Problems in Extremal and Probabilistic Combinatorics: Cubes, Squares and Permutations



Tom Johnston
Lady Margaret Hall
University of Oxford

Thesis
Doctor of Philosophy in Mathematics
Trinity Term 2021

Abstract

We begin by studying the possible intersection sizes of a k-dimensional linear subspace with the hypercube $\{0,1\}^n$. For a fixed k, the largest intersection size is 2^k and it was shown by Melo and Winter in [77] that the second largest intersection size is $2^{k-1}+2^{k-2}$. We show that all intersections sizes larger than 2^{k-1} are of the form $2^{k-1}+2^i$ or $35\cdot 2^{k-6}$, completely determining the "large" intersection sizes and disproving a conjecture of Melo and Winter. We also show that the only possible intersection size in the interval $(15\cdot 2^{k-5}, 2^{k-1})$ is $63\cdot 2^{k-7}$, disproving a second conjecture of Melo and Winter that all of the "small" values are present.

Next we consider Lipschitz bijections between functions from $\{0,1\}^n$ to $\{0,1\}^n$ and consider how much these bijections must "stretch" the edges of the hypercube. We answer four questions posed by Rao and Shinkar in [85]. (1) We show that there is no $O(1)$-bi-Lipschitz bijection from Dictator to XOR such that each output bit depends on $O(1)$-input bits. (2) We give a construction for a mapping from XOR to Majority which has average stretch $O(\sqrt{n})$, matching a previously known lower bound. (3) We give a 3-Lipschitz embedding $\phi : \{0,1\}^n \to \{0,1\}^{2n+1}$ such that $\mathrm{XOR}(x) = \mathrm{Majority}(\phi(x))$ for all $x \in \{0,1\}^n$. (4) We show that with high probability there is a $O(1)$-bi-Lipschitz mapping from Dictator to a uniformly random balanced function.

Moving from (hyper)cubes to squares, we consider a problem about zero-sum squares in $\{-1,1\}$-matrices. Given a matrix $M = (a_{i,j})$ a square is a 2×2 submatrix with entries $a_{i,j}$, $a_{i,j+s}$, $a_{i+s,j}$, $a_{i+s,j+s}$ for some $s \geq 1$, and a zero-sum square is a square where the entries sum to 0. Recently, Arévalo, Montejano and Roldán-Pensado [6] proved that all large $n\times n$ $\{-1,1\}$-matrices with discrepancy $|\sum a_{i,j}| \leq n$ contain a zero-sum square unless they are split, and conjectured that this can be improved to Cn for large enough n. We show that a quadratic bound holds: all large $n\times n$

$\{-1, 1\}$-matrices with discrepancy at most $n^2/4$ are either split or contain a zero-sum square.

We next consider another problem regarding squares, but this time squares in permutations. A square in a permutation is a factor $S = (S_1; S_2)$ where S_1 and S_2 have the same pattern, and a permutation is said to be square-free if it contains no non-trivial squares. The permutation is further said to be bicrucial with respect to squares if every extension to both the left and right contains a square. We completely classify for which n there exists a bicrucial square-free permutation, confirming two conjectures of Gent, Kitaev, Konovalov, Linton and Nightingale [49].

Two permutations σ and π are said to be ℓ-similar if they can be partitioned into subpermutations $\sigma^{(1)}, \ldots, \sigma^{(\ell)}$ and $\pi^{(1)}, \ldots, \pi^{(\ell)}$ respectively such that $\sigma^{(i)}$ and $\pi^{(i)}$ have the same pattern for all $i \in [\ell]$. Clearly, two permutations may only be n-similar, but what happens in the "average" case? It is easy to get an upper bound of $O(n^{1/2})$ using the Erdős-Szekeres Theorem, and a lower bound of $\Omega(n^{1/3})$ follows from the length of the longest twin in a random permutation. We prove that two independent uniformly random permutations are $O\big(n^{1/3}\log^{11/6}(n)\big)$-similar with high probability, matching the lower bound up to a polylogarithmic factor.

Finally, we consider creating uniformly random permutations using sequences of independent lazy transpositions and ask how long such a sequence must be. We give an improved construction which uses $(2/3)\binom{n}{2} + O(n\log(n))$ lazy transpositions, giving a negative answer to a problem of Angel and Holroyd [5]. The best-known lower bound of $\Omega(n\log(n))$ is based on the number of transpositions needed to "reach" each configuration with positive probability. We introduce similar reachability questions for smaller subsets and settle the case of subsets of size 2 exactly: the smallest sequence of transpositions such that each pair of distinct positions (x, y) has a subsequence mapping $(1, 2)$ to (x, y) has length $\lceil 3n/2 \rceil - 2$. We also consider the case where all transpositions are of the form $(1, \cdot)$ and use the structure in this case to provide the first separation between uniformity and reachability.

Acknowledgements

First and foremost, I'd like to thank my supervisor Alex Scott for all his support and guidance over the course of my DPhil. He has supplied me with more interesting problems than I have had time to consider, and it has been a pleasure to work on (and even solve) some of them with him.

I'd like to thank Carla for the many good times we've shared over the course of my DPhil, and I know she will be a lifelong friend. Not only have I enjoyed working on our numerous shared projects, my time in Oxford would not have been anywhere near as fun without her company away from maths and on our frequent trips to the pub. There are still many pubs left on the pub crawl poster, and I look forward to ticking off the rest of them (and then we'll have to branch out to pubs around the world)!

I have particularly enjoyed working on problems with others, and I must thank all of my coauthors for their time spent working with me. I'd also like to thank the whole combinatorics group; their advice and support has been invaluable, and the lessons I've learnt will be useful throughout the rest of my mathematical career and beyond. I've also appreciated the (sometimes virtual) socials outside of maths, especially during the lockdowns when there was little else going on.

I'd also like to offer my thanks to my friends and family for all they have done to support me and for many of the good memories I'll take with me from Oxford. In particular, I must thank my housemates Jake, Anna, Jess and Ben for putting up with me and my (perfectly sensible) projects. Outside of maths I have spent much of my time playing badminton with Oxford University Badminton Club, and I'd like to thank everyone who has made my time in the club so enjoyable, especially my frequent doubles (and pub) partner Tom.

Statement of Originality

Chapter 2 is based on joint work with Carla Groenland [53] and a version appeared in her DPhil thesis [51]. Chapter 3 is based on joint work with Alex Scott [64]. Chapter 5 is based on joint work with Carla Groenland [54]. Chapter 6 is based on joint work with Carla Groenland, Dániel Korándi, Alexander Roberts, Alex Scott and Jane Tan. Chapter 7 is based on joint work with Carla Groenland, Jamie Radcliffe and Alex Scott.

Contents

1 Introduction **1**

1.1 Intersection sizes of linear subspaces with the hypercube 1

1.2 Lipschitz bijections between boolean functions 2

1.2.1 Bijections between Dictator and Majority 3

1.2.2 Bijections between Dictator and XOR 4

1.2.3 Bijections between XOR and Majority 5

1.2.4 Bijections from Dictator to a random f 6

1.3 Zero-sum squares in $\{-1, 1\}$-matrices with low discrepancy 7

1.4 Squares and twins in permutations 9

1.4.1 Bicrucial square-free permutations 10

1.4.2 Decomposing permutations into order-isomorphic subpermutations 11

1.5 Uniform shuffling using lazy transpositions 15

1.5.1 Shorter transposition shuffles 16

1.5.2 Reachability 17

1.5.3 Star transposition shuffles 18

2 Intersection sizes of linear subspaces with the hypercube **20**

2.1 Framework and notation 21

2.2 Preliminaries 23

2.3 Large intersection sizes 24

2.4 Small intersection sizes 30

2.5 Computer search 33

2.6 Open problems 35

3 Lipschitz bijections between boolean functions **36**

3.1 Bijections from Dictator to XOR 38

3.2 Bijections from XOR to Majority 39

3.3 Bijections from Dictator to a random f . . . 41
3.4 Open Problems . . . 47

4 Zero-sum squares in $\{-1, 1\}$-matrices with low discrepancy 50
4.1 Proof of Theorem 4.2 . . . 51
4.2 Open problems . . . 62

5 Bicrucial square-free permutations 65
5.1 Preliminaries . . . 67
5.2 Constructions for large n . . . 68
5.3 Bicrucial permutations for small even n . . . 75
5.3.1 The non-existence of bicrucial permutations of length 38 . . . 76
5.3.2 Enumerating square-free permutations . . . 78
5.4 Open problems . . . 81

6 Decomposing permutations into twins 83
6.1 Proof of Theorem 6.2 . . . 84
6.2 Open problems . . . 90

7 Uniform shuffling with lazy transpositions 92
7.1 Preliminaries . . . 96
7.2 A new divide and conquer construction . . . 99
7.2.1 Splitting into two groups . . . 102
7.2.1.1 More positions than counters . . . 105
7.2.2 Splitting into three or more groups . . . 109
7.3 Reachability . . . 113
7.4 Star transposition shuffles . . . 118
7.5 Open problems . . . 124

Appendix A Bicrucial square-free permutations 126
A.1 Constructions from Theorem 5.5 . . . 126
A.2 Bicrucial permutations for small even n . . . 127

Appendix B Best known bounds on $U(n)$ 129

Bibliography 130

List of Figures and Tables

2.1 An example of the shape of a non-minimal map (left) and of a minimal map (right). 22
2.2 A $(3, 2)$-star with 2 edges. 26

4.1 The entries known from applying Lemma 4.3. 52
4.2 A subset of the regions used in the proof of Lemma 4.4. 53
4.3 The matrix M with the submatrices N and $B_1, \ldots, B_{11}$. 57
4.4 The three possible cases when C contains $a_{1,n}$ and $a_{1,n} = 1$. 61
4.5 The case where C contains $a_{n,n}$ and $a_{n,n} = 1$. 62
4.6 A 9×9 $\{-1, 1\}$-matrix with no zero-sum squares and discrepancy 31. Amongst all $\{-1, 1\}$ matrices which are neither split nor contain a zero-sum square this has the smallest (in magnitude) discrepancy. . . 63
4.7 An 8×8 $\{-1, 1\}$-matrix with no zero-sum squares and discrepancy 30. 63

5.1 A permutation and a possible right extension. 66
5.2 A bicrucial permutation of length 51 constructed as in the proof of Theorem 5.4. 69
5.3 A bicrucial permutation of length 72 constructed as in the proof of Theorem 5.5. 72
5.4 The number of square-free, left-crucial and bicrucial permutations of length n. 80

6.1 An example set of matched red and blue points in the unit square. . . 87
6.2 A sample minimax matching along with the bipartite graph for the colour $(0, 0)$. 88

7.1 The $(3, 6)$-shuffle with just 11 transpositions given in the proof of Lemma 7.17. 102
7.2 The $(6, 12)$-shuffle with just 45 transpositions given in the proof of Lemma 7.17. 102

7.3 A $(3, 12)$-shuffle using just 25 transpositions constructed as in Lemma 7.19. 106
7.4 A $(3, 15)$-shuffle using just 37 transpositions constructed as in the proof of Lemma 7.21. 109
7.5 A $(4, 12)$-shuffle using just 35 transpositions constructed as in the proof of Lemma 7.22. 111
7.6 The multigraph corresponding to the minimum 2-reachable sequence given in the proof of Theorem 7.7. 114
7.7 A $(4, 12)$-shuffle using only star transpositions constructed as in the proof of Lemma 7.23. 119

A.1 Bicrucial permutations of length 64, 66 and 68 constructed as in the proof of Theorem 5.5. 126
A.2 Bicrucial permutations of lengths 34, 36, 40, 42, 44 and 46 respectively. 128

Chapter 1

Introduction

1.1 Intersection sizes of linear subspaces with the hypercube

What are the possible sizes of the intersection of the hypercube $\mathbb{H}^n = \{0,1\}^n$ and a k-dimensional linear subspace S of $\mathbb{R}^n$? Since both contain the point 0, the smallest possible intersection size is 1, and largest is 2^k, but which numbers in between are possible? Motivated by a combinatorial observation related to problem in quantum information [81], Melo and Winter [77] initiated the study of the set $H(k,m)$ of possible intersection sizes a k-dimensional subspace can have with the $(m+k)$-dimensional hypercube. They made two conjectures about the structure of $H(k,\infty) = \bigcup_{m\in\mathbb{N}} H(k,m)$, depending on whether the intersection size is "large" (greater than 2^{k-1}) or "small" (at most 2^{k-1}). Let $H^+(k,m)$ be the large values of $H(k,m)$, i.e.

$$H^+(k,m) = H(k,m) \setminus \{1,\dots,2^{k-1}\},$$

and set $H^+(k,\infty) = H(k,\infty) \setminus \{1,\dots,2^{k-1}\}$. Melo and Winter proved the containment relation $H^+(k,\infty) \supseteq \{2^{k-1}+2^i : i \in \{0,1,\dots,k-1\}\}$ and conjectured that this is all of $H^+(k,\infty)$. As a first step towards this conjecture, they proved that the second largest value in $H^+(k,m)$ is $2^{k-1}+2^{k-2}$ as expected.

Conjecture 2.1 (Conjecture 3.2 in [77])**.** *For all $k \geq 1$,*

$$H^+(k,\infty) = \{2^{k-1}+2^i : i \in \{0,1,\dots,k-1\}\}.$$

In Chapter 2 we show that this conjecture is almost correct: there is one additional value in $H^+(k,\infty)$ for $k \geq 6$.

Theorem 2.2. *For $k \geq 6$,*

$$H^+(k,\infty) = \{2^{k-1} + 2^i : i \in \{0,\ldots,k-1\}\} \cup \{2^{k-1} + 2^{k-5} + 2^{k-6}\}.$$

In fact, by proving a precise structural result and combining this with a computer search, we determine $H^+(k,m)$ for every m from which the theorem above follows immediately. In particular, when restricted to the large intersections, we are also able to answer another question raised by Melo and Winter [77], namely, given $t \in H^+(k,\infty)$ what the smallest m such that $t \in H^+(k,m)$?

Let the small values of $H(k,m)$ be $H^-(k,m) = H(k,m) \cap \left[2^{k-1}\right]$, where we use the standard notation $[i] = \{1,\ldots,i\}$, and define $H^-(k,\infty)$ by $H^-(k,\infty) = H(k,\infty) \cap \left[2^{k-1}\right]$. Based on their observations for small values of k, Melo and Winter conjecture that every small value is present.

Conjecture 2.3 (Conjecture 3.3 in [77])**.**

$$H^-(k,\infty) = [2^{k-1}].$$

By extending the method used in the proof of Theorem 2.2 to work just below 2^{k-1}, we disprove the conjecture with the following theorem.

Theorem 2.4. *For $k \geq 8$ and any $m \geq 1$,*

$$H(k,m) \cap \left\{\tfrac{15}{16}2^{k-1},\ldots,2^{k-1}\right\} = \left\{\tfrac{15}{16}2^{k-1}, \tfrac{126}{128}2^{k-1}, 2^{k-1}\right\}.$$

The chapter on intersection sizes of linear subspaces with the hypercube is based on joint work with Carla Groenland.

1.2 Lipschitz bijections between boolean functions

The analysis of boolean functions, in particular their Fourier coefficients and noise stability, is widely-studied and has applications in many areas of mathematics including the theory of social choice, complexity theory and in the hardness of approximations (see for instance [24, 66, 74, 80, 23, 17, 36, 14, 65, 16]). A frequent theme in the literature is the analysis of the similarities and differences between boolean functions with different geometric or structural properties; for example, between functions such as Dictator that are determined by a small number of coordinates, and functions such as Majority or XOR that are not. One measure of similarity between functions is the existence of a Lipschitz mapping (with small constant) between them. In this

paper we continue the study of Rao and Shinkar [85] on Lipschitz mappings between the boolean functions Dictator, XOR, Majority and a uniformly random balanced function and answer several of the questions they pose.

Given two boolean functions $f, g : \{0,1\}^n \to \{0,1\}$ we say that a bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ is a *mapping from f to g* if, for every $x \in \{0,1\}^n$, we have $f(x) = g(\phi(x))$. Write a point $x \in \{0,1\}^n$ as $x = (x_1, \ldots, x_n)$ and, for $\phi : \{0,1\}^n \to \{0,1\}^n$, write $\phi = (\phi_1, \ldots, \phi_n)$. For $x, y \in \{0,1\}^n$, let $|x| = \sum_{i=1}^n x_i$ denote the Hamming weight of x and $\mathrm{dist}(x,y) = \sum_{i=1}^n |x_i - y_i|$ the Hamming distance between x and y.

A bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ is said to be *C-Lipschitz* if, for all $x, y \in \{0,1\}^n$, $\mathrm{dist}(\phi(x), \phi(y)) \leq C\,\mathrm{dist}(x,y)$, and ϕ is said to be *C-bi-Lipschitz* if both ϕ and ϕ^{-1} are C-Lipschitz. As a relaxation from being Lipschitz we define the *average stretch* of a mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ by

$$\mathrm{avgStretch}(\phi) = \mathbb{E}_{x,i}[\mathrm{dist}(\phi(x), \phi(x + e_i))]$$

where $x \in \{0,1\}^n$ and $i \in [n]$ are both chosen independently and uniformly at random.

Given a bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ we say that the jth output bit ϕ_j *depends on* the ith input bit x_i if there exists $x \in \{0,1\}^n$ such that $\phi_j(x) \neq \phi_j(x + e_i)$. If every output bit depends on at most k input bits, we say the map ϕ is *k-local*.

In Sections 3.1 and 3.2 we study mappings between three boolean functions Dictator, XOR and Majority which we define by

- $\mathrm{Dictator}(x) = x_1$,
- $\mathrm{XOR}(x) = \sum_{i=1}^n x_i \mod 2$,
- $\mathrm{Majority}(x) = 1$ if $|x| > n/2$ and $\mathrm{Majority}(x) = 0$ otherwise.

In Section 3.3 we consider a uniformly random balanced function $f : \{0,1\}^n \to \{0,1\}$ where we say a boolean function $g : \{0,1\}^n \to \{0,1\}$ is *balanced* if $g^{-1}(1)$ and $g^{-1}(0)$ are of the same size. Clearly, both XOR and Dictator are always balanced while Majority is only balanced when n is odd (and so a bijection from Dictator or XOR to Majority can only exist for odd n).

1.2.1 Bijections between Dictator and Majority

The functions Dictator and Majority are in many ways opposites of each other. For example, the first coordinate clearly has a large influence on the value of Dictator, while for Majority every coordinate has the same small amount of influence. There

are many results which show that functions which differ from Majority in some way must have influential coordinates and are therefore similar to Dictator functions. For example, the "Majority Is Stablest" theorem of Mossel, O'Donnell and Oleszkiewicz in [80] shows that if a balanced boolean function is essentially more noise-stable than Majority, then it must have an influential coordinate.

It is straightforward to see that no bijection ϕ from Dictator to Majority can be C-Lipschitz for any $C \leq n/2$. Indeed, let n be odd and suppose ϕ is such a bijection. Take $x \in \{0,1\}^n$ to be the point such that $\phi(x) = \sum_{i=1}^{n} e_i$. Clearly $y := x - e_1$ differs from x in the first coordinate so Dictator$(y) \neq$ Dictator(x). By the definition of ϕ, we must have Majority$(\phi(y)) = 0$ and $|\phi(y)|$ is less than $n/2$. Hence, $\text{dist}(\phi(x), \phi(y)) > n/2$ which gives a contradiction.

For maps from Majority to Dictator, the situation is better. Rao and Shinkar [85] showed that, for all odd n, there is a mapping ψ from Majority to Dictator which is 11-Lipschitz. As noted above, the function ψ^{-1} cannot be $O(1)$-Lipschitz. However, it has the weaker property that ψ^{-1} has $O(1)$-average stretch. Indeed, Rao and Shinkar's construction gives a Lipschitz function that is in fact an $O(1)$-bi-Lipschitz bijection from the upper half of $\{0,1\}^n$ to the half-cube with first coordinate 1; similarly ψ induces a $O(1)$-bi-Lipschitz bijection between the lower half of $\{0,1\}^n$ and the half-cube with first coordinate 0. As there are only 2^{n-1} edges between the two half-cubes, the average stretch of ψ^{-1} is $O(1)$.

1.2.2 Bijections between Dictator and XOR

Rao and Shinkar note that the map ϕ given by $\phi(x) = (\text{XOR}(x), x_2, x_3, \ldots, x_n)$ is a mapping from Dictator to XOR, and it is easy to check that ϕ is 2-bi-Lipschitz. This leads them to consider maps with stronger properties. In the above map, the first output bit depends on all n input bits and the map is not k-local for any $k < n$. However, one can easily find a map which is 2-local: Rao and Shinkar give the example $\phi(x) = (x_1 + x_2, x_2 + x_3, \ldots, x_{n-1} + x_n, x_n)$ which is 2-Lipschitz and 2-local, although the inverse of this map is not even $(n-1)$-Lipschitz. In [85] Rao and Shinkar construct a map ϕ which is 2-Lipschitz and 3-local and where the inverse is $O(\log(n))$-Lipschitz. This leads them to ask the following question.

Question 3.1 (Question 6.1 in [85])**.** *Is there a mapping from Dictator to XOR that is $O(1)$-local and $O(1)$-bi-Lipschitz?*

We answer this in the negative with the following theorem.

Theorem 3.2. *Let $\phi : \{0,1\}^n \to \{0,1\}^n$ be a mapping from* Dictator *to* XOR *which is C-Lipschitz and where each output bit depends on at most k input bits. Then there is a constant $\delta = \Omega(1/(k + \log(C)))$ such that the inverse map ϕ^{-1} is not $\delta \log(n)$-Lipschitz.*

Furthermore, if ϕ is a linear mapping, then we may take $\delta = \Omega(1/\log(C + k))$.

It follows that the map constructed by Rao and Shinkar in [85] is essentially best possible: if ϕ is a mapping from Dictator to XOR which is $O(1)$-Lipschitz and $O(1)$-local, then ϕ^{-1} cannot be $o(\log(n))$-Lipschitz.

1.2.3 Bijections between XOR and Majority

Since we have a 2-bi-Lipschitz map from Dictator to XOR, we might expect the maps between XOR and Majority to behave similarly to those between Dictator and Majority. Indeed, composing this 2-bi-Lipschitz map from Dictator to XOR with the $O(1)$-Lipschitz map from Majority to Dictator gives a $O(1)$-Lipschitz map from Majority to XOR, and the argument used to show that there is no $O(1)$-Lipschitz mapping from Dictator to Majority can also be applied to mappings from XOR to Majority to show that they cannot be C-Lipschitz for $C \leq n/2$. However, while there is a map from XOR to Dictator with constant average stretch, Rao and Shinkar [85] show that for any mapping ϕ from XOR to Majority the average stretch is $\Omega(\sqrt{n})$. They then ask if this lower bound is tight:

Question 3.3 (Question 6.3 in [85])**.** *Is there a mapping ϕ from* XOR *to* Majority *such that* $\text{avgStretch}(\phi) = O(\sqrt{n})$*?*

In Section 3.2 we show that the bound is indeed tight by giving a map which has average stretch $O(\sqrt{n})$. This result also follows from a recent result of Boczkowski and Shinkar [19] where they prove, for any two sets $A, B \subseteq \{0,1\}^n$ with $|A| = |B| = 2^{n-1}$, there is a mapping $\phi : A \to B$ with $\mathbb{E}[\text{dist}(x, \phi(x))] \leq \sqrt{2n}$.

Theorem 3.4. *For odd n there is a mapping ϕ from* XOR *to* Majority *such that*

$$\text{avgStretch}(\phi) = O\big(\sqrt{n}\big).$$

Given the need for such a mapping to have large average stretch, Rao and Shinkar also ask what happens if we relax the problem from finding a bijection to finding an embedding in a larger space.

Question 3.5 (Question 6.4 in [85])**.** *Is there a Lipschitz embedding $\phi : \{0,1\}^n \to \{0,1\}^{\text{poly}(n)}$ such that $\text{XOR}(x) = \text{Majority}(\phi(x))$ for all $x \in \{0,1\}^n$?*

We give a simple construction that gives a positive answer:

Theorem 3.6. *For every n there exists a 3-Lipschitz embedding $\phi : \{0,1\}^n \to \{0,1\}^{2n+1}$ such that* $\mathrm{XOR}(x) = \mathrm{Majority}(\phi(x))$ *for all $x \in \{0,1\}^n$.*

1.2.4 Bijections from Dictator to a random f

Rao and Shinkar also consider random functions. We say a function $f : \{0,1\}^n \to \{0,1\}$ is *balanced* if $f^{-1}(1)$ and $f^{-1}(0)$ are of the same size. Building on a construction of Håstad, Leighton and Newman [58], Rao and Shinkar show that with high probability there is a bijection from Dictator to a uniformly random balanced function f which has average stretch bounded by an absolute constant [85]. They ask whether we can in fact ask for more:

Question 3.7 (Question 6.2 in [85])**.** *Is it true that with high probability there is a $O(1)$-bi-Lipschitz mapping from* Dictator *to a uniformly random balanced function f?*

We not only give a positive answer to this question, we prove a stronger statement about the maximum distance between x and $\phi(x)$.

Theorem 3.8. *Let $f : \{0,1\}^n \to \{0,1\}$ be a balanced boolean function chosen uniformly at random. Then, with high probability, there exists a mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ from* Dictator *to f such that for all $x \in \{0,1\}^n$,*

$$\mathrm{dist}(x, \phi(x)) \leq C$$

where C is an absolute constant. In particular, ϕ is $(2C+1)$-bi-Lipschitz.

It is straightforward to extend this to bijections between two independent uniformly random balanced boolean functions f and g.

Corollary 3.9. *Let $f, g : \{0,1\}^n \to \{0,1\}$ be two balanced boolean functions chosen independently and uniformly at random. Then, with high probability, there exists a mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ from f to g such that for all $x \in \{0,1\}^n$,*

$$\mathrm{dist}(x, \phi(x)) \leq C$$

where C is an absolute constant. In particular, ϕ is $(2C+1)$-bi-Lipschitz.

Proof. By the above theorem there exists a mapping ϕ_f from Dictator to f such that $\text{dist}(x, \phi_f(x)) \leq c$ and a mapping ϕ_g from Dictator to g such that $\text{dist}(x, \phi_g(x)) \leq c$ for some absolute constant c, both with high probability. Using the triangle inequality, it is easy to see that $\phi = \phi_g \circ \phi_f^{-1}$ is a mapping from f to g such that $\text{dist}(x, \phi(x)) \leq 2c$. This proves the first part of the corollary with $C = 2c$.

The second part also follows easily. Fix $x \in \{0,1\}^n$ and let $y = x + e_i$ for some i. Then

$$\text{dist}(\phi(x), \phi(y)) \leq \text{dist}(x, \phi(x)) + \text{dist}(x, y) + \text{dist}(y, \phi(y)) \leq 2C + 1$$

and the triangle inequality shows ϕ is $(2C+1)$-Lipschitz. A similar argument shows that ϕ^{-1} is also $(2C+1)$-Lipschitz. □

The chapter on Lipschitz bijections between boolean functions is based on joint work with Alex Scott.

1.3 Zero-sum squares in $\{-1, 1\}$-matrices with low discrepancy

A *square* S in a matrix $M = (a_{i,j})$ is a 2×2 submatrix of the form

$$S = \begin{pmatrix} a_{i,j} & a_{i,j+s} \\ a_{i+s,j} & a_{i+s,j+s} \end{pmatrix}.$$

In 1996 Erickson [44] asked for the largest n such that there exists an $n \times n$ binary matrix M with no squares which have constant entries. An upper bound was first given by Axenovich and Manske [8], before the answer, 14, was determined by Bacjer and Eliahou in [9].

Recently, Arévalo, Montejano and Roldán-Pensado [6] initiated the study of a zero-sum variant of Erickson's problem. Here we wish to avoid *zero-sum squares*, squares with entries that sum to 0.

Zero-sum problems have been well-studied since the Erdős-Ginsburg-Ziv Theorem in 1961 [40], which says that any set of $2n-1$ integers must have a set of n which sum to 0 modulo n. Much of the research has been on zero-sum problems in finite abelian groups (see the survey [47] for details), but problems have also been studied in other settings such as on graphs (see e.g. [29, 27, 26, 18]). Of particular relevance is the work of Balister, Caro, Rousseau and Yuster in [11] on submatrices of integer valued matrices where the rows and columns sum to 0 mod p, and the work of

Caro, Hansberg and Montejano on zero-sum subsequences in bounded sum $\{-1,1\}$-sequences [28].

Given an $n \times m$ matrix $M = (a_{i,j})$ define the *discrepancy* of M as the sum of the entries, that is

$$\operatorname{disc}(M) = \sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} a_{i,j}.$$

We say a square S is a *zero-sum square* if $\operatorname{disc}(S) = 0$, or equivalently,

$$a_{i,j} + a_{i,j+s} + a_{i+s,j} + a_{i+s,j+s} = 0.$$

We will be interested in $\{-1,1\}$-matrices M which do not contain any zero-sum squares. Clearly, matrices with at most one -1 cannot contain a zero-sum square and, in general, there are many such matrices when the number of -1s is low. But what happens if the number of 1s and the number of -1s are similar? In particular, what happens if the matrix M is itself zero-sum?

An $n \times m$ $\{-1,1\}$-matrix $M = (a_{i,j})$ is said to be *t-split* for some $0 \le t \le n+m-1$ if

$$a_{i,j} = \begin{cases} 1 & i+j \le t+1, \\ -1 & i+j \ge t+2. \end{cases}$$

We say a matrix M is *split* if there is some t such that a t-split matrix N can be obtained from M by applying vertical and horizontal reflections. Split matrices are of particular interest since they can have low discrepancy, yet they never contain a zero-sum square. However, it is not hard to check that an $n \times n$ split matrix cannot have discrepancy 0, and it may still be the case that a zero-sum matrix M must contain a zero-sum square.

This was confirmed by Arévalo, Montejano and Roldán-Pensado in [6]. In fact, they proved that, except when $n \le 4$, every $n \times n$ non-split $\{-1,1\}$-matrix M with $|\operatorname{disc}(M)| \le n$ has a zero-sum square. They remark that it should be possible to extend their proof to give a bound of $2n$, and they conjecture that the bound Cn should hold for any $C > 0$ when n is large enough relative to C.

Conjecture 4.1 (Conjecture 4 in [6])**.** *For every $C > 0$ there is a integer N such that whenever $n \ge N$ the following holds: every $n \times n$ non-split $\{-1,1\}$-matrix M with $|\operatorname{disc}(M)| \le Cn$ contains a zero-sum square.*

Let $f(n)$ be the minimum discrepancy of a non-split $\{-1,1\}$-matrix with no zero-sum squares. Arévalo, Montejano and Roldán-Pensado proved that $f(n) \ge n+1$ for all $n \ge 5$, and the conjecture would imply that $f(n) = \omega(n)$. We improve the lower bound on f to $\lfloor n^2/4 \rfloor + 1$ (for all $n \ge 5$), showing that $f = \Omega(n^2)$.

Theorem 4.2. *Let $n \geq 5$. Every $n \times n$ non-split $\{-1, 1\}$-matrix M with $|\operatorname{disc}(M)| \leq n^2/4$ contains a zero-sum square.*

The best known construction for a non-split matrix with no zero-sum squares has discrepancy close to $n^2/2$, about twice the lower bound given here, and our computer experiments suggest that this construction is in fact optimal. Closing the gap between the upper and lower bounds remains a very interesting problem and we discuss it further in Section 4.2.

1.4 Squares and twins in permutations

Given a vector $a = (a_1, \ldots, a_n)$ consisting of n distinct real numbers, one can associate a unique permutation $\sigma \in S_n$ (which we write as a vector $(\sigma(1), \ldots, \sigma(n))$) by replacing the ith smallest element of a with i, and we call this σ the *pattern* of a. We say two vectors a and b are *order-isomorphic* if they have the same pattern, and in this case we write $a \sim b$. For example, the vector $(5, 2, 4, 10)$ has the pattern $(3, 1, 2, 4)$, or if we index from 0 instead, the pattern $(2, 0, 1, 3)$.

Patterns in permutations have been well-studied, particularly the enumeration of permutations which avoid certain patterns. The first such result is from MacMahon [75] who showed that the permutations which can be decomposed into two decreasing sequences (i.e. those that avoid the pattern $(1, 2, 3)$) are enumerated by the Catalan numbers. The Catalan numbers also enumerate the permutations which can be sorted using a stack, which are the permutations avoiding the pattern $(2, 3, 1)$ [68], and there has since been much work on which permutations can be sorted under different restrictions (see e.g. [22] for more details). We remark that we have just seen that the $(1, 2, 3)$-avoiding and $(2, 3, 1)$-avoiding permutations are enumerated by the Catalan numbers, and there are now many bijections between the two sets of permutations [33]. More generally, the question of when there the same number of permutations avoiding two different permutations is a common area of study [88, 87, 10]. Another important result in this area is the proof of the Stanley-Wilf conjecture which states that the size of every proper permutation class is singly exponential [76].

We will consider two questions related to patterns in permutations. First, we consider the existence of square-free permutations σ such that any extension of σ contains a square, and we completely classify the n for which such permutations exist. Second, we consider decomposing two uniformly random permutations into order-isomorphic subpermutations, and show that $O\big(n^{1/3} \log^{11/6}(n)\big)$ subpermutations suffice with high probability.

1.4.1 Bicrucial square-free permutations

A word W contains a *square* if it has two equal consecutive factors, that is, if $W = UXXV$ for some non-empty word X. For example, the word 102021 contains the square 0202 while the word 01213 is square-free. Squares in words have been well studied since Thue [91] proved that there is an infinite square-free word over an alphabet of size 3 (known as the Prouhet–Thue–Morse sequence), and it has seen applications in a variety of fields. The most famous application was in constructing infinite, finitely generated groups of exponent n (for n large) [82, 83, 84] which answers the bounded Burnside problem in the negative. For more detail about the history of square-free words we refer the reader to [2].

Although there are no squares in permutations when they are viewed as words, the definition naturally extends to permutations using the notion of a pattern defined above. A square in a permutation is two consecutive factors which follow the same pattern. Formally, a *square* of length 2ℓ is a factor

$$(S_1; S_2) = (\sigma_k, \ldots, \sigma_{k+\ell-1}; \sigma_{k+\ell}, \ldots, \sigma_{k+2\ell-1})$$

where $S_1 \sim S_2$, and we say a permutation is *square-free* if it contains no squares of length at least 4.

A word W is *left-crucial* (respectively *right-crucial*) with respect to squares if it does not itself contain any square, but any word obtained by prepending (respectively appending) a single letter (from the alphabet) contains a square. We say W is *bicrucial* (also known as *maximal* in the literature) if it is both left-crucial and right-crucial. Right-crucial words were characterised by Li in 1976 [72], and it follows from this characterisation that, up to permuting the alphabet, the smallest bicrucial word on an alphabet of size n is the nth *Zimin word* (where Zimin words are defined recursively as $W_1 = 1$ and $W_n = W_{n-1}nW_{n-1}$). It is also known that any square-free word is a factor of a bicrucial square-free word [12], and the result of Thue [91] shows there are infinitely many square-free words. These two results combine to show that there are arbitrarily long bicrucial square-free words.

By modifying the process of extending a word, the notion of crucial words also naturally extends to permutations. A permutation π of length $n+1$ is a *right-extension* of σ if $(\pi_0, \ldots, \pi_{n-1}) \sim \sigma$, or equivalently, if π can be formed by appending an entry $x \in \{0, \ldots, n\}$ to σ and replacing σ_i by σ_i+1 if $\sigma_i \geq x$. We say that the permutation σ is *right-crucial* if it is square-free, and every right-extension π contains a square, and we define left-extensions and left-crucial permutations similarly. We will be interested

in permutations which are simultaneously left-crucial and right-crucial which we call *bicrucial* permutations.

The first result showing the existence of bicrucial permutations was from Avgustinovich, Kitaev, Pyatkin and Valyuzhenich in 2011 [7] when they showed that bicrucial permutations exist for all odd lengths of the form $8k+1$, $8k+5$ or $8k+7$ where $k \geq 1$. Using the constraint solver Minion, bicrucial permutations of length $8k+3$ for $k = 2, 3$ and of even length 32 were found by Gent, Kitaev, Konovalov, Linton and Nightingale in 2015 [49], and they conjectured that larger permutations should exist.

Conjecture 5.1 (Conjecture 8 in [49])**.** *There exist bicrucial permutations of length* $8k+3$ *for all* $k \geq 2$*.*

Conjecture 5.2 (Conjecture 9 in [49])**.** *There exist arbitrarily long bicrucial permutations of even length.*

We confirm both these conjectures and further, completely classify the n for which there exist bicrucial permutations of length n.

Theorem 5.3. *A bicrucial permutation of length* n *exists if and only if* $n = 9$*,* $n \geq 13$ *is odd, or* $n \geq 32$ *is even and not* 38*.*

The proof of this result is split into several parts: we give a construction of bicrucial permutations of length $8k+3$ for every $k \geq 3$, which completes the classification of the odd n, and a construction for when $n \geq 48$ is even. The remaining small (even) cases are handled using computer searches, including the non-existence of bicrucial permutations of length 38.

The chapter on bicrucial square-free permutations is based on joint work with Carla Groenland [54].

1.4.2 Decomposing permutations into order-isomorphic subpermutations

Squares in permutations can be generalised in several ways. The most closely related is the concept of block twins. A *block twin* of length 2ℓ in σ is two factors S_1 and S_2 of length ℓ which are disjoint and have the same pattern, but need not be adjacent. That is, S_1 and S_2 form a block twin if

$$(S_1; S_2) = (\sigma_i, \ldots, \sigma_{i+\ell-1}; \sigma_j, \ldots, \sigma_{j+\ell-1})$$

where $j \geq i + \ell$ and $S_1 \sim S_2$. Clearly, squares are a special case of block twins and the existence of a square implies the existence of a block twin. However, unlike squares, all long permutations must contain long block twins. It was shown by Dudek, Grytczuk and Ruciński [37] that every permutation contains a block twin of length

$$(2 + o(1))\frac{\log(n)}{\log\log(n)},$$

and that a random permutation contains a block twin of length

$$(4 + o(1))\frac{\log(n)}{\log\log(n)}$$

with high probability.

Block twins can be further generalised by removing the condition that S_1 and S_2 are factors and only enforcing that they are disjoint. A *twin* $T = (T_1, T_2)$ of length 2ℓ in a permutation $\sigma \in S_n$ consists of two disjoint subpermutations $T_1 = \sigma_{i_1}, \ldots, \sigma_{i_\ell}$ and $T_2 = \sigma_{j_1}, \ldots, \sigma_{j_\ell}$ such that $T_1 \sim T_2$. These were introduced by Gawron [48] who proved an upper bound of $O(n^{2/3})$ on the longest twin guaranteed in a permutation of length n and conjectured that this should be tight (up to the constant). A simple lower bound of $\Omega(n^{1/2})$ can be obtained from the Erdős-Szekeres theorem, and this was improved to $\Omega(n^{3/5})$ by Bukh and Rudenko [25]. The upper bound comes from applying a first moment method to a uniformly random permutation, and so there has also been work on the lower bound for random permutations. A lower bound of $\Omega\big(n^{2/3}/\log^{1/3}(n)\big)$ was shown by Dudek, Grytczuk and Ruciński [37], and this was improved to $\Omega(n^{2/3})$ by Bukh and Rudenko [25].

We consider the related problem, proposed by Dudek, Grytczuk and Ruciński [37], of decomposing two permutations into order-isomorphic subpermutations. We say that two permutations σ and π are *ℓ-similar* if they can be decomposed into ℓ subpermutations $\sigma^{(1)}, \ldots, \sigma^{(\ell)}$ and $\pi^{(1)}, \ldots, \pi^{(\ell)}$ such that $\sigma^{(i)} \sim \pi^{(i)}$ for all $i \in [\ell]$, and we denote the least such ℓ by $U(\sigma, \pi)$. Clearly, in the worst case, two permutations might only be n-similar. For example, the identity permutation $(1, \ldots, n)$ and the reverse permutation $(n, \ldots, 1)$ contain no non-trivial order-isomorphic subpermutations. But what happens in the average case?

Problem 6.1 (Problem 4.3 in [37])**.** *Let σ and π be two permutations of length n chosen independently and uniformly at random. What is the expected value of $U(\sigma, \pi)$?*

One can prove a simple $O(\sqrt{n})$ upper bound by decomposing the permutations into increasing subpermutations as follows. The length of a longest increasing sequence in a uniformly random permutation is asymptotically $2\sqrt{n}$ [89], and this is

well concentrated [46, 21]. In particular, if L is the length of a longest increasing subpermutation in a uniformly random permutation, we have

$$\mathbb{P}\Big(|L - \mathbb{E}[L]| \geq n^{\frac{1}{3}}\Big) \leq \exp\big(-n^{\beta}\big)$$

for some $\beta > 0$. Hence, rather crudely, we have $L \leq 3\sqrt{n}$ with very high probability. By Dilworth's Theorem, this implies that a uniformly random permutation can be decomposed into at most $3\sqrt{n}$ decreasing subpermutations with very high probability. It follows easily that, with very high probability, a pair of uniformly random permutations can be simultaneously decomposed into at most $6\sqrt{n}$ decreasing subsequences of the same length.

Our main result improves the preceding simple bound to the following.

Theorem 6.2. *Let σ and π be two permutations of length n chosen independently and uniformly at random. Then σ and π are $O\big(n^{1/3}\log^{11/6}(n)\big)$-similar with probability at least $1 - o(n^{-1})$.*

Since two permutations are at worst n-similar, the following bound on $U(\sigma, \pi)$ is immediate

Corollary 6.3. *Let σ and π be two independent, uniformly random permutations of length n. Then*

$$\mathbb{E}[U(\sigma, \pi)] = O\Big(n^{1/3}\log^{11/6}(n)\Big).$$

A simple lower bound of $\Omega(n^{1/3})$ can be deduced from the average length of a longest twin as follows. Note that if one chooses a permutation σ of length $2n$ uniformly at random, then the patterns of the first and the second half of σ are independent uniformly random permutations of length n, and that order-isomorphic subpermutations of length ℓ between the first and second of half of σ give rise to a twin of length 2ℓ in σ. Hence, from Gawron's result that the longest twin in a permutation of length $2n$ is of length $O(n^{2/3})$ with high probability [48], it follows that the longest order-isomorphic subpermutations of two uniformly random permutations of length n are also of length $O(n^{2/3})$ with high probability. This means that $\Omega(n^{1/3})$ pieces are needed with high probability and Theorem 6.2 is tight up to a polylogarithmic factor.

