are independent probability distributions on G, and $g_i \in G$ is distributed according to A_i $(i = 1, 2)$, the distribution $A_1 A_2$ on $g_1 g_2$, the result of combining g_1 and g_2 using the binary operation of multiplication in

G, is given by

$$\Pr_{A_1A_2}(g) = \sum_{g_1,g_2:g_1g_2=g} \Pr_{A_1}(g_1)\Pr_{A_2}(g_2). \tag{4.1}$$

Thus $g \mapsto \Pr_{A_1A_2}(g)$ is the convolution of $g \mapsto \Pr_{A_1}(g)$ and $g \mapsto \Pr_{A_2}(g)$: that is, their product in the algebra $\mathbb{C}G$.

Suppose now that $D : \mathbb{C}G \to \bigotimes_{i=1}^{h} \mathbb{C}^{d_i}$ is a DFT for G. Then

$$D\left(g \mapsto \Pr_{A_1A_2}(g)\right) = D\left(g \mapsto \Pr_{A_1}(g)\right) D\left(g \mapsto \Pr_{A_2}(g)\right). \tag{4.2}$$

From this it follows by induction that if $A_1, \ldots, A_s$ are pairwise independent probability distributions on a finite group G, and D is a DFT for G, then the distribution $A_1 \ldots A_s$ on a product $g_1 \ldots g_s$, where g_i is distributed according to A_i $(1 \leq i \leq s)$, can be computed as

$$\left(g \mapsto \Pr_{A_1\ldots A_s}(g)\right) = D^{-1}\left(\prod_{i=1}^{s} D(g \mapsto \Pr_{A_1}(g))\right).$$

This can imply a considerable improvement in computational efficiency over straightforward calculation not using D. For example, for $s = 2$ and G supersolvable, the two vectors $D\left(g \mapsto \Pr_{A_1}(g)\right)$ can be computed in time $\approx |G| \ln |G|$, their product in time $\approx |G|$, and D^{-1} in time $\approx |G| \ln |G| + |G|$, giving an overall complexity of order $|G| \ln |G|$. This compares favourably with the calculation of (4.1) performed by looping over all values of $g_1, g_2 \in G$, and hence of complexity $|G|^2$.

Consider finally the case of a distribution A on $G = C_2$, the cyclic group of order 2, which we identify with the set $\{0, 1\}$ under addition mod 2. If we write the distribution as a function $f : i \mapsto p_i$ $(i = 0, 1)$, then the Walsh-Hadamard transform D (a DFT for G) maps this to the function

$$\begin{cases} 0 \mapsto p_0 + p_1 \\ 1 \mapsto p_0 - p_1 \end{cases}$$

Note that $p_0 + p_1 = 1$, so that the distribution A is characterised entirely by

$$p_0 - p_1 = E((-1)^g).$$

This explains the importance of expressions of the form $E((-1)^g)$ in the work presented in chapter 3.

4.1.1 The "Pile-up" Lemma Revisited

It follows from the preceding paragraph that the "Pile-up Lemma", lemma 1.19, is an immediate consequence of the Walsh-Hadamard transform D on C_2. For suppose that X and Y are independent distributions on $\{0,1\}$. For any distribution Z on $\{0,1\}$ denote by f_Z the function $\{0,1\} \to \mathbb{C}$ mapping $i \mapsto \Pr_Z(i)$. By equation (4.2)

$$D(g \mapsto \Pr_{X \oplus Y}(g)) = D(g \mapsto \Pr_X(g)) D(g \mapsto \Pr_Y(g))$$

with D as defined in the previous paragraph. In particular,

$$D(g \mapsto \Pr_{X \oplus Y}(g))(1) = \left[D(g \mapsto \Pr_X(g))(1)\right] \left[D(g \mapsto \Pr_Y(g))(1)\right],$$

or, in the notation of 1.18,

$$c(X \oplus Y) = c(X)c(Y).$$

It readily follows from this, by induction, that for m independent distributions $X_1, \ldots X_m$ on $\{0,1\}$

$$c(X_1 \oplus \cdots \oplus X_m) = c(X_1) \ldots c(X_m),$$

as required.

4.2 Induced Distributions and the DFT

An important consideration in chapter 3 was the result presented in section 3.4 concerning the transform of a "reduced distribution". Here we present this result in a more general context.