Twins and decompositions into twins have been studied before in many contexts. The corresponding question in graphs is the following: what is the largest $\ell(m)$ such

that any graph G with m edges has two edge-disjoint isomorphic subgraphs with at least ℓ edges? In 1987, Erdős, Pach and Pyber [43] showed that

$$cm^{2/3} \leq \ell(m) \leq Cm^{2/3}\frac{\log(m)}{\log\log(m)}$$

for some constants c and C, and they further proved bounds for r-uniform hypergraphs. The gap between the upper and lower bounds was closed up to a constant by Lee, Loh and Sudakov in 2013 [70] when they showed that $\ell = \Theta\big((m\log(m))^{2/3}\big)$. Surprisingly, in the case of trees there is a twin containing almost all of the edges [4]. Explicitly, there are two edge-disjoint isomorphic subtrees each containing

$$\frac{m}{2} - O\left(\frac{m}{\log\log(m)}\right)$$

edges. Determining whether the entire tree forms a twin (an *even tree*) was supposedly shown to be NP-hard (provided the trees are suitably coded) by Graham and Robinson in unpublished work. On the other hand there is a linear time algorithm to determine whether the tree forms a twin where each factor is a tree (a *bisectable tree*) [56]. This forms part of a series of papers on the factorisation of graphs into several isomorphic factors, and we refer the reader to [57] for a summary.

Of more relevance to the problem we consider is the decomposition of multiple graphs into isomorphic subgraphs. Given two graphs G and H with the same number of vertices and edges, we say that G and H are *ℓ-similar* if G and H can be decomposed into edge disjoint subgraphs $G_1, \ldots, G_\ell$ and $H_1, \ldots, H_\ell$ respectively such that $G_i \cong H_i$. What is the minimum $\ell(n)$ such that any two graphs with n vertices and the same number of edges are ℓ-similar? These are known as *U-decompositions*, named after Ulam who originally posed the problem [50]. The answer, determined by Chung, Erdős, Graham, Ulam and Yao [32], is $2n/3 + o(n)$. Chung, Erdős and Graham [30] later also studied the problem of simultaneously decomposing $k \geq 3$ graphs showing that the answer in this more general case is $3n/4 + o_k(n)$, so the coefficient of the leading term is independent of k. Given this, one might also ask what happens if we are interested in every graph on n vertices with the same number of edges. It turns out this is also $3n/4 + o(n)$ [31]. This question has also been studied for classes such as bipartite graphs, and we refer the reader to [50] for more information.

The chapter on decomposing pairs of permutations into order-isomorphic subpermutations is based on joint work with Carla Groenland, Dániel Korándi, Alexander Roberts, Alex Scott and Jane Tan.

1.5 Uniform shuffling using lazy transpositions

Let S_n be the symmetric group on n elements, and write $(a, b) \in S_n$ for the transposition that swaps a and b, and $1 \in S_n$ for the identity permutation. A *lazy transposition* $T = (a, b, p)$ is the random permutation

$$T = \begin{cases} (a, b) & \text{with probability } p, \\ 1 & \text{otherwise.} \end{cases}$$

The composition of a sequence of lazy transpositions is also a random permutation, and we will be interested in sequences which generate a uniformly random permutation. Let $U(n)$ denote the minimum length ℓ for which there exists a sequence of independent lazy transpositions $T_1, \ldots, T_\ell$ such that the composition $T_1 \cdots T_\ell$ is uniformly distributed over S_n. We will call such a sequence a *transposition shuffle*. These were first discussed on Stack Exchange in 2011 [45], and investigated in more depth by Angel and Holroyd in 2018 [5]. Both give constructions using $\binom{n}{2}$ lazy transpositions and ask if this is best possible.

Problem 7.1 (Fitzsimons [45], Angel and Holroyd [5])**.** *Does* $U(n) = \binom{n}{2}$ *for all* n*?*

We remark that it is important that the output permutation is exactly uniform: it was shown by Czumaj [34] that there are sequences of lazy transpositions of length $O(n \log(n))$ which are very close to uniform.

Theorem 7.2 (Czumaj [34])**.** *Let* c *be an arbitrary constant. There is a sequence of* $O(n \log(n))$ *independent lazy transpositions* $T_1, \ldots, T_\ell$*, all with probability* $1/2$*, such that the total variation distance between the distribution of* $T_1 \cdots T_\ell$ *and* $\mathrm{Unif}(S_n)$ *is* $O(n^{-c})$*.*

It will often be convenient to think of transpositions as moving n distinguishable "counters" across n "positions" which we label $1, \ldots, n$. A transposition (a, b) acts by switching the counters currently in positions a and b, so a lazy transposition (a, b, p) switches the counters currently in positions a and b with probability p, and does nothing with probability $1 - p$. In this language a transposition shuffle is a sequence of lazy transpositions such that the final order of the counters is uniformly distributed over all $n!$ orders.

1.5.1 Shorter transposition shuffles

The main result of this chapter is a new construction for shuffling with lazy transpositions which improves the constant in the upper bound of $U(n)$ and answers Problem 7.1 in the negative.

Theorem 7.3. *There exists an $\varepsilon > 0$ such that $U(n) \leq (1-\varepsilon)\binom{n}{2}$ for all $n \geq 6$. Moreover,*

$$U(n) \leq \frac{2}{3}\binom{n}{2} + O(n\log(n)).$$

As well as considering transposition shuffles, we will also be interested in the related problem of shuffling $t < n$ counters across n positions. Instead of starting with n counters, one in each position, we start with counters in the positions $1, \ldots, t$, and leave the other positions empty. A transposition (a, b) switches any counters in positions a and b, including switching a counter with an empty space. A sequence of lazy transpositions naturally defines a random variable by the positions of the counters at the end of sequence, and we can again ask that this be uniformly distributed over all $t!\binom{n}{t}$ options. If the output is uniformly distributed, we say the sequence achieves *t-uniformity on n points* and call the sequence a *(t, n)-shuffle* (so a transposition shuffle is an (n, n)-shuffle). Let $U_t(n)$ denote the minimum number of lazy transpositions in a (t, n)-shuffle. The special case of shuffling all elements is $U_n(n) = U(n)$. The constructions in Section 7.1 give an upper bound $U_t(n) \leq tn - \binom{t+1}{2}$ and, in light of Problem 7.1, it is natural to ask the following.

Problem 7.4. *Does $U_t(n) = tn - \binom{t+1}{2}$?*

When $t = n$, this is equivalent to Problem 7.1, and Theorem 7.3 shows that the bound can be improved by a constant factor, but what about other values of t? The constructions used in the proof of Theorem 7.3 give $(3, n)$- and $(4, n)$-shuffles which beat the upper bound by a constant factor, and we obtain the following theorem.

Theorem 7.5. *There exists a constant $\varepsilon > 0$ such that*

$$U_t(n) \leq (1-\varepsilon)\left(tn - \binom{t+1}{2}\right)$$

for all $3 \leq t \leq n$ and $n \geq 6$. Moreover, for a fixed $t = 3\ell + r$, we have

$$U_t(n) \leq (2\ell + r)n + O_\ell(\log(n)).$$

1.5.2 Reachability

What about lower bounds on $U(n)$? The only lower bounds known for this problem do not use the probabilistic nature of the problem at all and instead use only that each configuration can be "reached" with positive probability.

Let $T_1, \ldots, T_\ell$ be a transposition shuffle and suppose that $T_i = (a_i, b_i, p_i)$ for $i = 1, \ldots, \ell$. Since $T_1, \ldots, T_\ell$ is a transposition shuffle, it is certainly possible to end in, or "reach", any given permutation. That is, for every choice of $\sigma \in S_n$, there is some subsequence $T_{i_1}, \ldots, T_{i_{\ell'}}$ such that

$$\sigma = (a_{i_1}, b_{i_1}) \cdots (a_{i_{\ell'}}, b_{i_{\ell'}}).$$

Such sequences of transpositions are known as *permutation networks* or *non-blocking networks* and have been well-studied due to their usefulness in communication networks and distributed computing (see e.g. [35]). Of particular relevance is the following result which shows that permutation networks can be constructed with at most $\sum_{i=1}^{n} \lceil \log_2(i) \rceil$ transpositions, and this is tight.

Theorem 7.6 (Waksman [92], Beauquier and Darrot [13])**.** *There is a permutation network on n elements using $\sum_{i=1}^{n} \lceil \log_2(i) \rceil$ transpositions, and this is tight.*

Although this is a long way from the best known upper bounds in Theorem 7.3, this gives the best known lower bound for uniformity.

A natural question is what happens with fewer counters: how many transpositions are needed to ensure a specific subset can reach any configuration? Let $T_1, \ldots, T_\ell$ be a sequence of lazy transpositions and suppose that $T_i = (a_i, b_i, p_i)$ for $i = 1, \ldots, \ell$. We say the sequence achieves *t-reachability* if, for every choice of t distinct ordered points $x_1, \ldots, x_t$, there is some subsequence of $T_{i_1}, \ldots, T_{i_{\ell'}}$ such that the composition $(a_{i_1}, b_{i_1}) \cdots (a_{i_{\ell'}}, b_{i_{\ell'}})$ maps $i \mapsto x_i$ for every $i \leq t$. That is, there is positive probability of the counters ending in the positions $x_1, \ldots, x_t$ (unless $p_i \in \{0, 1\}$ for some i). Of course, t-reachability gives lower bounds on t-uniformity in the same way permutation networks can be used to get lower bounds for transposition shuffles.

Let us first consider the case of 1-uniformity where we want to shuffle 1 counter across n positions. By noting there are at most 2^ℓ possible outputs from a sequence of ℓ lazy transpositions, we immediately get the lower bound $\lceil \log_2(n) \rceil$ for the number of transpositions in a $(1, n)$-shuffle, but it is possible to do much better. Indeed, a transposition can only move the counter if one end of the transposition is a position where the counter may already be, and the number of "reachable" positions can only

increase by one for each transposition. Hence, $n-1$ transpositions are needed even for 1-reachability. This implies that a $(1,n)$-shuffle needs at least $n-1$ transpositions (and it is not hard to show that this can be achieved).

Looking for lower bounds in other cases, we exactly determine the lower bound for 2-reachability, which gives a lower bound for 2-uniformity.

Theorem 7.7. *Let $n \geq 2$. The smallest sequence of transpositions achieving 2-reachability on n points has length $\lceil 3n/2 \rceil - 2$.*

This lower bound is a long way from the best known constructions for 2-uniformity which use $2n-3$ transpositions. We also give a simple randomised construction which shows that t-reachability can be achieved using $(2+o_t(1))n$ transpositions. Surprisingly, the coefficient of the leading term is independent of t, which contrasts sharply with the current best known upper bounds for t-uniformity from Theorem 7.5.

Theorem 7.8. *Let $t \geq 3$. There is a sequence of transpositions of length at most $(2+o_t(1))n$ which achieves t-reachability on n points.*

1.5.3 Star transposition shuffles

Angel and Holroyd studied the problem of constructing transposition shuffles when all transpositions are of the form $(i, i+1)$, and they classified the minimum transposition shuffles using reduced words [5]. We consider the problem of transposition shuffles which only use transpositions of the form $(1,\cdot)$, which we call *star transpositions* (as they match the edges of a star graph).

While it is easy to see that any transposition shuffle using transpositions of the form $(a, a+1)$ requires $\binom{n}{2}$ transpositions just to reach the reverse permutation, the case of star transpositions is more difficult. In fact, the minimum star transposition shuffles must use $\Theta(U(n))$ transpositions: by replacing the general transposition (a,b,p) by the star transpositions $(1,a,1),(1,b,p),(1,a,1)$, any transposition shuffle can easily be converted to one using star transpositions with only triple the number of transpositions. In particular, showing that the minimum number of transpositions in this highly restrictive model is $\Omega(n^2)$ would show that the general case is $\Omega(n^2)$ as well.

One might hope that the analogue of Problem 7.1 is true when restricted to star transpositions: if all transpositions are of the form $(1,\cdot)$, does every transposition shuffle have at least $\binom{n}{2}$ transpositions? Unfortunately, the following theorem shows this is not the case.

Theorem 7.9. *Let $t = 4\ell + r$ for some $0 \leq r < 3$. Let $\widetilde{U}_t(n)$ be the minimum length of a (t, n)-shuffle in which every transposition is a star transposition. Then*

$$\widetilde{U}_t(n) \leq \left(\frac{7\ell}{2} + r\right)n + O_\ell(\log(n)).$$

Moreover,

$$\widetilde{U}(n) \leq \frac{7}{8}\binom{n}{2} + O(n\log(n)).$$

A modification of Theorem 7.7 shows that the minimum number of transpositions needed for 2-reachability is again around $3n/2$ even when restricting to star transpositions.

Theorem 7.10. *The shortest transposition sequence achieving 2-reachability on n points in which each transposition is a star transposition, contains $\lceil 3(n-1)/2 \rceil$ transpositions.*

However, using the restrictive nature of the transpositions, we are able to show that achieving 2-uniformity is strictly harder than achieving 2-reachability, the first result of this kind.

Theorem 7.11. *Any sequence of lazy star transpositions achieving 2-uniformity on n points has length at least $1.6n - C$ for some absolute constant C.*

The chapter on uniform shuffling with independent lazy transpositions is based on joint work with Carla Groenland, Jamie Radcliffe and Alex Scott.

Chapter 2

Intersection sizes of linear subspaces with the hypercube

In this chapter we continue the study of the possible intersection sizes that a k-dimensional linear subspace can have with the $(k+m)$-dimensional hypercube, and we denote the set of all such possible intersection sizes by $H(k,m)$. These were first studied by Melo and Winter [77] after they encountered a similar problem in quantum information theory [81]. In particular, Melo and Winter were interested in the second largest intersection size. The largest intersection size is clearly 2^k, but what it is the next largest? Is it possible to intersect in 2^k-1 points? Melo and Winter [77] showed that the second largest intersection size is not 2^k-1 but $2^{k-1}+2^{k-2}$, and they made a conjecture on what other "large" intersection sizes should be possible. Let $H^+(k,m) = H(k,m) \setminus \{1,\dots,2^{k-1}\}$ be the large values of $H(k,m)$ and set $H^+(k,\infty) = \cup_{m\geq 1} H^+(k,m)$. Melo and Winter [77] showed that $H^+(k,\infty) \supseteq \{2^{k-1}+2^i : i \in \{0,1,\dots,k-1\}\}$, and they conjectured that equality holds. We show this conjecture is almost true: there is only one additional value.

Theorem 2.2. *For $k \geq 6$,*

$$H^+(k,\infty) = \{2^{k-1}+2^i : i \in \{0,\dots,k-1\}\} \cup \{2^{k-1}+2^{k-5}+2^{k-6}\}.$$

In fact, in Theorem 2.13 we determine $H^+(k,m)$ for every $m \geq 1$ from which the theorem above follows.

Denote the small values by $H^-(k,m) = H(k,m) \cap \left[2^{k-1}\right]$ and set $H^-(k,\infty) = \cup_{m\geq 1} H^-(k,m)$. After checking the set $H^-(k,\infty)$ for several small values of k, Melo and Winter [77] conjectured that $H^-(k,\infty) = [2^{k-1}]$. We disprove this by proving that a positive fraction of the small values are missing.

Theorem 2.4. *For $k \geq 8$ and any $m \geq 1$,*

$$H(k,m) \cap \left\{\tfrac{15}{16}2^{k-1},\dots,2^{k-1}\right\} = \left\{\tfrac{15}{16}2^{k-1}, \tfrac{126}{128}2^{k-1}, 2^{k-1}\right\}.$$

Chapter outline In Section 2.1 we establish some notation and definition used throughout the rest of this chapter. We include some of the relevant results from the literature along with their proofs in Section 2.2. In Section 2.3 we prove Theorem 2.2, and in Section 2.4 we prove Theorem 2.4. Both of these proofs rely on exhaustive computer searches which we describe in Section 2.5. We end the chapter with a discussion of several open problems in Section 2.6

This chapter is based on joint work with Carla Groenland [53].

2.1 Framework and notation

Let $e_i \in \mathbb{R}^k$ denote the vector with a 1 in position i and 0s elsewhere, so that $\{e_1, \dots, e_k\}$ is the standard basis for $\mathbb{R}^k$. Given two vectors a and b, we write (a, b) for their concatenation. After permuting coordinates as necessary, any linear subspace $S \subseteq \mathbb{R}^n$ of dimension k can be parametrised as $S = \{(v, L(v)) : v \in \mathbb{R}^k\}$ for some linear map $L : \mathbb{R}^k \to \mathbb{R}^m$ where $m = n - k$ (see Lemma 2.7). Using this we have

$$S \cap \{0,1\}^n = \big\{(v, L(v)) : v \in \{0,1\}^k,\ L(v) \in \{0,1\}^{n-k}\big\}.$$

Melo and Winter [77] introduced the equivalent formulation

$$H(k,m) = \big\{t : \exists L : \mathbb{R}^k \to \mathbb{R}^{n-k} \text{ linear},\ t = \big|\{0,1\}^k \cap L^{-1}\{0,1\}^{n-k}\big|\big\}$$

that we will also use. Their original motivation came from quantum information theory: if one also enforces that L is an isometry, the largest two elements in the set above are 2^k and 2^{k-1}, and this was used to show certain quantum circuits yield 'ε-approximate unitary t-designs' in [81].

Throughout the text, $L : \mathbb{R}^k \to \mathbb{R}^m$ is a linear map and $m = n - k$. Any linear map is determined by the values it takes on the basis elements $\{e_1, \dots, e_k\}$. We use

$$I(L) = L^{-1}(\{0,1\}^m) \cap \{0,1\}^k$$

for the *intersection pattern* associated with L and $t(L) = |I(L)|$ for its size.

The *support* $S(L) = \{i \in \{1, \dots, k\} : L(e_i) \neq 0\}$ denotes the indices of the basis vectors on which L is non-zero.

Observation 2.1. *After permuting coordinates as necessary, there exists a* $J(L) \subseteq \{0,1\}^{|S(L)|}$ *such that* $I(L) = J(L) \times \{0,1\}^{k-|S(L)|}$.

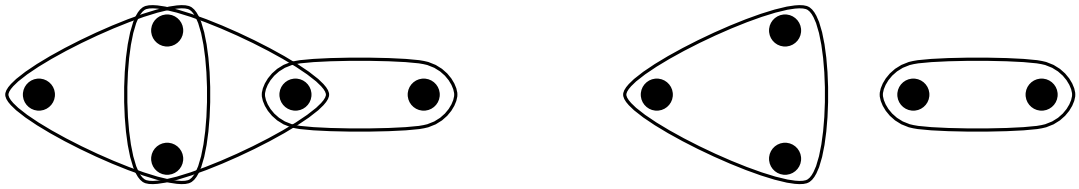


Figure 2.1: An example of a shape of a non-minimal map (left) and of a minimal map (right). Note that on the left-hand side, the support of one of the edges is completely contained in the combined support of the other edges.

For $A \subseteq \{1, \ldots, m\}$, define the linear map $L_A : \mathbb{R}^k \to \mathbb{R}^{|A|}$ by $L_A = \pi_A \circ L$, where π_A is the projection map onto the coordinates in the set A, for example $\pi_{\{1,3\}}(x_1, \ldots, x_m) = (x_1, x_3)$. We use the shorthand notation $L_i := L_{\{i\}}$. Note that $L(x) \in \{0,1\}^m$ if and only if $L_i(x) \in \{0,1\}$ for $i = 1, \ldots, m$ and that adding conditions can only decrease the intersection size:

$$I\big(L_{A\cup\{i\}}\big) = \bigcap_{j \in A\cup\{i\}} I(L_j) \subseteq \bigcap_{j\in A} I(L_j) = I(L_A)$$

for each $A \subseteq [m]$ and $i \in [m]$.

Definition 2.5. *A linear map $L : \mathbb{R}^k \to \mathbb{R}^m$ is called minimal if $S(L_i) \not\subseteq \bigcup_{j\neq i} S(L_j)$ and $t(L_i) < 2^k$ for all $i \in \{1, \ldots, m\}$.*

We will see that only specific 'shapes' of the sets $S(L_i)$ can 'cover' $[m]$ and still give a large intersection value $t(L)$.

Definition 2.6. *The shape of the linear map $L : \mathbb{R}^k \to \mathbb{R}^m$ is the hypergraph (V, E) with vertex set $V = \{1, \ldots, k\}$ and edge set $\{S(L_i) : i \in [m]\}$.*

Examples of a minimal and non-minimal shape are given in Figure 2.1. Two specific shapes will be of particular interest.

- A *(2,1)-star* is a 2-uniform hypergraph (V, E) for which there is a centre $c \in V$ such that $E = \{\{c, v\} : v \in V \setminus \{c\}\}$.
- A *(3,2)-star* is a 3-uniform hypergraph (V, E) for which there exist distinct $c_1, c_2 \in V$ such that $E = \{\{c_1, c_2, v\} : v \in V \setminus \{c_1, c_2\}\}$.

We will write "the sets $S(L_i)$ form a (3,2)-star" to mean that the the shape of L is a (3,2)-star.

2.2 Preliminaries

Lemma 2.7 ([77]). *Let V be a k-dimensional linear subspace. There is a linear map $L : \mathbb{R}^k \to \mathbb{R}^{n-k}$ such that $S = \{(v, L(v)) : v \in \mathbb{R}^k\}$ up to permuting coordinates.*

Proof. Take an arbitrary basis $\{v_1, ..., v_k\}$ of S and let M be the $k \times n$ matrix with rows $v_1, \ldots, v_k$. The row span of M is exactly the linear subspace S. Convert the matrix into reduced row echelon form using Gaussian elimination and let the new matrix be M' with rows $v'_1, \ldots, v'_k$. Gaussian elimination does not change the row span, so the row span is still the linear subspace S. Let the first non-zero entry of v'_i be in position j_i and arbitrarily relabel the coordinates such that $j_i \mapsto i$. Writing $v_i = (v_{i,1}, \ldots, v_{i,n})$, we define the linear map $L : \mathbb{R}^k \to \mathbb{R}^{n-k}$ by setting $L(e_i) = (v_{i,k+1}, \ldots, v_{i,n})$ so that $v'_i = (e_i, L(e_i))$.

Clearly the vectors $v'_1 \ldots v'_k$ are linearly independent and form a basis for S so one can write a vector $w = (w_1, \ldots, w_n) \in S$ uniquely in terms of the rows of M' as

$$w = w_1 v'_1 + w_2 v'_2 + \cdots + w'_k v'_k = (w_1 e_1, w_1 L(e_1)) + \cdots + (w_k e_k, w_k L(e_k)).$$

Hence, $w = (v, L(v))$ where $v = (w_1, \ldots, w_n)$. Clearly, any vector of this form is also in the row span of M' and in S. □

In order to get some intuition about the problem and the notation, this subsection also includes some results shown by Melo and Winter [77].

Lemma 2.8 ([77]). *$H(k, m) \subseteq H(k', m')$ for all $m' \geq m$ and $k' \geq k$.*

Proof. Let $L : \mathbb{R}^k \to \mathbb{R}^m$ be a linear map. It suffices to prove that there exist linear maps $L' : \mathbb{R}^k \to \mathbb{R}^{m+1}$ and $L'' : \mathbb{R}^{k+1} \to \mathbb{R}^m$ with the same intersection pattern (ignoring issues with the differing dimensions). These maps will then in particular have the same intersection size, i.e. $t(L) = t(L') = t(L'')$.

To this end, define $L' : \mathbb{R}^k \to \mathbb{R}^{m+1}$ by $L'(v) = (L(v), 0)$ and $L'' : \mathbb{R}^{k+1} \to \mathbb{R}^m$ by

$$L''(e_i) = \begin{cases} L(e_i) & \text{for } i \leq k, \\ \sum_{i=1}^{k} |L(e_i)| + 2 & \text{for } i = k+1. \end{cases}$$

The first map can be viewed as adding an empty constraint which is always satisfied. The second map can be viewed as adding an extra coordinate in a way that ensures the subspace never intersects the hypercube when the new coordinate is 1. □

Melo and Winter proved a slight variation on the following proposition, which implies the containment relation $H^+(k, \infty) \supseteq \{2^{k-1} + 2^i : i \in \{0, 1, \ldots, k-1\}\}$. Indeed, this shows that $2^{k-1} + 2^i \in H(k, k-1)$.

Proposition 2.9 ([77]). *Let* $0 \le t_0 < t_1 < \cdots < t_a$. *Then* $\sum_{i=0}^{a} 2^{t_i} \in H(t_a + a, t_a)$.

Proof. We need to construct a linear map $L : \mathbb{R}^{t_a+a} \to \mathbb{R}^{t_a}$ with $t(L) = \sum_{i=0}^{a} 2^{t_i}$. We define $L(e_i) = e_i$ for $i \le t_a$ and, for $i \ge t_a + 1$, we define $L(e_i) = \sum_{j=t_{i'}+1}^{t_a} e_j$ where $i' = i - t_a - 1$.

We count the number $t(L)$ of $v \in \{0,1\}^{t_a+a}$ for which $L(v) \in \{0,1\}^{t_a}$ by conditioning on the last a values of v, which we denote by y. Denote the first t_a values of v by x. If $y = 0$, then x can take any value, contributing 2^{t_a} to $t(L)$. Note that the vector y cannot have more than a single 1, else the last entry of $L(v)$ is at least 2. Suppose $y = e_1$, the last $t_a - t_0$ entries of x must be 0 else there is a coordinate where $L(v) = 2$. This contributes 2^{t_0} to $t(L)$. More generally, if $y = e_i$, the last $t_a - t_i$ entries of x must be 0 and this contributes 2^{t_i} to $t(L)$. Hence, $t(L) = 2^{t_a} + \cdots + 2^{t_0}$ as required. □

We will depend on the following observation in the next section.

Lemma 2.10 ([77]). *If* $t(L) > 2^{k-1}$, *then* $L(\{e_1, \ldots, e_k\}) \subseteq \{-1, 0, 1\}^m$.

Proof. Suppose $\alpha = L_1(e_1) \notin \{-1, 0, 1\}$. For any $x \in \{0\} \times \{0,1\}^{k-1}$, if $L_1(x) \in \{0,1\}$, then $L_1(x + e_1) = L_1(x) + \alpha \notin \{0,1\}$. Hence either x or $x + e_1$ is not in $I(L_1)$, which implies $t(L) \le t(L_1) \le 2^{k-1}$. □

2.3 Large intersection sizes

We first investigate the set $H^+(k, 1)$ which is the set of intersection sizes $t(L) > 2^{k-1}$ where $L : \mathbb{R}^k \to \mathbb{R}$ a linear map. As mentioned in the preliminaries, we must have $L(e_i) \in \{-1, 0, 1\}$ since $t(L) > 2^{k-1}$.

Lemma 2.11. *Suppose* $L : \mathbb{R}^k \to \mathbb{R}$ *is linear with* $L(e_i) \in \{-1, 0, 1\}$ *for all* $i \in [k]$ *and define*

$$a := |\{i : L(e_i) = 1\}|, \quad b := |\{i : L(e_i) = -1\}|, \quad c := |\{i : L(e_i) = 0\}|.$$

Then

$$\left|\left\{x \in \{0,1\}^k : L(x) = j\right\}\right| = 2^c \binom{a+b}{b+j}.$$

Proof. Since $L(x) = x_1L(e_1) + \cdots + x_kL(e_k)$, we find

$$\begin{aligned}\big|\{x \in \{0,1\}^k : L(x) = j\}\big| &= \sum_{i=0}^{a} 2^c \binom{a}{i}\binom{b}{i-j} \\ &= 2^c \sum_{i=0}^{b+j} \binom{a}{i}\binom{b}{b+j-i} \\ &= 2^c \binom{a+b}{b+j},\end{aligned}$$

where we have used the Chu-Vandermonde identity in the last line. Alternatively this can be shown with less calculation in the following self-contained argument. We may renumber the e_i such that $e_1, \ldots, e_b$ are mapped to -1 and $e_{b+1}, \ldots, e_{a+b}$ are mapped to 1. Any element $x \in \{0,1\}^k$ is defined by its *support*, $\operatorname{supp}(x) := \{i \in [k] : x_i \neq 0\}$. The smallest value $L(x)$ can take is $-b$, which is achieved by $\operatorname{supp}(x) = \{1, \ldots, b\}$. If an element $i > b$ is added to $\operatorname{supp}(x)$ or an element $i \in [b]$ is removed from $\operatorname{supp}(x)$, then the value of $L(x)$ will increase by 1. In order to get to 0, b such changes have to be made which can be done in exactly $\binom{a+b}{b}$ ways. The coordinates mapping to 0 (the x_i where $i \geq a+b+1$) have no effect on the value of $L(x)$ so each can individually be included or not; this gives the factor 2^c. Similarly, the value j is achieved by making $b+j$ switches which can be done in $\binom{a+b}{b+j}$ ways, giving $2^c\binom{a+b}{b+j}$ possible $x \in \{0,1\}^k$ with $L(x) = j$. □

Corollary 2.12. *For $k \geq 6$,*

$$H^+(k,1) = \big\{2^{k-1} + 2^{k-5} + 2^{k-6}, 2^{k-1} + 2^{k-3}, 2^{k-1} + 2^{k-2}, 2^k\big\}.$$

Proof. Let $L : \mathbb{R}^k \to \mathbb{R}$ be a linear map. As noted above, we may assume that $L(e_i) \in \{-1, 0, 1\}$ for all $i \in [k]$. Define a, b and c as in the previous lemma. Then

$$t(L) = \big|\big\{x \in \{0,1\}^k : L(x) \in \{0,1\}\big\}\big| = 2^c\left(\binom{a+b}{b} + \binom{a+b}{b+1}\right) = 2^c\binom{a+b+1}{b+1}.$$

Note that $t(L) = 2^c t(L')$ where $L' : \mathbb{R}^{k-c} \to \mathbb{R}$ is the projection of L onto the support of L. We therefore first look for triples $(a, b, 0)$ with $a + b = k$ which give an intersection $t \geq 2^{k-1} + 1$ (and then multiply the result by 2^c). We first show that for such triples to exist, we must have $a + b \leq 7$ for which we can check all possible options. The largest binomial coefficient for a fixed number of elements is the central binomial coefficient so

$$t(L) = \binom{a+b+1}{b+1} \leq \binom{k+1}{\lfloor (k+1)/2 \rfloor}.$$

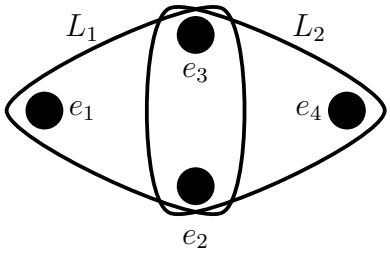


Figure 2.2: A $(3,2)$-star with 2 edges.

The ratio $2^{-n}\binom{n}{\lfloor n/2 \rfloor}$ is non-increasing in n, since

$$\frac{2^{-(2n+1)}\binom{2n+1}{n}}{2^{-2n}\binom{2n}{n}} = \frac{2n+1}{2(n+1)} < 1 \text{ and } \frac{2^{-(2n-1)}\binom{2n-1}{n-1}}{2^{-2n}\binom{2n}{n}} = \frac{2n}{2n} = 1.$$

Since $2^{-(n+1)}\binom{n+1}{\lfloor (n+1)/2 \rfloor} \leq \frac{1}{4}$ for $n = 8$, it follows that $t(L) \leq \binom{k+1}{\lfloor (k+1)/2 \rfloor} \leq 2^{k-1}$ for all $k = a + b \geq 8$. Checking all possible $a + b \leq 7$ (by hand or computer) gives the possible values for $t > 2^{k-1}$, which are given in the table below.

a	1	2	2	2	3	3	3	4	4
b	1	0	1	2	1	2	3	2	3
t	3	3	6	10	10	20	35	35	70

Hence, if $t(L) \geq 2^{k-1} + 1$, then $t(L)$ is of the form $2^c t'$ for some t' in the table above and the appropriate c. □

Let $L : \{0,1\}^3 \to \mathbb{R}$ be a linear map with $t(L) \geq 2^2$. Then we know $L(e_i) \in \{-1, 0, 1\}$ for all $i = 1, \ldots, n$ and by filling in all the 8 possible elements from $\{0,1\}^3$, we observe the following.

Observation 2.2. *Let $L : \{0,1\}^3 \to \mathbb{R}$ with $|t(L)| \geq 2^2$. Then $L(e_1), L(e_2)$ and $L(e_3)$ must be 1, 1 and -1 in some order. If $L(e_1) = L(e_2) = 1$ and $L(e_3) = -1$, then $I(L) = \{000, 010, 011, 100, 101, 111\}$.*

An important feature of the set $I = \{000, 010, 011, 100, 101, 111\}$ is that if $x \in I$ starts 00, then we know $x = 000$ but if $x \in I$ begins with 01, then x can be either 010 or 011. We will see below that this is an important part of calculating the intersection size of a shape.

Suppose that the shape of $L : \{0,1\}^4 \to \mathbb{R}^2$ is a (3,2)-star with two edges as shown in Figure 2.2 and suppose also that $t(L) \geq 2^3$. We must have $t(L_1) \geq t(L) \geq 2^3$. Since $L_1(e_4) = 0$, we find that the map $L_1' : \mathbb{R}^3 \to \mathbb{R}$ defined by $L_1'(x) = L_1(x, 0)$

for $x \in \mathbb{R}^3$ has $t(L_1') = t(L_1)/2 \geq 2^2$. From Observation 2.2, we know that $L'(e_1)$, $L'(e_2)$ and $L'(e_3)$ are 1, 1 and -1 in some order, and we can work out the intersection pattern if we know the order. To calculate the intersection pattern $I(L)$ we condition on the values that the maps L_1 and L_2 take on the shared basis elements e_2 and e_3.

If they both take the value 1 on both e_2 and e_3 and $x = (x_1, x_2, x_3, x_4)$ is such that $L(x) \in \{0,1\}^2$, then $x_2x_3 = 00$ implies $x = 0000$ (since L_1 forces $x_1 = 0$ and L_2 forces $x_4 = 0$). If we are given $x_2x_3 = 01$ instead, then x_1 could be 0 or 1 and independently x_4 could be 0 or 1 so there are four options for x. Continuing like this, we see that the intersection has size

$$2 \times 2 + 2 \times 2 + 1 \times 1 + 1 \times 1 = 10.$$

Suppose now instead that L_1 takes the value 1 on both e_2 and e_3 but L_2 takes the value -1 on e_2 and 1 on e_3. If $x_2x_3 = 00$, we again know that $x_1 = 0$ but now x_4 could be 0 or 1 giving two options for x. If $x_2x_3 = 01$, then x_1 could be 0 or 1 but now x_4 must be 0 so we only get two options for x. Completing the calculation shows that the intersection size is

$$1 \times 2 + 2 \times 1 + 2 \times 1 + 1 \times 2 = 8.$$

Combining the above with a symmetry argument shows that an intersection size strictly greater than half the cube is only possible if the two maps 'agree' on whether their common intersection has a -1, since this corresponds to agreeing on which values of the shared coordinates have many extensions.

This illustrates the general principle that for large intersection sizes we need the points where each map can extend to a large number of different options to agree to some extent. While this works on all shapes, it quickly becomes very tedious and we will use a computer to calculate the largest intersection sizes of different shapes.

Many shapes can be quickly ruled out; for example, if the shape of $L : \{0,1\}^7 \to \mathbb{R}^3$ is a '(3,1)-star with three edges', say $S(L_1) = \{1,2,3\}$, $S(L_2) = \{1,4,5\}$ and $S(L_3) = \{1,6,7\}$, then counting the number of $x \in I(L) = \cap_{j=1}^3 I(L_j)$ with $x_1 = 0$ and those with $x_1 = 1$, we find

$$t(L) \leq 3 \times 3 \times 3 + 3 \times 3 \times 3 = \left(\frac{3}{4}\right)^3 2^7 < \frac{1}{2} \times 2^7.$$

Similar calculations can rule out many other shapes such as those with three disjoint edges.

Our main result for large intersections is the following theorem, from which Theorem 2.2 follows directly.

Theorem 2.13. *For $k \geq 6$,*

$$\begin{aligned} H^+(k,1) &= \left\{2^{k-1}+2^{k-5}+2^{k-6},\ 2^{k-1}+2^{k-3},\ 2^{k-1}+2^{k-2}, 2^k\right\}, \\ H^+(k,2) &= H^+(k,1) \cup \left\{2^{k-1}+2^{k-4}\right\}, \\ H^+(k,i) &= H^+(k,i-1) \cup \left\{2^{k-1}+2^{k-(i+1)}\right\} \text{ for } i = 3,\ldots,k-1, \\ \text{and } H^+(k,i) &= H^+(k,k-1) \text{ for } i \geq k. \end{aligned}$$

In particular, note that $H(k,2) = H(k,3)$ for $k \geq 6$. We also note that it is easy to check that the conjecture of Melo and Winter is true for $k \in \{1,\ldots,5\}$ using Lemma 2.11 and a computer search.

To prove Theorem 2.13 we require two more auxiliary lemmas, the first of which describes the structure of L.

Lemma 2.14. *Suppose $L : \mathbb{R}^k \to \mathbb{R}^m$ is a minimal linear map with $|S(L_i)| > 1$ for all $i \in [m]$, $|S(L)| = k \geq 8$ and $t(L) > 2^{k-1}$. Then the sets $S(L_i)$ form either a (2,1)-star or a (3,2)-star.*

Proof. As seen in the proof of Corollary 2.12, the size of $S(L_i)$ must be in $\{2,\ldots,7\}$ for all i or $t(L) \leq t(L_i) \leq 2^{k-1}$. As we have assumed that $|S(L)| \geq 8$, we must have $m \geq 2$. Checking the possible minimal shapes (as defined in Definitions 2.5 and 2.6) the maps L_1 and L_2 can form the following options (up to switching L_1 and L_2).

$\lvert S(L_1)\rvert$	$\lvert S(L_2)\rvert$	$\lvert S(L_1) \cap S(L_2)\rvert$
2	2	0
2	2	1
2	3	0
2	3	1
3	3	0
3	3	1
3	3	2

In all of these options $|S(L_{\{1,2\}})| \leq 6$ so in fact $m \geq 3$. This can now be repeated for L_1, L_2 and L_3 and a computer search shows the sets $S(L_i)$ for $i = 1,2,3$ form either a (2,1)-star or a (3,2)-star if $t(L) > 2^{k-1}$. By the symmetry of $\{0,1\}^m$, this holds for all distinct i_1, i_2, i_3 and the result follows. □

Lemma 2.15. *Let $L : \mathbb{R}^k \to \mathbb{R}^m$ be a linear map with $k \geq 8$ and $t(L) \geq 2^{k-1} + 1$. Let $X_i = \{x \in \{0,1\}^k : x_i = 0\}$. Then one of the following statements holds:*

(a) $m \geq k - 1$ and $t(L) = 2^{k-1} + 1$,

(b) there exists an $i \in [k]$ such that $|L^{-1}(\{0,1\}^m) \cap X_i| = \frac{1}{2}t(L)$.

Proof. If $|S(L)| < k$, then there is some coordinate i such that $L(e_i) = 0$. This means the value $L(x)$ does not depend on whether x_i is 0 or 1, so that i satisfies (b). Thus we may assume that $|S(L)| = k$. Choose $A \subseteq [m]$ such that $L' = L_A$ is minimal and it satisfies $|S(L')| = k$. All assumptions of Lemma 2.14 are then satisfied and so the sets $S(L'_i)$ form either a $(2,1)$-star or a $(3,2)$-star.

Suppose the $S(L'_i)$ form a (2,1)-star. Then, as $|S(L')| = k$, $|A| = k - 1$ and $m \geq k - 1$. The maximum intersection size of a (2,1)-star is $2^{k-1} + 1$ so, since we have assumed $t(L) \geq 2^{k-1} + 1$, (a) is satisfied.

Suppose the $S(L'_i)$ form a $(3,2)$-star. It is easy to see that $t(L') = 2^{k-1} + 2$ and that (b) holds for L' for every i. Either L and L' have the same intersection pattern (that is, $I(L) = I(L')$) and we are done, or $m \geq |A| + 1 \geq k - 1$ and $t(L) \leq t(L') - 1 \leq 2^{k-1} + 1$ so (a) is satisfied. In either case at least one of (a) or (b) is satisfied for L. □

Proof of Theorem 2.13. Using the above lemmas we prove the theorem using induction on k. A computer search shows that the claim holds for $k = 6, 7, 8$, so assume the theorem holds for some $k \geq 8$.

For any linear map $L : \mathbb{R}^k \to \mathbb{R}^m$, we can define a map $L' : \mathbb{R}^{k+1} \to \mathbb{R}^m$ by $L'(e_i) = L(e_i)$ for $i = 1, \ldots, k$ and $L'(e_{k+1}) = 0$. Then $t(L') = 2t(L)$, so that $H^+(k+1, i)$ contains the claimed elements for $i \in \{1, \ldots, k-1\}$. This only leaves the case of $H(k+1, k)$ for which we are missing the value $2^k + 1$. The map $L : \mathbb{R}^{k+1} \to \mathbb{R}^k$ defined by

$$L_j(e_i) = \begin{cases} 1 & \text{if } i = j \in \{1, \ldots, k\} \text{ or } i = k+1 \\ 0 & \text{otherwise} \end{cases}$$

satisfies $t(L) = 2^k + 1$, which implies that $H^+(k+1, m)$ contains the claimed elements for all $m \in \mathbb{N}$.

It remains to show that if $L : \mathbb{R}^{k+1} \to \mathbb{R}^m$ is a linear map, then $t(L)$ is in the claimed set. Lemma 2.15 shows that either $t(L) \leq 2^k + 1$ and $m \geq k$ in which case we are done, or there is an i such that $|L^{-1}(\{0,1\}^m) \cap X_i| = \frac{1}{2}t(L)$. In the latter case,

the map $L' : \mathbb{R}^k \to \mathbb{R}^m$ with

$$L'(e_j) = \begin{cases} L(e_j) & \text{for } j = 1, \dots, i-1 \\ L(e_{j+1}) & \text{for } j = i, \dots, k \end{cases}$$

has $\left|(L')^{-1}(\{0,1\}^m) \cap \{0,1\}^k\right| = |L^{-1}(\{0,1\}^m) \cap X_i| = \frac{1}{2}t(L) \geq 2^{k-1} + 1$. It follows that $\frac{1}{2}t(L) \in H^+(k,m)$ so t must be of the claimed form by the induction hypothesis. □

2.4 Small intersection sizes

To determine the large intersection sizes we made extensive use of the fact that L must map each of the basis vectors into $\{-1,0,1\}^m$. This is not true in general for small intersection sizes but we show that it is true on the interval $\big((15/16) \cdot 2^{k-1}, 2^{k-1}\big)$. This allows us to use a structural approach similar to the large case.

To see this suppose $L(\{e_1, \dots, e_k\}) \not\subseteq \{-1,0,1\}^m$. Without loss of generality we can assume that, for some $\ell \geq 1$, $L_1(e_i) \neq 0$ for $i = 1, \dots, \ell$, $\alpha = L_1(e_\ell) \notin \{-1,0,1\}$ and that $L_1(e_i) = 0$ for $i \geq \ell + 1$. Call an element $y \in \{0,1\}^{\ell-1} \times \{0\}$ *good* if either $L_1(y + e_\ell) \in \{0,1\}$ or $L_1(y) \in \{0,1\}$. As $\alpha \notin \{-1,0,1\}$, it can't be the case that both $L_1(y + e_\ell)$ and $L_1(y)$ are in $\{0,1\}$ and hence

$$t(L_1) = \#\text{good elements} \cdot 2^{k-\ell}.$$

An element $y \in \{0,1\}^{\ell-1} \times \{0\}$ is good if and only if $L_1(y) \in \{0, 1, 1-\alpha, -\alpha\}$ so we can bound the maximum intersection size using a Littlewood-Offord style result [41, 73].

Lemma 2.16. *Let $a_1, \dots, a_\ell \in \mathbb{R} \setminus \{0\}$ and $B \subseteq \mathbb{R}$ of size $|B| = 4$. Then*

$$\left| \left\{ I \subseteq [\ell] : \sum_{i \in I} a_i \in B \right\} \right| \leq \frac{15}{16} 2^\ell$$

for all $\ell \geq 4$.

This lemma can be proved by associating the set $\{I \subseteq [\ell] : \sum_{i \in I} a_i \in B\}$ with the (disjoint) union of four antichains in $\mathcal{P}([\ell])$. Oliver Riordan pointed out the following more direct argument. Given $a_1, \dots, a_4 \in \mathbb{R} \setminus \{0\}$, the set $\{\sum_{i \in I} a_i : I \subseteq [4]\}$ contains at least 5 possible values. Hence, at most 15 of the 16 possible subsets $I \subseteq [4]$ satisfy $\sum_{i \in I} a_i \in B$. Since this also holds when $\sum_{i \in J} a_i$ is added to all terms for some $J \subseteq \{5, \dots, \ell\}$, this proves the lemma.

The lemma shows that the intersection size is small if many e_i are mapped to non-zero elements, something we have seen before in Corollary 2.12. In fact, it can be shown that any map with at least two non-zero entries can have at most 3/4 of the hypercube mapping to any set of two elements.

The following result shows if there are a small number of non-zero elements, adding (non-redundant) constraints must reduce the intersection by a large amount, which will cause a gap in the values that can be achieved.

Lemma 2.17. *Let $L : \mathbb{R}^k \to \mathbb{R}^{m+1}$ be a linear map such that the last condition is not redundant, i.e. $t(L) < t(L_{[m]})$. Then*

$$t(L) \leq \max\{\tfrac{3}{4}t(L_{[m]}), t(L_{[m]}) - 2^{k-s-1}\}$$

where $s = |S(L_{[m]})|$.

Proof. Permuting the coordinates as necessary we can write $I(L_{[m]}) = J(L_{[m]}) \times \{0,1\}^{k-s}$. Let L' be the restriction of L_{m+1} to the last $k-s$ coordinates.

- Suppose L' maps at least two basis vectors to non-zero values. Then, as observed above, L' can map at most 3/4 of the hypercube $\{0,1\}^{k-s}$ to any set of two real numbers. But for any $(x,z) \in I(L_{[m]})$, we have $(x,z) \in I(L)$ if and only if $L'(z) \in \{-L_{m+1}(x), 1 - L_{m+1}(x)\}$. This implies that $t(L) \leq (3/4)t(L_{[m]})$.

- Suppose L' maps one basis vector e_i to a non-zero value. Since the last condition is not redundant, there exist $x \in J(L_{[m]})$ and $z \in \{0,1\}^{k-s}$ such that $L(x,z) \notin \{0,1\}$. Every vector $y \in \{0,1\}^{k-s}$ that agrees with z in position i (that is, $y_i = z_i$) satisfies $L_{m+1}(x,y) = L_{m+1}(x,z) \notin \{0,1\}$. Since there are 2^{k-s-1} such vectors, it follows that $t(L) \leq t(L_{[m]}) - 2^{k-s-1}$.

- The case where L' maps everything to zero is similar to the second case (but with the slightly stronger bound $t(L) \leq t(L_{[m]}) - 2^{k-s}$).

In all three cases, the statement holds. □

Lemma 2.18. *If $t(L) > (15/16)2^{k-1}$ and $t(L) \neq 2^{k-1}$, then $L(\{e_1, \ldots, e_k\}) \subseteq \{-1,0,1\}^m$.*

Proof. Suppose $L(\{e_1, \ldots, e_k\}) \not\subseteq \{-1,0,1\}^m$ and as before assume that, for some $\ell \geq 1$, $L_1(e_1), \ldots, L_1(e_{\ell-1}) \neq 0$, $\alpha = L_1(e_\ell) \notin \{-1,0,1\}$ and that $L_1(e_{\ell+1}) = L_1(e_{\ell+2}) = \cdots = L_1(e_k) = 0$.

By Lemma 2.16, the set of good elements has size at most $(15/16)2^{\ell-1}$ for $\ell-1 \geq 4$. Hence, if $\ell \geq 5$,

$$t(L) \leq t(L_1) \leq \frac{15}{16}2^{\ell-1}2^{k-\ell} = \frac{15}{16}2^{k-1}.$$

Suppose $\ell \leq 4$, then either $t(L_1) = 2^{k-1}$ or

$$t(L) \leq t(L_1) \leq \frac{7}{8}2^{k-1} \leq \frac{15}{16}2^{k-1}.$$

If $t(L_1) = 2^{k-1}$, then either $t(L) = t(L_1) = 2^{k-1}$ or, by Lemma 2.17,

$$t(L) \leq \max\left\{\frac{3}{4}2^{k-1}, \left(1-2^{-\ell}\right)2^{k-1}\right\} \leq \frac{15}{16}2^{k-1}. \qquad \square$$

Armed with the lemmas above we are now ready to prove the main result of this section: for $k \geq 8$ and any $m \geq 1$,

$$H(k,m) \cap \left\{\tfrac{15}{16}2^{k-1}, \ldots, 2^{k-1}\right\} = \left\{\tfrac{15}{16}2^{k-1}, \tfrac{63}{64}2^{k-1}, 2^{k-1}\right\}.$$

Proof of Theorem 2.4. Lemma 2.11 shows that the claimed values are present for $m = 1$ and hence for all $m \geq 1$. Indeed, $H(k,1)$ contains the value $2^c\binom{a+b+1}{b+1}$ for all choices of $a, b, c \geq 0$ such that $a+b+c = 8$, and we have $120 = 2^3\binom{6}{2}$, $126 = \binom{9}{5}$ and $128 = 2^7\binom{2}{2}$. Hence, it only remains to show that no other values are present and we assume that $(15/16)2^{k-1} < t(L) < 2^{k-1}$.

By Lemma 2.18, we may assume that L maps the basis vectors e_i to $\{-1,0,1\}^m$ and, without loss of generality, we can also assume that no conditions are redundant, i.e. there is no $i \in [m]$ such that $t(L) = t(L_{[m]\setminus\{i\}})$. If $|S(L)| < k$, then the result for L follows from the result for L restricted to $S(L)$. Hence, we only need to prove the result for the maps L such that $|S(L)| = k$. Although the theorem is only stated for $k \geq 8$, it naturally extends to $k < 8$ by removing entries which are not integers. The cases where $k \leq 7$ can be checked with a computer, and so we can assume that $k \geq 8$.

Having made these assumptions we now proceed as in the large case and look for the shapes which can have a sufficiently large intersection size. Suppose $S(L_i) = e_1$, then either $I(L_i) = \{0,1\}^k$ and the condition is redundant, or $I(L_i) = \{0\} \times \{0,1\}^{k-1}$. Removing the first coordinate, we get a map $L' : \mathbb{R}^{k-1} \to \mathbb{R}^m$ with intersection size $(15/16)2^{k-1} < t(L') < 2^{k-1}$ and no such map exists. Hence, $|S(L_i)| \geq 2$ for all i and by checking binomial coefficients, we get $|S(L_i)| \leq 9$ for all i as well. A computer search shows that there are 18 options for the shape formed by $S(L_1)$ and $S(L_2)$ that have a sufficiently large intersection size, but all of them have $|S(L_{1,2})| \leq 6$. As

we have assumed $|S(L)| = k \geq 8$, we must have $m \geq 3$ and checking the possible shapes for $m = 3$ shows that we must have $m \geq 4$ to have $k \geq 8$. The search shows that either $t(L) \leq t(L_{[4]}) \leq (15/16)2^{k-1}$ or the sets $S(L_1), \ldots, S(L_4)$ form one of six different shapes. These shapes are

- the (2,1)-star with 4 edges,
- the (3,2)-star with 4 edges,
- the (2,1)-star with 3 edges plus a second copy of one of the edges,
- the (3,2)-star with 3 edges plus a second copy of one of the edges,
- the (3,2)-star with 3 edges where there is an edge containing just the two central vertices,
- the (3,2) with 2 edges where there are 2 edges containing just the two central vertices.

Every subset of 4 edges must form one of the above shapes and so it follows that the only shapes with $m \geq 4$ and sufficiently large intersection size are extensions of the above formed by adding extra edges to the star. By conditioning on the value over the centre of the star, we find that the largest intersection size smaller than 2^{k-1} is bounded above by $2^{k-2} + 4$ which, for $k \geq 8$, is less than $(15/16)2^{k-1}$. □

2.5 Computer search

This section gives a brief overview of the computer searches that were instrumental in proving the preceding theorems. The aim of each search is to identify the possible shapes $\{S_i : i \in [m]\}$ for which there is a corresponding map with a large intersection size. This is done by starting with the shapes for $m = 1$ with sufficiently large intersection (these are easily found by checking binomial coefficients) and incrementally adding adding new edges.

As adding conditions does not increase the proportion of the hypercube that is covered, only shapes which cover a sufficiently high proportion need to be kept. Isolated points in the hypergraph do not affect the proportion of points covered by a shape so are ignored by the computation. We do a breadth-first search over the possible shapes where, at each step, every way of adding another condition is checked. It is clearly possible to add the conditions in any order and we make use of this by adding conditions in order of non-increasing support size. This reduces the

computation required and removes some isomorphic shapes, although we make no other attempt to eliminate isomorphic shapes and these must be removed by hand later. In particular, the code for the proof of Theorem 2.13 will output ten shapes, but they fall into only six isomorphism classes.

A shape does not uniquely define a map and different maps with different intersection sizes can have the same shape. For each shape that is considered, the maximum intersection size is computed by checking all assignments of ± 1 to every appropriate $L_i(e_j)$ (where $L_i(e_j) \neq 0$ if and only if the shape dictates that $j \in S(L_i)$). Suppose that the coordinates that are in multiple edges are $1, \ldots, \ell$ and the coordinates $\ell+1, \ldots, k$ are in exactly one edge. We consider every point $x \in \{0,1\}^\ell$ and, for each edge L_i, we calculate $L_i(x)$. Given this value, we can use Lemma 2.11 to calculate the number of ways of extending x to the other coordinates in $S(L_i)$. This is done for every map and the total number of ways of extending x to $[k]$ is then the product of the number of ways of extending each map. For a given map, the number of ways of extending x depends only on the number of 1s (and the number of -1s) and not which coordinates they correspond to. This is used to reduce the computation and speed up the search.

Note that removing any set S_i from a minimal shape leaves another minimal shape (although for a smaller domain size k). This means that we only need to keep minimal shapes at each step of the search.

For the last proof, we cannot restrict to minimal shapes, and we instead require that there are no redundant conditions (i.e. there is no i such that $t(L) = t(L_{[m]\setminus\{i\}})$). For simplicity, the search does the following which may allow some assignments with redundant conditions. Suppose we are attempting to add a set S_{m+1} to the shape with sets $S_1, \ldots, S_m$ which has some previously computed maximum intersection size t (when no conditions are redundant). When calculating the maximum intersection size of the new shape, any assignments that achieve an intersection of size at least t are ignored as these must have at least one redundant condition, although it is possible that the redundant condition is not the last one. The shape is kept if the largest intersection size, amongst those which were not discarded, is sufficiently large.

The code can be found attached to the arXiv submission [52] on which this chapter is based.

2.6 Open problems

From our results, one can also draw conclusions about the structure of the sets $H(n) = \bigcup_{k \leq n} H(k, n-k)$, that is, the possible intersections sizes of a fixed cube with a linear subspace. For example, using Corollary 2.12, and recalling that the maximum intersection size of k-dimensional subspace is 2^k,

$$\begin{aligned} H(n) \cap \left[\tfrac{1}{4}2^n, 2^n\right] &= (H(n,0) \cup H(n-1,1) \cup H(n-2,2)) \cap \left[\tfrac{1}{4}2^n, 2^n\right] \\ &= \left\{2^n, 2^{n-2}+2^{n-6}+2^{n-7}, 2^{n-2}+2^{n-4}, 2^{n-2}+2^{n-3}, 2^{n-1}, 2^{n-2}\right\} \\ &= \left\{\frac{1}{4}2^n, \frac{35}{128}2^n, \frac{5}{16}2^n, \frac{3}{8}2^n, \frac{1}{2}2^n, 2^n\right\}. \end{aligned}$$

While we have shown that the conjecture that $[1, 2^{k-1}] \subseteq H(k, \infty)$ for all $k \in \mathbb{N}$ is false, there is still much that is not known about the small intersections. It seems likely that, as the size of the cubes increases to infinity, there are values missing in the lower 2^{-M} fraction for any fixed M. In fact, we make the stronger conjecture that there is a positive proportion of such values missing.

Conjecture 2.19. *For every fixed $M \in \mathbb{N}$, there exist $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ such that $\left|\left[0, 2^{k-M}\right] \cap H(k, \infty)\right| \leq (1-\varepsilon)2^{k-M}$ for all $k \geq k_0$.*

If this conjecture is true, then another interesting question is whether a positive proportion of the values is missing in $[\alpha 2^k, \beta 2^k]$ for any fixed $0 < \alpha < \beta < 1$ and all k sufficiently large (depending on α and β).

When trying to 'save' the spirit of the original conjecture, the following problem seems natural.

Problem 2.20. *Let $k \in \mathbb{N}$. What is the largest $\ell = \ell(k)$ such that $[1, \ell] \subseteq H(k, \infty)$?*

Any integer $x \in [2^k]$ can be written as $\sum_{i=0}^{a} 2^{t_i}$ for some $0 \leq t_0 < t_1 < \cdots < t_a \leq k$ with $a \leq k$ and is hence in $H(2k, \infty)$ by Proposition 2.9. This shows the lower bound $\Omega(2^{k/2})$.

Chapter 3

Lipschitz bijections between boolean functions

Given two functions $f, g : \{0,1\}^n \to \{0,1\}$ we say that a bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ is a *mapping from f to g* if, for every $x \in \{0,1\}^n$, we have $f(x) = g(\phi(x))$. Write a point $x \in \{0,1\}^n$ as $x = (x_1, \ldots, x_n)$ and, for $\phi : \{0,1\}^n \to \{0,1\}^n$, write $\phi = (\phi_1, \ldots, \phi_n)$. Recall that, for $x, y \in \{0,1\}^n$, we use $|x| = \sum_{i=1}^n x_i$ to denote the Hamming weight of x and $\mathrm{dist}(x,y) = \sum_{i=1}^n |x_i - y_i|$ to denote the Hamming distance between x and y.

A bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ is said to be *C-Lipschitz* if, for all $x, y \in \{0,1\}^n$, $\mathrm{dist}(\phi(x), \phi(y)) \leq C\,\mathrm{dist}(x,y)$, and ϕ is said to be *C-bi-Lipschitz* if both ϕ and ϕ^{-1} are C-Lipschitz. We say that the jth output bit ϕ_j *depends on* the ith input bit x_i if there exists $x \in \{0,1\}^n$ such that $\phi_j(x) \neq \phi_j(x + e_i)$. If every output bit depends on at most k input bits, we say the map ϕ is *k-local*.

In Sections 3.1 and 3.2 we study mappings between three common boolean functions Dictator, XOR and Majority which we define by

- $\mathrm{Dictator}(x) = x_1$,
- $\mathrm{XOR}(x) = \sum_{i=1}^n x_i \mod 2$,
- $\mathrm{Majority}(x) = 1$ if $|x| > n/2$ and $\mathrm{Majority}(x) = 0$ otherwise.

In Section 3.3 we consider a uniformly random balanced function $f : \{0,1\}^n \to \{0,1\}$ where we say a boolean function $g : \{0,1\}^n \to \{0,1\}$ is *balanced* if $g^{-1}(1)$ and $g^{-1}(0)$ are of the same size.

In [85] Rao and Shinkar study mappings from Dictator to XOR and give a construction for a mapping which is 2-Lipschitz and 3-local and where the inverse is $O(\log(n))$-Lipschitz. This leads them to ask if there is a mapping from Dictator to

XOR which is both $O(1)$-local and $O(1)$-bi-Lipschitz. In Section 3.1 we prove that no such mapping exists with the following theorem.

Theorem 3.2. *Let $\phi : \{0,1\}^n \to \{0,1\}^n$ be a mapping from* Dictator *to* XOR *which is C-Lipschitz and where each output bit depends on at most k input bits. Then there is a constant $\delta = \Omega(1/(k + \log(C)))$ such that the inverse map ϕ^{-1} is not $\delta \log(n)$-Lipschitz.*

Furthermore, if ϕ is a linear mapping, then we may take $\delta = \Omega(1/\log(C + k))$.

Rao and Shinkar also consider the *average stretch* of a mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ which is defined

$$\text{avgStretch}(\phi) = \mathbb{E}_{x,i}[\text{dist}(\phi(x), \phi(x + e_i))]$$

where $x \in \{0,1\}^n$ and $i \in [n]$ are both chosen independently and uniformly at random. Rao and Shinkar [85] show that for any mapping ϕ from XOR to Majority the average stretch is $\Omega(\sqrt{n})$ and ask if this lower bound is tight. We show that this is indeed the case with the following theorem.

Theorem 3.4. *For odd n there is a mapping ϕ from* XOR *to* Majority *such that*

$$\text{avgStretch}(\phi) = O\big(\sqrt{n}\big).$$

Since any mapping from XOR to Majority must have large average stretch, Rao and Shinkar ask if there is a $O(1)$-Lipschitz embedding from $\{0,1\}^n$ to $\{0,1\}^{\text{poly}(n)}$ such that $\text{XOR}(x) = \text{Majority}(\phi(x))$. We give a simple construction that gives a positive answer:

Theorem 3.6. *For every n there exists a 3-Lipschitz embedding $\phi : \{0,1\}^n \to \{0,1\}^{2n+1}$ such that $\text{XOR}(x) = \text{Majority}(\phi(x))$ for all $x \in \{0,1\}^n$.*

In [85] Rao and Shinkar show that there is a constant C such that, with high probability, there is a mapping from Dictator to a random balanced boolean function f with average stretch at most C. In Section 3.3 we show that this result extends to the worst-case stretch.

Theorem 3.8. *Let $f : \{0,1\}^n \to \{0,1\}$ be a balanced boolean function chosen uniformly at random. Then, with high probability, there exists a mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ from* Dictator *to f such that for all $x \in \{0,1\}^n$,*

$$\text{dist}(x, \phi(x)) \leq C$$

where C is an absolute constant. In particular, ϕ is $(2C + 1)$-bi-Lipschitz.

Chapter outline In Section 3.1 we prove Theorem 3.2, in Section 3.2 we prove Theorem 3.4 and Theorem 3.6. We then prove Theorem 3.8 in Section 3.3 and end with some open problems in Section 3.4.

This chapter is based on joint work with Alex Scott [64].

3.1 Bijections from Dictator to XOR

Given a mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ from Dictator to XOR we define the *dependency graph* D_ϕ as follows. Let D_ϕ be the bipartite graph with vertex sets $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$ where there is an edge $a_i b_j$ if and only if the jth output bit of ϕ depends on the ith input bit. The following lemma shows that if there is an output bit ϕ_j which is at a large distance from a_1 in the dependency graph, then changing ϕ_j must cause many input bits to change.

Lemma 3.10. *Suppose ϕ is a mapping from* Dictator *to* XOR *such that the distance between a_1 and b_j in D_ϕ is at least d. Then, for any $y \in \{0,1\}^n$,*

$$\operatorname{dist}\big(\phi^{-1}(y), \phi^{-1}(y + e_j)\big) \geq \frac{d+1}{2}.$$

Proof. Suppose that $\phi^{-1}(y) = x$ and $\phi^{-1}(y + e_j) = x'$ and let the set of coordinates in which they differ be X. Now let D'_ϕ be the subgraph of D_ϕ induced by $A_X := \{a_i : i \in X\}$ and its neighbours. Since $\phi(x)$ and $\phi(x')$ differ in the jth bit, one of the neighbours of b_j must be in A_X so that $b_j \in V(D'_\phi)$. As ϕ is a mapping from Dictator to XOR and $\mathrm{XOR}(y) \neq \mathrm{XOR}(y+e_j)$, we must have $\mathrm{Dictator}(x) \neq \mathrm{Dictator}(x')$ and so $a_1 \in A_X$. If a_1 and b_j are in the same component of D'_ϕ, then D'_ϕ must contain a path from a_1 to b_j, which must have length at least d. Since D'_ϕ is bipartite, $|X| \geq (d+1)/2$ and we are done.

Otherwise, let the component of D'_ϕ containing a_1 have vertices $A_Y \cup B_{Y'}$, where $Y, Y' \subset [n]$. Consider the input $x + e_Y$ where $e_Y = \sum_{i \in Y} e_i$: we will show that $\phi(x + e_Y) = \phi(x)$. If $y' \in Y'$, then the output bit $\phi_{y'}$ depends only on the input bits in Y and input bits that are not in X so $\phi_{y'}(x + e_Y) = \phi_{y'}(x') = \phi_{y'}(x)$ (since $\phi(x)$ and $\phi(x')$ differ only in the jth bit and $j \notin Y'$). Changing the value of x on Y can only change $\phi(x)$ in the output bits in Y' and so $\phi_i(x) = \phi_i(x + e_Y)$ for every $i \notin Y'$. This means that $\phi(x) = \phi(x + e_Y)$, which gives a contradiction as ϕ is a bijection. □

Now that we have related distance in the dependency graph D_ϕ to the Hamming distance of the inverse map, we prove Theorem 3.2 by showing that the conditions on ϕ imply that there is a vertex at least logarithmically far from a_1 in D_ϕ.

Proof of Theorem 3.2. As each output bit of $\phi(x)$ depends on at most k input bits of x, the degree of a vertex b_j in D_ϕ is at most k.

Let us now show that the degrees of the a_i are also bounded. Fix a vertex a_i in D_ϕ and let it have neighbours $b_{j_1}, \ldots, b_{j_m}$. The value of the output bit j_1 depends on i and at most $k-1$ other bits which we denote X. By definition, there is some assignment x of the $k-1$ bits X such that $\phi(x)$ and $\phi(x+e_i)$ differ in the bit j_1. Let $U \in \{0,1\}^n$ be a bit string chosen uniformly at random: then the probability that this is equal to x on X is 2^{1-k} and, hence, the probability that $\phi(U)$ and $\phi(U+e_i)$ differ in bit j_1 is at least 2^{1-k}. The same holds for bits $j_2, \ldots, j_m$ and so, by the linearity of expectation, the expected number of bits in which $\phi(U)$ and $\phi(U+e_i)$ differ is at least $m2^{1-k}$. There must be some value for U for which the number of bits that differ is at least $m2^{1-k}$ and, as ϕ is C-Lipschitz, we must have $m2^{1-k} \leq C$. In particular, the degree of a vertex a_i is bounded by $C2^{k-1}$.

Define $\Delta := \max\{k, C2^{k-1}\}$ so that every vertex has degree at most Δ. The number of vertices at distance at most l from a_1 is at most $(\Delta+1)^l$ and hence, if the distance from a_1 to any vertex is at most d, $(\Delta+1)^d \geq 2n$ and $d \geq (\log 2n)/\log(\Delta+1)$. As the graph is bipartite, there must be a vertex b_j at distance at least $d-1$ from a_1 and hence, by Lemma 3.10, ϕ is not $\delta\log(n)$-Lipschitz for $\delta = 1/(2\log(\Delta+1))$.

If ϕ is a linear map, then $\phi(x)$ and $\phi(x+e_i)$ differ in the bits $j_1, \ldots, j_m$ for every x. Hence, the degree of a vertex a_i is bounded by C (compared to $C2^{k-1}$ in the general case), and the bound on δ follows from bounding the maximum degree by $C+k$. □

The construction presented by Rao and Shinkar in [85] gives a linear map which shows $\delta(2,k) = O(1/\log(k))$. Theorem 3.2 shows that this is tight for linear maps, but only gives the bound $\delta(2,k) = \Omega(1/k)$ for general maps. This raises the following question: Does there exist a map ϕ from Dictator to XOR which is C-Lipschitz, where each output bit depends on at most k input bits and such that ϕ^{-1} is $o(\log(n)/\log(k))$-Lipschitz as $n, k \to \infty$?

3.2 Bijections from XOR to Majority

It was shown by Rao and Shinkar in [85] that any mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ from XOR to Majority must have average stretch $\Omega(\sqrt{n})$ in any direction. In this section we prove Theorem 3.4 by constructing a map with average stretch $O(\sqrt{n})$. Our strategy is to map each $x \in \{0,1\}^n$ for which $\mathrm{XOR}(x) = \mathrm{Majority}(x)$ to itself, and otherwise to swap x with a y chosen such that $|x| + |y| = n$. The problem is then to find a matching, so that when the matched elements are switched, they are

not too far from their neighbours on average. We do this by matching according to a symmetric chain decomposition so the Hamming distance between switched elements is minimised. This is similar to the proof of Theorem 1.2 in [85] which defines a mapping from Majority to Dictator by permuting elements along the chains in a particular symmetric chain decomposition. Recall that a *symmetric chain* is a path $C = \{c_k, \ldots, c_{n-k}\}$ in the hypercube such that $|c_i| = i$. A *symmetric chain decomposition* is a partition of the hypercube into symmetric chains. It is well known that these decompositions exist for all n (see e.g. [20]).

Proof of Theorem 3.4. Suppose we have a symmetric chain decomposition of $\{0,1\}^n$ and, for a point $x \in \{0,1\}^n$, let y_x be the unique point in the chain containing x such that $|x| + |y_x| = n$. We note that $\mathrm{XOR}(x) \neq \mathrm{XOR}(y_x)$ and $\mathrm{Majority}(x) \neq \mathrm{Majority}(y_x)$ so the bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ defined by

$$\phi(x) = \begin{cases} x & \text{if } \mathrm{XOR}(x) = \mathrm{Majority}(x) \\ y_x & \text{if } \mathrm{XOR}(x) \neq \mathrm{Majority}(x) \end{cases}$$

is a mapping from XOR to Majority.

Claim 3.1. *Suppose $x \in \{0,1\}^n$ has Hamming weight m. Then, for any i,*

$$\mathrm{dist}(\phi(x), \phi(x + e_i)) \leq |n - 2m| + 3$$

Proof of Claim 3.1. The proof of this claim is a case analysis over the values $\mathrm{XOR}(x)$ and $\mathrm{Majority}(x)$, whether $e_i \leq x$ and, where necessary, $\mathrm{Majority}(x + e_i)$. Let us fix i and use x' to denote $x + e_i$.

First consider the case where $\mathrm{XOR}(x) = 1,\ \mathrm{Majority}(x) = 1$ and $x_i = 0$. Then we have $|x'| = m + 1$, $\mathrm{XOR}(x') = 0$ and $\mathrm{Majority}(x') = 1$. Hence, ϕ keeps x constant and switches x' to $y_{x'}$. We know that $y_{x'} \leq x'$ and is of weight $n - m - 1$ so x and $y_{x'}$ must agree in at least $n - m - 2$ places. Hence, the distance between x and $y_{x'}$ is bounded by $1 + m - (n - m - 2) = 2m - n + 3$.

Now consider the case $\mathrm{XOR}(x) = 1,\ \mathrm{Majority}(x) = 1$ and $x_i = 1$. In this case it is possible for $\mathrm{Majority}(x') = 1$ or $\mathrm{Majority}(x') = 0$ so we must consider these cases separately. Suppose $\mathrm{Majority}(x') = 1$. Then ϕ keeps x constant and switches x' to $y_{x'}$ where $|y_{x'}| = n - m + 1$. We also have $y_{x'} \leq x + e_i \leq x$ so the stretch is $2m - n - 1$. If $\mathrm{Majority}(x + e_i) = 0$, then ϕ keeps both x and $x + e_i$ constant and the stretch is 1.

The other cases follow similarly. □

For a uniformly chosen x and an arbitrarily chosen direction we have

$$\mathbb{E}_{x,i}[\mathrm{dist}(\phi(x), \phi(x+e_i))] \leq 3 + 2^{-n} \sum_{m=0}^{n} |2m-n| \binom{n}{m}. \tag{3.1}$$

Let S_n be the position a simple random walk on $\mathbb{Z}$ after n steps (i.e. $S_n = \sum_{i=1}^{n} X_i$ where the X_i are independent and take the value 1 with probability $1/2$ and the value -1 otherwise). Then $\mathbb{E}[|S_n|] = 2^{-n} \sum_{m=0}^{n} |2m-n| \binom{n}{m}$, and applying standard results we get

$$\mathbb{E}_{x,i}[\mathrm{dist}(\phi(x), \phi(x+e_i))] \leq 3 + \frac{n+1}{2^{n+1}} \binom{n+1}{(n+1)/2} \sim \sqrt{\frac{2n}{\pi}}.$$

□

Given that the average stretch for a mapping from XOR to Majority must be $\Omega(\sqrt{n})$, Rao and Shinkar ask if it is possible to relax the problem by increasing the size of the codomain and asking instead for a $O(1)$-Lipschitz injection. In the proof of Theorem 3.6 below we give one such example.

Proof of Theorem 3.6. Define $\phi(x) : \{0,1\}^n \to \{0,1\}^{2n+1}$ by

$$\phi(x) = (x^c, \mathrm{XOR}(x), x)$$

where x^c is the bitwise complement of x (and where we have written (a, b, c) for the concatenation of vectors a, b and c). As we can read off x from the last section, this is clearly a one-to-one mapping. If x has Hamming weight m, x^c has Hamming weight $n - m$, so overall $\phi(x)$ has Hamming weight $n + \mathrm{XOR}(x)$. Hence, $\mathrm{Majority}(\phi(x)) = \mathrm{XOR}(x)$ for all $x \in \{0,1\}^n$.

Suppose $x, y \in \{0,1\}^n$ are distinct. Then

$$\begin{aligned} \mathrm{dist}(\phi(x), \phi(y)) &= \mathrm{dist}(x^c, y^c) + \mathrm{dist}(\mathrm{XOR}(x), \mathrm{XOR}(y)) + \mathrm{dist}(x, y) \\ &= 2\,\mathrm{dist}(x, y) + \mathrm{XOR}(x+y) \\ &\leq 3\,\mathrm{dist}(x, y) \end{aligned}$$

and the map is 3-Lipschitz. □

3.3 Bijections from Dictator to a random f

Let f be a boolean function. We say a point $x \in \{0,1\}^n$ is a *1 of f* if $f(x) = 1$ and similarly say x is a *0 of f* if $f(x) = 0$. We say a set B is *balanced* if B contains an equal number of 1s and 0s of f, is *positively imbalanced* if B contains more 1s of f

than 0s of f and *negatively imbalanced* if B contains more 0s of f than 1s of f. The *imbalance* of f over a set B is the unsigned difference between the number of 1s of f and $|B|/2$.

Before we prove Theorem 3.8, let us first sketch the ideas behind the proof without worrying about the technicalities.

Sketch proof of Theorem 3.8. We start by partitioning the hypercube $\{0,1\}^n$ into "blocks" which contain a large polynomial number of points (say n^a), have bounded diameter and contain an equal number of 1s and 0s of Dictator. These blocks are small in comparison to the cube so, in a given block B, we expect the number of 1s of f to have a distribution similar to that of a binomial random variable with $|B|$ trials and success probability $1/2$. This means we expect every block to have an imbalance not much more than $n^{a/2}$ and a large proportion (tending to 1) of the blocks to have an imbalance not much less than $n^{a/2}$. We call a sequence of distinct blocks $B_1, \ldots, B_k$ a *block path* if, for $i = 1, \ldots, k-1$, there is an edge between a vertex of B_i and a vertex of B_{i+1}. In our construction, we take a large set of random block paths in the hypercube and use these to move the imbalance. If a block B_1 has too many 1s of f, we find a block path $B_1, \ldots, B_k$ from B_1 to a block B_k with too many 0s and map a 0 of Dictator from B_1 to a 0 of f from B_2, a 0 of Dictator from B_2 to a 0 of f from B_3, and so on. We also do the same with the 1s but in the opposite direction. That is, we map a 1 of Dictator from B_2 to a 1 of f from B_1, map a 1 of Dictator from B_3 to a 1 of f from B_2, and so on. By doing this along enough paths, we can even out the sets and then arbitrarily match within them.

For this to work we need to make sure that we don't try to map too many points to a block, which means ensuring there aren't too many paths through any single block. However, we also need to make sure there are enough paths between the blocks to spread the imbalance around. Once we have chosen our paths (at random) we construct a bipartite graph between the positive and negative blocks and find a subgraph with suitable degrees. This corresponds to the paths that will actually be used to move the imbalance around. The imbalance is not much more than $n^{a/2}$ in any block, so we don't need too many paths; at the same time, the imbalance is not much less than $n^{a/2}$ in most blocks, which ensures that there are enough random paths between imbalanced sets. □

Our proof of Theorem 3.8 uses the fact that, with high probability, there is a perfect matching in our random bipartite graph. The following lemma is an imme-

diate consequence of a result of Erdős and Rényi [42]. (The lemma can easily be strengthened, but this form is sufficient for our case.)

Lemma 3.11. *Let A be a multi-set of at least $2n\log(n)$ pairs (i,j) each chosen uniformly and independently at random from $[n]\times[n]$. Let G be the random bipartite graph with vertex set $\{v_1^+,\dots,v_n^+,v_1^-,\dots,v_n^-\}$, where $v_i^+v_j^-$ is an edge if and only if $(i,j)\in A$. Then the probability that G contains a perfect matching tends to 1 as $n\to\infty$.*

We will also make use of the following Chernoff bounds; see [78] for a discussion and derivation of these bounds.

Lemma 3.12. *Let $X\sim \mathrm{Bin}(n,p)$ and $t>0$. Then*

$$\mathbb{P}(X\geq np+t)\leq e^{-2t^2/n}.$$

Also, for $0\leq\varepsilon\leq 1$ we have

$$\mathbb{P}(X\geq(1+\varepsilon)np)\leq e^{-\varepsilon^2 np/4}$$

and

$$\mathbb{P}(X\leq(1-\varepsilon)np)\leq e^{-\varepsilon^2 np/2}.$$

We are now ready to prove the main theorem. Throughout this proof we will use the constants a, b, c and d to which we assign explicit values only at the end of the proof.

Proof of Theorem 3.8. Suppose first that we have a mapping $\phi:\{0,1\}^n\to\{0,1\}^n$ from Dictator to f such that $\mathrm{dist}(x,\phi(x))\leq C$ for all $x\in\{0,1\}^n$. Fix $x\in\{0,1\}^n$ and let $y=x+e_i$ for some i. Then

$$\mathrm{dist}(\phi(x),\phi(y))\leq \mathrm{dist}(x,\phi(x))+\mathrm{dist}(x,y)+\mathrm{dist}(y,\phi(y))\leq 2C+1$$

and the triangle inequality shows ϕ is $(2C+1)$-Lipschitz. A similar argument shows that ϕ^{-1} is also $(2C+1)$-Lipschitz and this proves the latter part of the theorem. Thus it remains to show that such a ϕ exists with high probability.

We shall assume for the rest of the argument that n is large enough for our estimates to hold. We start by partitioning the hypercube $\{0,1\}^n$ into "blocks" with diameter at most $4a$ which contain between n^{a-2} and $4n^a$ elements, exactly half of which are 1s of Dictator. Let B be a maximal set of points in $\{0,1\}^{n-1}$ such that

the pairwise distance between any two points is at least $2a$. For each point $v \in B$, we define the set of points A_v to be the points in $\{0,1\}^{n-1}$ which are closer to v than to any other point in B (settling ties arbitrarily). As the set B is maximal, the radius of A_v must be less than $2a$, or else we could add a point to B. As the distance from v to any other point in B is at least $2a$, A_v contains a Hamming ball of radius $a-1$ centred at v and at least n^{a-2} points. If a set contains N elements with $N > 2n^a$, we arbitrarily split the set into $\lfloor N/n^a \rfloor$ sets each containing between n^a and $2n^a$ points. For each of our sets A, we define a corresponding "block" in $\{0,1\}^n$ to be $\{0,1\} \times A$. Clearly this gives a set $\mathcal{B}$ of blocks with the desired properties. Note that $2^n/(4n^a) \leq |\mathcal{B}| \leq 2^n/n^{a-2}$.

The following claim shows that, with probability tending to 1, no block $B \in \mathcal{B}$ has an imbalance much more than $n^{a/2}$.

Claim 3.2. *Provided that $b > 1/2$, with probability tending to 1, none of the sets $B \in \mathcal{B}$ have an imbalance of more than $n^{a/2+b}$.*

Proof. Say that a block B is *bad* under a function f if B has an imbalance between the number of 1s and 0s of f greater than $n^{a/2+b}$. Let $g : \{0,1\}^n \to \{0,1\}$ be a random boolean function where $g(x)$ is chosen independently and uniformly at random for each $x \in \{0,1\}^n$. Then the probability that B is bad under f is the same as the probability that B is bad under g conditioned on the event that g is balanced, so

$$\mathbb{P}(B \text{ bad under } f) = \mathbb{P}(B \text{ bad under } g | g \text{ balanced}) \leq \frac{\mathbb{P}(B \text{ bad under } g)}{\mathbb{P}(g \text{ balanced})}.$$

The number of 1s of g in B is a binomial random variable so the first Chernoff bound in Lemma 3.12 shows the probability that B is bad under g is at most $2e^{-2n^{a+2b}/N} \leq 2e^{-n^{2b}/2}$. The probability that g is balanced is $2^{-n}\binom{2^n}{2^{n-1}} \sim (\pi 2^{n-1})^{-1/2}$. Using the union bound over all blocks $B \in \mathcal{B}$, the probability that at least one of the sets B is bad is at most $2^{3n/2}e^{-n^{2b}/2}$ for all sufficiently large n, and this is $o(1)$ provided $b > 1/2$. □

We will also need that most sets have an imbalance not much less than $n^{a/2}$.

Claim 3.3. *Provided that $c > 2$, with probability tending to 1, the sum over $B \in \mathcal{B}$ of the imbalance of B is at least $2^n/(8n^{a/2+c})$.*

Proof. Fix B and let $|B| = N$. We can then write down the probability that the imbalance of B is at most $n^{a/2-c}$ as

$$\sum_{i=-n^{a/2-c}}^{n^{a/2-c}} \frac{\binom{N}{N/2+i}\binom{2^n-N}{2^{n-1}-N/2-i}}{\binom{2^n}{2^{n-1}}}.$$

Using $\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}$ and that $N = o(2^n)$, the largest term ($i = 0$) is at most $O(N^{-1/2})$. As $N \geq n^{a-2}$, each term is $O(n^{-(a/2-1)})$ and the sum is $O(1/n^{c-1})$. Let X be the number of blocks with imbalance at most $n^{a/2-c}$. Then,

$$\mathbb{E}[X] = O\left(\frac{|\mathcal{B}|}{n^{c-1}}\right)$$

and so, by Markov's Inequality, $\mathbb{P}(X \geq |\mathcal{B}|/n) = O(1/n^{c-2})$. This is $o(1)$ provided $c > 2$, and in this case, with probability tending to 1, the sum of the imbalances is at least $(2^n/(4n^a)) \cdot (1 - 1/n) \cdot n^{a/2-c} \geq 2^n/(8n^{a/2+c})$ for large n. □

Now suppose that every block has an imbalance of at most $n^{a/2+b}$ and that the total imbalance is at least $2^n/(8n^{a/2+c})$.

Independently sample $K = \lfloor 2^n/n^d \rfloor$ uniformly random pairs of vertices in the hypercube and a random shortest path between them.

Claim 3.4. *Provided that $a + 1 - d > 1$, with high probability, the number of paths in K which intersect a given block $B \in \mathcal{B}$ is at most $8n^{a+1-d}$.*

Proof. By symmetry every point is equally likely to be on a random shortest path and, as any shortest path has at most n vertices, the probability that a given vertex is on each sampled path is at most $n/2^n$. Using the union bound the probability that a given random path goes through a fixed $B \in \mathcal{B}$ is at most $n|B|/2^n$. As each path is sampled independently the number of paths through a fixed B is dominated by a binomial distribution with K trials and success probability $4n^{a+1}/2^n$. Applying the second Chernoff bound from Lemma 3.12 with $\varepsilon = 1$, the probability that there are more than $8n^{a+1-d}$ paths through a given block is at most $\exp(-n^{a+1-d}/2)$ and, taking the union bound, the probability that any B intersects more than $8n^{a+1-d}$ paths is at most $\exp(-n^{a+1-d}/2) \cdot 2^n/n^{a-2}$. This tends to 0 as $n \to \infty$ provided that $a + 1 - d > 1$. □

We now form a random bipartite graph G with vertex sets V^+ and V^- as follows. If $B \in \mathcal{B}$ has a positive imbalance of i, we add i *positive vertices* $v_1^B, \ldots, v_i^B$ to V^+ and similarly if B has a negative imbalance of i, we add i *negative vertices* to V^-. We generate the edge set of G by examining each of our random paths in turn. Given a path, say between $v_1 \in B_1$ and $v_2 \in B_2$, we do the following:

- Discard the path with probability $1 - n^{2a-4}/(|B_1| \cdot |B_2|)$.

- If the path hasn't been discarded, independently choose two numbers i, j in $[n^{a/2+b}]$ and say that the edge $v_i^{B_1} v_j^{B_2}$ is present if both vertices exist and have opposite signs.