Suppose G is a finite group with DFT D, and $H \leqslant G$ is a subgroup. Given a distribution A on G, there is a natural induced distribution $\bar{A}$ on G defined by averaging the probabilities of elements in the same coset of G. For convenience we will consider left cosets for the remainder of this section, but similar considerations apply to right cosets: it turns out that the result is simplest when G is abelian, and then we do not need to distinguish between the two.

Explicitly, $\bar{\mathrm{A}}$ is defined, for each $g \in G$, by

$$\Pr_{\bar{A}}(g) = \frac{1}{|H|} \sum_{h \in H} \Pr_A(gh).$$

We can now compute

$$\begin{aligned}
& D\left(\sum_{g \in G} \Pr_{\bar{A}}(g) g\right) \\
&= D\left(\sum_{g \in G} \frac{1}{|H|} \sum_{h \in H} \Pr_A(gh) g\right) \\
&= \frac{1}{|H|} D\left(\sum_{g \in G} \sum_{h \in H} \Pr_A(g) g h^{-1}\right) \\
&= \left[D\left(\sum_{g \in G} \Pr_A(g) g\right)\right] \left(\frac{1}{|H|} \sum_{h \in H} D(h)\right).
\end{aligned}$$

Moreover for $x \in G$ and any $h_1 \in H$,

$$\left(\sum_{h \in H} D(h)\right)(x)$$

$$= \left(\sum_{h\in H} D(hh_1)\right)(x)$$
$$= \left[\left(\sum_{h\in H} D(h)\right)(x)\right][D(h_1)(x)];$$

thus

$$\left[\left(\sum_{h\in H} D(h)\right)(x)\right][D(h_1)(x) - I] = 0.$$

In the case where G is abelian, so that $D : \mathbb{C}G \to \mathbb{C}^{|G|}$, this implies that either $\left(\sum_{h\in H} D(h)\right)(x) = 0$ or $D(h_1)(x) = 1$ for all $h_1 \in H$. Hence

$$D\left(\sum_{g\in G} \Pr_{\bar{A}}(g)g\right)(x)$$
$$= \begin{cases} D\left(\sum_{g\in G} \Pr_A(g)g\right)(x) & \text{if } D(h)(x) = 1 \text{ for all } h \in H \\ 0 & \text{otherwise.} \end{cases}$$

4.3 The DFT and Order 2 Information

In this section we revisit the ideas behind the proof of the conjecture for order 2 information presented in section 2.4.

Suppose that X is the uniform distribution on an abelian group G and that f is a probabilistic function $G \to [0,1]$ i.e. a map $f : G \to \{\text{distributions on } \{0,1\}\}$. Our objective is to express $I_2(X; f(X))$ in terms of a DFT D on G (c.f. definition 1.8). To help do this, first define three maps f_0, f_1 and $\tilde{f} : G \to \mathbb{C}$ by the rules

$$f_0(g) := \Pr(f(g) = 0)$$
$$f_1(g) := \Pr(f(g) = 1)$$
$$\tilde{f}(g) := \Pr(f(g) = 0) - \Pr(f(g) = 1)$$

for any $g \in G$. Denote the average values of these maps over all $g \in G$ by ϕ_0, ϕ_1 and ϕ respectively, so that

$$\begin{aligned} \phi_0 + \phi_1 &= 1 \\ \phi_0 &= (1+\phi)/2 \\ \phi_1 &= (1-\phi)/2 \end{aligned}$$

First we observe that

$$\begin{aligned} D(\tilde{f})(0) &= \sum_{h \in G} \tilde{f}(h) f_h(0) \qquad \text{by definition 1.8} \\ &= \sum_{h \in G} \tilde{f}(h) \qquad \text{since each } f_h(0) = f_h(0) f_h(0) \text{ and } |f_h(0)| = 1 \\ &= |G|\, \phi. \end{aligned}$$