Claim 3.5. *The graph G has a perfect matching M with high probability provided*

$$a/2 > 3b + 2c + d + 7. \tag{3.2}$$

Proof. The probability that a random path gives any particular edge between V^+ and V^- is $p = 2(n^{a/2-2-b}/2^n)^2$ and, in particular, every edge is equally likely. Hence, the number of edges in the multi-set induced by the sampled random paths has a binomial distribution with K trials and success probability $|V^+|^2 p$. By assumption $|V^+| \geq 2^n/(16n^{a/2+c})$ and so $|V^+|^2 p \geq (128 n^{2b+2c+4})^{-1}$. Using the third Chernoff bound in Lemma 3.12, the probability the multi-set contains at most $2^{n-9}/n^{4+2b+2c+d}$ pairs is at most $\exp(-2^{n-11}/n^{4+2b+2c+d})$. There are at most $2^n/n^{a-2}$ blocks each with imbalance at most $n^{a/2+b}$ so $|V^+| \leq 2^n/n^{a/2-b-2}$. Hence, $2|V^+| \log |V^+| \leq 2^{n+1}/n^{a/2-b-3}$. From (3.2), for large enough n,

$$2^{n+1}/n^{a/2-b-3} \leq 2^{n-9}/n^{4+2b+2c+d},$$

and we have that G contains a perfect matching M with high probability by Lemma 3.11. □

Given the matching M we can greedily construct the map ϕ. We start with all vertices as both unset and unused. During the construction we will let $\phi(x) = y$ for some $x, y \in \{0,1\}^n$ and we will then say that x is *set* and y is *used*. The vertices for which we still have to define ϕ are exactly the unset vertices, and the vertices which are not yet in the image of ϕ are exactly the unused vertices.

Each edge in M corresponds to a path in the hypercube and this induces a walk between two blocks in the block decomposition. By removing any loops, we can say the edge corresponds to a path $P = B_{i_1} \dots B_{i_k}$ in the block decomposition where B_{i_1} has positive imbalance and B_{i_k} has negative imbalance. For $j = 1, \dots, k-1$, we choose any point $x \in B_{i_j}$ with $\text{Dictator}(x) = 0$ which is currently unset and any point $y \in B_{i_{j+1}}$ with $f(y) = 0$ which is currently unused and set $\phi(x) = y$, setting x and using y. We also choose any point in $x' \in B_{i_{j+1}}$ with $\text{Dictator}(x) = 1$ which is currently unset and any point $y' \in B_{i_j}$ with $f(y) = 1$ and set $\phi(x') = y'$. In order to guarantee there will always be enough points to do the above, we will enforce that $8n^{a+1-d} \leq n^{a-2}$.

If B has positive imbalance i, there are i paths starting at B (those corresponding to the edges in the matching using $v_1^B, \ldots, v_i^B$) and hence, after processing all the paths which start at B, the number of unset 1s of Dictator in B is still $|B|/2$ while the number of 0s is now $|B|/2 - i$. Similarly the number of unused 1s of f in B is now $|B|/2$ while the number of 0s is still $|B|/2 - i$. The paths for which B is an internal block reduce all four quantities by 1 so, after processing all the paths, the number of 1s and 0s of Dictator which are unset and the number of 1s and 0s of f which are unused, are equal. This means we can greedily complete ϕ inside B by taking any unset x and any unused y with $\text{Dictator}(x) = f(y)$ and setting $\phi(x) = y$. A similar argument works when B is balanced or has negative imbalance.

It is clear that, provided nothing goes wrong, this construction gives a bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ which maps Dictator to f. If $x \in B$, then either $\phi(x) \in B$ or $\phi(x) \in B'$ for some B' for which there is an edge between B and B' in the hypercube. As the diameter of a block is bounded by $4a$, the distance between x and $\phi(x)$ is bounded by $8a + 1$.

This construction can fail if any block has too large an imbalance, the total imbalance is too small, there are too many paths through a block, there aren't enough edges in the multi-set or there isn't a matching in G; but all of these events happen with probability tending to 0 (for suitable choices of a, b, c and d).

To complete the proof it remains to find suitable values for the constants a, b, c and d. From Claim 3.2 we must have $b > 1/2$ and from Claim 3.3 we must have $c > 2$. We will also need $8n^{a+1-d} \leq n^{a-2}$, which follows for large enough n provided $d > 3$. Finally, we require $a \geq 2$, $a+1-d > 1$ and, from Claim 3.4, that $a/2 > 3b+2c+d+7$. All these constraints can be satisfied by taking $a = 42$, $b = 1$, $c = 3$ and $d = 4$. □

3.4 Open Problems

We saw in Theorem 3.2 that it is not possible for a map from Dictator to XOR to be $O(1)$-Lipschitz, $O(1)$-local and also have an inverse that is $O(1)$-Lipschitz. On the other hand we saw that maps exist if we drop the condition on the inverse or on the locality. Can we drop the first condition while keeping the other two?

Problem 3.13. *Is there a mapping $\phi : \{0,1\}^n \to \{0,1\}^n$ from* Dictator *to* XOR *such that each output bit depends on $O(1)$ input bits and such that ϕ^{-1} is $O(1)$-Lipschitz?*

Corollary 3.9 shows that, with high probability, if $A, B \subseteq \{0,1\}^n$ are two independent uniformly random sets with $|A| = |B| = 2^{n-1}$, there exists a $O(1)$-bi-Lipschitz

bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ such that $\phi(A) = B$, but we made no attempt to optimise the constant. We expect a much smaller constant to hold in this corollary (and in Theorem 3.8), possibly even as small as 2 or 3. How small could the constant be?

Problem 3.14. *Let $A, B \subseteq \{0,1\}^n$ be two independent uniformly random sets with $|A| = |B| = 2^{n-1}$. What is the smallest constant C such that, with high probability, there is a C-bi-Lipschitz bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ such that $\phi(A) = B$?*

There are many interesting variations of this problem. For example, what happens if we let $|A| = |B| = f(n)$ for some function $f(n)$?

Problem 3.15. *Let $A, B \subseteq \{0,1\}^n$ be two independent uniformly random sets with $|A| = |B| = f(n)$. For what functions $f : \mathbb{N} \to \mathbb{N}$ does there exist a C such that, with high probability, there is a C-bi-Lipschitz bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ with $\phi(A) = B$? For example, what happens when $f(n) = \Theta(2^n/n)$?*

A closely related problem concerns colourings. If we view the partition of $\{0,1\}^n$ into 2 parts (A and A^c) as a partition into 2 colour classes, then we can view mappings as relabellings of the cube which respect two balanced 2-colourings. What happens if we instead colour the cube with k colours? If k is a constant, then a modification of the argument used to prove Theorem 3.8 will show that there is an $O(1)$-bi-Lipschitz mapping with high probability. But what happens if we let $k \to \infty$ as $n \to \infty$?

We say that a partition of $\{0,1\}^n$ into k parts $A_1, \dots, A_k$ is a *balanced partition* if $|A_1| \leq |A_2| \leq \cdots |A_k| \leq |A_1| + 1$.

Problem 3.16. *Let $A_1, \dots, A_k$ and $B_1, \dots, B_k$ be two independent uniformly random balanced partitions of $\{0,1\}^n$ into $k(n)$ parts. For what functions $k : \mathbb{N} \to \mathbb{N}$ does there exist a constant C such that, with high probability, there is a C-bi-Lipschitz bijection $\phi : \{0,1\}^n \to \{0,1\}^n$ with $\phi(A_i) = B_i$ for all i? How small can C be?*

The problems above have been concerned with functions on the hypercube but we can ask similar questions for functions on other graphs and, in particular, Cayley graphs. For example, what happens for $\mathbb{Z}_n \times \mathbb{Z}_n$ generated by the elements $(1,0)$ and $(0,1)$? In this case, we say that a map $\phi : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n \times \mathbb{Z}_n$ is *C-Lipschitz* if, for every $x, y \in V$, $d(\phi(x), \phi(y)) \leq Cd(x, y)$ where d is the graph distance in the Cayley graph. The proof of Theorem 3.8 relied on a decomposition of $\{0,1\}^n$ into "blocks", each with bounded radius but containing a large polynomial number of points. In $\mathbb{Z}_n \times \mathbb{Z}_n$ any subset with bounded radius must contain a bounded number of points so the "blocks" won't tend towards being relatively balanced.

Problem 3.17. *Let $A, B \subseteq \mathbb{Z}_n \times \mathbb{Z}_n$ be two independent uniformly random sets with $|A| = |B| = \lceil n^2/2 \rceil$. What is the best $C = C(n)$ such that, with high probability, there is a C-bi-Lipschitz bijection $\phi : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n \times \mathbb{Z}_n$ with $\phi(A) = B$?*

In a different direction, we observe that there is a natural symmetric measure of the difference between two boolean functions. Define the *Lipschitz constant* of a mapping ϕ by

$$\|\phi\|_{\text{Lip}} = \inf\{C \geq 0 : \|\phi(x) - \phi(y)\| \leq C\|x - y\| \text{ for all } x, y \in \{0,1\}^n\},$$

and define the *dissimilarity* between two boolean functions f and g by

$$\text{diss}(f, g) := \inf\big\{\|\phi\|_{\text{Lip}} \cdot \|\phi^{-1}\|_{\text{Lip}} : \phi \text{ is a mapping from } f \text{ to } g\big\}.$$

The 2-bi-Lipschitz mapping between Dictator and XOR we saw earlier implies that $\text{diss}(\text{Dictator}, \text{XOR}) \leq 4$, and Theorem 3.8 states that, if f and g are uniformly random balanced boolean functions, $\text{diss}(f, g) = O(1)$ with high probability. On the other hand we know that any mapping from Dictator to Majority is not C-Lipschitz for any $C < n/2$ so $\text{diss}(\text{Dictator}, \text{Majority}) = \Omega(n)$. Rao and Shinkar [85] give a $O(1)$-Lipschitz mapping from Majority to Dictator, so we can strengthen this to $\text{diss}(\text{Dictator}, \text{Majority}) = \Theta(n)$. In general we know that every mapping is n-Lipschitz which gives a trivial upper bound of n^2, but is it possible to do better?

Problem 3.18. *Is there a constant $\alpha < 2$ such that $\text{diss}(f, g) = O(n^\alpha)$ for all balanced boolean functions f and g?*

Finally, we note that the paper by Rao and Shinkar is motivated in part by a problem from Benjamini, Cohen and Shinkar in [15] which remains open and seems very interesting.

Problem 3.19. *Is there a set $A \subseteq \{0,1\}^n$ of size $|A| = 2^{n-1}$ such that any bijection from $\{0,1\}^{n-1}$ to A has average stretch $\omega(1)$?*

Chapter 4

Zero-sum squares in $\{-1, 1\}$-matrices with low discrepancy

Recall that a *square* in a matrix $M = (a_{i,j})$ is a 2×2 submatrix of the form

$$S = \begin{pmatrix} a_{i,j} & a_{i,j+s} \\ a_{i+s,j} & a_{i+s,j+s} \end{pmatrix},$$

and a *zero-sum square* is a square where the entries sum to 0. We will be interested in $\{-1, 1\}$-matrices M which do not contain any zero-sum squares. Given an $n \times m$ matrix $M = (a_{i,j})$ define the *discrepancy* of M as the sum of the entries, that is

$$\operatorname{disc}(M) = \sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} a_{i,j}.$$

A common question in the study of zero-sum problems is to ask if it is possible for a large structure to be zero-sum without containing a small zero-sum structure. In this case: is it possible for the entries of an $n \times n$ $\{-1, 1\}$-matrix M to sum to 0 (i.e. $\operatorname{disc}(M) = 0$), but the matrix to not contain a zero-sum square? This question was recently answered in the negative by Arévalo, Montejano and Roldán-Pensado [6]. In fact, they showed that all $n \times n$ matrices M with $|\operatorname{disc}(M)| \le n$ must be either split or contain a zero-sum square, and a basic calculation shows that no split matrix has discrepancy 0. Here an $n \times m$ $\{-1, 1\}$-matrix $M = (a_{i,j})$ is said to be *t-split* for some $0 \le t \le n + m - 1$ if

$$a_{i,j} = \begin{cases} 1 & i + j \le t + 1, \\ -1 & i + j \ge t + 2, \end{cases}$$

and is said to be *split* if there is some rotation of M which is t-split for some t. These clearly have no zero-sum squares, yet they can have low discrepancy.

Based on their proof, Arévalo, Montejano and Roldán-Pensado conjectured that their result should extend to all $n \times n$ matrices M with $|\operatorname{disc}(M)| \leq Cn$ when n is large enough relative to C.

Conjecture 4.1 (Conjecture 4 in [6]). *For every $C > 0$ there is a integer N such that whenever $n \geq N$ the following holds: every $n \times n$ non-split $\{-1, 1\}$-matrix M with $|\operatorname{disc}(M)| \leq Cn$ contains a zero-sum square.*

Not only do we confirm this conjecture, we prove a much stronger result: a matrix with no zero-sum squares is either split, or has discrepancy greater than $n^2/4$. We remark that proving the conjecture only requires a bound which is $\omega(n)$, while we prove a quadratic bound (and this is clearly the best possible order).

Theorem 4.2. *Let $n \geq 5$. Every $n \times n$ non-split $\{-1, 1\}$-matrix M with $|\operatorname{disc}(M)| \leq n^2/4$ contains a zero-sum square.*

This conjecture is still is a long way from the best known construction which has discrepancy about $n^2/2$, and we discuss this in Section 4.2.

Chapter outline We prove Theorem 4.2 in Section 4.1, and we end the chapter with some open problems in Section 4.2.

4.1 Proof of Theorem 4.2

For $p \leq r$ and $q \leq s$ define the *consecutive submatrix* $M[p : r, q : s]$ by

$$M[p : r, q : s] = \begin{pmatrix} a_{p,q} & a_{p,q+1} & \cdots & a_{p,s} \\ a_{p+1,q} & a_{p+1,q+1} & \cdots & a_{p+1,s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r,q} & a_{r+1,q} & \cdots & a_{r,s} \end{pmatrix}.$$

Throughout the rest of this chapter, we will assume that all submatrices except squares are consecutive submatrices.

We start by stating the following lemma from [6] which, starting from a small t'-split submatrix M', determines many entries of the matrix M. An example application is shown in Figure 4.1.

Lemma 4.3 (Claims 4 and 5 in [6]). *Let M be an $n \times n$ $\{-1, 1\}$-matrix with no zero-sum squares, and suppose that there is an $s \times s$ submatrix $M' = M[p : p + s - 1, q : q + s - 1]$ which is t'-split for some $2 \leq t' \leq 2s - 3$. Let $t = t' + p + q - 2$, $T = \lfloor 3t/2 \rfloor$ and suppose $t \leq n$.*

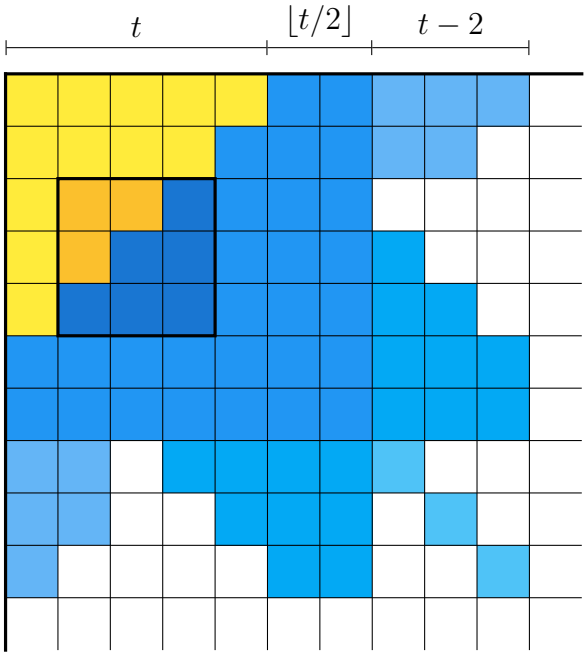


Figure 4.1: The entries known from applying Lemma 4.3. The yellow squares represent -1s and the blue squares represent 1s. The submatrix M' is shown in a darker shade.

1. *The submatrix*

$$N = M[1 : \min\{T, n\}, 1 : \min\{T, n\}]$$

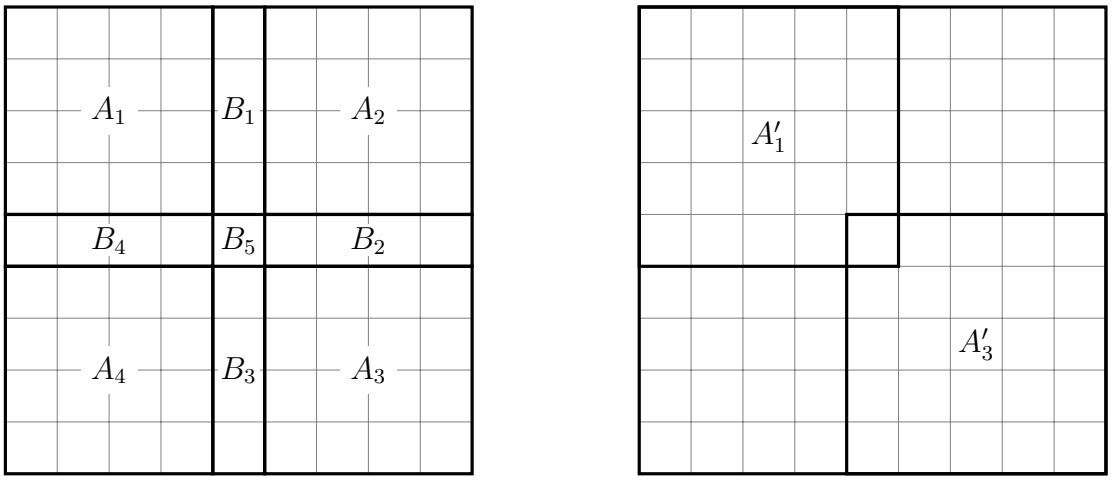
is t-split.

Furthermore, both $a_{i,j} = 1$ and $a_{j,i} = 1$ whenever $T < j \leq T + t - 2$ and one of the following holds:

2. $j - t < i \leq T$,
3. $i \leq \left\lfloor \frac{T+t+1-j}{2} \right\rfloor$,
4. $i = j$.

Note that we can apply this lemma even when it is a reflection of M' which is t-split; we just need to suitably reflect M and potentially multiply by -1, and then undo these operations at the end. The matrix N will always contain at least one of $a_{1,1}$, $a_{1,n}$, $a_{n,1}$ and $a_{n,n}$, and if N contains two, then M is split.

We will also make use of the following observation. This will be used in conjunction with the above lemma to guarantee the existence of some additional 1s, which allows us to show a particular submatrix has positive discrepancy.



(a) (b)

Figure 4.2: A subset of the regions used in the proof of Lemma 4.4.

Observation 4.1. *Let M be an $n \times n$ $\{-1,1\}$-matrix with no zero-sum squares, and suppose that $a_{i,i} = 1$ for every $i \in [n]$. Then at least one of $a_{i,j}$ and $a_{j,i}$ is 1. In particular, $a_{i,j} + a_{j,i} \geq 0$ for all $1 \leq i, j \leq n$.*

The final lemma we will use to prove Theorem 4.2 is a variation on Claims 1 and 2 from [6]. The main difference between Lemma 4.4 and the result used by Arévalo, Montejano and Roldán-Pensado is that we will always find a square submatrix which simplifies the proof of Theorem 4.2.

Lemma 4.4. *For $n \geq 8$, every $n \times n$ $\{-1,1\}$-matrix M with $|\operatorname{disc}(M)| \leq n^2/4$ has an $n' \times n'$ submatrix M' with $|\operatorname{disc}(M')| \leq (n')^2/4$ for some $(n-1)/2 \leq n' \leq (n+1)/2$.*

Proof. We only prove this in the case n is odd as the case n is even is similar, although simpler. Partition the matrix M into 9 regions as follows. Let the four $(n-1)/2 \times (n-1)/2$ submatrices containing $a_{1,1}$, $a_{1,n}$, $a_{n,n}$ and $a_{n,1}$ be $A_1, \ldots, A_4$ respectively. Let the $(n-1)/2 \times 1$ submatrix between A_1 and A_2 be B_1 and define B_2, B_3 and B_4 similarly. Finally, let the central entry be B_5. The partition is shown in Figure 4.2a.

As these partition the matrix M, we have

$$\operatorname{disc}(M) = \operatorname{disc}(A_1) + \cdots + \operatorname{disc}(A_4) + \operatorname{disc}(B_1) + \cdots + \operatorname{disc}(B_5). \tag{4.1}$$

Let the overlapping $(n+1)/2 \times (n+1)/2$ submatrices containing $a_{1,1}$, $a_{1,n}$, $a_{n,n}$ and $a_{n,1}$ be $A'_1, \ldots, A'_4$, as indicated in Figure 4.2b. The submatrices $B_1, \ldots, B_4$

each appear twice in the A'_i and B_5 appears four times and, by subtracting these overlapping regions, we obtain a second equation for $\operatorname{disc}(M)$:

$$\operatorname{disc}(M) = \operatorname{disc}(A'_1) + \cdots + \operatorname{disc}(A'_4) \\ - \operatorname{disc}(B_1) - \cdots - \operatorname{disc}(B_4) - 3\operatorname{disc}(B_5). \quad (4.2)$$

If any of the A_i or A'_i have $|\operatorname{disc}(A_i)| \leq (n-1)^2/16$ or $|\operatorname{disc}(A'_i)| \leq (n+1)^2/16$ respectively, we are done, so we may assume that this is not the case. First, suppose that $\operatorname{disc}(A_i) > (n-1)^2/16$ and $\operatorname{disc}(A'_i) > (n+1)^2/16$ for all $i = 1, 2, 3, 4$. Since $n-1$ is even and $\operatorname{disc}(A_i) \in \mathbb{Z}$, we must have $\operatorname{disc}(A_i) \geq (n-1)^2/16 + 1/4$, and similarly, $\operatorname{disc}(A'_i) \geq (n+1)^2/16 + 1/4$. Adding the equations (4.1) and (4.2) we get the bound

$$n^2/2 \geq 2\operatorname{disc}(M) \geq (n+1)^2/4 + (n-1)^2/4 + 2 - 2\operatorname{disc}(B_5),$$

which reduces to $\operatorname{disc}(B_5) \geq 5/4$. This gives a contradiction since B_5 is a single square. Similarly we get a contradiction if, for every i, both $\operatorname{disc}(A_i) < -(n-1)^2/16$ and $\operatorname{disc}(A'_i) < -(n+1)^2/16$.

This only leaves the case where two of the 8 submatrices have different signs. If $A'_i > (n+1)^2/16$, then, for $n \geq 8$,

$$A_i > (n+1)^2/16 - n > -(n-1)^2/16,$$

and either $|\operatorname{disc}(A_i)| \leq (n-1)^2/16$, a contradiction, or $\operatorname{disc}(A_i) > 0$. By repeating the argument when $\operatorname{disc}(A'_i)$ is negative, it follows that A_i and A'_i have the same sign for every i. In particular, two of the A'_i must have different signs, and we can apply an interpolation argument as in [6].

Without loss of generality, we can assume that $\operatorname{disc}(A'_1) > (n+1)^2/16$ and $\operatorname{disc}(A'_2) < -(n+1)^2/16$. Consider the sequence of matrices $N_0, \ldots, N_{(n-1)/2}$ where

$$N_i = M[1 : (n+1)/2, 1+i : i+(n+1)/2].$$

We claim that there is a j such that $|\operatorname{disc}(N_j)| \leq (n+1)^2/16$, which would complete the proof of the lemma. By definition, $N_0 = A'_1$ and $N_{(n-1)/2} = A'_2$ so there must be some j such that $\operatorname{disc}(N_{j-1}) > 0$ and $\operatorname{disc}(N_j) \leq 0$. Since the submatrices N_{j-1} and N_j share most of their entries $|\operatorname{disc}(N_{j-1}) - \operatorname{disc}(N_j)| \leq n+1$ and, as $(n+1)^2/8 > (n+1)$, it cannot be the case that $\operatorname{disc}(N_{j-1}) > (n+1)^2/16$ and $\operatorname{disc}(N_j) < -(n+1)^2/16$. This means there must be some j such that $|\operatorname{disc}(N_j)| \leq (n+1)^2/16$, as required. □

Armed with the above results, we are now ready to prove our main result, but let us first give a sketch of the proof which avoids the calculations in the main proof.

Sketch proof of Theorem 4.2. Assume we have an $n \times n$ $\{-1, 1\}$-matrix M with no zero-sum squares and which has $|\operatorname{disc}(M)| \leq n^2/4$. We will prove the result by induction, so we assume that the result is true for $5 \leq n' < n$.

Applying Lemma 4.4 gives a submatrix M' with low discrepancy. Since M' also contains no zero-sum squares, we know that it is split by the induction hypothesis. Applying Lemma 4.3 then gives a lot of entries M and, in particular, a submatrix N with high discrepancy. Since we are assuming that M has low discrepancy, the remainder $M \setminus N$ of M not in N must either have low discrepancy or negative discrepancy. In both cases we will find B, a submatrix of M with low discrepancy. When the discrepancy of $M \setminus N$ is low, we use an argument similar to the proof of Lemma 4.4, and when the discrepancy of $M \setminus N$ is negative, we find a positive submatrix using Observation 4.1 and use an interpolation argument.

By the induction hypothesis, B must also be split and we can apply Lemma 4.3 to find many entries of M. By looking at specific $a_{i,j}$, we will show that the two applications of Lemma 4.3 contradict each other. □

We now give the full proof of Theorem 4.2, complete with all the calculations. To start the induction, we must check the cases $n < 30$ which is done using a computer. The problem is encoded as a SAT problem using PySAT [62] and checked for satisfiability with the CaDiCaL solver. The code to do this is attached to the arXiv submission [63] on which this chapter is based.

Proof of Theorem 4.2. We will use induction on n. A computer search gives the result for all $n < 30$, so we can assume that $n \geq 30$ and that the result holds for all $5 \leq n' < n$.

Suppose, towards a contradiction, that M is an $n \times n$ matrix with no zero-sum squares and $|\operatorname{disc}(M)| \leq n^2/4$. By Lemma 4.4, we can find an $n' \times n'$ submatrix $M' = M[p : p + n' - 1, q : q + n' - 1]$ with $(n-1)/2 \leq n' \leq (n+1)/2$ and $|\operatorname{disc}(M')| \leq (n')^2/4$. By the induction hypothesis and our assumption that M doesn't contain a zero-sum square, the matrix M' must be split. By reflecting M and switching -1 and 1 as necessary, we can assume that the submatrix M' is t'-split for some t', and that $t := t' + p + q - 2 \leq n$.

We will want to apply Lemma 4.3, for which we need to check $2 \leq t' \leq 2n' - 3$. If $t' \leq 1$ or $t' \geq 2n' - 2$, then the discrepancy of M' is

$$|\operatorname{disc}(M')| \geq (n')^2 - 1 > (n')^2/4,$$

which contradicts our choice of M'. In fact, since $\operatorname{disc}(M') \leq (n')^2/4$ and $\operatorname{disc}(M') \leq (n')^2 - t'(t'+1)$ we find

$$t \geq t' \geq \frac{1}{2}\Big(\sqrt{3(n')^2+1} - 1\Big) \approx 0.433n. \tag{4.3}$$

If $t + \lfloor t/2 \rfloor \geq n$, the matrix M is t-split and we are done, so we can assume that this is not the case, and that $t \leq 2n/3$. We will also need the following bound on $2t + \lfloor t/2 \rfloor - 2$, which follows almost immediately from (4.3).

Claim 4.1. *We have*

$$2t + \lfloor t/2 \rfloor - 2 \geq n - 1.$$

Proof. Substituting $n' \geq (n-1)/2$ into (4.3) gives the following bound on t:

$$t \geq \frac{1}{4}\Big(\sqrt{3n^2 - 6n + 7} - 2\Big).$$

We now lower bound $\lfloor t/2 \rfloor$ by $(t-1)/2$ to find

$$\begin{aligned} 2t + \lfloor t/2 \rfloor - 2 &\geq 2t + \frac{t-5}{2} \\ &\geq \frac{5}{8}\sqrt{3n^2 - 6n + 7} - \frac{15}{4}. \end{aligned}$$

The right hand side grows like $\frac{\sqrt{75}}{8}n$ asymptotically, which is faster than n, so the claim is certainly true for large enough n. In fact, the equation $\frac{5}{8}\sqrt{3n^2 - 6n + 7} - \frac{15}{4} \geq n - 1$ can be solved explicitly to obtain the following the bound on n:

$$n \geq \frac{1}{11}\Big(251 + 20\sqrt{166}\Big) \approx 46.2.$$

This still leaves the values $30 \leq n \leq 46$ for which the bounds above are not sufficient. These cases can be checked using a computer. □

Let $k = \lceil 5n/6 \rceil$ and let $N = M[1:k, 1:k]$ be the $k \times k$ sub-matrix in the top left corner which contains $a_{1,1}$. We will apply Lemma 4.3 and Observation 4.1 to guarantee lots of 1s in N, and therefore ensure N has large discrepancy. This will mean that the rest of M which is not in N must have low discrepancy, and we can find another split submatrix B.

Claim 4.2. *There is an $(n-k) \times (n-k)$ submatrix B which is disjoint from N and with $|\operatorname{disc}(B)| \leq (n-k)^2/4$.*

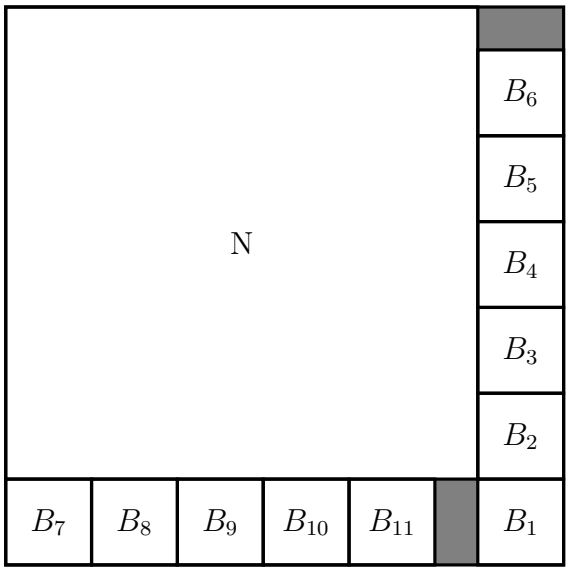


Figure 4.3: The matrix M with the submatrices N and $B_1, \ldots, B_{11}$. The entries of M which are not in any of the submatrices are shown in grey.

Proof. Consider the 11 $(n-k) \times (n-k)$ disjoint submatrices $B_1, \ldots, B_{11}$ of M given by

$$B_i = \begin{cases} M[k:n, n-ik:n-(i-1)k] & \text{if } i \leq 6 \\ M[(i-7)(n-k):(i-6)(n-k), k:n] & \text{if } i > 6, \end{cases}$$

and shown in Figure 4.3. The submatrix B_1 contains $a_{n,n}$ and sits in the bottom right of M, while the others lie along the bottom and right-hand edges of M.

If one of the B_i satisfies $|\operatorname{disc}(B_i)| \leq (n-k)^2/4$, we are done, so suppose this is not the case.

We start by using Observation 4.1 to show that $\operatorname{disc}(B_1) > 0$. Let the entries of B be $b_{i,j}$ where $1 \leq i, j \leq n-k$. By Claim 4.1, $2t + \lfloor t/2 \rfloor - 2 \geq n-1$ and, applying Lemma 1, $b_{i,i} = 1$ for all $i \leq n-k-1$. Further, by Observation 4.1, we have $b_{i,j} + b_{j,i} \geq 0$ for all $1 \leq i, j \leq n-k-1$. This means

$$\operatorname{disc}(B_1) \geq (n-k-1) - (2(n-k)-1) = -(n-k).$$

For $(n-k) \geq 5$, $(n-k) < (n-k)^2/4$ so we must have $\operatorname{disc}(B_1) > (n-k)^2/4$.

As $\operatorname{disc}(B_1) > 0$, if $\operatorname{disc}(B_i) < 0$ for any $i \neq 1$, we can use an interpolation argument as in Lemma 4.4 to find the claimed matrix. The argument only requires

$$2(n-k) < \frac{(n-k)^2}{2},$$

which is true for $(n-k) > 4$.

We must now be in the case where $\operatorname{disc}(B_i) > (n-k)^2/4$ for every i. The bulk of the work in this case will be bounding the discrepancy of the matrix N, and then the discrepancy of M. There are $2n(n-k) - 12(n-k)^2 \leq 10(n-k)$ entries of M in the gaps between the B_i, or in other words, there are at most $10(n-k)$ entries $a_{i,j}$ which are not contained in either N or one of the B_i. In particular, we have

$$\begin{aligned}\operatorname{disc}(M) &\geq \operatorname{disc}(N) + \operatorname{disc}(B_1) + \cdots + \operatorname{disc}(B_{11}) - 10(n-k) \\ &> \operatorname{disc}(N) + 11(n-k)^2/4 - 10(n-k). \end{aligned} \tag{4.4}$$

Let $s = \min\{k, t + \lfloor t/2 \rfloor\}$ so that $M[1:s, 1:s]$ is t split, and let $r = k - s$ be the number of remaining rows in N. Let $a_1, \ldots, a_4$ be the number of 1s in N guaranteed by Lemma 4.3, and let a_5 be the number of additional 1s guaranteed by also applying Observation 4.1. This guarantees that at least one of $a_{i,j}$ and $a_{j,i}$ is 1 for all $s+1 \leq i, j \leq k$, and $a_5 \geq r(r-1)/2$.

We have the following bounds.

$$\begin{aligned} a_1 &= s^2 - \frac{t(t+1)}{2}, \\ a_2 &= 2\sum_{i=1}^{r}(t-i), \\ a_3 &= 2\sum_{i=1}^{r}\left\lfloor \frac{t+1-i}{2} \right\rfloor, \\ a_4 &= r, \\ a_5 &\geq \frac{r(r-1)}{2}. \end{aligned}$$

Let us first consider the case where $s = k$, so that N is t-split. In this case $a_2 = \cdots = a_5 = 0$, and we can easily write down the discrepancy of N as $k^2 - t(t+1)$. Since $k \geq 5n/6$, we get the bound

$$\operatorname{disc}(N) \geq \frac{25n^2}{36} - t(t+1).$$

Substituting this into (4.4) and using the bounds $(n-5)/6 \leq n-k \leq n/6$ we get

$$\begin{aligned}\operatorname{disc}(N) &> \frac{25n^2}{36} - t(t+1) + \frac{11}{4}\left(\frac{n-5}{6}\right)^2 - \frac{10n}{6} \\ &= \frac{1}{144}\big(111n^2 - 350n - 144t^2 - 144t + 275\big).\end{aligned}$$

For $n \geq 4$, the righthand side is greater than $n^2/4$ whenever

$$t < \frac{1}{12}\Big(\sqrt{75n^2 - 350n + 311} - 6\Big) \approx 0.722n + o(n).$$

Since we have assumed $t \leq 2n/3$, we get a contradiction for all sufficiently large n. In fact, we get a contradiction for all $n \geq 40$. The remaining cases need to be checked using exact values for the floor and ceiling functions which we do with the help of a computer.

Now we consider the case where $s = t + \lfloor t/2 \rfloor$ which is very similar, although more complicated. To be in this case, we must have $t + \lfloor t/2 \rfloor \leq k$ which implies

$$t + \frac{t-1}{2} \leq \frac{5(n+1)}{6},$$

and $t \leq (5n+8)/9 \approx 0.556n$.

Start by using the bounds $(t-1)/2 \leq \lfloor t/2 \rfloor$ and $(t-i)/2 \leq \lfloor (t+1-i)/2 \rfloor$ to get

$$\begin{aligned} a_1 + \cdots + a_5 &\geq \left(t + \frac{t-1}{2}\right)^2 - \frac{t(t+1)}{2} + r(2t-r-1) + \frac{r(2t-r-1)}{2} \\ &\quad + r + \frac{r(r-1)}{2} \\ &= \frac{7t^2}{4} - 2t - r^2 + 3rt - r + \frac{1}{4}. \end{aligned}$$

By definition, $r = k - t - \lfloor t/2 \rfloor$, so we get the bounds $5n/6 - t - t/2 \leq r \leq 5(n+1)/6 - t - (t-1)/2$, and substituting these in gives

$$\begin{aligned} a_1 + \cdots + a_5 &\geq \frac{7}{4}t^2 - 2t + \frac{1}{4} - \left(\frac{5(n+1)}{6} - t - \frac{t-1}{2}\right)^2 + 3t\left(\frac{5n}{6} - t - \frac{t}{2}\right) \\ &\quad - \left(\frac{5(n+1)}{6} - t - \frac{t-1}{2}\right) \\ &= \frac{1}{36}\big(-25n^2 + 180nt - 110n - 180t^2 + 126t - 103\big) \end{aligned}$$

Plugging this into (4.4) and using the bounds $5n/6 \leq k \leq 5(n+1)/6$ we get

$$\begin{aligned} \operatorname{disc}(M) &> 2(a_1 + \cdots a_5) - \left(\frac{5(n+1)}{6}\right)^2 + \frac{11}{4}\left(\frac{n-5}{6}\right)^2 - \frac{10n}{6} \\ &\geq \frac{1}{144}\big(-289n^2 + 1440nt - 1430n - 1440t^2 + 1008t - 649\big). \end{aligned}$$

When $n \geq 27$, this is greater than $n^2/4$ whenever

$$\begin{aligned} &\frac{1}{120}\Big(60n + 42 - \sqrt{350n^2 - 9260n - 4726}\Big) < \\ &\qquad t < \frac{1}{120}\Big(60n + 42 + \sqrt{350n^2 - 9260n - 4726}\Big), \end{aligned}$$

or approximately,

$$0.344n < t < 0.656n.$$

We have the bounds

$$\frac{1}{4}\Big(\sqrt{3n^2 - 6n + 7} - 2\Big) \leq t \leq \frac{5n+8}{9},$$

and so, for $n \geq 44$, $\operatorname{disc}(M) > n^2/4$.

This again leaves a few cases which we check with the help of a computer. □

Given a submatrix B as in the above claim we apply the induction hypothesis, noting that $(n-k) \geq 5$ since $n \geq 30$, to find that B is split. Let C be the split submatrix obtained from applying Lemma 4 to B, and let C be ℓ-split up to rotation. Note that $\ell \geq 3$ as $(n-k) \geq 5$ and $|\operatorname{disc}(B)| \leq (n-k)^2/4$, and we can assume $\ell \leq 2n/3$ as M is not split.

Hence, C contains exactly one of $a_{1,1}$, $a_{1,n}$, $a_{n,1}$ and $a_{n,n}$, and we will split into cases based on which one C contains. We will also sometimes need to consider cases for whether the entry is 1 or -1, but in all cases we will find a contradiction.

From Lemma 4.3 applied to M' and Claim 4.1, we already know some of the entries and we highlight some important entries in the following claim.

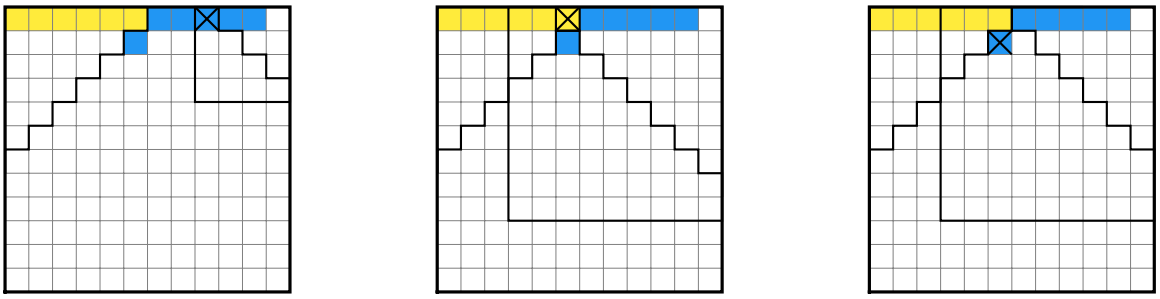
Claim 4.3. *We have*

1. $a_{j,1} = a_{1,j} = \begin{cases} 1 & t+1 \leq j \leq n-1, \\ -1 & 1 \leq j \leq t, \end{cases}$
2. $a_{2,t} = a_{t,2} = 1$,
3. $a_{i,i} = 1$ *for all* $(t+2)/2 \leq i \leq n-1$.

Suppose the submatrix C contains $a_{1,1}$ so sits in the top-left corner. Since $M[1 : t + \lfloor t/2 \rfloor, 1 : t + \lfloor t/2 \rfloor]$ is t-split, C must also be t-split. As C was found by applying Lemma 4.3 to B, it must contain a -1 from B. Hence, $t \geq 5n/6$ which is a contradiction as we assumed that $t \leq 2n/3$.

Suppose instead that C contains $a_{1,n}$ so sits in the top-right corner. Since $\ell \geq 3$, if the corner entry is -1, so is the entry $a_{1,n-1}$, but this contradicts Claim 4.3. Suppose instead that the corner entry is 1. Since C is ℓ-split up to rotation we have, for all $1 \leq i, (n-j+1) \leq \ell + \lfloor \ell/2 \rfloor$,

$$a_{i,j} = \begin{cases} -1 & i + (n-j+1) \geq \ell + 2, \\ 1 & \text{otherwise.} \end{cases} \tag{4.5}$$



(a) $n - \ell > t$ (b) $n - \ell < t$ (c) $n - \ell = t$

Figure 4.4: The three cases when C contains $a_{1,n}$ and $a_{1,n} = 1$. The yellow squares represent some of the $a_{i,j}$ which are known to be -1 from Claim 4.3 and the blue squares those which are 1. The square which gives the contradiction is marked with a cross.

If $n - \ell > t$, then $a_{1,n-\ell} = -1$ by (4.5) and $a_{1,n-\ell} = 1$ by Claim 4.3. Suppose $n - \ell < t$. Then $a_{1,t} = 1$ by (4.5) and $a_{1,t} = -1$ as $M[1 : t, 1 : t]$ is t-split. Finally, when $n - \ell = t$, we have $a_{2,t} = -1$ by (4.5) and $a_{2,t} = 1$ from Claim 4.3. Some illustrative examples of these three cases are shown in Figure 4.4.

The case where C contains $a_{n,1}$ is done in the same way with the rows and columns swapped.

This leaves the case where C contains $a_{n,n}$. Since $\ell \geq 3$, if the entry $a_{n,n}$ equals -1, so does the entry $a_{n-1,n-1}$, and this contradicts Claim 4.3. If instead $a_{n,n} = 1$, we consider the entry $a_{i,i}$ where $i = n + 1 - \lceil (l + 2)/2 \rceil$, which must be -1. However, since $\ell \leq 2n/3$,

$$n + 1 - \lceil (l + 2)/2 \rceil \geq \frac{2n}{3} - \frac{1}{2} > \frac{n}{3} + 1 \geq \frac{t + 2}{2},$$

and $a_{i,i} = 1$ by Claim 4.3. This final contradiction is shown in Figure 4.5. □

We remark that it should be possible to improve the bound $n^2/4$ using a similar proof provided one can check a large enough base case. Indeed, we believe that all the steps in the above proof hold when the bound is increased to $n^2/3$, but only when n is large enough. For example, Claim 4.1 fails for $n = 127$ and our proof of Claim 4.2 fails for $n = 67$. Checking base cases this large is far beyond the reach of our computer check, and some new ideas would be needed here.

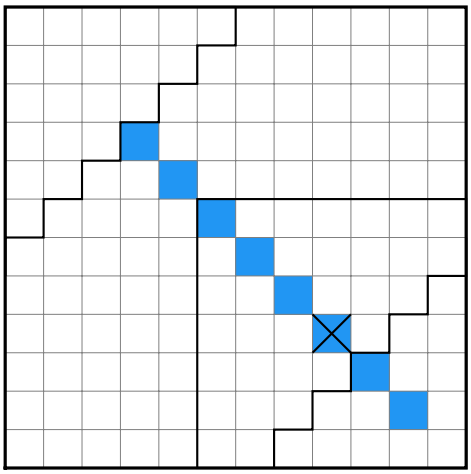


Figure 4.5: The case where C contains $a_{n,n}$ and $a_{n,n} = 1$. The square marked with a cross gives a contradiction.

4.2 Open problems

The main open problem is to determine the correct lower bound for the discrepancy of a non-split $\{-1, 1\}$-matrix with no zero-sum squares. We have improved the lower bound to $\lfloor n^2/4 \rfloor + 1$, but this does not appear to be optimal.

The best known construction is the following example by Arévalo, Montejano and Roldán-Pensado [6]. Let $M = (a_{i,j})$ be given by

$$a_{i,j} = \begin{cases} -1 & i \text{ and } j \text{ are odd,} \\ 1 & \text{otherwise.} \end{cases}$$

This has discrepancy $n^2/2$ when n is even and $(n-1)^2/2 - 1$ when n is odd. With the help of a computer we have verified that this construction is best possible when $9 \leq n \leq 32$, and we conjecture that this holds true for all $n \geq 9$. In fact, our computer search shows that the above example is the unique non-split matrix avoiding zero-sum squares with minimum (in magnitude) discrepancy, up to reflections and multiplying by -1. An example when $n = 9$ is given in Figure 4.6.

We note that the condition $n \geq 9$ is necessary, as shown by the 8×8 matrix with discrepancy 30 given in Figure 4.7.

Conjecture 4.5. *Let $n \geq 9$. Every $n \times n$ non-split $\{-1, 1\}$-matrix M with*

$$|\operatorname{disc}(M)| \leq \begin{cases} \frac{n^2}{2} - 1 & n \text{ is even} \\ \frac{(n-1)^2}{2} - 2 & n \text{ is odd} \end{cases}$$

contains a zero-sum square.

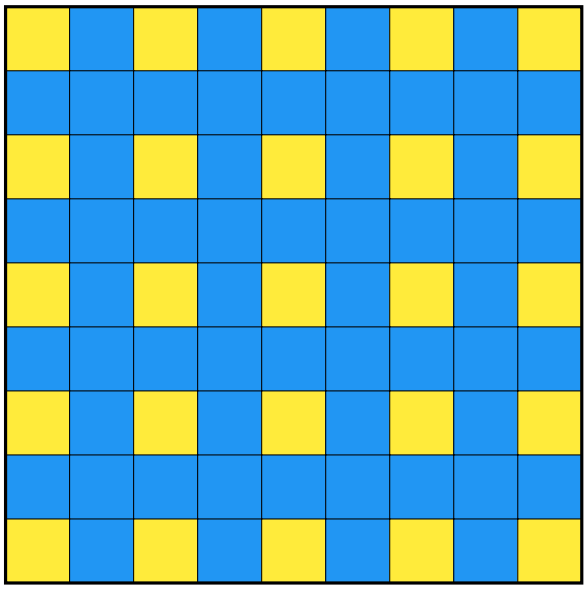


Figure 4.6: A 9×9 $\{-1, 1\}$-matrix with no zero-sum squares and discrepancy 31. Amongst all $\{-1, 1\}$ matrices which are neither split nor contain a zero-sum square this has the smallest (in magnitude) discrepancy. The yellow squares represent a -1 and the blue squares represent a 1.

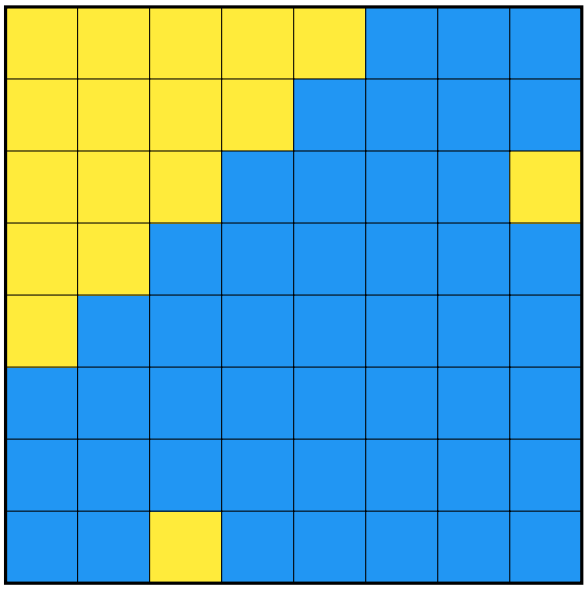


Figure 4.7: An 8×8 $\{-1, 1\}$-matrix with no zero-sum squares and discrepancy 30. The yellow squares represent a -1 and the blue squares represent a 1.

Arévalo, Montejano and Roldán-Pensado prove their result for both $n \times n$ and $n \times (n+1)$ matrices, and computational experiments suggest that Theorem 4.2 holds for $n \times (n+1)$ matrices as well. More generally, what is the best lower bound for a general $n \times m$ matrix when n and m are large?

Problem 4.6. *Let $f(n,m)$ be the minimum $d \in \mathbb{N}$ such that there exists an $n \times m$ non-split $\{-1,1\}$ matrix M with $|\operatorname{disc}(M)| \leq d$. What are the asymptotics of $f(n,m)$?*

Chapter 5

Bicrucial square-free permutations

In this chapter we consider the first of two problems on patterns in permutations. Recall that, given a vector $a = (a_0, \ldots, a_{n-1})$ consisting of n distinct real numbers, one can associate a unique permutation $\sigma \in S_n$ (which we write as a vector $(\sigma(0), \ldots, \sigma(n-1))$) by replacing the ith smallest element of a with i, and we call this σ the *pattern* of a. We say two vectors a and b are *order-isomorphic* if they have the same pattern, and in this case we write $a \sim b$. Throughout this chapter we will work with permutations over the set $\{0, 1 \ldots, n-1\}$, instead of the the usual $\{1, \ldots, n\}$.

We define a square in a permutation as two consecutive factors which follow the same pattern. Formally, a *square* of length 2ℓ is a factor

$$(S_1; S_2) = (\sigma_k, \ldots, \sigma_{k+\ell-1}; \sigma_{k+\ell}, \ldots, \sigma_{k+2\ell-1})$$

where $S_1 \sim S_2$, and we say a permutation is *square-free* if it contains no squares of length at least 4. For each $i \in \{0, \ldots, \ell-1\}$, the entry σ_{k+i} has the same position in the pattern as the entry $\sigma_{k+\ell+i}$ and we will say that the entry σ_{k+i} *corresponds* to the entry $\sigma_{k+\ell+i}$.

A permutation π of length $n+1$ is a *right-extension* of σ if $(\pi_0, \ldots, \pi_{n-1}) \sim \sigma$, or equivalently, if π can be formed by appending an entry $x \in \{0, \ldots, n\}$ to σ and replacing σ_i by $\sigma_i + 1$ when $\sigma_i \geq x$. We say that the permutation σ is *right-crucial* if it is square-free, and every right-extension π contains a square, and we define left-extensions and left-crucial permutations similarly. We will be interested in permutations which are simultaneously left-crucial and right-crucial which we call *bicrucial* permutations.

The first result showing the existence of bicrucial permutations was from Avgustinovich, Kitaev, Pyatkin and Valyuzhenich [7] who showed that bicrucial permutations

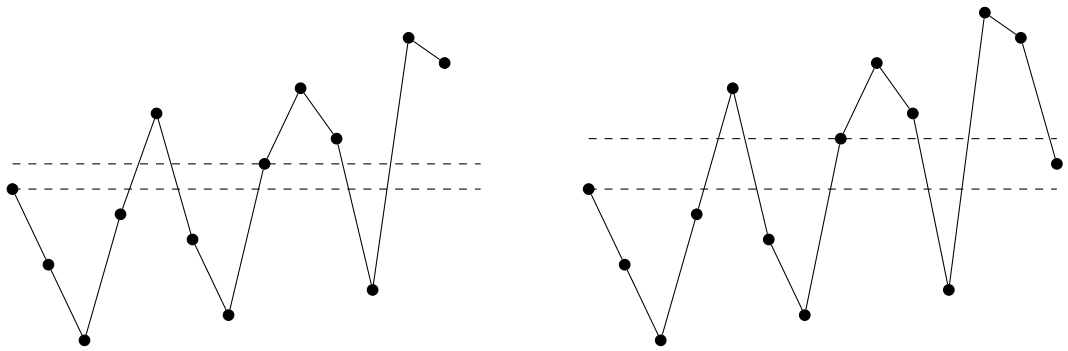


Figure 5.1: A permutation and a possible right extension.

exist for all odd lengths of the form $8k+1$, $8k+5$ or $8k+7$ where $k \geq 1$. Using the constraint solver Minion, bicrucial permutations of length $8k+3$ for $k = 2, 3$ and of even length 32 were found by Gent, Kitaev, Konovalov, Linton and Nightingale in 2015 [49], and they conjectured that larger permutations should exist.

Conjecture 5.1 (Conjecture 8 in [49])**.** *There exist bicrucial permutations of length* $8k+3$ *for all* $k \geq 2$*.*

Conjecture 5.2 (Conjecture 9 in [49])**.** *There exist arbitrarily long bicrucial permutations of even length.*

We confirm both these conjectures and further, completely classify the n for which there exist bicrucial permutations of length n.

Theorem 5.3. *A bicrucial permutation of length* n *exists if and only if* $n = 9$*,* $n \geq 13$ *is odd, or* $n \geq 32$ *is even and not* 38*.*

Chapter outline The proof of this result is split into several parts: In Section 5.2 we give a construction of bicrucial permutations of length $8k+3$ for every $k \geq 3$, which completes the classification of the odd n, and a construction for when $n \geq 48$ is even. The remaining small (even) cases are handled using computer searches, and we give constructions for the remaining possible n in Section 5.3. To show that there are no bicrucial permutations of length 38 we use an exhaustive computer search that we detail in Section 5.3.1. The code used to exhaustively search for such permutations can also be used to enumerate square-free, left-crucial and bicrucial permutations with only minor modifications, and we do this in Section 5.3.2.

This chapter is based on joint work with Carla Groenland [54].

5.1 Preliminaries

Up-up-down-down condition For $i \in \{0, \ldots, n-2\}$, we say a permutation $\sigma \in S_n$ goes *up* from position i to position $i+1$ if $\sigma_i < \sigma_{i+1}$ and goes *down* otherwise. The permutation σ satisfies the *up-up-down-down condition* if for every $i \in \{0, \ldots, n-4\}$, σ goes up from position i to position $i+1$ if and only if it goes down from $i+2$ to $i+3$. This forces the permutation to alternate between going up twice and going down twice. If a permutation σ violates the up-up-down-down condition and goes up from position i to $i+1$ and also up from position $i+2$ to $i+3$ say, then $(\sigma_i, \sigma_{i+1}; \sigma_{i+2}, \sigma_{i+3})$ is a square and the permutation is not square-free. In fact, the up-up-down-down condition is equivalent to σ not containing any squares of length 4.

In any permutation which satisfies the up-up-down-down condition, the entries fall naturally into three categories. We say the entry i is a *high entry* if σ goes up from $i-1$ to i and down from i to $i+1$, a *low entry* if σ goes down from $i-1$ to i and up from i to $i+1$ and a *medium entry* otherwise. This definition extends to the initial and final entries in the obvious way.

High-medium-low construction We now give a construction due to Avgustinovich, Kitaev, Pyatkin and Valyuzhenich [7] which, starting from a square-free permutation π of length n, constructs a square-free permutation σ of length $2n$. Let σ be any permutation such that $\sigma_i < \lceil n/2 \rceil$ when $i \equiv 1 \mod 4$, $\sigma_i = \pi_i + \lceil n/2 \rceil$ when i is even and $\sigma_i \geq n + \lceil n/2 \rceil$ when $i \equiv 3 \mod 4$. It is not hard to see that this permutation is square-free. Indeed, the permutation satisfies the up-up-down-down condition, so there are no squares of length 4 and any square must be of length at least 8. Restricting a square in σ to only the terms of even index (in σ) gives a square of length at least 4 in π. Since π is square-free, σ must also be square-free.

This simple high-medium-low construction is very useful. By suitably choosing the pattern of the prefix and the suffix of our permutation, we can ensure the permutation has a square when extending to the left and to the right. The difficulty is then in joining the prefix and suffix with a long middle section while ensuring that the permutation is square-free. To do this we construct the middle section using the high-medium-low construction above with additional constraints on π. The key to doing this is the observation that every high entry is higher than every medium entry, which are all in turn higher than every low entry, while this not going to be true in the chosen prefix/suffix.

To prove that our constructions give square-free permutations, we will use the following simple observation.

Observation 5.1. *Suppose σ_i is contained in a square and the corresponding entry in the other half of the square is σ_j. Then $|i - j| \leq n/2$.*

In particular, entries in our short prefix can only correspond to entries at most a little over halfway through the permutation and, provided n is large enough, cannot correspond to entries in the suffix. This means we only need to avoid our prefix corresponding to part of the high-medium-low construction. When n is small, we will check the construction using a computer.

5.2 Constructions for large n

We first show the existence of bicrucial permutations when n is of the form $8k + 3$ and $k \geq 3$. Since constructions are already known when n is of the form $8k + 1$, $8k + 5$ or $8k + 7$ ($k \geq 1$) [7] and the number of bicrucial permutations is known when $n \leq 19$ (OEIS A238935 [90]), this completely classifies the odd n for which bicrucial permutations exist.

Theorem 5.4. *Let $n = 8k+3$ where $k \geq 3$. Then there exists a bicrucial permutation of length n.*

Proof. Throughout this proof we shall assume that $n > 44$. This is chosen so that $11+n/2 < n-11$ and $n-14-n/2 > 8$ which will allow us to assume that the relevant corresponding entries come from the middle of the construction. When $n \leq 44$, the proof we give here does not apply. Instead, we have have verified by computer that there exists a bicrucial permutation by using the construction below.[1].

We split the set $\{0, \ldots, n-1\}$ into five different regions $R_0, \ldots, R_4$ where $R_i = [r_i, r_{i+1})$ for a specific choice of the values $0 = r_0 \leq r_1 \leq \cdots \leq r_4 \leq r_5 = n$. The values of the r_i could be determined explicitly from the construction below if desired, but we will leave them only implicitly defined. We begin with a carefully chosen prefix in region R_0, followed by a middle section which uses the high-medium-low construction from Section 5.1 across the regions R_1, R_2 and R_3. Finally, we add a carefully chosen suffix in region R_4.

[1]We repeatedly apply the high-medium-low construction to construct a single permutation π, and then check that the construction below yields a bicrucial permutation for that particular choice of π.

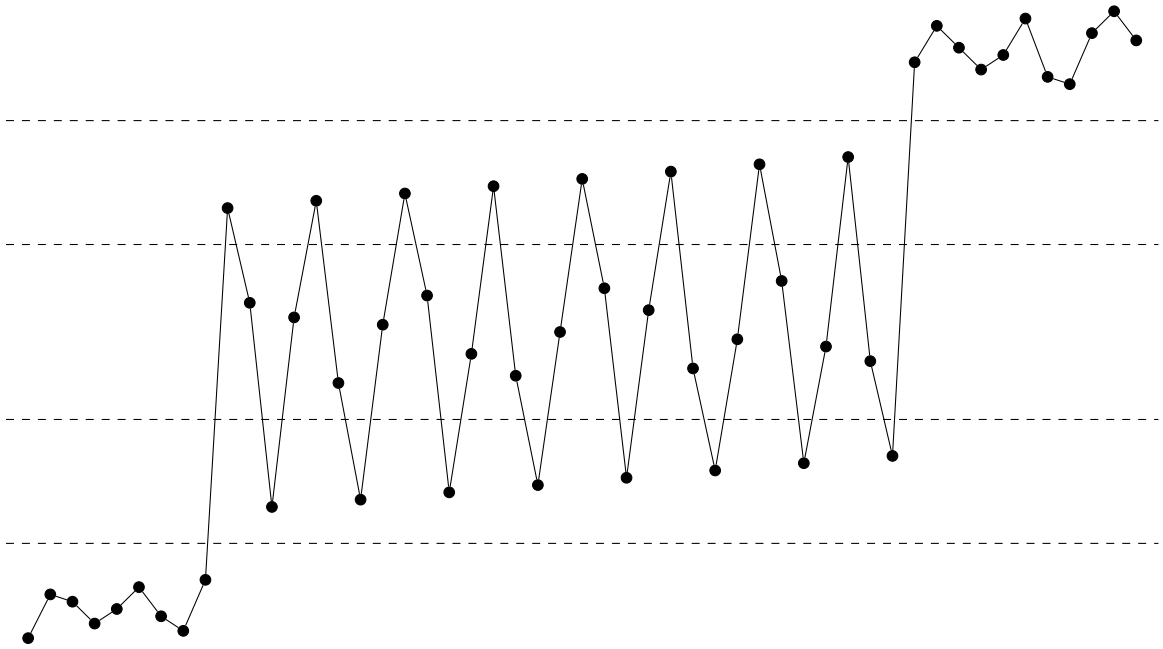


Figure 5.2: A bicrucial permutation of length 51 constructed as in the proof of Theorem 5.4. The dashed lines separate the different regions.

Let $m = (n-21)/2$ and let π be any square-free permutation of length m which begins down-down (i.e. with the pattern $(2,1,0)$). Since every square-free permutation follows the up-up-down-down pattern and trimming the ends of a square-free permutation leaves another square-free permutation, it is trivial to construct π. We now construct σ as follows. Start with the permutation $(0,6,5,2,4,7,3,1,8)$, and for $0 \leq i \leq 2m$, define σ_{9+i} by

$$\sigma_{9+i} = \begin{cases} r_3 + i/4 & i \equiv 0 \mod 4, \\ r_2 + \pi_{(i-1)/2} & i \equiv 1,3 \mod 4, \\ r_1 + (i-2)/4 & i \equiv 2 \mod 4. \end{cases}$$

It is clear the permutation constructed so far follows the up-up-down-down pattern as required and, since $2m \equiv 2 \bmod 4$, the permutation currently ends down-down. Finally, append the vector

$$(r_4+3, r_4+8, r_4+5, r_4+2, r_4+4, r_4+9, r_4+1, r_4+0, r_4+7, r_4+10, r_4+6),$$

and note that σ still follows the required up-up-down-down pattern.

By our choice of π, we have

$$(\sigma_0, \ldots, \sigma_{15}) \sim (0,6,5,2,4,7,3,1,8,14,13,9,12,15,11,10),$$

and this is left-crucial. Indeed, if we extend to the left by prepending $x \geq 1$, the permutation begins down-up-down and contains a square of length 4. When prepending 0, the permutation begins with the pattern

$$(0, 1, 7, 6, 3, 5, 8, 4, 2, 9, 15, 14, 10, 13, 16, 12, 11),$$

which is a square as

$$(0, 1, 7, 6, 3, 5, 8, 4) \sim (0, 1, 6, 5, 2, 4, 7, 3) \sim (2, 9, 15, 14, 10, 13, 16, 12).$$

By construction, σ ends with the pattern $(3, 8, 5, 2, 4, 9, 1, 0, 7, 10, 6)$ which is right-crucial. Hence, any extension of σ to the left or to the right contains a square, and it only remains to check σ is square-free.

Let $S = (S_1; S_2)$ be a square in σ and note that S_1 has length at least 4. Since the middle section is constructed using the high-medium-low construction, it is square-free and it must be the case that S_1 contains one of the first 9 entries or that S_2 contains one of the last 11 entries.

Suppose first that S_1 contains some of the first 9 entries. If S_1 has length 4, then the square is contained entirely in the first 16 elements (the worst case being $(\sigma_8, \ldots, \sigma_{11}; \sigma_{12}, \ldots, \sigma_{15})$). However, we know

$$(\sigma_0, \ldots, \sigma_{15}) \sim (0, 6, 5, 2, 4, 7, 3, 1, 8, 14, 13, 9, 12, 15, 11, 10),$$

and this is square-free. Similarly, S_1 cannot be the first 8 entries. Hence, we can assume that S_1 contains σ_8 and has length at least 8.

If S_1 contains $(\sigma_5, \sigma_6, \sigma_7, \sigma_8)$, then it contains the pattern $(2, 1, 0, 3)$. The corresponding elements are $(\sigma_{j+5}, \sigma_{j+6}, \sigma_{j+7}, \sigma_{j+8})$ for some $4 \leq j \leq n/2$. Since $n > 44$, we have $n/2 + 8 < n - 11$ and the corresponding entries must come from the middle section, but the middle section follows the high-medium-low construction and does not contain any such pattern as the first, high entry is lower than the last medium entry. If S_1 does not contain $(\sigma_5, \sigma_6, \sigma_7, \sigma_8)$, then it must contain both σ_8 which is a medium element and σ_{11} which is a low element, but $\sigma_8 < \sigma_{11}$. The corresponding entries must again be in the middle section by our choice of n, but every low element is lower than any medium element by construction. This gives a contradiction and S_1 does not intersect the first 9 elements.

We now turn our attention to the case where S_2 intersects the last 11 entries and apply a very similar argument. Since $m \equiv 3 \bmod 4$ and π is a square-free permutation

of length m which begins down-down, π must end down-down. In particular, the last 18 entries in our construction are entirely determined and form the square-free pattern

$$(5, 4, 0, 3, 6, 2, 1, 10, 15, 12, 9, 11, 16, 8, 7, 14, 17, 13).$$

In particular, the last $11 + 7$ elements contain no square of length 4. This means we can assume that S_2 has length at least 8. It cannot be the case that S_2 contains the factor $(\sigma_{n-14}, \ldots, \sigma_{n-11})$ as this has the pattern $(2, 1, 0, 3)$ and the corresponding entries must be in the middle section where this pattern never occurs. Since the medium entry σ_{n-13} is lower than the low entry σ_{n-8}, S_2 cannot contain both σ_{n-13} and σ_{n-8}. Similarly, S_2 cannot contain σ_{n-12} and σ_{n-5}. This means that S_2 must be contained entirely within the last 11 entries and there is only one possible location for the square which is not contained in the last 18 entries (which we know do not contain a square). We must have

$$(S_1; S_2) = (\sigma_{n-19}, \ldots, \sigma_{n-12}; \sigma_{n-11}, \ldots, \sigma_{n-4}).$$

This does not form a square because $\sigma_{n-8} > \sigma_{n-5}$ while we have $\sigma_{n-16} < \sigma_{n-13}$. Hence, S_2 does not intersect with the last 11 entries and we are done. □

We now turn our attention to the even lengths for which we use a similar but more complicated construction. We will use different length suffixes depending on the parity of n modulo 8. The suffixes are also more complicated and we will have to split into 7 regions to guarantee that construction is square-free.

Theorem 5.5. *There exist bicrucial permutations for all even $n \geq 48$.*

Proof. We will assume that $n > 94$. When $n \leq 94$, this proof does not apply and instead we (again) use a computer to check the construction.

We split the space $\{0, \ldots, n-1\}$ into 7 different regions $R_0, \ldots, R_6$ where $R_i = [r_i, r_{i+1})$ for some specific choice of the values $0 = r_0 \leq r_1 \leq \cdots \leq r_6 \leq r_7 = n$ determined implicitly by the construction. Let $m = 32 + 2k - 1$ where $k \geq 5$ is equal to 1 mod 4. We will prove that there exist bicrucial permutations of lengths $m+7, \ldots, m+13$ by extending the same left-crucial permutation τ of length m. Since the permutations all begin and end with fixed patterns which we have chosen to be left-crucial and right-crucial respectively, the only difficulty will be ensuring that the permutations are square-free. First, we need to find an arbitrarily long square-free permutation which satisfies certain conditions on the endpoints.

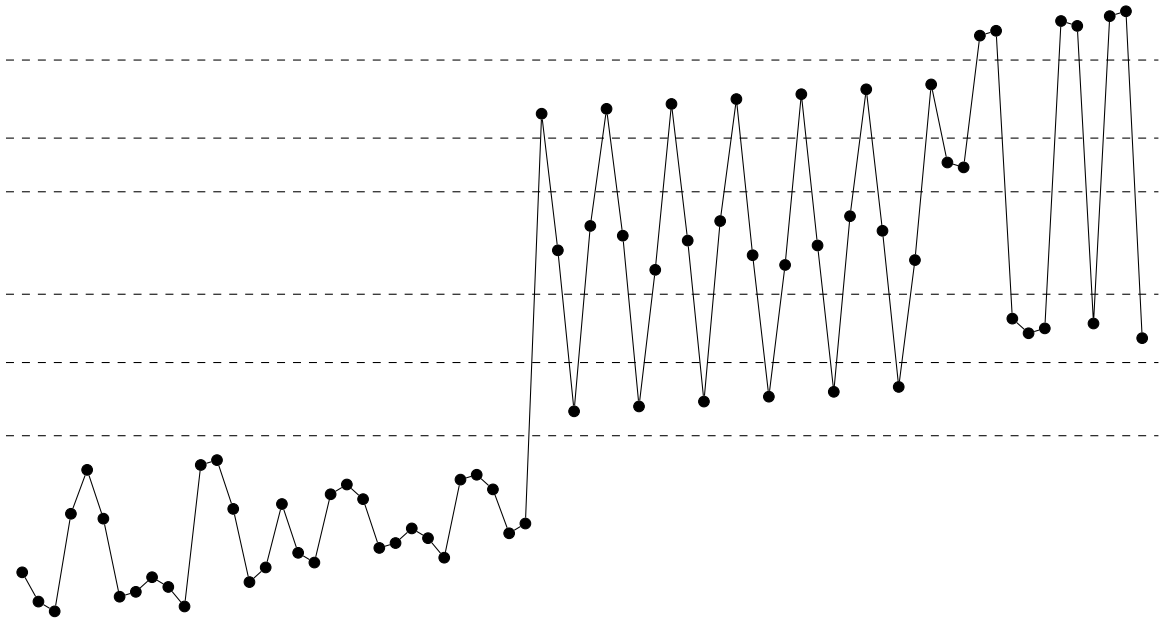


Figure 5.3: A bicrucial permutation of length 72 constructed as in the proof of Theorem 5.5. The dashed lines divide $\{0, \ldots, 71\}$ into the 7 different regions used in the construction.

Claim 5.1. *There exists a square-free permutation π of length k such that $\pi_{k-1} = k-1$ and $\max\{\pi_0, \pi_1, \pi_2\} = \pi_1$.*

Proof. Let π' be a square-free permutation of length $(k-1)/2$. We form the first $k-1$ entries of π by using a variation on the high-medium-low construction where we start with a 'medium' entry and then we go to a 'high' entry. We then append $k-1$. This is clearly a permutation where $\max\{\pi_0, \pi_1, \pi_2\} = \pi_1$ and $\pi_{k-1} = k-1$, so it remains to check that it is square-free. By construction any possible square must end at π_{k-1} and, since π follows the up-up-down-down construction, any square must be of length at least 8. In particular, the square must end in the pattern $(2, 1, 0, 3)$ which is not found anywhere else in the permutation. □

We now choose π as in Claim 5.1 and construct the permutation τ of length m. Start with the left-crucial permutation of $\{0, \ldots, 31\}$ given by

$$(8, 2, 0, 20, 29, 19, 3, 4, 7, 5, 1, 30, 31, 21, 6, 9, 22, 12, 10, 24, 26, 23, 13, 14, 17, 15, 11$$
$$27, 28, 25, 16, 18).$$

For $0 \leq i \leq 2k-2$, we define τ_{32+i} by

$$\tau_{32+i} = \begin{cases} r_5 + i/4 & i \equiv 0 \mod 4, \\ r_3 + \pi_{(i-1)/2} & i \equiv 1, 3 \mod 4, \\ r_1 + (i-2)/4 & i \equiv 2 \mod 4, \end{cases}$$

and we call this the *middle section* of σ. Note that so far we have only used the first $k-1$ entries of π and not π_{k-1}, but we will make use of the fact that the whole of π was square-free when we check that σ is square-free.

Let σ be formed by appending one of the following vectors, and denote the length of the chosen suffix by ℓ.

$$(r_4+1, r_4, r_6+2, r_6+4, r_6+1, r_6, r_6+3)$$
$$(r_4+1, r_4, r_6+4, r_6+6, r_6+3, r_6+1, r_6+2, r_6+6, r_6)$$
$$(r_4+1, r_4, r_6+3, r_6+4, r_6+2, r_6, r_6+1, r_6+7, r_6+6, r_6+5, r_6+8)$$
$$(r_4+1, r_4, r_6, r_6+1, r_2+4, r_2+1, r_2+2, r_6+3, r_6+2, r_2+3, r_6+4, r_6+5, r_2)$$

These have the patterns

$$(1,0,4,6,3,2,5)$$
$$(1,0,6,8,5,3,4,7,2)$$
$$(1,0,5,6,4,2,3,9,8,7,10)$$
$$(6,5,7,8,4,1,2,10,9,3,11,12,0)$$

respectively. While the top two of these are right-crucial, the last two are not. From our construction, we know the last 7 entries of τ except for τ_{m-6}, τ_{m-4} and τ_{m-2}. However, since $k \equiv 1 \bmod 4$ and π begins up-down, we know that π ends down-down-up and that $\tau_{m-6} > \tau_{m-4} > \tau_{m-2}$. Since these are the only entries in R_3 in the last 7 entries of τ and no entries in the suffixes are in R_3, this is enough to determine the pattern of the last $\ell+7$ entries of σ. The possible patterns are

$$(0,4,7,3,1,2,8,6,5,11,13,10,9,12)$$
$$(0,4,7,3,1,2,8,6,5,13,15,12,10,11,14,9)$$
$$(0,4,7,3,1,2,8,6,5,12,13,11,9,10,16,15,14,17)$$
$$(0,9,12,8,1,7,13,11,10,14,15,6,3,4,17,16,5,18,19,2),$$

and these are all right-crucial. Hence, any left or right extension of σ contains a square and it only remains to argue that σ is square-free.

First, note that by construction σ follows the up-up-down-down pattern so does not contain any squares of length 4 and any square must have length at least 8. Let $S = (S_1; S_2)$ be a square. Since the middle section was constructed using the high-medium-low construction and is square-free, it must either be the case that S_1 contains some of the first 32 entries, or that S_2 contains some of the last ℓ.

Let us first consider the case where S_1 contains some of the first 32 entries, and further that S_1 is contained entirely within the first 32 entries. It is easy to check that

the prefix does not contain the pattern $(3, 1, 0, 2)$ so it cannot possibly be the case that S_2 contains $(\tau_{32}, \dots \tau_{35})$. Hence, the square must be contained entirely within the first 35 entries, which is fixed for any choice of π and does not contain a square. Suppose instead that S_1 is not contained entirely within the first 32 entries. If S_1 contains $(\tau_{31}, \tau_{32}, \tau_{33}, \tau_{34})$, then S_2 must contain the pattern $(0, 3, 2, 1)$. However, since $n > 94$, the corresponding entries in S_2 must come from the middle section of σ and this never contains the pattern $(0, 3, 2, 1)$ as the first, medium element is lower than the last, low element. Hence, S_1 must end with either τ_{32} or τ_{33}.

Suppose S_1 ends at τ_{33}. It is easy to check that the square cannot be of length 8 as then S_1 would be order-isomorphic to $(0, 1, 3, 2)$ while S_2 is order-isomorphic to $(0, 2, 3, 1)$. Hence, the square must be of length at least 16 and contains the pattern $(0, 4, 5, 3, 1, 2, 7, 6)$, but the third, high entry is lower than the final, medium entry. This pattern never occurs in the middle section of σ, while our choice of n guarantees that the corresponding entries in S_2 come from the middle section. The case where S_1 ends at τ_{32} is similar. Explicitly, the square cannot be of length 8 as then S_1 is order-isomorphic to $(2, 0, 1, 3)$ while S_2 is order isomorphic to $(1, 0, 2, 3)$. If the square is of length at least 16, then S_1 contains the pattern $(1, 0, 5, 6, 4, 2, 3, 7)$ which cannot occur in the middle section as the initial, medium value is lower than the sixth, low entry. Hence, S_1 does not intersect the first 32 entries.

Suppose S_2 intersects with the last ℓ entries. First, consider the case that S_2 is entirely contained in the last ℓ entries. Since $\ell \leq 13$, we know S_2 has length at most 8. In particular, since we know the pattern of the last $\ell + 7$ entries and it is square-free, we know S_2 must be the first 8 entries of the suffix (and $\ell \geq 8$). However, this means we have one of the following patterns.

$$(1, 0, 5, 7, 4, 2, 3, 6)$$
$$(1, 0, 5, 6, 4, 2, 3, 7)$$
$$(4, 3, 5, 6, 2, 0, 1, 7)$$

In each of these patterns, there is a medium entry which is lower than a low entry so these patterns never occur in the middle section of σ and we get a contradiction.

Suppose instead that S_2 is not entirely contained in the last ℓ entries, but does intersect them. Then S_2 must contain at least one of the following.

$$(\tau_{m-1}, \tau_m, \tau_{m+1}, \tau_{m+2})$$
$$(\tau_{m-2}, \tau_{m-1}, \tau_m, \tau_{m+1})$$
$$(\tau_{m-3}, \tau_{m-2}, \tau_{m-1}, \tau_m)$$

From our construction, we know that

$$(\tau_{m-1}, \tau_m, \tau_{m+1}, \tau_{m+2}) = (r_6 - 1, r_4 + 1, r_4, r_6 + x)$$

for some choice of x. In particular, the last, medium entry is higher than the first, high entry. For the second case we note that $(\tau_{m-2}, \tau_{m-1}, \tau_m, \tau_{m+1}) = (r_3 + y, r_6 - 1, r_4 + 1, r_4)$ for some suitable choice of y, and the first, medium entry is lower than the last, low entry. Finally, we have the case in which S_2 contains $(\tau_{m-3}, \tau_{m-2}, \tau_{m-1}, \tau_m) = (r_3-1, r_3+y, r_6-1, r_4+1)$ for which we use the fact that $\pi_{k-1} = k-1$. The permutation $(\tau_{32}, \ldots, \tau_{m-1})$ follows the high-medium-low construction. If we continued using the high-medium-low construction, the entry τ_m would correspond to the entry π_{k-1}. Since we have chosen $\pi_{k-1} = k - 1$, τ_m would be smaller than all the high entries, but greater than every other entry in $\{\tau_{32}, \ldots, \tau_{m-1}\}$. In fact, this is exactly how we have chosen τ_m, and $(\tau_{32}, \ldots, \tau_m)$ follows the high-medium-low construction as well. Hence, σ is square-free. □

5.3 Bicrucial permutations for small even n

It was shown by Gent, Kitaev, Konovalov, Linton and Nightingale [49] that there are no bicrucial permutations of length n when n is even and $n < 32$, but that there is a bicrucial permutation when $n = 32$. Theorem 5.5 shows that there are bicrucial permutations of length n for all even $n \geq 48$, so to prove Theorem 5.3 it only remains to consider the cases where n is at least 34 and at most 46.

To find examples of bicrucial permutations for $n = 34$ it is enough to consider permutations which begin with the initial pattern used in the proof of Theorem 5.5 and to try different ways of extending the permutation to the right, ensuring it is always square-free and checking if the resulting permutation is right-crucial. When $n = 40$ or $n = 42$, it suffices to use the construction from Theorem 5.5 with the fixed permutation used in the computer search. To find the example for $n = 46$, we looked for permutations which begin with the prefix pattern from Theorem 5.5 and also end with the relevant suffix pattern, but we did not enforce any other conditions such as how the prefix and suffix interleave. There is no bicrucial permutation of length 36 which begins with the pattern from Theorem 5.5. Instead, we tried extending other left-crucial permutations of length 32 until we found the example given below. The permutations can be seen in Figure A.2 in Appendix A.2.

34 : (8, 2, 0, 21, 30, 20, 3, 4, 7, 5, 1, 31, 32, 22, 6, 9, 23, 12, 10, 25, 27, 24, 13, 14, 18, 15, 11, 28, 29, 26, 17, 19, 33, 16).

36 : (27, 33, 35, 22, 21, 26, 32, 31, 28, 30, 34, 19, 12, 20, 29, 25, 11, 18, 24, 9, 7, 10, 17, 16, 13, 15, 23, 4, 2, 5, 14, 8, 1, 3, 6, 0).

40 : (8, 2, 0, 20, 29, 19, 3, 4, 7, 5, 1, 30, 31, 21, 6, 9, 22, 12, 10, 24, 26, 23, 13, 14, 17, 15, 11, 27, 28, 25, 16, 18, 34, 33, 32, 37, 39, 36, 35, 38).

42 : (8, 2, 0, 20, 29, 19, 3, 4, 7, 5, 1, 30, 31, 21, 6, 9, 22, 12, 10, 24, 26, 23, 13, 14, 17, 15, 11, 27, 28, 25, 16, 18, 34, 33, 32, 39, 41, 38, 36, 37, 40, 35).

44 : (8, 2, 0, 20, 29, 19, 3, 4, 7, 5, 1, 30, 31, 21, 6, 9, 22, 12, 10, 24, 26, 23, 13, 14, 17, 15, 11, 27, 28, 25, 16, 18, 34, 33, 32, 38, 42, 37, 35, 36, 41, 40, 39, 43).

46 : (8, 2, 0, 25, 34, 24, 3, 4, 7, 5, 1, 35, 36, 26, 6, 9, 27, 12, 10, 29, 31, 28, 13, 14, 17, 15, 11, 32, 33, 30, 16, 23, 39, 38, 37, 40, 41, 22, 19, 20, 43, 42, 21, 44, 45, 18).

5.3.1 The non-existence of bicrucial permutations of length 38

To show that there are no bicrucial permutations of length 38 we used an exhaustive computer search which we describe in this section. The code can be found attached to the arXiv submission [54] on which this chapter is based.

Let σ be a bicrucial permutation of length $n = 38$. Then the *reverse* σ' of σ defined by $\sigma'_i = \sigma_{n-1-i}$ and the *complement* $\widetilde{\sigma}$ given by $\widetilde{\sigma}_i = n - 1 - \sigma_i$ are both also bicrucial. Since n is even and a square-free permutation must follow the up-up-down-down condition, either the start or end of σ must follow the pattern up-up or down-down. By taking the complement if necessary we can assume the pattern is up-up, and by reversing the permutation if necessary, we can assume the permutation begins up-up.

The up-up-down-down condition also means that any square in an extension of σ must either be of length 4 or a multiple of 8, and so every square in an extension of σ has length at most 32. Hence, the prefix of length 31 must be left-crucial. It is possible to enumerate all 180,319,733 left-crucial permutations of length 31 that begin up-up, but simply extending these to find all such permutations of length 38 and then checking if the extended permutation is right-crucial is too computationally expensive. Instead, we relax our condition by discarding k of the initial entries and try to extend

the remaining entries to a right-crucial permutation of length $38-k$. As long as $k \leq 7$, the extended permutations will have length at least 31 and therefore capture if the permutation can be right-crucial. Certainly, if there is a bicrucial permutation we will find its ending as a right-crucial permutation, but the existence of such a right-crucial permutation does not necessarily imply that there exists a bicrucial permutation of length 38 as we would still need to check for squares including the discarded entries. Fortunately, there turn out to be no such right-crucial permutations.

The main advantage in this method is that many of the left-crucial permutations end in the same pattern. Indeed, there are only 3,814,838 unique endings of length 24 from the 180,319,733 left-crucial permutations of length 31 that begin up-up. We are free to choose the length of the prefixes we generate and the value of k, although we only consider the case where $k = 7$. There are only 1,052,494,945 left-crucial permutations of length 32, a relatively modest increase of only 5.8 times. It turns out that nearly 80% (844,852,238) of these are in fact bicrucial and there are only 1,350,082,610 left-crucial permutations of length 33 (which begin up-up), an increase of only 1.3 times. This pattern does not hold and there are 34,609,211,226 such permutations of length 34, an increase of 25.6 times. We therefore generate all left-crucial permutations of length 33 which begin up-up and discard the first 7 entries of each to get 24,763,327 unique endings, each of which we try to extend to a right-crucial permutation of length 31. Since we find no such extensions, there cannot be a bicrucial permutation of length 38.

We generate permutations using a depth-first search. The children of a permutation σ of length n are the permutations of length $n+1$ which begin with the pattern σ. This is easily done in practice. Given a permutation σ of length n, the first child is formed by appending the entry n, and to move from the ith child to the next child we simply swap the entries $n-i+1$ and $n-i$. Once we reach the last child, we can step backwards by removing the last entry and decrementing each entry.

The advantage of this method of generating permutations is that we can easily prune unnecessary subtrees. For example, when we first visit a permutation we check if it contains a square, in which case we do not need to consider any children of the permutation and we can prune the subtree starting at this permutation.

When we first visit a permutation σ of length $n \geq 3$ we also calculate a crude lower bound on the length of a left-crucial permutation τ which begins with the given pattern (if such a permutation exists), and we can reject any permutation where we know τ is of length at least 39. To calculate the bound we consider extending the permutation σ to the left by prepending the entry i for each choice of i and checking

if the extended permutation $\sigma^{(i)}$ contains a square. If it does, we move on to $i+1$. If there is not a square, we look for a "partial square" in the permutation and compute a lower bound based on completing this square. Let $(n+1)/2 < k < n$ be the smallest multiple of 4 such that $(\sigma_0^{(i)}, \ldots, \sigma_{k-1}^{(i)})$ (obtained by extending $(\sigma_0, \ldots, \sigma_{k-2})$ to the left by prepending i) is order-isomorphic to $(\sigma_k^{(i)}, \ldots, \sigma_n^{(i)})$ (where we extend the definition of order-isomorphic to permutations of different lengths by replacing the larger permutation by its prefix of the appropriate length). Then the shortest square-free permutation beginning with σ which contains a square when prepending i is of length $2k-1$, and we can improve the lower bound to $2k-1$. If there is no such k, then the first part S_1 of a square (S_1, S_2) must contain all of $\sigma^{(i)}$ and any square created by prepending i must be of length at least $2m$ where $m = 4\lceil (n+1)/4 \rceil$ is the smallest multiple of 4 at least $n+1$. This gives a lower bound of $2m-1$.

Suppose there are $i \neq j$ which both fall into this last case. The squares created by prepending i and j cannot both be of length $2m$, as there is a single value for S_2, the second half of the square, (for the given m) and it would need to be order-isomorphic to two different choices of S_1. Hence, we can improve the lower bound to $2(m+4)-1$. More generally, we can count the number of times t that S_1 must contain all of σ', and we get the lower bound $2m + 8(t-1) - 1$.