Now we can compute

$$\begin{aligned} & I_2(X; f(X)) \\ &= \left[\sum_{\substack{g \in G \\ i=0,1}} \Pr(X=g) \Pr(f(X)=i) \left(\frac{\Pr(X=g, f(X)=i)}{\Pr(X=g)\Pr(f(X)=i)} \right)^2 \right] - 1 \\ &= \frac{1}{|G|} \left(\sum_{\substack{g \in G \\ i=0,1}} \frac{\Pr(f(g)=i)^2}{\Pr(f(X)=i)} \right) - 1 \\ &= \frac{1}{|G|} \left[\sum_{g \in G} \left(\frac{\left((1+\tilde{f}(g))/2\right)^2}{\phi_0} + \frac{\left((1-\tilde{f}(g))/2\right)^2}{\phi_1} \right) \right] - 1 \\ &= \frac{1}{|G|} \left[\sum_{g \in G} \left(\frac{1 + 2\tilde{f}(g) + \tilde{f}(g)^2}{4\phi_0} + \frac{1 - 2\tilde{f}(g) + \tilde{f}(g)^2}{4\phi_1} \right) \right] - 1 \\ &= \frac{1}{|G|} \sum_{g \in G} \left(\frac{1 + \tilde{f}(g)^2 + 2(\phi_1 - \phi_0)\tilde{f}(g) - 4\phi_0\phi_1}{4\phi_0\phi_1} \right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1 + \left(\frac{1}{|G|}\sum_{g\in G}\tilde{f}(g)^2\right) - 2\phi^2 - (1-\phi^2)}{1-\phi^2} \qquad \text{since } 4\phi_0\phi_1 = (1-\phi^2) \\
&= \frac{-\phi^2 + \frac{1}{|G|}\sum_{g\in G}\tilde{f}(g)^2}{1-\phi^2} \\
&= \frac{-\phi^2 + \frac{1}{|G|^2}\sum_{g\in G}D(\tilde{f})(g)^2}{1-\frac{1}{|G|^2}D(\tilde{f})(0)^2} \\
&= \frac{\sum_{g\in G\setminus\{0\}} c_g^2}{1-c_0^2}
\end{aligned}
$$

where we have written $c_g := \frac{1}{|G|}D(\tilde{f})(g)$ for each $g \in G$.

References

[1] Apostol, T.M. (1974). *Mathematical Analysis*. Addison-Wesley, second edition.

[2] Brynielsson, L. (1989). Below the Unicity Distance. *Proceedings of the E.I.S.S. (European Institute for System Security) Workshop on Stream Ciphers* held at Karlsruhe, Germany.

[3] Canteaut, A. and Trabbia, M. (2000). Improved Fast Correlation Attacks Using Parity-Check Equations of Weight 4 and 5. *Advances in Cryptology — Eurocrypt 2000*, pp. 573–588.

[4] Clausen, M. and Baum, U. (1993). *Fast Fourier Transforms*. BI-Wissenschaftsverlag, first edition.

[5] Dodd, M.W. (1995). Simultaneous Correlation to Many Linear Functionals: a New Cryptanalytic Technique which Can Almost Halve the Effective Key Size of Certain Stream Ciphers. *Proc. 4th IMA Conf. on Cryptography and Coding, Cirencester, 1993* (published by the IMA as *Codes and Cyphers: Cryptography and Coding IV*, ed. P.G. Farrell, 1995).

[6] Gallager, R.G. (1962). Low-Density Parity-Check Codes. *IRE Transactions on Information Theory*, IT-8, pp. 21–8.

[7] Hagenauer, J., Offer, E. and Papke, L. (1996). *Iterative decoding of binary block and convolutional codes*. IEEE Trans. Inform. Theory, vol. 42, no. 2, pp. 429–445.

[8] MacWilliams, F.J. and Sloane, N.J.A. (1977). *The Theory of Error-Correcting Codes*. North Holland, first edition, pp. 406–32.

[9] Meier, W. and Staffelbach, O. (1989). Fast correlation attacks on certain stream ciphers. *Journal of Cryptology*, vol. 1, no. 3, pp. 159–176.

[10] Mund, S., Gollman, D., and Beth, T. (1987). Some remarks on the cross correlation analysis of pseudo-random generators. *Advances in Cryptology — Eurocrypt '87*, pp. 25–35.

[11] Shannon, C.E. (1949). Communication Theory of Secrecy Systems. *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715.

[12] Siegenthaler, T. (1984). Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Transaction on Information Theory*, IT-30, no. 5, pp. 776–80.

[13] Siegenthaler, T. (1985). Decrypting a class of stream ciphers using ciphertext only. *IEEE Transaction on Computers*, C-34, no. 1, pp. 81–5.