Let us illustrate how we calculate the lower bound with an example. Let $\sigma = (0, 4, 5, 2, 1, 3)$. We first consider appending the entry 0 to get the permutation $(0, 1, 5, 6, 3, 2, 4)$, which has a square $(0, 1; 5, 6)$ and does not improve the lower bound. If we append the entry 1, we get the permutation $(1, 0, 5, 6, 3, 2, 4)$. This does not contain a square of length 4, but $(1, 0, 5) \sim (3, 2, 4)$, and we get the lower bound 7. Prepending 2, 3 or 4 works similarly, and also give the lower bound 7. Consider appending the entry 5 to get $\sigma^{(5)} = (5, 0, 4, 6, 2, 1, 3)$. The only possible choice for k is 4, but $(5, 0, 4) \not\sim (2, 1, 3)$. Hence, if τ is square-free but contains a square when prepending 4, then the length of τ is at least 15. There is also no viable k when prepending 6, and so we get the improved lower bound 23.

5.3.2 Enumerating square-free permutations

The code used to exhaustively generate all left-crucial permutations in the previous section can also be used to enumerate square-free, left-crucial and bicrucial permutations with only small modifications, and the results are shown in Table 5.4. If σ is square-free (or bicrucial), then the reverse, the complement and the reverse complement of σ are also square-free (resp. bicrucial). When n is even, this means we can

count the number of square-free permutations by counting the square-free permutations which begin up-up and multiplying by 4. When n is odd, there is not such a nice classification and it is harder to handle the reverse of the permutation. To avoid this we check only that each permutation in the depth-first search is square-free until the permutation is of length at least $n/2$, and only then do we consider exploiting the symmetry in the problem. Note that any extra work done generating square-free permutations of length $n/2$ is negligible when compared with the work to generate the permutations of length n. When appending an element to grow the permutation to length $(n+1)/2+k$, we consider the factor

$$\left(\sigma_{(n+1)/2-k}, \ldots, \sigma_{(n+1)/2+k}\right) \sim \pi$$

and check if π, the reverse of π, the complement of π or the reverse complement of π is lexicographically least. If π is not jointly lexicographically least, then we prune the permutation from our search. If π is the unique minimum, then we do not perform this check again when appending new entries. Multiplying the number of permutations found this way by 4 over counts permutations σ where the reverse and the complement are equal, and we must subtract off twice the number of such permutations (once for the reverse and once for the reverse complement).

Enumerating left-crucial permutations is slightly simpler: only the complement of a left-crucial permutation need be left-crucial, so we can handle the symmetry by only counting the left-crucial permutations which begin by going up and multiplying this result by 2.

n	Square-free	Left-crucial	Bicrucial
1	1	0	0
2	2	0	0
3	6	0	0
4	12	0	0
5	34	0	0
6	104	0	0
7	406	60	0
8	1 112	140	0
9	3 980	518	54
10	15 216	1 444	0
11	68 034	8 556	0
12	312 048	31 992	0
13	1 625 968	220 456	69 856
14	8 771 376	984 208	0
15	53 270 068	7 453 080	2 930 016
16	319 218 912	39 692 800	0
17	2 135 312 542	289 981 136	40 654 860
18	14 420 106 264	1 467 791 790	0
19	109 051 882 344	14 316 379 108	162 190 472
20	815 868 128 288	86 001 855 074	0
21	6 772 099 860 398	949 804 475 890	312 348 610 684
22	56 501 841 264 216	6 494 842 788 046	0
23	519 359 404 861 294	73 636 377 696 714	29 202 730 580 288

Table 5.4: The number of square-free, left-crucial and bicrucial permutations of length n. These are sequences A221989, A221990 and A238935 respectively in the OEIS [90].

5.4 Open problems

Recently, the notion of bicrucial square-free words has been extended to the notion of *extremal words*, words which are square-free but inserting any letter in any position introduces a square [55], and this naturally extends to *extremal permutations*. Since square-free permutations must follow the up-up-down-down pattern, any insertion which breaks this pattern must introduce a square and so any insertion except in positions 0, 1, $n-1$ or n introduces a square. Extremal permutations were studied by Gent, Kitaev, Konovalov, Linton and Nightingale in [49] where they showed that there exist small extremal permutations. In particular, they showed that there exist extremal permutations of lengths 17 and 21 but not for any other $n \leq 22$, and they conjectured that arbitrarily long extremal permutations exist. Since one only needs to consider inserting an element close to the beginning or end as in bicrucial permutations, it seems likely that an approach similar to that used in Theorem 5.4 and Theorem 5.5 may give arbitrarily long extremal permutations.

Conjecture 5.6 (Conjecture 10 in [49])**.** *There exist arbitrarily long extremal square-free permutations.*

It was shown by Grytczuk, Kordulewski and Niewiadomski [55] that there are arbitrarily long extremal square-free words over an alphabet of size 3, and the n for which there exist extremal words of length n was completely classified Mol and Rampersad [79]. In particular, there are extremal ternary words for all $n \geq 87$. There are no known extremal words over an alphabet of size 4 (or greater), and it is conjectured that none exist.

Conjecture 5.7 (Conjecture 12 in [55])**.** *There are no extremal square-free words over an alphabet of size 4.*

It was recently shown that there are no extremal square-free words over an alphabet of size 17 (or greater) [59], but the question remains open for smaller alphabets.

Another interesting problem suggested by Kitaev [67] is to replace squares by higher powers. A *kth power* in a permutation $\sigma \in S_n$ is a factor

$$(S_1; S_2; \ldots; S_k) = (\sigma_s, \ldots, \sigma_{s+\ell-1}; \ldots; \sigma_{s+(k-1)\ell}, \ldots, \sigma_{s+k\ell-1})$$

with $\ell \geq 2$ and $S_i \sim S_j$ for all $i, j \in [k]$. The definition of bicrucial with respect to squares easily extends to bicrucial with respect to containing a kth power. That is, we say the permutation σ is *bicrucial with respect to containing a kth power* if σ

does not contain a kth power, but any left or right extension of σ does contain a kth power.

Problem 5.8. *Are there arbitrarily long permutations which are bicrucial with respect to containing a kth power?*

In a slightly different direction, there are tight twins. A *tight twin* $T = (T_1, T_2)$ of length 2ℓ is a pair of disjoint subpermutations T_1 and T_2 of length ℓ contained in a block of length 2ℓ such that $T_1 \sim T_2$. That is, we relax the condition that the two halves of the square must be consecutive and instead require that the entries of the twin are consecutive. These were studied by Dudek, Grytczuk and Ruciński [37] where they showed that, for every $n \geq 1$, there exists a permutation avoiding tight twins of length at least 26. It may be possible to improve this further, even showing that there are arbitrarily long permutations avoiding tight twins of length at least 6. Every permutation of length 6 contains a tight twin of length 4, but computer searches have found reasonably long ($n \geq 36$) permutations avoiding tight twins of length at least 6.

Problem 5.9 (Problem 4.2 in [37])**.** *What is the minimum k such that there are arbitrarily long permutations avoiding tight twins of length at least k? Is it possible to avoid tight twins of lengths at least 6?*

Recently, Dudek, Grytczuk and Ruciński [38] introduced the study of multiple tight twins. A *tight r-twin* $T = (T_1, T_2, \ldots T_r)$ of length $r\ell$ is a set of r permutations of length ℓ contained in a block of length $r\ell$ such that $T_i \sim T_j$ for all $i, j \in [r]$. There are several interesting open questions related to this, particularly the following.

Problem 5.10 (Problem 7.1 in [38])**.** *Is there an absolute constant c such that there are arbitrarily long permutations which avoid tight r-twins of length $r\ell$ for all $r \geq c$?*

Problem 5.11 (Problem 7.4 in [38])**.** *Let $f(r, \ell)$ be the minimum n such that all permutations of length n contain a tight r-twin of length $r\ell$. For $r \geq 4$, is $f(r, 2) = \infty$? If not, what are the asymptotics of $f(r, 2)$?*

Chapter 6

Decomposing permutations into twins

In this chapter we consider a second problem on patterns in permutations. In order to keep the chapter self-contained, we again recall the relevant definitions. Given a vector $a = (a_1, \ldots, a_n)$ consisting of n distinct real numbers, one can associate a unique permutation $\sigma \in S_n$ (which we write as a vector $(\sigma(1), \ldots, \sigma(n))$) by replacing the ith smallest element of a with i, and we call this σ the *pattern* of a. We say two vectors a and b are *order-isomorphic* if they have the same pattern, and in this case we write $a \sim b$.

In this chapter we will be interested in decomposing permutations into subpermutations which are order-isomorphic. We say that two permutations σ and π are *ℓ-similar* if they can be decomposed into ℓ subpermutations $\sigma^{(1)}, \ldots, \sigma^{(\ell)}$ and $\pi^{(1)}, \ldots, \pi^{(\ell)}$ such that $\sigma^{(i)} \sim \pi^{(i)}$ for all $i \in [\ell]$, and we denote the least such ℓ by $U(\sigma, \pi)$. Clearly, there are pairs of permutations (such as the identity and the reverse permutation) which are only n-similar, but what happens on average?

Problem 6.1 (Problem 4.3 in [37])**.** *Let σ and π be two permutations of length n chosen independently and uniformly at random. What is the expected value of $U(\sigma, \pi)$?*

As we saw in Section 1.4.2, an upper bound of $O(\sqrt{n})$ follows from considering the longest increasing sequence in a random permutation. We also saw that the longest order-isomorphic subpermutations between two uniformly random permutations is $\Theta(n^{2/3})$ with high probability [48, 25], and so there is a lower bound of $\Omega(n^{1/3})$. The main result of this section is the following theorem which shows that $O\big(n^{1/3} \log^{11/6}(n)\big)$ subpermutations suffice, and the easy lower bound is tight up to a polylogarithmic factor.

Theorem 6.2. *Let σ and π be two permutations of length n chosen independently and uniformly at random. Then σ and π are $O\big(n^{1/3}\log^{11/6}(n)\big)$-similar with probability at least $1 - o(n^{-1})$.*

The following bound on $U(\sigma, \pi)$ is immediate.

Corollary 6.3. *Let σ and π be two independent, uniformly random permutations of length n. Then*

$$\mathbb{E}[U(\sigma, \pi)] = O\Big(n^{1/3}\log^{11/6}(n)\Big).$$

Chapter outline We give the proof of Theorem 6.2 in Section 6.1. We discuss some open problems in Section 6.2.

This chapter is based on joint work with Carla Groenland, Dániel Korándi, Alexander Roberts, Alex Scott and Jane Tan.

6.1 Proof of Theorem 6.2

Let us start with an outline of the proof.

We will generate the two random permutations σ and π from random points in the unit square $[0, 1]^2$. Given n uniformly random points in the unit square $[0, 1]^2$, ordered by their x-coordinate, the y-coordinates form a uniformly random permutation. We define the permutation σ from a set of n "red" points and the permutation π from a set of n "blue" points. If the two sets of points are independent and uniformly distributed, then the two random permutations σ and π are also independent and uniformly distributed.

We begin by choosing a matching between the two sets of points, and we only consider decompositions into subpermutations where the matched points are corresponding entries in the decomposition. We split $[0, 1]^2$ into an $M \times M$ grid with equally spaced rows and columns (where $M \approx n^{2/3}\log^{-1/3}(n)$). Given a subset of the red points such that there is at most one point in each row and column, the pattern formed by this subpermutation of σ is determined entirely by the squares each point is in, and we need not know exactly where each point is. We will only look for subpermutations where this is the case and we can "forget" the exact locations of the points. Suppose we have such a subset of points P determining a subpermutation of σ. The subpermutation of π determined by the matched blue points has the same pattern if the squares of the matched points form the same pattern as the squares for the points P. This could mean the squares are exactly the same and trivially

have the same pattern, or maybe the squares are all offset by the same amount, or even something more complicated. We will only consider the second case and we will label the red points based on the relative position (measured in grid squares) of the matched blue point. We will treat each class of labels separately, and so we will want to avoid having too many different labels. To this end, we use a minimax matching between the two sets of points, and bound (with high probability) the maximum distance between two matched points using the following result of Leighton and Shor.

Theorem 6.4 (Leighton, Shor [71])**.** *Suppose A and B are two independent sets of n points uniformly distributed in $[0,1]^2$. Then there is a function $\alpha = \alpha(n) = \Omega(\sqrt{\log(n)})$ and an absolute constant C such that, with probability at least $1 - n^{-\alpha}$, there is a perfect matching between A and B for which the maximum distance between two matched points is at most $C \log^{3/4}(n)/\sqrt{n}$.*

For each distinct label, we will split the points with the given label into sets where each set contains at most one point from each row and column. This can be done by computing an edge colouring of an auxillary bipartite multigraph. Indeed, let G be the graph where we have a vertex for each row and a vertex for each column, and for each point we add an edge between its row and its column. Two edges share an end vertex if and only if the two corresponding points are in the same row or column. Hence, we wish to partition the edges into sets such that no two edges in the same set share an end vertex, and this is exactly a proper edge colouring. We therefore want to bound the edge-chromatic number of these bipartite multigraphs for which it suffices to upper bound the number of points in different regions. If we were to simply generate n points uniformly at random, then the number of points in different regions is not independent, even if the regions are disjoint. Instead, we will generate points using a Poisson process with mean $2n$. This is highly likely to generate more than n points, and so we will delete points at random until there are only n points remaining, which leaves us with n uniformly distributed points. However, we can bound the number of our n points in a given region by the number of points of the Poisson process in that region, and this is independent for disjoint regions.

The idea of using an auxiliary bipartite multigraph has been used by Dudek, Grytczuk and Ruciński [39] and later by Bukh and Rudenko [25] to study twins in a single random permutation. The idea of sampling random permutations using a Poisson process also appears in [25].

As well as the result of Leighton and Shor above, we will also make use of the following well-known bounds on the tail probabilities of the binomial and Poisson distributions.

Lemma 6.5 (Follows from Theorem 4.4 in [78]). *Let $X \sim \mathrm{Bin}(n,p)$ and $\varepsilon > 0$. Then*

$$\mathbb{P}(X \geq (1+\varepsilon)np) \leq \exp\left(-\frac{\varepsilon^2 np}{2+\varepsilon}\right).$$

Lemma 6.6 (Theorem 5.4 in [78]). *Let $X \sim \mathrm{Poi}(\lambda)$ and $x > \lambda$. Then*

$$\mathbb{P}(X \geq x) \leq \frac{(e\lambda)^x e^{-\lambda}}{x^x}.$$

Proof of Theorem 6.2. Given n points $\{(a_i, b_i) : i \in [n]\}$ in the unit square $[0,1]^2$ such that $a_1 < a_2 < \cdots < a_n$, we can define a permutation σ by setting $\sigma(i) = j$, where b_i is the jth smallest element of the set $\{b_1, \ldots, b_n\}$. If the n points are chosen uniformly at random in the unit square, this process generates a uniformly random permutation.

Let R be a 2D Poisson process with rate $2n$ and let N_R denote the number of points of R in $[0,1]^2$. Then $N_R \sim \mathrm{Poi}(2n)$ and

$$\mathbb{P}(N_R \leq n) \leq \exp\left(-\frac{n^2}{2n}\right).$$

Conditional on the number of points in $[0,1]^2$ being $k \geq n$, the k points are distributed uniformly. After removing $k-n$ points uniformly at random, we have n uniformly distributed points, which we call the *red points*, and we can use this to get a uniformly random permutation $\sigma \in S_n$. Similarly, we can use an independent Poisson process B to define a set of n uniformly distributed *blue points* and a corresponding permutation $\pi \in S_n$.

We now match up the n red points with the n blue points, using Theorem 6.4.

Let $M = \left\lceil n^{2/3} \log^{-1/3}(n) \right\rceil$ and split the square $[0,1]^2$ into an $M \times M$ grid with evenly spaced rows and columns. Let $S_{i,j}$ denote the square in the ith row and jth column. That is,

$$S_{i,j} = \left[\frac{j-1}{M}, \frac{j}{M}\right) \times \left[\frac{i-1}{M}, \frac{i}{M}\right).$$

Suppose that we have a set of red points such that no two red points are in the same row or column. Then the pattern of the corresponding subpermutation is entirely determined by the squares, and we do not need to know the particular points. If the squares for the matched blue points follow the same pattern, then the red points and blue points correspond to order-isomorphic subpermutations of σ and π.

To find such sets of points, we will look for matchings in which all edges have the same label in a suitable edge-labelled bipartite multigraph G constructed as follows. Let G have vertex set $\{i_1, \ldots, i_M\} \cup \{j_1, \ldots, j_M\}$. Given a red point r in row r_{row} and column r_{col} matched with a blue point b in row b_{row} and column b_{col}, add an edge

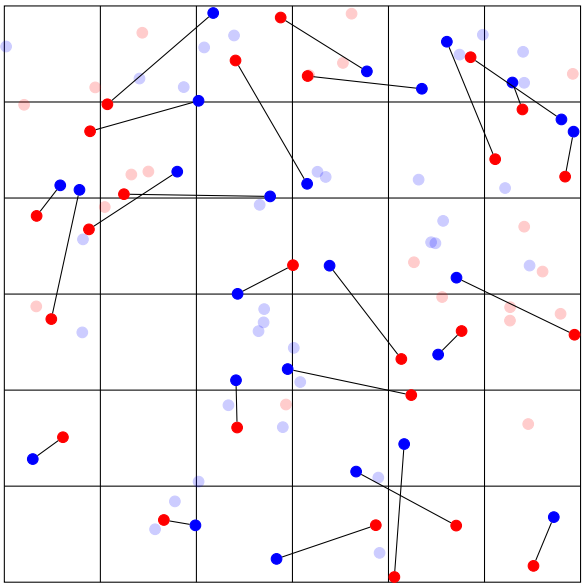


Figure 6.1: The square $[0,1]^2$ split into an $M \times M$ grid. The red and blue points have been sampled from a Poisson point process with mean 50 and 25 of each have been selected uniformly at random to form the red and blue permutations respectively. The selected points have been matched using a minimax matching.

from row $i_{r_{\text{row}}}$ to column $j_{r_{\text{col}}}$ in the graph and label it $(b_{\text{row}} - r_{\text{row}},\, b_{\text{col}} - r_{\text{col}})$, i.e. label it with the relative position of b from r. Note that a matching in G represents red points in different rows and columns, so a matching in which all edges have the same label, as explained above, corresponds to order-isomorphic subpermutations of σ and π. It will therefore suffice to decompose the edges of G into such matchings. Equivalently, we wish to separately properly edge-colour the edges for each class of labels. Note that as each matched edge has length at most $C\log^{3/4}(n)/\sqrt{n}$, there are at most

$$\left(2M\frac{C\log^{3/4}(n)}{\sqrt{n}} + 2\right)^2 = O\left(\frac{M^2\log^{3/2}(n)}{n}\right) \tag{6.1}$$

possible labels. We will handle each label separately.

It is well-known that a bipartite multigraph can be edge-coloured using Δ colours where Δ is the maximum degree of the graph. In fact, this can be done efficiently (see e.g. [86]). We bound the degree in two steps: first we bound the number of distinct edges labelled (c_1, c_2) that a vertex sees, then we bound the maximum number of repetitions of an edge.

There is an edge from row i to column j of label (c_1, c_2) if and only if there is a red point in the square $S_{i,j}$ matched to a blue point in the square $S_{i+c_1,j+c_2}$. In particular, this means that there must be a point from the Poisson process R in $S_{i,j}$

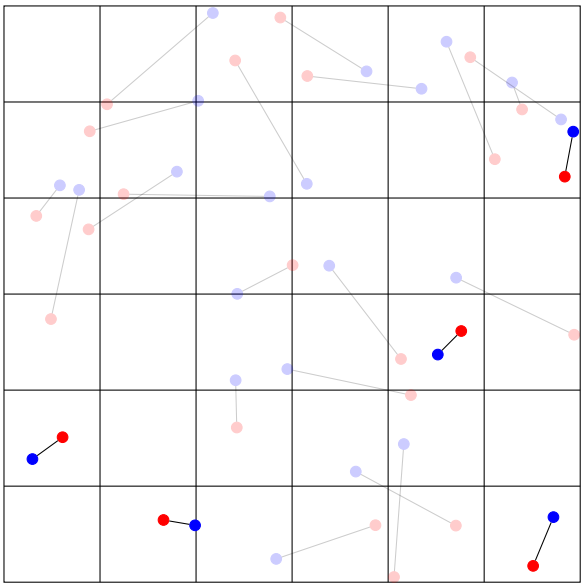


Figure 6.2: A sample minimax matching along with the bipartite graph for the colour $(0, 0)$. The 5 edges can be coloured using 2 colours, and so the subpermutations of length 5 can be decomposed into 2 parts.

and from the process B in the square $S_{i+c_1,j+c_2}$. This happens with probability

$$\left(1 - \exp\left(-\frac{2n}{M^2}\right)\right)^2 \leq \frac{4n^2}{M^4}.$$

As the number of points of a Poisson process in a particular square is independent of the numbers in the other squares, the number of distinct edges labelled (c_1, c_2) that are incident with a given vertex in G, is dominated by a binomial random variable with M trials and success probability $4n^2/M^4$. Let $X \sim \text{Bin}(M, 4n^2/M^4)$. Then

$$\mathbb{E}[X] = 4n^2/M^3 \leq \frac{4n^2}{\left(n^{2/3}\log^{-1/3}(n)\right)^3} = 4\log(n).$$

We also have

$$4n^2/M^3 \geq \frac{4n^2}{\left(n^{2/3}\log^{-1/3}(n) + 1\right)^3} \geq 3\log(n)$$

for $n \geq 64$. Hence, using a Chernoff bound (see Lemma 6.5),

$$\begin{aligned}
\mathbb{P}(X \geq 12\log(n)) &\leq \mathbb{P}\left(X \geq 12n^2/M^3\right) \\
&\leq \exp\left(-4n^2/M^3\right) \\
&\leq \exp(-3\log(n)) \\
&= n^{-3}.
\end{aligned}$$

There are $2M$ vertices in G and there are $O\big(M^2 \log^{6/4}(n)/n\big)$ labels (see (6.1)), so the probability that any vertex is incident with at least $12\log(n)$ distinct edges of the same label is at most

$$n^{-3} \cdot 2M \cdot O\left(\frac{M^2 \log^{3/2}(n)}{n}\right) = O\left(\frac{\log^{1/2}(n)}{n^2}\right) = o(n^{-1}).$$

We now look to bound the number of times a particular edge appears in the graph. For there to be 8 edges from row i to column j, the square $S_{i,j}$ must contain at least 8 points of the Poisson process R. The number of such points in any given square is a Poisson random variable with mean

$$2n/M^2 \leq \frac{2\log^{2/3}(n)}{n^{1/3}}.$$

Hence, using the bound from Lemma 6.6, the probability that there are at least 8 points in $S_{i,j}$ is at most

$$\frac{(2e)^8 \log^{16/3}(n)}{8^8 n^{8/3}} \exp\left(-\frac{2\log^{2/3}(n)}{n^{1/3}}\right) \leq \frac{\log^{16/3}(n)}{n^{8/3}}.$$

Taking the union bound over all M^2 squares, the probability that any square contains 8 points is at most $O\big(\log^{14/3}(n)/n^{4/3}\big) = o(n^{-1})$.

Consider the subgraph of G formed by the edges with label (c_1, c_2). With high probability, every edge is repeated at most 7 times and every vertex sees at most $12\log(n)$ distinct edges. Hence, the maximum degree is bounded by $84\log(n)$, and so we can cover the edges with $84\log(n)$ matchings. We can do this separately for every distinct label and cover all edges in at most

$$O\left(\frac{M^2 \log^{6/4}(n)}{n}\right) \cdot 84\log(n) = O\Big(n^{1/3}\log^{11/6}(n)\Big)$$

matchings, each consisting of edges with the same label as required. □

We remark that this proof gives a simple, efficient algorithm to decompose two permutations σ and π into twins, although the decomposition may not be optimal. Given a permutation σ of length n, generate n red points in $[0,1]^2$ by first sampling n values $x_1, \ldots, x_n$ uniformly in $[0,1]$ to be the x coordinates and then sampling n values $y_1, \ldots, y_n$ uniformly in $[0,1]$ to be the y coordinates. By relabelling the values as necessary, we can assume that $x_1 < \cdots < x_n$ and $y_1 < \cdots < y_n$. Let the n red points in $[0,1]^2$ be $(x_i, y_{\sigma(i)})$, and similarly choose n blue points in $[0,1]^2$ for

π. We then find a minimax matching between the red and blue points. This can easily be done in $O\big(n^{5/2}\log(n)\big)$ time by using a binary search over the maximum allowed edge weight (from the $\binom{n}{2}$ possibilities) and checking if there is a perfect matching using the allowed edges in $O(n^{5/2})$ time [60]. Crudely, there are at most n possible labels, and for each label we need to edge-colour a bipartite multigraph with at most n edges. Each colouring can be done in $O(n\log(n))$ time [86, 3]. The above process decomposes the permutations into order-isomorphic subpermutations in $O\big(n^{5/2}\log(n)\big)$ time. Although the decomposition produced may not be optimal, the proof of Theorem 6.2 shows that the algorithm uses on average $O\big(n^{1/3}\log^{11/6}(n)\big)$ subpermutations, which is not far from the simple lower bound of $\Omega(n^{1/3})$.

6.2 Open problems

Theorem 6.2 is tight up to the polylogarithmic factor and it would be interesting to know if some polylogarithmic factor is necessary, or if two uniformly random permutations are $O(n^{1/3})$-similar. Even if a polylogarithmic factor is necessary, it is likely that the $\log^{11/6}(n)$ we have shown can be improved. For example, we have used a minimax matching which guarantees that the number of colours is small and assumed that every colour is equally bad, but it is likely that some of the colours have fewer edges and the edges can be covered using fewer matchings. Indeed, if we replace the minimax matching with a matching that minimises transportation distance, the average distance between a red point and the matched blue point is $\Theta\big(\log^{1/2}(n)/\sqrt{n}\big)$, saving a factor of $\log^{1/4}(n)$. This suggests that most edges should be spread over fewer colours and the polylogarithmic factor can be improved.

Problem 6.7. *Are two permutations of length n chosen independently and uniformly at random $O(n^{1/3})$-similar with high probability?*

We can also ask what the best bound is when we wish to decompose k permutations, as was asked for decomposing graphs. We say a set of permutations $\{\sigma_1,\dots,\sigma_k\}$ is *ℓ-similar* if each permutation σ_j can be split into ℓ subpermutations $\sigma_j^{(1)},\dots,\sigma_j^{(\ell)}$ such that $\sigma_j^{(i)}\sim\sigma_{j'}^{(i)}$ for all choices of i, j and j'. What is the minimum $\ell_k(n)$ such a set of k permutations of length n chosen independently and uniformly at random are ℓ_k-similar? There is again a simple upper bound of $O_k(\sqrt{n})$ which follows from decomposing the permutations into increasing subpermutations. Indeed, as we argued when considering two permutations, each permutation can be split into at most $3\sqrt{n}$ increasing subpermutations with high probability. By splitting at the end points of

every subpermutation across all k permutations, we have simultaneously decomposed the k permutations into at most $3k\sqrt{n}$ increasing subpermutations, and the set of permutations is $O_k(\sqrt{n})$-similar.

Problem 6.8. *Let $\{\sigma_1, \ldots, \sigma_k\}$ be a set of permutations of length n chosen independently and uniformly at random. What is the minimum $\ell_k(n)$ such that the set is ℓ_k-similar with high probability?*

The other main problem related to (general) twins in permutations is to determine the maximum size of a twin that one is guaranteed to find in a permutation of length n. The best lower bound, due to Bukh and Rudenko [25], is currently $\Omega(n^{3/5})$, while the best upper bound is $O(n^{2/3})$ obtained by applying a first moment calculation to a random permutation [48].

Problem 6.9 ([48])**.** *Does every permutation of length n contain a twin of length $\Omega(n^{2/3})$?*

Chapter 7

Uniform shuffling with lazy transpositions

Let S_n be the symmetric group on n elements. Write $(a, b) \in S_n$ for the transposition that swaps a and b, and $1 \in S_n$ for the identity permutation. A *lazy transposition* $T = (a, b, p)$ is the random permutation

$$T = \begin{cases} (a, b) & \text{with probability } p, \\ 1 & \text{otherwise.} \end{cases}$$

We will view a permutation $\sigma \in S_n$ as n distinguishable counters $c_1, \ldots, c_n$ sitting in n positions $1, \ldots, n$ by saying σ corresponds to having the counter c_i in position $\sigma(i)$. A transposition (a, b) acts by switching the counters currently in positions a and b, so a lazy transposition (a, b, p) switches the counters currently in positions a and b with probability p, and does nothing with probability $1 - p$. A *transposition shuffle* is a sequence of independent lazy transpositions such that the composition $T_1 \cdots T_\ell$ is uniformly distributed over S_n, or equivalently, a sequence which shuffles the counters so they are in a uniformly random order order. Let $U(n)$ denote the minimum length ℓ of a transposition shuffle over S_n. There are several known constructions giving transposition shuffles using $\binom{n}{2}$ lazy transpositions, and the main problem is to determine if this the correct value of $U(n)$.

Problem 7.1 (Fitzsimons [45], Angel and Holroyd [5])**.** *Does* $U(n) = \binom{n}{2}$ *for all* n*?*

The problem can also be generalised to the case where we only have $t < n$ counters. Instead of starting with n counters, one in each position, we start with counters in the positions $1, \ldots, t$ and leave the rest empty. A transposition (a, b) switches any counters in positions a and b, including switching a counter with an empty space. A sequence of lazy transpositions naturally defines a random variable by the positions

of the counters at the end of sequence, and we can again ask that this be uniformly distributed over all $t!\binom{n}{t}$ choices of t ordered positions. If the output is uniformly distributed, we say the sequence achieves *t-uniformity on n points* and call the sequence a *(t, n)-shuffle* (so a transposition shuffle is an (n, n)-shuffle). Let $U_t(n)$ denote the minimal number of lazy transpositions in a (t, n)-shuffle. The special case of shuffling all elements is $U_n(n) = U(n)$. Problem 7.1 naturally generalises to the following.

Problem 7.4. *Does $U_t(n) = tn - \binom{t+1}{2}$?*

In this chapter we introduce a new divide and conquer construction which gives better constructions for all $t \geq 3$ and $n \geq 6$, and improves on the best known bounds by a constant factor.

Theorem 7.3. *There exists an $\varepsilon > 0$ such that $U(n) \leq (1 - \varepsilon)\binom{n}{2}$ for all $n \geq 6$. Moreover,*

$$U(n) \leq \frac{2}{3}\binom{n}{2} + O(n \log(n)).$$

Theorem 7.5. *There exists a constant $\varepsilon > 0$ such that*

$$U_t(n) \leq (1 - \varepsilon)\left(tn - \binom{t+1}{2}\right)$$

for all $3 \leq t \leq n$ and $n \geq 6$. Moreover, for a fixed $t = 3\ell + r$, we have

$$U_t(n) \leq (2\ell + r)n + O_\ell(\log(n)).$$

We also consider the problem of reachability. Let $(T_i)_{i=1}^{\ell} = T_1, \ldots, T_\ell$ be a sequence of lazy transpositions and suppose that $T_i = (a_i, b_i, p_i)$ for $i = 1, \ldots, \ell$. We say the sequence $(T_i)_{i=1}^{\ell}$ achieves *t-reachability* if, for every choice of t distinct ordered points $x_1, \ldots, x_t$, there is some subsequence $T_{i_1}, \ldots, T_{i_{\ell'}}$ such that the composition $(a_{i_1}, b_{i_1}) \cdots (a_{i_{\ell'}}, b_{i_{\ell'}})$ maps $i \mapsto x_i$ for every $i \leq t$. If $(T_i)_{i=1}^{\ell}$ is a (t, n)-shuffle, then the probability the counters end in positions $x_1, \ldots, x_t$ is exactly $\left(t!\binom{n}{t}\right)^{-1}$, and there must be at least one subsequence which reaches the positions $x_1, \ldots, x_t$. Hence, lower bounds on t-reachability provide lower bounds on t-uniformity, and this approach actually provides the best known bound on $U(n)$. It is straightforward to construct $(1, n)$-shuffles using $n-1$ lazy transpositions, and 1-reachability shows that this is best possible. Motivated by these lower bounds, we consider the problem of 2-reachability on n points, and exactly determine the minimum number of transpositions. However, the lower bound is still a long way from the best known upper bounds and whether $U_2(n) = 2n - 3$ remains open.

Theorem 7.7. *Let $n \geq 2$. The smallest sequence of transpositions achieving 2-reachability on n points has length $\lceil 3n/2 \rceil - 2$.*

We also give a simple randomised construction which shows that t-reachability can be achieved using $(2 + o_t(1))n$ transpositions. Surprisingly, the coefficient of the leading term is independent of t, which contrasts sharply with the current best known upper bounds for t-uniformity from Theorem 7.5.

Theorem 7.8. *Let $t \geq 3$. There is a sequence of transpositions of length at most $(2 + o_t(1))n$ which achieves t-reachability on n points.*

Another approach to lower bounds is to first consider transposition shuffles with a particular structure and then show that a general transposition shuffle cannot be much better. For example, if every transposition is of the form $(1, \cdot)$ then one might hope it is easier to prove a lower bound, and yet such a shuffle requires at most $3U(n)$ transpositions. We call such transpositions *star transpositions* since they match the edges of a star graph. The constructions in Section 7.1 show that $\binom{n}{2}$ star transpositions suffice for n-uniformity, and one might be tempted to suggest that this should be optimal. Unfortunately, the following theorem shows that this is not the case.

Theorem 7.9. *Let $t = 4\ell + r$ for some $0 \leq r < 3$. Let $\widetilde{U}_t(n)$ be the minimum length of a (t, n)-shuffle in which every transposition is a star transposition. Then*

$$\widetilde{U}_t(n) \leq \left(\frac{7\ell}{2} + r\right)n + O_\ell(\log(n)).$$

Moreover,

$$\widetilde{U}(n) \leq \frac{7}{8}\binom{n}{2} + O(n \log(n)).$$

Clearly, n-reachability can still be achieved in $O(n \log(n))$ star transpositions and a modification of Theorem 7.7 shows that the minimum number of transpositions needed for 2-reachability is again around $3n/2$.

Theorem 7.10. *The shortest transposition sequence achieving 2-reachability on n points in which each transposition is a star transposition, contains $\lceil 3(n-1)/2 \rceil$ transpositions.*

However, using the restrictive nature of the transpositions, we are able to show that achieving 2-uniformity is strictly harder than achieving 2-reachability, the first result of this kind.

Theorem 7.11. *Any sequence of lazy star transpositions achieving 2-uniformity on n points has length at least $1.6n - C$ for some absolute constant C.*

Sorting networks A closely related problem is that of sorting networks, where the n counters are labelled $1, \ldots, n$, transpositions are replaced by comparators and the goal is to ensure a sorted output. A *comparator* between a pair of positions $a < b$ containing counters x and y switches the counters if $x > y$ and does nothing if $x \leq y$. A *sorting network* is a fixed sequence of comparators such that, for every initial ordering, the counters are sorted in descending order at the end of the network. The problem of minimising the number of comparators has been studied at least since 1954 when Armstrong, Nelson and O'Connor obtained the US Patent 3029413 for some examples of optimal networks when $n \leq 8$. After a long sequence of improvements over many years the asymptotics were determined up to a constant factor in 1983 by Ajtai, Komlós and Szemerédi [1] who showed there are sorting networks using $O(n \log(n))$ comparators, matching the trivial counting lower bound. For a more detailed history and discussion of sorting networks see e.g. [69].

One variation of a sorting network is a selection network, where the goal is to select the t largest counters in order, and these are similar to (t, n)-shuffles. Formally, a *t-selection network* is a sequence of comparators such that, for any input order of counters labelled $1, \ldots, n$, the count $n - i$ ends in position $i + 1$ for $0 \leq i \leq t - 1$. The fewest comparators in a 1-selection network is $n - 1$, and for 2-selection the best possible is $2n - 3$. The situation is already more complicated for 3-selection networks, but was shown to be $2n + \log_2(n) + O(1)$ by Yao in 1980 [93]. Yao also showed that the minimum number of comparators in a t-selection network is asymptotically $\lceil \log_2(t + 1) \rceil n + o(n)$. For more information on selection networks and their variants see e.g. [69].

Chapter outline In Section 7.1 we give some existing constructions which achieve the trivial bound and will be useful throughout the rest of the chapter. In Section 7.2 we give our new construction and prove Theorem 7.3 and Theorem 7.5. We prove Theorem 7.7, settling the case of 2-reachability, in Section 7.3. In Section 7.4 we consider transposition shuffles using star transpositions and prove Theorem 7.9. We also provide a separation between 2-reachability and 2-uniformity when restricting to star transpositions, which follows from Theorem 7.10 and Theorem 7.11. We finish with some open problems in Section 7.5.

This chapter is based on joint work with Carla Groenland, Jamie Radcliffe and Alex Scott.

7.1 Preliminaries

In this section we give simple "sweeping" and "divide and conquer" constructions for (t, n)-shuffles, all of which give upper bounds of $U_t(n) \leq tn - \binom{t+1}{2}$. Combining these constructions gives the following lemma mentioned (but not proven) by Angel and Holroyd [5] which we prove at the end of this section.

Lemma 7.12 ([5]). *Suppose that $U(n_0) \leq (1-\varepsilon)\binom{n_0}{2}$. Then there exists $\delta > 0$ such that*

$$U(n) \leq (1-\delta)\binom{n}{2}$$

for all $n \geq n_0$. Explicitly, we may take

$$\delta = \frac{n_0 - 1}{4n_0}\varepsilon.$$

We start by giving a simple sweeping construction which extends a $(t, n-1)$-shuffle to a (t, n)-shuffle by adding t transpositions. Note that there are other variations of the sweeping construction, and the variation here was chosen as all transpositions are star transpositions (provided an appropriate $(t, n-1)$-shuffle is chosen).

Lemma 7.13 ([45, 5]). *Let $n \geq t+1$. Then*

$$U_t(n) \leq U_t(n-1) + t.$$

Proof. We give a construction which extends any $(t, n-1)$-shuffle to a (t, n)-shuffle.

1. Start with the transpositions $(1, 2, 1/2), (1, 3, 1/3), \ldots, (1, t, 1/t)$ and the transposition $(1, n, t/n)$.
2. Append a $(t, n-1)$-shuffle.

Let the t distinguishable counters be $c_1, \ldots, c_t$. After the first $t-1$ transpositions, the counter in position 1 is uniformly distributed amongst the t possible counters. With probability t/n, the tth transposition "fires" and a counter is placed in position n, with each counter appearing there with probability $1/n$ as required. The remaining $t-1$ counters are in the first t positions so are then uniformly distributed by the $(t, n-1)$-shuffle. If the tth transposition does not fire, the $(t, n-1)$-shuffle distributes the t counters uniformly over the positions $1, \ldots, n-1$, and this happens with probability $(n-t)/n$ as required. □

Clearly, shuffling $n-1$ counters over n positions is the same as shuffling all n counters, so this construction gives the bound

$$U(n) = U_{n-1}(n) \leq n - 1 + U_{n-1}(n-1). \tag{7.1}$$

This alone can be used to show $U(n) \leq \binom{n}{2}$, and starting with any such (t,t)-shuffle gives the trivial upper bound

$$U_t(n) \leq \binom{t}{2} + t(n-t) = tn - \binom{t+1}{2}. \tag{7.2}$$

As well as a sweeping construction which introduces extra positions, there is a simple divide and conquer construction which introduces extra counters. Substituting in the trivial bound (7.2) for $U_{t-m}(n-m)$ and $U_m(n)$ in this construction, gives the trivial bound for $U_t(n)$.

Lemma 7.14. *Let $0 \leq m \leq t$. Then*

$$U_t(n) \leq U_{t-m}(n-m) + U_m(n)$$

Proof. We construct a (t,n)-shuffle in two stages. In the first stage, shuffle the counters in positions $m+1,\ldots,t$ across the positions $m+1,\ldots,n$. Then shuffle the counters in positions $1,\ldots,m$ across all n positions $1,\ldots,n$. It is easiest to see this works by considering the reverse of the sequence. The reverse of the second stage chooses m (ordered) positions from all n options uniformly at random, while the first stage then chooses another $t-m$ from the remaining $n-m$ options. Hence, together they choose t ordered positions from all n options (uniformly at random), and the sequence forms a (t,n)-shuffle. □

The last construction in this section is another divide and conquer approach which gives the bound $U_t(n) \leq 2U(t) + t + U_t(n-t)$. In particular, combining this with Lemma 7.14, gives the upper bound

$$U(n) \leq 2U(t) + 2U(n-t) + t. \tag{7.3}$$

Lemma 7.15 ([45, 5])**.** *Let $t \leq n/2$. Then*

$$U_t(n) \leq 2U(t) + t + U_t(n-t).$$

Sketch proof. We give a construction with 3 stages.

1. Uniformly shuffle the counters in positions $1,\ldots,t$.

2. Append the transpositions $(1, t+1, p_1), \ldots, (t, 2t, p_t)$ where the probabilities $p_1, \ldots, p_t$ are chosen such that

$$\prod_{i=1}^{t}(1+p_i x) = \sum_{i=0}^{t} \frac{\binom{n-t}{i}\binom{t}{t-i}}{\binom{n}{t}}(1+x)^i. \tag{7.4}$$

   Such probabilities were shown to exist by Hui and Park [61].

3. Uniformly shuffle the positions $1, \ldots, t$ and, separately, uniformly shuffle the positions $t+1, \ldots, n$.

Since the counters are shuffled uniformly in stage 1, the order of the counters after this stage is uniformly distributed. This means that stages 2 and 3 only need to uniformly distribute indistinguishable counters. In stage 2 we move some counters from the positions $1, \ldots, t$ to the positions $t+1, \ldots, n$. Stage 3 then ensures that, given there are i counters across the positions $t+1, \ldots, 2t$ (and therefore $t-i$ counters across the positions $1, \ldots, t$), every set of positions for the counters is equally likely. In order to make sure every possible set of positions for the t counters is equally likely, we only need to ensure the probability that i counters move in stage 2 is correct. There are $\binom{n-t}{i}$ ways of distributing the i counters that have moved across and $\binom{t}{t-i}$ ways of distributing the $t-i$ counters which have not moved, so the probability that i counters move should be $\binom{n-t}{i}\binom{t}{t-i}/\binom{n}{t}$, and follows a hypergeometric distribution. If W is the (random) number of counters moved in stage 2, it is straightforward to recover the condition (7.4) by considering $\mathbb{E}\big[(1+x)^W\big]$. □

Using these constructions we now prove Lemma 7.12.

Proof of Lemma 7.12. Let $n = 2^a n_0 + b$ where $a, b \in \mathbb{N}$ and $0 \leq b < 2^a n_0$. Start with an (n_0, n_0)-shuffle using $U(n_0) \leq (1-\varepsilon)\binom{n_0}{2}$ transpositions and apply (7.3) to see that

$$U(2n_0) \leq 4U(n_0) + n_0 \leq \binom{2n_0}{2} - 4\varepsilon\binom{n_0}{2}.$$

Applying (7.3) a further $a-1$ times, we find

$$U(2^a n_0) \leq \binom{2^a n_0}{2} - 4^a\varepsilon\binom{n_0}{2}.$$

Next, apply (7.1), the sweeping construction, a total of b times to get an (n, n)-transposition shuffle using at most

$$\binom{n}{2} - 4^a\varepsilon\binom{n_0}{2} = \left(1 - \frac{4^a\varepsilon n_0(n_0-1)}{n(n-1)}\right)\binom{n}{2}$$

transpositions. For a fixed value of a, the worst constant is obtained when $b = 2^a n_0 - 1$ and substituting in $n = 2^{a+1} n_0 - 1$ gives

$$U(n) \leq \left(1 - \frac{4^a \varepsilon n_0 (n_0 - 1)}{(2^{a+1} n_0 - 1)(2^{a+1} n_0 - 2)}\right) \binom{n}{2}.$$

The result now follows since

$$\frac{4^a n_0 (n_0 - 1)}{(2^{a+1} n_0 - 1)(2^{a+1} n_0 - 2)} \geq \frac{n_0 - 1}{4 n_0}$$

for all $n_0 \geq 1$. □

We remark that we have not tried to optimise the constant δ in this lemma and it may be possible to improve it by combining the constructions above in a better way. However, any bound from such a lemma will always be quite crude as it only takes into account a single improved construction for a single value of n_0.

7.2 A new divide and conquer construction

We start by outlining and verifying a divide and conquer strategy which produces a (t, mk)-shuffle by splitting the n positions into m groups of k positions each (where $k \geq t$). As in the previous constructions, it does not specify every transposition and instead focuses on the conditions which need to be satisfied at the end of each stage. To prove the first part of Theorem 7.3, one can fill in this outline for a suitable choice of the parameters t, m and k, and then apply Lemma 7.12. In Section 7.2.1, we do this in the simplest case where $m = 2$ and $t = k$, which gives several efficient $(k, 2k)$-shuffles. To prove Theorem 7.5 and the second part of Theorem 7.3, we will use the construction to generate efficient $(3, n)$-shuffles. Using Lemma 7.14 we can build efficient transposition shuffles by repeatedly mixing in 3 counters at a time.

Construction 7.1 is similar to the construction used in the proof of Lemma 7.15, particularly when $m = 2$ and $t = k$. In this case, the main difference is the introduction of a fourth stage which shuffles counters between the groups. This final stage allows us to relax condition (7.4) in the previous construction. If W is the number of counters that move between the two groups, then condition (7.4) is equivalent to

$$\mathbb{P}(W = w) = \frac{\binom{k}{w}\binom{k}{k-w}}{\binom{2k}{k}}.$$

Indeed, the proof of Lemma 7.15 uses this form of the condition, and (7.4) follows by considering $\mathbb{E}\big[(1+x)^W\big]$. Using Construction 7.1, we can replace (7.4) by the weaker

condition

$$\mathbb{P}(W = w) + \mathbb{P}(W = k - w) = \frac{2\binom{k}{w}\binom{k}{k-w}}{\binom{2k}{k}},$$

at the expense of some extra transpositions later on. This will allow us to use fewer transpositions when k is small.

Construction 7.1. *Let $n = mk$ for two integers $m, k \geq 2$ and fix $t \leq k$. The construction consists of four stages.*

1. *Uniformly shuffle the t counters in positions $1, \ldots, t$.*

2. *Perform any sequence of lazy transpositions which satisfies the following condition: let W_i be the number of counters in positions $(i-1)k+1, \ldots, ik$ at the end of the chosen sequence, and let $W = (W_1, \ldots, W_m)$; then*

$$\sum_{\sigma \in S_m} \mathbb{P}\big(W = (w_{\sigma(1)}, \ldots, w_{\sigma(m)})\big) = m! \frac{\prod_{i=1}^{m} \binom{k}{w_i}}{\binom{n}{t}} \tag{7.5}$$

   for every choice of $(w_1, \ldots, w_m)$ with $w_1 + \cdots + w_m = t$.

3. *For each $i = 1, \ldots, m$, uniformly shuffle the counters in positions $(i-1)k+1$, $(i-1)k+2, \ldots, ik$ across the positions $(i-1)k+1, (i-1)k+2 \ldots, ik$.*

4. *For each $j = 1, \ldots, k$, uniformly shuffle the counters in positions $j, k+j$, $\ldots, (m-1)k+j$ across the positions $j, k+j, \ldots, (m-1)k+j$.*

We remark that the sequence of independent lazy transpositions is fixed, so stages 3 and 4 cannot be changed depending on the previous stages or the locations of the counters. However, this does not present a major problem; we can use an arbitrary minimum (a, b)-shuffle (for a suitable choices of a and b) to shuffle the subset of positions which could possibly contain a counter (over the relevant positions). The main difficulty is choosing transpositions for stage 2, and in fact, it is not necessarily possible to choose a suitable sequence of transpositions for every choice of t, m and k. We now show that any sequence of transpositions following this construction gives a valid (t, mk)-shuffle, and we will show that we can make suitable choices for each stage in the later sections.

Theorem 7.16. *If a sequence $T_1, \ldots, T_\ell$ of transpositions satisfies the conditions in Construction 7.1, then the sequence forms a (t, mk)-shuffle.*

Proof. After stage 1, we may view the counters as indistinguishable. Indeed, stage 1 defines a uniformly random permutation σ and stages 2, 3 and 4 define a (random) injective map F from $\{1,\dots,t\}$ to $\{1,\dots,n\}$. The probability that the counters end in the ordered positions $(v_1,\dots,v_t)$ is given by

$$\sum_{f:\mathrm{Im}(f)=\{v_1,\dots,v_t\}} \mathbb{P}(F=f)\mathbb{P}(f^{-1}(v_i)=\sigma(i) \text{ for all } i),$$

and $\mathbb{P}(f^{-1}(v_i)=\sigma(i) \text{ for all } i) = 1/t!$. Hence, the probability only depends on the set $\{v_1,\dots,v_t\}$, and we may view the counters as indistinguishable.

We now show that every set of positions appears with equal probability. Denote the random number of counters in positions $j, k+j, \dots, (m-1)k+j$ at the start of stage 4 by $\widetilde{W}_j$, and let $\widetilde{W} = \left(\widetilde{W}_1,\dots,\widetilde{W}_k\right)$. Since the last step shuffles the counters across $j, k+j, \dots, (m-1)k+j$ uniformly for every j, any two patterns which have the same number of counters in the positions $j, k+j, \dots, (m-1)k+j$ for every j, occur with equal probability. Hence, to ensure that every set of positions appears with equal probability we need

$$\mathbb{P}\Big(\widetilde{W} = (\widetilde{w}_1,\dots,\widetilde{w}_k)\Big) = \frac{\prod_{i=1}^k \binom{m}{\widetilde{w}_i}}{\binom{n}{t}}$$

for all $\widetilde{w} = (\widetilde{w}_1,\dots,\widetilde{w}_k)$ such that $\widetilde{w}_1 + \dots + \widetilde{w}_k = t$.

By construction, the distribution of $\widetilde{W}$ is entirely determined by the positions of the counters at the end of stage 2. In fact, by our choice of stage 3, it is determined by the vector $W = (W_1,\dots,W_n)$ where W_i is number of counters across the positions $(i-1)k+1,\dots,ik$ at the end of stage 2. Further, due to the symmetry of the m sets of positions, the distribution is determined by the unordered multiset $\{W_i : i = 1,\dots,m\}$. This means, we can guarantee that $\widetilde{W}$ has the correct distribution, and therefore that every set of t positions appears with equal probability, provided the unordered multiset has the correct distribution. That is, we require

$$p_w := \mathbb{P}\big(W \in \{(w_{\sigma(1)},\dots,w_{\sigma(m)}) : \sigma \in S_m\}\big) \tag{7.6}$$

to have the correct value for all $w = (w_1,\dots,w_m)$ such that $w_1 + \dots w_m = t$.

We claim that this is implied by the condition (7.5). To do so we calculate the correct value for p_w by considering the case where we have the stronger condition that

$$\mathbb{P}(W = (w_1,\dots,w_m)) = \frac{\prod_{i=1}^m \binom{k}{w_i}}{\binom{n}{t}}$$

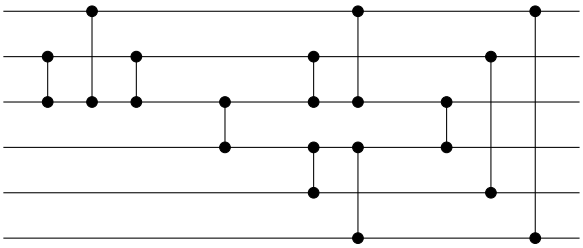


Figure 7.1: The $(3, 6)$-shuffle with just 11 transpositions given in the proof of Lemma 7.17.

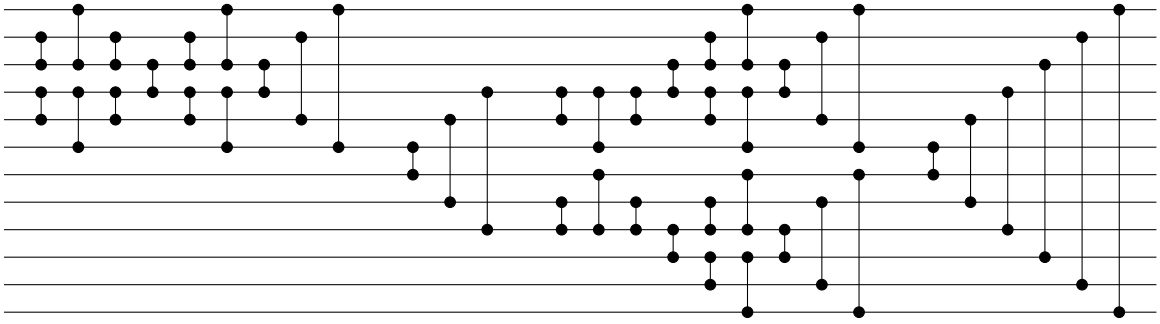


Figure 7.2: The $(6, 12)$-shuffle with just 45 transpositions given in the proof of Lemma 7.17. Note the repeated use of the efficient $(3, 6)$-shuffle from Figure 7.1.

for every w such that $w_1 + \cdots + w_m = t$. After shuffling the counters in stage 3, the counters are uniformly distributed over all n positions. Clearly, applying any lazy transposition to the uniform distribution also gives the uniform distribution, and the counters are also uniformly distributed at the end of stage 4. This implies that we should take

$$p_w = \left|\{(w_{\sigma(1)}, \ldots, w_{\sigma(m)}) : \sigma \in S_n\}\right| \frac{\prod_{i=1}^{m} \binom{k}{w_i}}{\binom{n}{t}},$$

and this is equivalent to the condition (7.5). □

7.2.1 Splitting into two groups

Construction 7.1 only gives conditions on the output of each stage and it does not specify the exact transpositions. The following lemma shows how to use this construction in the simple case where $m = 2$ and $t = k$ to get improved bounds on $U_k(2k)$ for some small choices for k. Examples for $U_3(6)$ and $U_6(12)$ constructed as in Lemma 7.17 are shown (without probabilities) in Figure 7.1 and Figure 7.2 respectively.

Lemma 7.17. *The following upper bounds hold.*

1. $U_3(6) \leq 11$,

2. $U_4(8) \leq 22$,

3. $U_5(10) \leq 31$,

4. $U_6(12) \leq 45$.

Proof. All of these bounds are found by considering Construction 7.1 in the case where $m = 2$ and $t = k$, and are therefore shuffles by Theorem 7.16.

Let $n = 2k$. In stage 1, use any transposition shuffle on k elements with $U(k)$ lazy transpositions. In stage 2, we will use the lazy transpositions $T_i = (k - i + 1, k + i, p_i)$ for $i = 1, \ldots, \lfloor k/2 \rfloor$. The condition (7.5) becomes

$$\mathbb{P}(W_2 = w) + \mathbb{P}(W_2 = k - w) = \frac{2\binom{k}{w}^2}{\binom{2k}{k}}$$

for $w = 0, \ldots, \lfloor k/2 \rfloor$, and this can be further simplified to

$$\mathbb{P}(W_2 = w) = \begin{cases} \dfrac{2\binom{k}{w}^2}{\binom{2k}{k}} & w \neq k/2, \\ \dfrac{\binom{k}{w}^2}{\binom{2k}{k}} & w = k/2. \end{cases} \tag{7.7}$$

To compute the values of p_i, we use the same approach as Hui and Park in [61] and consider the generating function $\mathbb{E}\big[(1+x)^{W_2}\big]$. As the sum of independent Bernoulli random variables we have

$$\mathbb{E}\big[(1+x)^{W_2}\big] = \prod_{i=1}^{\lfloor k/2 \rfloor} (1 + p_i x).$$

We may also write

$$\mathbb{E}\big[(1+x)^{W_2}\big] = \sum_{w=0}^{\lfloor k/2 \rfloor} \mathbb{P}(W_2 = w)(1+x)^w,$$

where the probabilities are as given in (7.7). We can compare the roots of these polynomials to find the values of p_i. For example, in the case $k = 5$ we have

$$(1 + p_1 x)(1 + p_2 x) = \frac{126 + 225x + 100x^2}{126},$$

so comparing roots we get

$$\left\{ -\frac{1}{p_1}, -\frac{1}{p_2} \right\} = \left\{ -\frac{6}{5}, -\frac{21}{20} \right\},$$

and we can take $p_1 = 5/6$ and $p_2 = 20/21$.

In stage 3, use two copies of a minimal $(\lfloor k/2 \rfloor, k)$-shuffle, one to shuffle the positions $k - \lfloor k/2 \rfloor + 1, \ldots, k$ into $1, \ldots, k$ and one to shuffle the positions $k+1, \ldots, k+\lfloor k/2 \rfloor$ into $k+1, \ldots, 2k$. Finally, use the transpositions $(k, k+1, 1/2), \ldots, (1, 2k, 1/2)$ in stage 4. This gives the bound (for $k = 3, 4, 5, 6$)

$$U_k(2k) \leq U(k) + \lfloor k/2 \rfloor + 2U_{\lfloor k/2 \rfloor}(k) + k,$$

and substituting in appropriate values will complete the proof. In the cases $k = 3$, $k = 4$ and $k = 5$ we do not know of any $(\lfloor k/2 \rfloor, k)$-shuffles which do better than the trivial upper bound. For the case $k = 6$, we can use the efficient $(3, 6)$-shuffle given by this proof, which only uses 11 transpositions. □

Applying Lemma 7.14 in the case $t = n = 2k$ and $m = k$, we get

$$U(2k) \leq U(k) + U_k(2k).$$

Substituting in the values from the lemma above and the best known values of $U(k)$ immediately gives the following corollary, and the first part of Theorem 7.3 follows from applying Lemma 7.12.

Corollary 7.18. *The following upper bounds hold.*

1. $U(6) \leq 14$,
2. $U(8) \leq 28$,
3. $U(10) \leq 41$,
4. $U(12) \leq 59$.

Unfortunately, the probabilities required in stage 2 do not appear to exist for $k \geq 7$ and the construction does not immediately extend. However, using the transpositions $T_i = (k - i + 1, k + i, p_i)$ only for $i = 1, \ldots, \lfloor k/2 \rfloor$ is the best one could possibly hope for. If instead we replaced $\lfloor k/2 \rfloor$ with k, then we get a less efficient shuffle but we can find the probabilities. In fact, we obtain the divide and conquer approach in Lemma 7.15, but with the addition of stage 4. Using a computer search, we have found constructions where we can use $\lfloor k/2 \rfloor < j < k$ transpositions for $7 \leq k \leq 18$, and some of these lead to efficient constructions. However, these are less efficient than the constructions we consider in the next section and we do not take this any further. Of course, there is no need to restrict to disjoint transpositions, and it may be possible to find shuffles using overlapping transpositions.

7.2.1.1 More positions than counters

We have already seen an improved bound for $U_3(6)$, and combining this with Lemma 7.13 shows that $U_3(n) \leq 3n-7$ for all $n \geq 6$, beating the upper bound by an additive factor. Applying Lemma 7.14 with $m = 3$ shows that $U_t(n) \leq U_3(n) + U_{t-3}(n-3)$, and we can again beat the trivial upper bound by an additive factor for all $t \geq 3$ and $n \geq 6$. In order to improve this to a constant factor as required by Theorem 7.5, we will use Construction 7.1 in the case where $k > t$ and $m = 2$ to construct efficient $(3, n)$- and $(4, n)$-shuffles when n is large. The difficulty is again in finding suitable lazy transpositions for use in stage 2, and we again use the simplest possible construction. This works for any k when shuffling 3 or 4 counters, but when shuffling 5 or 6 counters, the probabilities only exist for very small values of k.

Suppose we are shuffling 3 counters. When $k = 3$, the positions $1, 2$ and 3 can be split into 2 groups, the positions 1 and 2 which are guaranteed to contain a counter and the position 3 which may or may not contain a counter. We use a $(1, 3)$-shuffle to shuffle the position 3 into positions $1, 2, 3$. When $k > 3$, there are no longer only 2 kinds of position. There are 2 positions which definitely contain a counter, a position which may or may not contain a counter and then $k-3$ positions which definitely do not contain a counter. We could use a $(3, k)$-shuffle to shuffle the positions $1, 2, 3$ into $1, 2, \ldots, k$, but this is slightly stronger than necessary as there are 2 counters that are already indistinguishable. Indeed, if we can choose an optimal $(3, k)$ shuffles which starts by uniformising the two indistinguishable counters, we can remove this initial transposition. To capture this slight improvement, we will consider the minimum number of transposition to shuffle t indistinguishable counters over n positions, which we denote by $\widehat{U}_t(n)$.

Clearly, we can create a (t, n)-shuffle from a shuffle which only shuffles t indistinguishable counters over n positions by first shuffling the t counters amongst themselves, which takes $U(t)$ transpositions. Hence, we get the bound

$$U_t(n) \leq U(t) + \widehat{U}_t(n).$$

This means efficient shuffles for indistinguishable counters immediately lead to efficient shuffles for distinguishable counters.

Lemma 7.19. *If $k \geq 3$, then*

$$\widehat{U}_3(2k) \leq 2k + 2 + \widehat{U}_3(k),$$

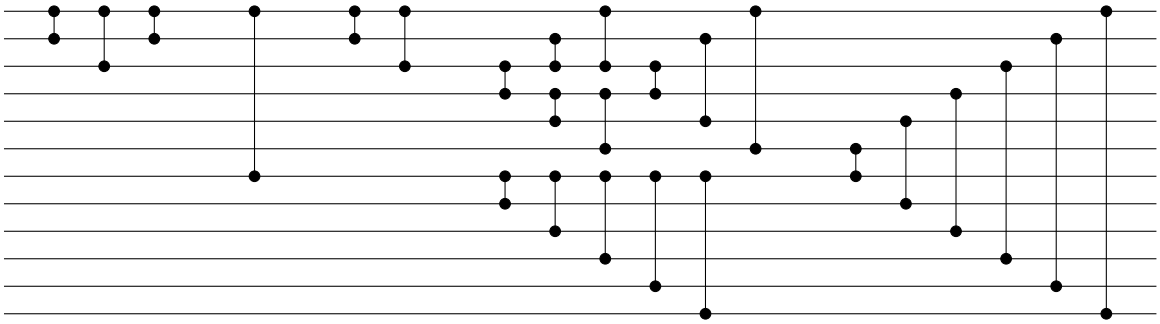


Figure 7.3: A $(3, 12)$-shuffle using just 25 transpositions constructed as in Lemma 7.19.

and if $k \geq 4$,

$$\widehat{U}_4(2k) \leq 3k + 4 + \widehat{U}_4(k).$$

Proof. We proceed much as in the proof of Lemma 7.17 and again use Construction 7.1. Since we are starting with indistinguishable counters, stage 1 is unnecessary and we start with stage 2. We only consider the case where $t = 4$ since the case $t = 3$ is similar, although slightly easier.

Start stage 2 with the transpositions $(1, k+1, p_1)$ and $(2, k+2, p_2)$. The condition (7.7) becomes

$$\mathbb{P}(W_2 = w) = \begin{cases} \frac{2\binom{k}{w}\binom{k}{t-w}}{\binom{2k}{t}} & w \neq 2, \\ \frac{\binom{k}{w}^2}{\binom{2k}{t}} & w = 2. \end{cases}$$

for $w = 0, \ldots, 2$, and so we are interested in the roots of the polynomial

$$\mathbb{E}\big[(1+x)^W\big] = 1 + \frac{k(5k-7)}{(2k-1)(2k-3)}x + \frac{3k(k-1)}{2(2k-1)(2k-3)}x^2.$$

This has determinant

$$\frac{k(k-2)(k^2+4k-9)}{(2k-3)^2(2k-1)^2}$$

which is positive for all $k > 2$. Hence, the polynomial has 2 roots, and they must lie in $(-\infty, -1]$ as the polynomial is clearly positive for $x \geq -1$. If the two roots are r_1 and r_2, take $p_1 = -1/r_1$ and $p_2 = -1/r_2$.

In stage 3, we first shuffle the positions 1 and 2 into the positions $1, 2, 3, 4$ using $U_2(4) = 5$ transpositions. Given there are w counters left in the positions $1, 2, 3, 4$,

they are in a uniformly random order and in uniformly random positions. Hence, we can view the positions as containing 4 indistinguishable counters and shuffle the counters across the positions $1, 2, \ldots, k$ using $\widehat{U}_4(k)$ transpositions. Shuffle any counters in positions $k+1$ and $k+2$ across the positions $k+1, \ldots, 2k$ using $U_2(k)$ transpositions.

Finally, we end with the k transpositions $(k, k+1, 1/2), \ldots, (1, 2k, 1/2)$ in stage 4. This shuffles 4 indistinguishable counters across $2k$ positions using

$$7 + \widehat{U}_4(k) + U_2(k) + k$$

transpositions, and the bound follows using $U_2(k) \leq 2k - 3$. □

When shuffling 5 or 6 counters this construction can be used up to $k = 6$ and $k = 8$ respectively, giving shuffles which beat the trivial bounds. However, more efficient shuffles follow from the other constructions and the lemmas in Section 7.1.

If we do not worry about the small savings obtained by noting that some positions with counters are already indistinguishable, we can replace stage 3 with a (t, k)-shuffle and a $(\lfloor t/2 \rfloor, k)$-shuffle. This gives the simple bounds,

$$U_3(2k) \leq 2k + 3 + U_3(k),$$
$$U_4(2k) \leq 3k + 5 + U_4(k).$$

Substituting in the trivial bounds for $U_3(k)$ and $U_4(k)$ gives the following.

$$U_3(2k) \leq 5k - 3,$$
$$U_4(2k) \leq 7k - 5.$$

Although these bounds can be improved by repeated applications of the construction, they already guarantee a constant-factor improvement whenever n is even, which is a remarkably weak condition.

Corollary 7.20. *We have*

$$\widehat{U}_3(n) \leq 2n + 5\log_2(n),$$
$$\widehat{U}_4(n) \leq 3n + 8\log_2(n).$$

Therefore,

$$U_3(n) \leq 2n + 5\log_2(n) + 3,$$
$$U_4(n) \leq 3n + 8\log_2(n) + 6.$$

Since we are primarily interested in the asymptotics of the $U_k(n)$ as n tends to infinity, we have not attempted to optimise the lower order terms. In fact, the bounds are quite weak for small n and start off considerably worse than the trivial bounds.

Proof. We only prove the first case where $t = 3$ as the other case is similar (and less useful). Note that the bound is trivially true for $n \leq 30$ as $2n+5\log_2(n) \leq 3(n-9)$, the trivial upper bound for $\widehat{U}_3(n)$. If n is odd, start with the transpositions $(1, n, 3/n)$, followed by the transpositions $(1, 2, 1/3)$ and $(1, 3, 1/2)$. The latter transpositions shuffle the position 1 into positions $1, 2, 3$, making the positions again indistinguishable. It is now sufficient to distribute the three indistinguishable positions $1, 2, 3$ over $1, 2, \ldots, n-1$. Let $\lfloor n/2 \rfloor = k$ so that $n = 2k + x$ for some $0 \leq x < 2$. Then, by induction,

$$\begin{aligned}
\widehat{U}_3(n) &\leq 3 + \widehat{U}_3(2k) \\
&\leq 3 + 2k + 2 + \widehat{U}_3(k) \\
&\leq 4k + 5\log_2(k) + 5 \\
&= 4k + 5\log_2(2k) \\
&\leq 2n + 5\log_2(n).
\end{aligned}$$

The second part of the theorem follows immediately from the bound $U_t(n) \leq U(t) + \widehat{U}_t(n)$ and the trivial bound $U(t) \leq \binom{t}{2}$. □

We are now armed with the results needed to prove the second part of Theorem 7.3, which we do by repeatedly shuffling in 3 counters at a time.

Proof of Theorem 7.3. The first part of the theorem follows from combining Lemma 7.12 with the bound $U(6) \leq 14$ from Corollary 7.18.

We first bound $U_{3\ell}(n)$. Applying Lemma 7.14 and Corollary 7.20 we find

$$\begin{aligned}
U_{3\ell}(n) &\leq U_3(n - 3(\ell-1)) + U_{3(\ell-1)}(n) \\
&\leq 2(n - 3(\ell-1)) + 5\log_2(n) + 3 + U_{3(\ell-1)}(n).
\end{aligned}$$

Solving this recursion relation we find

$$U_{3\ell}(n) \leq 2\ell n + 5\ell\log_2(n) + 6\ell - 3\ell^2.$$

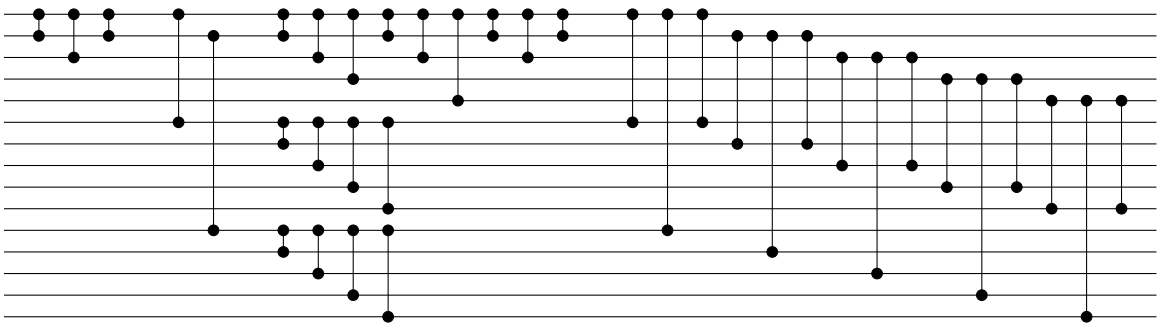


Figure 7.4: A $(3, 15)$-shuffle constructed as in the proof of Lemma 7.21 in the case $m = 3$ and $k = 5$. This uses 37 transpositions, beating the trivial upper bound of 39.

Let $\ell = \lfloor n/3 \rfloor$ and write $n = 3\ell + r$ where $0 \leq r < 3$. Then

$$\begin{aligned} U(n) &\leq U_r(r) + U_{3\ell}(n) \\ &\leq 1 + 2\ell n + 5\ell \log_2(n) + 6\ell - 3\ell^2 \\ &\leq \frac{n^2}{3} + \frac{5n \log_2(n)}{3} + 2n + 1 \\ &= \frac{2}{3}\binom{n}{2} + O(n \log(n)). \end{aligned}$$

□

The second part of Theorem 7.5 will be shown in a similar way. We could prove the first part using the results in this section, but instead we use the efficient $(3, 3k)$-shuffle given in the next section. Although the shuffle is less efficient than the shuffles in this section, it makes the proof marginally easier as the parity modulo 3 is preserved throughout the proof.

7.2.2 Splitting into three or more groups

In this section we consider Construction 7.1 in the case where $m \geq 3$ to generate efficient $(3, n)$- and $(4, n)$-shuffles. Although the construction in Lemma 7.19 leads to more efficient shuffles in general, both of these constructions will be useful: the efficient $(3, 3k)$-shuffles given in Lemma 7.21 will be used in the proof of Theorem 7.5, and the efficient $(4, 3k)$-shuffles given in Lemma 7.22 will be used in Section 7.4 to give efficient shuffles consisting entirely of star transpositions. Examples of the shuffles are given in Figure 7.4 and Figure 7.5 respectively.

Lemma 7.21. *Let $m, k \geq 3$. Then*

$$U_3(mk) \leq 2k + 3 + U_3(k) + kU_3(m),$$

and in particular,

$$U_3(3k) \leq 5k + 3 + U_3(k).$$

Proof of Lemma 7.21. We use Construction 7.1 in the case where $t = 3$. Start by shuffling the 3 counters using 3 transpositions. In stage 2 we use the transpositions $(1, k+1, x)$ and $(2, 2k+1, y)$. There are only 3 possible options for the unordered multisets of the W_i, namely $\{3, 0, \ldots, 0\}$, $\{2, 1, 0, \ldots, 0\}$ and $\{1, 1, 1, 0, \ldots, 0\}$. Consider the first case, where $w_1 = 3$ and $w_2 = \cdots = w_m = 0$. Then

$$\begin{aligned}\sum_{\sigma \in S_m} \mathbb{P}\big(W = (w_{\sigma(1)}, \ldots, w_{\sigma(m)})\big) &= \sum_{\substack{\sigma \in S_m \\ \sigma(1)=1}} \mathbb{P}(W_1 = 3) \\ &= (m-1)! \cdot \mathbb{P}(W_1 = 3) \\ &= (m-1)! \cdot (1-x)(1-y).\end{aligned}$$

The other cases are similar, and the condition (7.5) becomes the following.

$$(1-x)(1-y) = \frac{m\binom{k}{3}}{\binom{mk}{3}}, \tag{7.8}$$

$$x(1-y) + (1-x)y = \frac{m(m-1)\binom{k}{2}\binom{k}{1}}{\binom{mk}{3}}, \tag{7.9}$$

$$xy = \frac{\binom{m}{3}k^3}{\binom{mk}{3}}. \tag{7.10}$$

This reduces to a quadratic and can be solved explicitly. Since $xy > 0$ and $(1-x)(1-y) > 0$, we necessarily have $0 < x, y < 1$ for any solution and x and y will be valid choices for a probability. Since both sides will sum to 1, one of the conditions is redundant, say (7.9). Rearranging (7.10) gives

$$y = \frac{k^2(m-2)(m-1)}{(km-2)(km-1)x}.$$

Substituting this into (7.8) and rearranging shows

$$(km-1)(km-2)x^2 + k(3km + m - k - 3)x - k^2(m-1)(m-2) = 0.$$

The determinant of this quadratic is

$$(k-1)k^2(m-1)(5km - k - m - 7)$$

which is positive for $k, m \geq 3$. Hence, there are always two roots and there is a suitable choice for x. In fact, y satisfies the same equation and will be the other root.

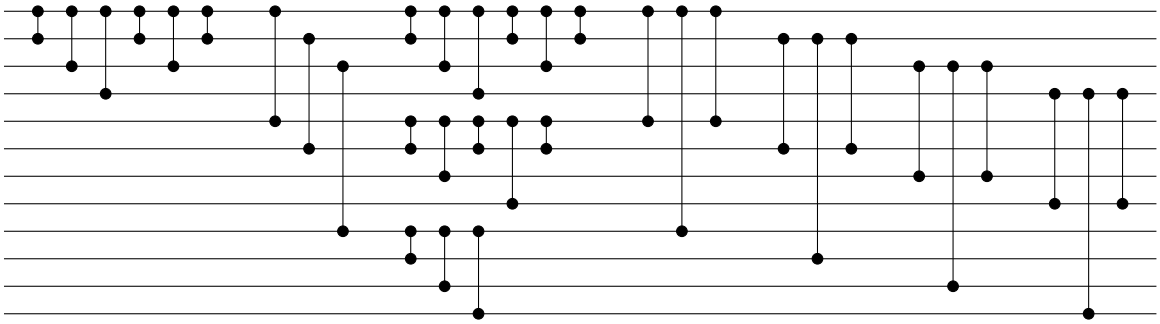


Figure 7.5: A $(4, 12)$-shuffle using just 35 transpositions constructed as in the proof of Lemma 7.22.

In stage 3, we use $U_3(k)$ transpositions to shuffle the positions $1, 2, 3$ across the positions $1, \ldots, k$. We use $k-1$ transpositions to shuffle the position $k+1$ into the positions $k+1, \ldots, 2k$ and another $k-1$ transpositions to shuffle the position $2k+1$ into $2k+1, \ldots, 3k$. Since the counters can only be in positions $1, \ldots, 3k$, stage 4 can be completed using k copies of a $(3, m)$-shuffle. Explicitly, shuffle $i, k+i$ and $2k+i$ into $i, k+i, \ldots, (m-1)k+i$ for every $1 \leq i \leq k$. □

We can also use Construction 7.1 to construct efficient shuffles when $t = 4$, but only in the special case $m = 3$. When $m = 3$, there are only 4 different unordered multisets for the W_i, namely $\{4, 0, 0\}$, $\{3, 1, 0\}$, $\{2, 2, 0\}$ and $\{2, 1, 1\}$. In particular, we avoid the multiset of $\{1, 1, 1, 1, 0 \ldots 0\}$ which appears to clash with $\{2, 2, 0, \ldots, 0\}$ when using the simplest constructions for stage 2. We give the details below.

Lemma 7.22. *Let $k \geq 4$. Then*

$$U_4(3k) \leq 4k + 8 + U_4(k) + U_2(k).$$

Proof. We apply Construction 7.1 when $t = 4$ and $m = 3$. Start by shuffling the counters using 6 transpositions, before using the transpositions $(1, k{+}1, x)$, $(2, k{+}2, y)$

and $(3, 2k+1, z)$. Condition (7.5) becomes

$$\begin{aligned}
\frac{3\binom{k}{4}}{\binom{3k}{4}} &= (1-x)(1-y)(1-z),\\
\frac{6k\binom{k}{3}}{\binom{3k}{4}} &= x(1-y)(1-z) + y(1-x)(1-z) + z(1-x)(1-y),\\
\frac{3\binom{k}{2}^2}{\binom{3k}{4}} &= xy(1-z),\\
\frac{3k^2\binom{k}{2}}{\binom{3k}{4}} &= x(1-y)z + (1-x)yz + xyz.
\end{aligned}$$

which has solutions for $k \geq 4$. This can be seen by substituting in the value

$$z = \frac{19k^2 - 11k - k\sqrt{37k^2 - 94k + 49}}{27k^2 - 27k + 6}$$

and proceeding as in the proof of Lemma 7.21.

In stage 4 shuffle the first 4 positions into the first k positions using $U_4(k)$ transpositions. Then shuffle $k+1$ and $k+2$ into $k+1, \ldots, 2k$ using $U_2(k)$ transpositions and $2k+1$ into $2k+1, \ldots, 3k$ using $k-1$ transpositions. Finally, end by shuffling the positions $i, k+i, 2k+i$ for every $1 \leq i \leq k$. □

We end this section by proving Theorem 7.5 using the constructions from Section 7.2.1 and Section 7.1 as well as Lemma 7.21.

Proof of Theorem 7.5. We use the simple bound $U_3(3k) \leq 8k - 3$ for $k \geq 3$ which follows from combining the trivial bound $U_3(k) \leq 3k - 6$ with Lemma 7.21. Lemma 7.14 gives

$$\begin{aligned}
U_{3\ell}(3k) &\leq U_3(3k - 3(\ell-1)) + U_{3(\ell-1)}(3k)\\
&\leq 8(k - \ell + 1) - 3 + U_{3(\ell-1)}(3k).
\end{aligned}$$

In particular, if we denote $U_{3\ell}(3k) = 9k\ell - \binom{3\ell+1}{2} - a_\ell$, we get the recursion relation

$$a_\ell \geq (k - \ell + 1) - 3 + a_{\ell-1},$$

with the initial condition $a_1 = k - 3$. Hence,

$$a_\ell \geq \frac{1}{2}\ell(2k - \ell - 5).$$

Now let $n = 3k + r$ for $0 \leq r < 3$ and $t = 3\ell + s$ for $0 \leq s < 3$ where $n \geq 9$. Using Lemma 7.13 and Lemma 7.14, we have

$$U_t(n) \leq tn - \binom{t+1}{2} - \frac{1}{2}\ell(2k - \ell - 5).$$

In particular, $U_t(n) \leq (1-\delta)\big(tn - \binom{t+1}{2}\big)$ for all

$$\delta \leq \frac{\ell(2k - \ell - 5)}{2\big((3\ell+2)(3k+2) - \binom{3\ell+2}{2}\big)}.$$

This is an increasing function of k so the worst case is when $k = \ell$ where it takes the value

$$\frac{(\ell - 5)\ell}{9\ell^2 + 9\ell + 2}.$$

This is greater than $3/190$ for all $\ell \geq 6$. The cases where $\ell \leq 5$ follow from the second part of the theorem. Indeed, the second part of the theorem guarantees the existence of a constant δ_t such that $U_t(n) \leq (1-\delta_t)\big(tn - \binom{t+1}{2}\big)$ for large enough n, and we already know that we can beat the bound by an additive factor for all $t \geq 3$ and $n \geq 6$.

The second part of the theorem follows almost immediately from the proof of Theorem 7.3 where we proved the following bound.

$$U_{3\ell}(n) \leq 2\ell n + 5\ell \log_2(n) + 6\ell - 3\ell^2.$$

If $t = 3\ell + r$ for some $0 \leq r < 2$, we have

$$\begin{aligned} U_{3\ell+r}(n) &\leq U_r(n - 3\ell) + U_{3\ell}(n) \\ &\leq r(n - 3\ell) - \binom{r+1}{2} + U_{3\ell}(n) \\ &\leq (2\ell + r)n + 5\ell \log_2(n) + (6 - 3r)\ell - 3\ell^2 - \binom{r+1}{2} \\ &= (2\ell + r)n + O_\ell(\log(n)). \end{aligned}$$

□

7.3 Reachability

Recall that we say a sequence of transpositions $T_1, \ldots, T_\ell$ achieves t-reachability across n positions if for every choice of t ordered positions $x_1, \ldots, x_t$ from $[n]$, there is a subsequence of $T_1, \ldots, T_\ell$ which maps i to x_i for all $i \leq t$. Clearly, any sequence

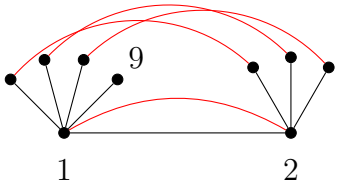


Figure 7.6: The multigraph corresponding to the minimum 2-reachable sequence given in the proof of Theorem 7.7. All vertices except 1 and 9 have discrepancy 0.

which achieves t-uniformity also achieves t-reachability, and so a lower bound on t-reachability gives a lower bound on t-uniformity. The number of transpositions needed for 1-reachability is $n-1$ showing that the upper bound for 1-uniformity is tight, while the number needed for n-reachability is $\Theta(n\log(n))$, a long way from the current best known upper bound for n-uniformity which is $n^2/3 + O(n\log(n))$. We start by showing that the minimum number of transpositions for 2-reachability is $\lceil 3n/2\rceil - 2$, which gives a non-trivial lower bound on the case of 2-uniformity, although there remains a substantial gap between the upper and lower bounds.

Proof of Theorem 7.7. We first give an upper bound construction. Let $n \geq 2$ be given. Our sequence of transpositions starts with the following transpositions, in order.

1. $(1, 2)$.
2. $(1, x)$ for all odd $3 \leq x \leq n$.
3. $(2, y)$ for all even $4 \leq y \leq n$.
4. $(x, x+1)$ for all odd $3 \leq x \leq n-1$.

If n is odd, we also add the transposition $(1, 2)$ at the end, which will be needed to reach the position $(1, n)$. This defines a sequence of transpositions of length $\lceil 3n/2\rceil - 2$, and it is straightforward to check that this sequence achieves 2-reachability.

The main work in this proof is in showing the lower bound. Suppose that $\sigma_1, \ldots, \sigma_\ell$ is a shortest sequence of transpositions that achieves 2-reachability on $n \geq 2$ points. We will consider a process on an auxiliary multigraph G with vertex set $\{1, \ldots, n\}$. We initialise G as the empty graph and process the transpositions in order, adding an edge for each. When we process a transposition (a, b), we add an edge $\{a, b\}$ to G. Since our sequence of transpositions is minimal, we will never encounter a transposition (a, b) where neither a nor b could possibly contain a counter (when we

process (a, b)). If the first time a or b can get a counter is via the transposition (a, b), then we colour the edge black. Otherwise, we colour the edge red. The black edges always form a forest consisting of at most two trees each containing one of the starting positions.

The process begins by growing two trees of black edges from the vertices 1 and 2, which we denote T_1 and T_2 respectively. We do not preclude the trees containing some internal red edges, although this would mean the sequence is not minimal. Let (a, b) where $a \in V(T_1)$ and $b \in V(T_2)$ be the first time the two trees meet. Before this time, the counter c_1 is contained in $V(T_1)$ and the counter c_2 is contained in $V(T_2)$, so immediately after (a, b) the counter c_1 is contained in $V(T_1) \cup \{b\}$ and the second counter c_2 is contained in $V(T_2) \cup \{a\}$. We replace the initial sequence of transpositions, up to and including (a, b), by the following sequence which can reach every pair (x, y) the original sequence could reach, and yet uses at most as many transpositions. We start with the transpositions $(1, v)$ for all $v \in V(T_1) \setminus \{a\}$, then the transpositions $(1, u)$ for all $u \in V(T_2) \setminus \{b\}$.

Lastly, we do $(1, 2)$, $(1, a)$ and $(2, b)$. We will colour the edge $(1, 2)$ black, so it now makes sense to view the black edges as forming a tree rooted at the pair of vertices $\{1, 2\}$.

We now prove that the sum of the red degrees is at least $n - 2$, which shows that the number of edges in G (and transpositions in our sequence) is at least $n - 1 + \lceil (n-2)/2 \rceil = \lceil 3n/2 \rceil - 2$. We do this by considering the "deficit" of a subtree and showing that this is always at most 1. A *subtree* rooted at v is the set of vertices u (including v) where the unique path from u to $\{1, 2\}$ in the black tree passes through v. Any vertex in the subtree rooted at v is said to be *above* v. The *deficit* of a vertex v is the number of vertices in the subtree rooted at v minus the sum of the red degrees of the vertices in the subtree. Note that the red edges do not need to be wholly contained in the subtree in order to count towards the red degree.

Since we rooted our black tree at $\{1, 2\}$, each vertex $v \notin \{1, 2\}$ has a parent which is the first vertex after v on the unique path from v to $\{1, 2\}$, and we denote it by P_v. We show the following claim inductively, which we will extend to $v \in \{1, 2\}$ to finish the proof.

Claim 7.1. *Every vertex $v \notin \{1, 2\}$ has deficit at most 1. Moreover, if the deficit of v is equal to 1, then there is a vertex ℓ such that the only way for a counter to reach ℓ is to enter v using the black edge $P_v v$ and then to follow the path of black edges from v to ℓ.*

Proof. We prove the claim by induction on the height of the subtree rooted at $v \notin \{1, 2\}$. If the height is 0, that is, v is a leaf of the black tree, then the deficit is at most 1 and equals 1 if and only if v can only be reached via the black edge from its parent P_v.

Assume that the claim has been shown up to height $h \geq 0$, and suppose the subtree rooted at v has height $h + 1$.

We first show that there are at least as many red edges incident to v as there are neighbours above v with deficit 1, which implies that the deficit at v is at most 1. If a vertex u of deficit 1 is above v and connected to v via a black edge, then we define ℓ_u to be a vertex in the subtree rooted at u which can only be reached using black edges from v. Let z be the last neighbour of v which is connected with a black edge and with deficit 1. Then there is no way to get a counter into both v and ℓ_z unless there is a red edge after (v, z). Similarly, if v has black edges to vertices x and y that are above v and both have deficit 1, then there has to be a red edge incident with v which occurs between the transpositions (v, x) and (v, y) else there is no way to reach (ℓ_x, ℓ_y).

Suppose now that v has deficit exactly 1. We need to find a vertex ℓ which can only be reached using a path of black edges from the parent of v. If none of the neighbours above v have deficit 1, then the deficit of v implies that there are no red edges incident to v, and so we can take $\ell = v$. Otherwise, let x be the first neighbour above v which has deficit 1. Since, v has deficit exactly 1, our argument above shows that there cannot be a red edge incident to v corresponding to a transposition which is earlier than (v, x). Hence, the only way of getting a counter into ℓ_x is using the black edge to v, the edge (v, x) and the path of black edges from x to ℓ_x, and we take $\ell = \ell_x$. □

We further claim that the vertices 1 and 2 have deficit at most 1 as well. The difficulty in this case is that it is possible to put a counter on 1 (say) using the black edge $(1, 2)$. This means that we can put a counter on ℓ_x and ℓ_y without requiring a red edge between them. However, this only works for indistinguishable counters, not for distinguishable counters, and we need to modify the argument from the claim slightly to make use of this.

Let the counters starting in positions 1 and 2 be c_1 and c_2 respectively, and suppose there is a neighbour x of 1 which is above 1 and has deficit 1. Then there is a vertex ℓ_x such that the only way for a counter to reach ℓ_x is to follow the path of black edges from 1 to x to ℓ_x. If the transpositions $(1, x)$ occurs before the black edge $(1, 2)$, then there is no way to move the counter c_2 to ℓ_x. Hence, every neighbour above 1 and

before $(1,2)$ has deficit 0. The black edge $(1,2)$ is the only way to put a counter on 1 without using a red edge, so after this transposition we can apply a proof similar to that of the claim above to see the deficit of 1 is at most 1.

Since every vertex is in either the subtree rooted at 1 or the subtree rooted at 2 and the deficit of these two subtrees is at most 1, the sum of the red degrees must be at least $n-2$. Hence there are at least $\lceil (n-2)/2 \rceil$ red edges and at least $\lceil 3n/2 \rceil - 2$ transpositions in the sequence. □

We now consider the problem of t-reachability, and give a probabilistic construction which shows that at most $(2+o_t(1))n$ transpositions are necessary. This is much lower than the best known constructions for uniformity and, surprisingly, the coefficient of the leading term is bounded. We remark that all the transpositions used in this construction are star transpositions.

Proof of Theorem 7.8. Recall that the counters start in positions $\{1,\dots,t\}$ and let $M = \{t+1,\dots,n\}$ be the remaining positions. Let $L = \lfloor n^{1-\varepsilon} \rfloor$ for some sufficiently small ε depending only on t. Let G be a random bipartite graph with vertex set $A \cup B = \{a_{t+1},\dots,a_n\} \cup \{b_1,\dots,b_L\}$ constructed by adding two uniformly random edges from a_i to $\{b_1,\dots,b_L\}$ for each i. The random transposition sequence begins with L phases. In the ith phase, we add the transpositions $(1,j)$ for $j \in \{2,\dots,t\}$ followed by $(1,j)$ for the j such that a_j is adjacent to b_i. Finally, the sequence ends with a t-reachable sequence over the first t positions.

Fix t positions for the t counters to end in and denote the position where the ith counter ends by x_i. Let the x_i which are in M be $x_{i_1},\dots,x_{i_m} = y_1,\dots,y_m$ and suppose there is a matching in G containing the vertices $a_{y_1},\dots,a_{y_m}$, say $\{a_{y_j}b_{s_j} : j \in [m]\}$. We will put the counter $i_j \in [t]$ in position $y_j = x_{i_j}$ in the phase s_j using the transpositions $(1,i_j)$ and $(1,y_j)$. The remaining counters can easily be positioned using the t-reachable sequence at the end, and so we only need to show that there is some choice for G such that every set of $s \le t$ vertices from A is in a matching.

The probability that a given set of s vertices from A has at most $s-1$ neighbours in B is at most

$$\binom{L}{s-1}\left(\frac{s-1}{L}\right)^{2s} \sim \frac{(s-1)^{2s}}{(s-1)!}n^{(1-\varepsilon)(-1-s)}$$

Hence, by the union bound, the probability that there is a set of s vertices from A with at most $s-1$ neighbours is at most

$$\sum_{s=3}^{t}\binom{n}{s}\binom{L}{s-1}\left(\frac{s-1}{L}\right)^{2s} \sim \sum_{s=3}^{t}\frac{(s-1)^{2s}}{(s-1)!s!}n^{-1+(s+1)\varepsilon} = O_t\big(n^{-1+(t+1)\varepsilon}\big).$$

If $\varepsilon < 1/(t+1)$, this is $o(1)$ and less than 1 for n sufficiently large, which implies there exists a suitable choice for the graph G. Note that for $j \in M$ the vertex a_j has degree exactly 2 and the transposition $(1, j)$ appears exactly twice in the sequence. Hence, this construction gives t-reachability using at most

$$(t-1)L + 2(n-t) + \binom{t+1}{2} = 2n + o_t(n)$$

transpositions. □

7.4 Star transposition shuffles

We now consider the problem of generating efficient shuffles which only use star transpositions. Although star transpositions are much more restrictive, any transposition shuffle can be converted to a shuffle which only uses star transpositions by replacing the lazy transposition (a, b, p) with the transpositions $(1, a, 1)(1, b, p)(1, a, 1)$. This means that all the constructions above can be converted to ones using only star transpositions, but it turns out only the construction from Lemma 7.22 leads to an efficient shuffle. All of the efficient constructions we have given make use of several "subshuffles" which shuffle certain positions together, and these can be taken to use transpositions of the form $(x, \cdot)$ for some suitable value of x. We can "switch" the central position of the star to position x using the transposition $(1, x, 1)$, use the subshuffle with x replaced by 1 and then use $(1, x, 1)$ again to switch the central position back to 1. If we carefully keep track of which positions have been switched, we can often do better and avoid switching the central position back to 1 the majority of the time.

Although the construction in Lemma 7.22 is not the most efficient for general transpositions, it uses relatively few "subshuffles" and we only need to switch the central position relatively few times. Indeed, plugging in trivial bounds for the subshuffles, the construction uses around $10k$ transpositions instead of $12k$, a saving of $2k$, and only uses around k subshuffles, most requiring only 1 extra transposition to switch the central position. This gives a total of around $11k$ transpositions, beating the trivial bound by a constant factor.

Let $\widetilde{U}_t(n)$ be the minimum number of lazy transpositions in a (t, n)-shuffle where all transpositions are star transpositions. Lemma 7.13 shows that the trivial upper bound of $tn - \binom{t+1}{2}$ holds for $\widetilde{U}_t(n)$, but the following theorem shows it is again possible to beat the upper bound by a constant factor when $t = 4$ and n is large

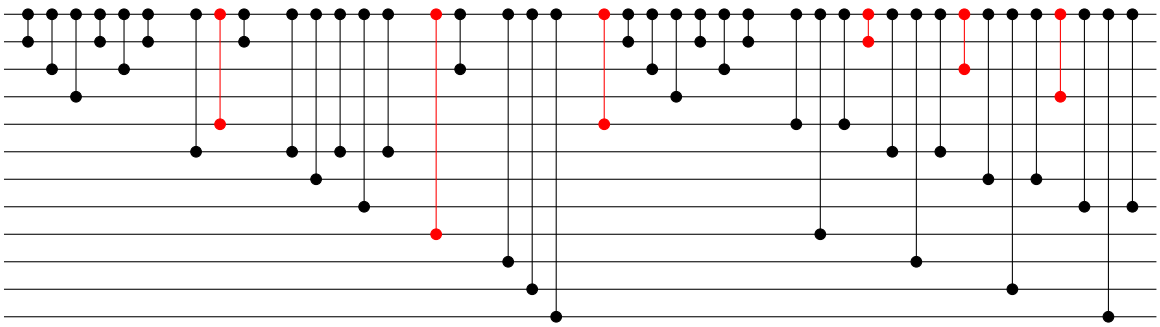


Figure 7.7: A $(4, 12)$-shuffle constructed as in the proof of Lemma 7.23. The red transpositions indicate the additional transpositions inserted to "switch" the central position and ensure all transpositions are of the form $(1, \cdot)$. These fire with probability 1. Note that for this small value of k, this shuffle is worse than the trivial upper bound.

enough. By repeatedly shuffling in 4 counters at a time, we can construct efficient shuffles for all $t \geq 4$.

Lemma 7.23. *For every $k \geq 3$, we have*

$$\widetilde{U}_4(3k) \leq 5k + 10 + \widetilde{U}_2(k) + \widetilde{U}_4(k).$$

In particular,

$$\widetilde{U}_4(3k) \leq 11k - 3.$$

Proof. We follow the construction given in the proof of Lemma 7.22 but with some additional transpositions which always fire. Start by shuffling the 4 counters using 6 star transpositions. In order to minimise the number of times we use deterministic transpositions to "switch" the central position, we will interleave parts of stages 2 and 3. Begin with the transpositions $(1, k+2, x)$, $(1, k+1, 1)$ and $(1, 2, y)$ where x and y are as given in the proof of Lemma 7.22. Shuffle the positions 1 and $k+2$ over the positions $1, k+2, k+3, \ldots, 2k$ using $\widetilde{U}_2(k)$ star transpositions. Next, use the transposition $(1, 2k+1, 1)$ followed by $(1, 3, z)$ where z is also as in the proof of Lemma 7.22. Shuffle the position 1 over $1, 2k+2, \ldots, 3k$ using $k-1$ star transpositions. To finish stages 2 and 3, use the transposition $(1, k+1, 1)$ to return the centre to its original position and shuffle the positions $1, 2, 3, 4$ over the positions $1, 2, \ldots, k$.

Begin stage 4 by shuffling the positions $1, k+1$ and $2k+1$. For each $2 \leq i \leq k$, use the transposition $(1, i, 1)$ and then the transpositions $(1, k+i)$, $(1, 2k+i)$ and $(1, k+i)$ to shuffle the positions $1, k+i$ and $2k+i$. □

To prove Theorem 7.9, we can mimic the proofs used for general transpositions for which we only need versions of Lemma 7.13 and Lemma 7.14. The former extends immediately to give

$$\widetilde{U}_t(n) \leq t + \widetilde{U}_t(n-1),$$

but the latter does not immediately extend. Instead, we need to add a single transposition to switch the centre, and we get the bound

$$\widetilde{U}_t(n) \leq \widetilde{U}_{t-m}(n-m) + \widetilde{U}_m(n) + 1.$$

The proof is now very similar to the proofs of Theorem 7.3 and Theorem 7.5 given in the general case, and is omitted.

We now turn our attention back to the problem of 2-reachability and prove Theorem 7.10, which shows that restricting to star transpositions leads to only a constant difference in the number of transpositions needed for 2-reachability.

Let us first consider the upper bound. The idea is simple: we will start with the transposition $(1, 2)$ so the two counters are indistinguishable, then we will sweep through the even numbers and potentially load one of them with a counter. Finally, we will sweep through every position. That is,

1. $(1, 2)$,
2. $(1, 4), (1, 6), \cdots (1, 2\lfloor n/2 \rfloor)$,
3. $(1, 2), (1, 3), (1, 4), \ldots, (1, n)$.

It is easy to see that the counters can reach (x, y) when x and y are not both odd, simply place one counter in an even position in the first sweep and place the other counter in the second sweep. To put a counter on x and y when both are odd and $x, y \neq 1$, we load $(x + 1)$ in the first sweep and use this to "reload" 1 before the transposition $(1, y)$. When $x = 1$, this sequence works when n is even and we can use $(1, n)$ to leave a counter on 1, but it breaks down when n is odd. This could be fixing by appending the transposition $(1, 2)$ say, but it is possible to do slightly better with the following "twisted" sequence. Let $n = 2m + 1$. The following sequence of transpositions is 2-reachable.

1. $(1, 2)$.
2. $(1, 4), (1, 6), \cdots (1, 2m)$.
3. $(1, 3), (1, 2), (1, 5), (1, 4) \ldots (1, 2m + 1), (1, 2m)$.

Proof of Theorem 7.10. The upper bound is given in the discussion above, so we only need to prove a matching lower bound. For this we proceed much as in the proof of Theorem 7.7. We colour the transpositions $(1, x)$ black if it is the only occurrence, red if it has occurred previously and blue otherwise (i.e. if it is a first occurrence, but will occur later as well). There are $n-1$ transpositions that are either black or blue, and it suffices to show that there are at least $(n-1)/2$ red transpositions.

Let the two counters be c_1 and c_2, and suppose first that $(1, 2)$ is black. In order to end with the counter c_2 in position y where $y \geq 3$, the transposition $(1, y)$ needs to happen after the transposition $(1, 2)$. Moreover, in order to put c_1 in position x and c_2 in position y where $x, y \geq 3$, the transposition $(1, x)$ needs to happen before $(1, 2)$. Hence, we have used $2(n-2)+1$ transpositions, and we can assume that $(1, 2)$ is not black.

If $(1, x)$ is the last black transposition, there needs to be a red transposition after $(1, x)$ in order to leave a counter in 1 and in x. Similarly, in between any pair of black transpositions there has to a red transposition. This shows the number of red transpositions is at least the number of black transpositions. By definition the number of red transpositions is at least the number of blue transpositions. Since there is a total of $n-1$ black and blue transpositions, there must be at least $(n-1)/2$ of one of them and hence, there are at least $(n-1)/2$ red transpositions. □

The lower bound for 2-reachability above immediately gives a lower bound for the more difficult problem of achieving 2-uniformity with star transpositions but, using the restrictive nature of the transpositions, we can show that the number of star transpositions needed for 2-uniformity is at least $1.6n - C$, a constant factor higher. This confirms, for this specific case, that the problem of uniformity is 'truly more difficult' than reachability.

Proof of Theorem 7.11. We will use a discharging argument to show that all but a constant number of the transpositions are used at least 1.6 times on average. Let $\sigma_1, \ldots \sigma_m$ be the subsequence of star transpositions $(1, x)$ with $x \geq 3$ which are used exactly once. Start by assigning to the transposition $(1, a)$ a weight equal to the number of times $(1, a)$ is used in the sequence of lazy transpositions. We transfer weight from the transpositions which are used multiple times to the transpositions which are used exactly once according to the following rules.

- If there are at least two transpositions which appear for the last time between σ_i and σ_{i+1}, then they each transfer transfer 0.3 to σ_i and 0.1 to σ_{i+1}.

- If there is only one transposition which appears for the last time between σ_i and σ_{i+1}, it transfers 0.4 to σ_i.
- Each transposition which appears between σ_i and σ_{i+1} for neither the first time nor the last time transfers 0.6 to σ_i and 0.2 to σ_{i+1}.

A transposition used exactly twice transfers 0.4 and ends with weight 1.6, and a transposition used more than twice transfers $0.8(i-2)+0.4$ so ends with weight $1.6+0.2(i-2)$. Hence, we only need to show that all but a constant number of the σ_i end with weight at least 1.6.

Let $\sigma_{i-1}=(1,a)$ and $\sigma_i=(1,b)$. In order to end with counters in both a and b, there must a transposition between σ_{i-1} and σ_i which can place a counter in position 1. In particular, there is either a transposition which is not appearing for the first time, or $(1,2)$ appearing for the first time. This immediately shows that all but one of $\sigma_1,\dots\sigma_{m-1}$ have weight at least 1.4. We claim that in between either σ_{i-1} and σ_i or between σ_i and σ_{i+1} one of the following must hold:

1. there are at least two transpositions appearing for the last time,
2. there is a transposition appearing for neither the first nor the last time,
3. there is the transposition $(1,2)$ appearing for the first time.

If one of the first two cases occurs, then σ_i ends with weight at least 1.6, while the last case can only occur twice. This analysis does not apply to σ_1 and σ_m, giving a total of at most 4 of the σ_i which may have weight less than 1.6. Accounting for the possibility that $(1,2)$ appears once, this will prove the statement with the constant $C=3$.

We now prove the above claim. Let $\sigma_{i-1}=(1,a)$, $\sigma_i=(1,b)$ and $\sigma_{i+1}=(1,c)$ and assume that none of the conditions above hold. In between σ_{i-1} and σ_i, there can be only transpositions used for the first time and a single transposition $(1,\ell)$ used for the last time. We may write the sequence as

$$(1,a),(1,f_1),\dots,(1,f_k),(1,\ell),(1,f_{k+1}),\dots,(1,f_s),(1,b) \tag{7.11}$$

where the f_j are distinct, not equal to 2 and are appearing for the first time.

The only way to end the sequence of lazy transpositions with the counters in $\{a,b\}$, $\{a,\ell\}$ or $\{b,\ell\}$ is to start the sequence (7.11) with the two counters in $\{1,\ell\}$. Let p_j

be the probability associated with the transposition $(1, j)$ in (7.11). The probability that this sequence ends with counters in $\{a, b\}$ and $\{a, \ell\}$ must be equal, so

$$p_a(1 - p_\ell) = p_a p_\ell (1 - p_{f_{k+1}}) \cdots (1 - p_{f_s}) p_b.$$

In particular, if $q = (1 - p_{f_{k+1}}) \cdots (1 - p_{f_s})$, then $1 - p_\ell = p_\ell q p_b$ which implies $p_\ell = 1/(qp_b + 1) \geq 1/2$. We now claim that the probability a counter ends in ℓ is strictly higher than the probability a counter ends in b unless $qp_b = 1$. We condition on which of 1 and ℓ contain a counter before the sequence (7.11) and calculate the probability a counter ends in each of the positions ℓ and b.

Starting positions	{1,ℓ}	$\{\ell\}$	$\{1\}$	$\varnothing$
Counter ends in ℓ	1	$1 - p_\ell$	p_ℓ	0
Counter ends in b	qp_b	$p_\ell q p_b$	$(1 - p_\ell) q p_b$	0

In every case, the probability that a counter ends in ℓ is at least the probability that a counter ends in b, and the first one is strict unless $qp_b = 1$.

Now we consider the transpositions between $(1, b)$ and $(1, c)$ below.

$$(1, b), (1, f'_1), \ldots, (1, f'_k), (1, \ell'), (1, f'_{k+1}), \ldots, (1, f'_s), (1, c)$$

Since the transposition $(1, b)$ fires with probability 1, there is no way for the counters to end in both ℓ' and c, giving a contradiction and proving the claim. □

We have seen in Theorem 7.10 that the shortest sequence of star transpositions achieving 2-reachability has length approximately $3n/2$. Moreover, it is possible to construct a 3-reachable sequence using $5n/3 + C$ such transpositions. Using a similar argument to the proof of Theorem 7.11, we offer the following lower bound for arbitrary t when using star transpositions, which is tight up to additive constants when $t = 2, 3$.

Proposition 7.24. *For each $t \geq 1$, there is a constant C_t such that any sequence of star transpositions that achieves t-reachability on n points has length at least $(2 - 1/t)n - C_t$.*

Proof. We will show that all but at most $2t$ transpositions are used at least $2 - 1/t$ times on average. Let $\sigma_1, \ldots, \sigma_m$ be the sequence of transpositions $(1, x)$, where $x \notin [t]$, which are used exactly once. We start by assigning to the transposition $(1, a)$ a weight equal to the number of times $(1, a)$ is used, and we transfer weight from a transposition used multiple times according to the following simple rules.

- For each transposition between σ_i and σ_{i+1} used for the last time (but not the first), transfer $1/t$ to σ_i.
- For each transposition between σ_i and σ_{i+1} used for neither the first time nor the last time, transfer 1 to σ_i.

Clearly, any transposition which is used multiple times ends with weight $2 - 1/t$ as required. If there is a transposition between σ_i and σ_{i+1} used for neither the first nor the last time, then σ_i has weight at least 2. We will ignore any interval which contains the first instance of some transposition $(1, x)$ where $x \in [t]$ and note that there are at most $t - 1$ of them. Now consider the section between $\sigma_i = (1, a)$ and $\sigma_{i+1} = (1, b)$

$$(1, a)(1, s_1) \cdots (1, s_\ell)(1, b).$$

We assume that all of these are used for either the first or last time and, if $(1, s_j)$ is used for the first time, then $s_j \notin [t]$. Let $s'_1, \ldots, s'_{\ell'}$ be the transpositions used for the last time and suppose that $\ell' \leq t - 2$. It must be possible to reach the positions $\{a, s'_1, \ldots, s'_{\ell'}, b\}$, but this isn't the case. Indeed, the transposition $\sigma_i = (1, a)$ must be used (as this is the only way to put a counter on a) and then the only way for a counter to 'reload' position 1 before σ_{i+1} is using a transposition $(1, s'_j)$ used before. But then no counter can end in position s'_j. □

7.5 Open problems

While we have given constructions which show that the upper bound for $U(n)$ can be improved by a constant factor, the new upper bounds are still a long way from the best known lower bound and the biggest open problem is to determine the correct asymptotic for $U(n)$.

Problem 7.25. *What is correct asymptotic for $U(n)$? Is it true that $U(n) = o(n^2)$?*

We also considered constructing transposition shuffles where all transpositions are of the form $(1, \cdot)$. The minimum number of transpositions in such a shuffle is at most $3U(n)$, but the restrictive nature of the transpositions might make it easier to prove a lower bound.

Problem 7.26. *What is the minimum number of lazy transpositions in a transposition shuffle if all transpositions are of the form $(1, \cdot)$? Is it $o(n^2)$?*

We also introduced the study of (t, n)-shuffles and showed that the trivial upper bound can be improved by a constant factor for all $t \geq 3$ and $n \geq 6$. However, we have been unable to find any $(2, n)$-shuffles using fewer than $2n - 3$ transpositions, and we conjecture that this is optimal (i.e. $U_2(n) = 2n - 3$). We remark that, if this were true, it would match nicely with selection networks. As a first case, it would be interesting to close the gap when restricting to star transpositions and confirm the conjecture in this case.

Conjecture 7.27. *For $n \geq 2$, $U_2(n) = 2n - 3$.*

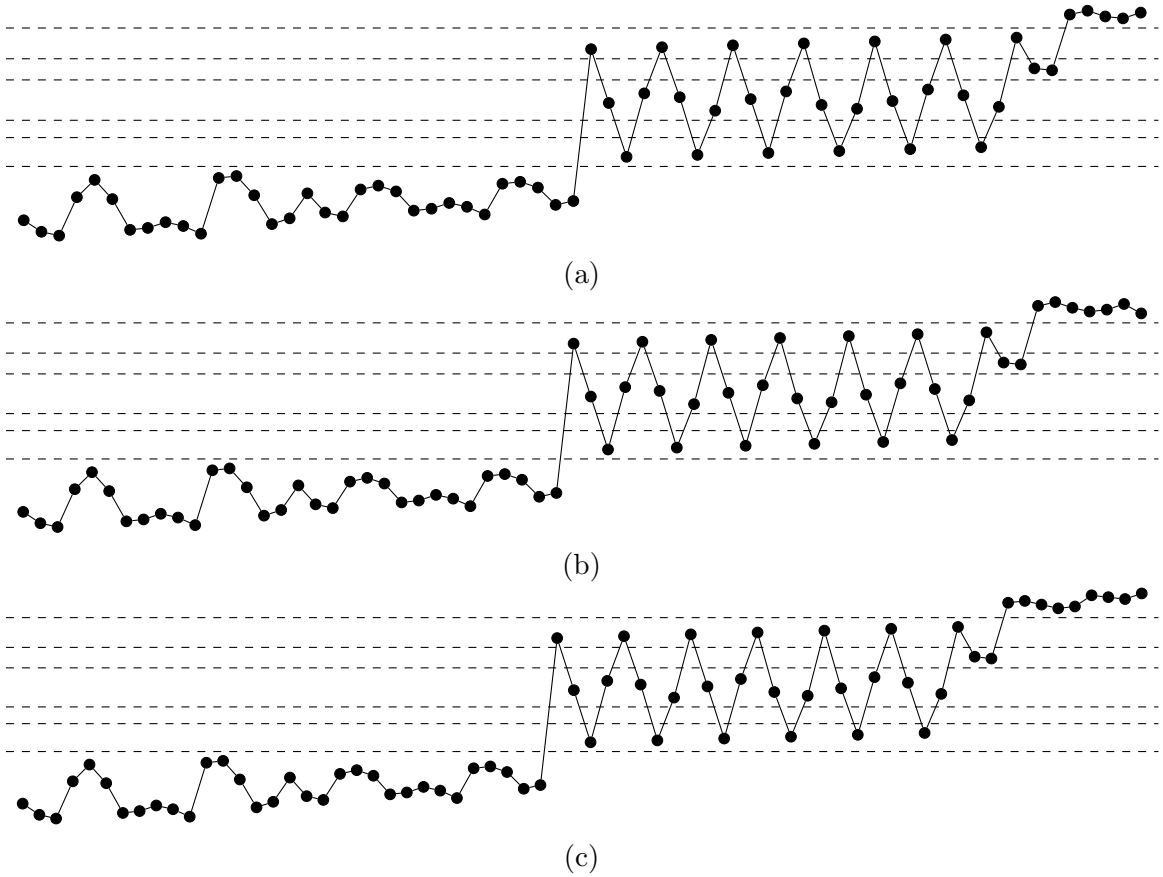
We gave a lower bound for 2-uniformity by exactly determining the minimum number of transpositions in a 2-reachable sequence of transpositions. We also gave upper and lower bounds for general t-reachability using star transpositions, although there is still a small gap between them.

Problem 7.28. *What is the minimum number of transpositions needed in a sequence of transpositions which is t-reachable? What if all of the transpositions are of the form $(1, \cdot)$?*

Appendix A

Bicrucial square-free permutations

A.1 Constructions from Theorem 5.5



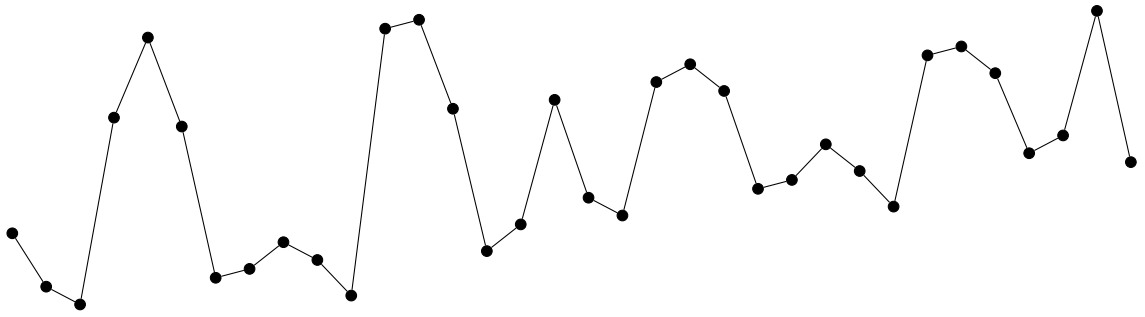
(a)

(b)

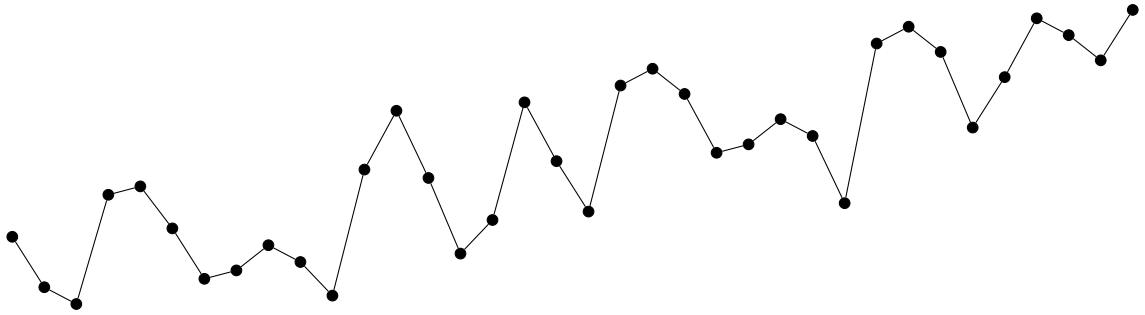
(c)

Figure A.1: Bicrucial permutations of length 64, 66 and 68 constructed as in the proof of Theorem 5.5. The dashed lines divide $\{0, \ldots, n-1\}$ into the 7 different regions used in the construction.

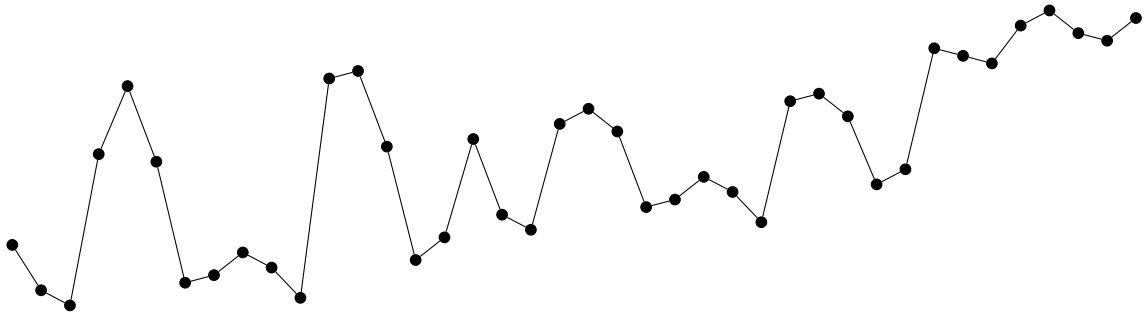
A.2 Bicrucial permutations for small even n



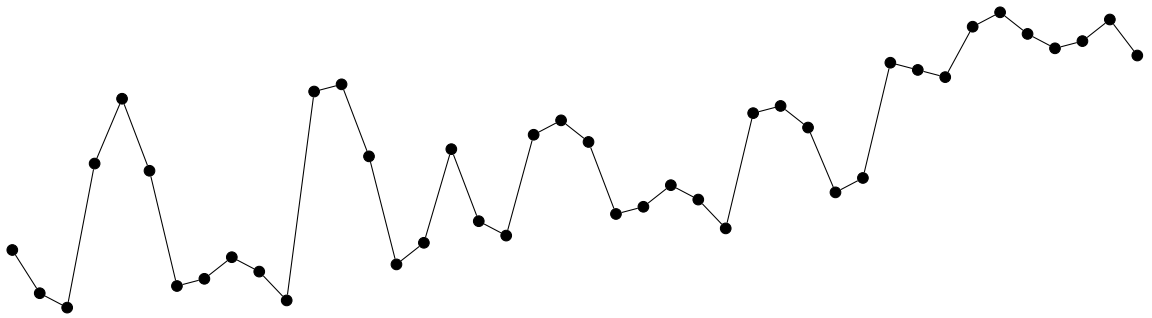
(a)



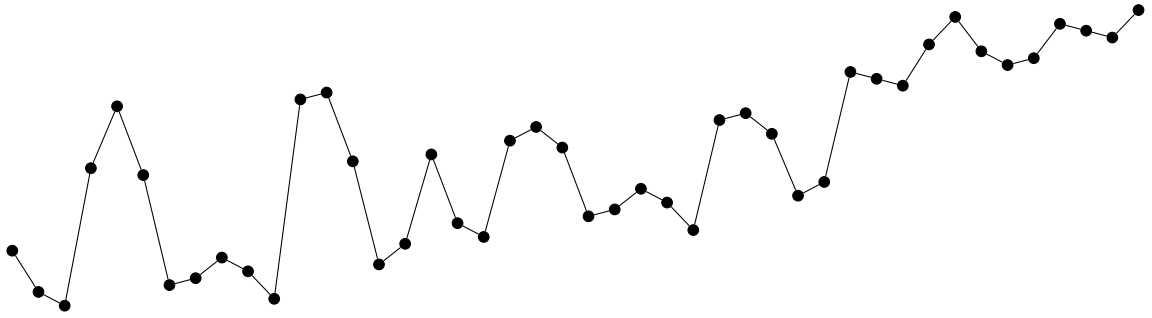
(b)



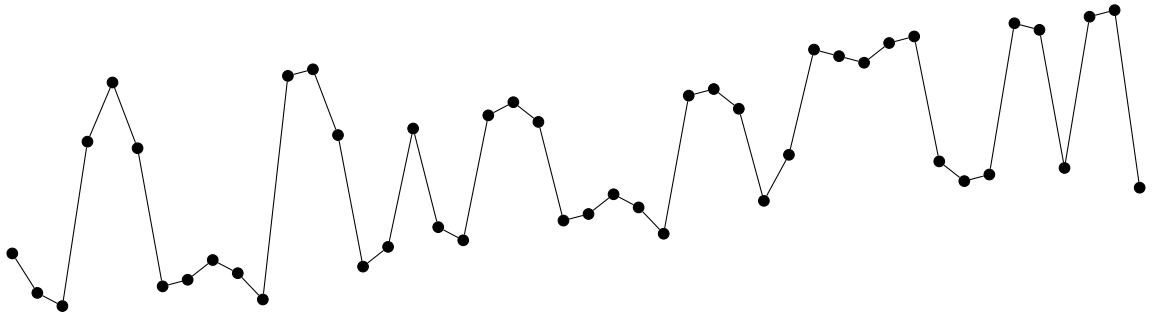
(c)



(d)



(e)



(f)

Figure A.2: Bicrucial permutations of lengths 34, 36, 40, 42, 44 and 46 respectively.

Appendix B

Best known bounds on $U(n)$

n	$\binom{n}{2}$	Upper bound	Proportion of $\binom{n}{2}$
2	1	1	1.000
3	3	3	1.000
4	6	6	1.000
5	10	10	1.000
6	15	14	0.933
7	21	20	0.952
8	28	26	0.929
9	36	33	0.917
10	45	41	0.911
11	55	50	0.909
12	66	58	0.879
13	78	69	0.885
14	91	80	0.879
15	105	91	0.867
16	120	103	0.858
17	136	117	0.860
18	153	130	0.850
19	171	145	0.848
20	190	160	0.842

Table B.1: The best known upper bounds for $U(n)$. The upper bound is known to be correct for $n = 2, 3, 4$ and computer experiments suggests it holds for $n = 5$ as well [5].

Bibliography

[1] M. Ajtai, J. Komlós and E. Szemerédi. Sorting in $c\log(n)$ parallel steps. *Combinatorica*, **3**(1):1–19, 1983.

[2] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge university press, 2003.

[3] N. Alon. A simple algorithm for edge-coloring bipartite multigraphs. *Information Processing Letters*, **85**(6):301–302, 2003.

[4] N. Alon, Y. Caro and I. Krasikov. Bisection of trees and sequences. *Discrete Mathematics*, **114**(1):3–7, 1993.

[5] O. Angel and A. E. Holroyd. Perfect shuffling by lazy swaps. *Electronic Communications in Probability*, **23**, 2018.

[6] A. R. Arévalo, A. Montejano and E. Roldán-Pensado. Zero-sum squares in bounded discrepancy $\{-1, 1\}$-matrices. *arXiv preprint arXiv:2005.07813*, 2020.

[7] S. Avgustinovich, S. Kitaev, A. Pyatkin and A. Valyuzhenich. On square-free permutations. *Journal of Automata, Languages and Combinatorics*, **16**(1):3–10, 2011.

[8] M. Axenovich and J. Manske. On monochromatic subsets of a rectangular grid. *Integers*, **8**(1):A21, 2008.

[9] R. Bacher and S. Eliahou. Extremal binary matrices without constant 2-squares. *Journal of Combinatorics*, **1**(1):77–100, 2010.

[10] J. Backelin, J. West and G. Xin. Wilf-equivalence for singleton classes. *Advances in Applied Mathematics*, **38**(2):133–148, 2007.

[11] P. Balister, Y. Caro, C. Rousseau and R. Yuster. Zero-sum square matrices. *European Journal of Combinatorics*, **23**(5):489–497, 2002.

[12] D. Bean, A. Ehrenfeucht and G. McNulty. Avoidable patterns in strings of symbols. *Pacific Journal of Mathematics*, **85**(2):261–294, 1979.

[13] B. Beauquier and E. Darrot. On arbitrary size Waksman networks and their vulnerability. *Parallel Processing Letters*, **12**(03n04):287–296, 2002.

[14] M. Ben-Or and N. Linial. Collective coin flipping. *Advances in Computing Research*, **5**:91–115, 1989.

[15] I. Benjamini, G. Cohen and I. Shinkar. Bi-Lipschitz bijection between the boolean cube and the Hamming ball. *Israel Journal of Mathematics*, **212**:677–703, 2016.

[16] I. Benjamini, G. Kalai and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, **90**(1):5–43, 1999.

[17] A. Bernasconi. Mathematical techniques for the analysis of boolean functions. *Bulletin of the European Association for Theoretical Computer Science*, **65**:228–230, 1998.

[18] A. Bialostocki. Zero sum trees: A survey of results and open problems. In N. W. Sauer, R. E. Woodrow and B. Sands, editors, *Finite and Infinite Combinatorics in Sets and Logic*, pages 19–29. Springer, 1993.

[19] L. Boczkowski and I. Shinkar. On mappings on the hypercube with small average stretch. *arXiv preprint arXiv:1905.11350*, 2019.

[20] B. Bollobás. *Combinatorics: Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability*. Cambridge University Press, 1986.

[21] B. Bollobás and G. Brightwell. The height of a random partial order: concentration of measure. *The Annals of Applied Probability*, **2**(4):1009–1018, 1992.

[22] M. Bóna. A survey of stack-sorting disciplines. *The Electronic Journal of Combinatorics*, **9**:A1, 2002.

[23] R. B. Boppana. The average sensitivity of bounded-depth circuits. *Information processing letters*, **63**(5):257–261, 1997.

[24] J. Bourgain. On the distribution of the Fourier spectrum of boolean functions. *Israel Journal of Mathematics*, **131**(1):269–276, 2002.

[25] B. Bukh and O. Rudenko. Order-isomorphic twins in permutations. *SIAM Journal on Discrete Mathematics*, **34**(3):1620–1622, 2020.

[26] Y. Caro, A. Hansberg, J. Lauri and C. Zarb. On zero-sum spanning trees and zero-sum connectivity. *arXiv preprint arXiv:2007.08240*, 2020.

[27] Y. Caro, A. Hansberg and A. Montejano. Zero-sum K_m over $\mathbb{Z}$ and the story of K_4. *Graphs and Combinatorics*, **35**(4):855–865, 2019.

[28] Y. Caro, A. Hansberg and A. Montejano. Zero-sum subsequences in bounded-sum $\{-1, 1\}$-sequences. *Journal of Combinatorial Theory, Series A*, **161**:387–419, 2019.

[29] Y. Caro and R. Yuster. On zero-sum and almost zero-sum subgraphs over $\mathbb{Z}$. *Graphs and Combinatorics*, **32**(1):49–63, 2016.

[30] F. Chung, P. Erdős and R. L. Graham. Minimal decompositions of graphs into mutually isomorphic subgraphs. *Combinatorica*, **1**(1):13–24, 1981.

[31] F. Chung, P. Erdős and R. L. Graham. Minimal decomposition of all graphs with equinumerous vertices and edges into mutually isomorphic subgraphs. In *Finite and Infinite Sets*, volume 37 of *Colloquia Mathematica Societatis János Bolyai*, pages 171–179. North-Holland, 1984.

[32] F. Chung, P. Erdős, R. L. Graham, S. Ulam and F. Yao. Minimal decompositions of two graphs into pairwise isomorphic subgraphs. In *Proceedings of the Tenth Southeastern Conference on Combinatorics, Graph Theory and Computing*, pages 3–18, 1979.

[33] A. Claesson and S. Kitaev. Classification of bijections between 321-and 132-avoiding permutations. *arXiv preprint arXiv:0805.1325*, 2008.

[34] A. Czumaj. Random permutations using switching networks. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 703–712. Association for Computing Machinery, 2015.

[35] W. J. Dally and B. P. Towles. *Principles and Practices of Interconnection Networks*. Elsevier, 2004.

[36] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, pages 439–485, 2005.

[37] A. Dudek, J. Grytczuk and A. Ruciński. Variations on twins in permutations. *arXiv preprint arXiv:2001.05589*, 2020.

[38] A. Dudek, J. Grytczuk and A. Ruciński. Tight multiple twins in permutations. *arXiv preprint arXiv:2105.02591*, 2021.

[39] A. Dudek, J. Grytczuk and A. Ruciński. Variations on twins in permutations. *Electronic Journal of Combinatorics*, **28**(3):P3.19, 2021.

[40] P. Erdős, A. Ginzburg and A. Ziv. Theorem in the additive number theory. *Bulletin of the Research Council of Israel F*, **10**:41–43, 1961.

[41] P. Erdős and A. C. Offord. On the number of real roots of a random algebraic equation. *Proceedings of the London Mathematical Society*, **s3-6**(1):139–160, 1956.

[42] P. Erdős and A. Rényi. On random matrices. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, **8**:455–461, 1963.

[43] P. Erdős, J. Pach and L. Pyber. Isomorphic subgraphs in a graph. In *Combinatorics (Eger, 1987). Colloquia Mathematica Societatis János Bolyai*, volume 52, pages 553–556, 1988.

[44] M. Erickson. *Introduction to Combinatorics*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 1996.

[45] J. Fitzsimons. What is the most efficient way to generate a random permutation from probabilistic pairwise swaps? https://cstheory.stackexchange.com/questions/5321/what-is-the-most-efficient-way-to-generate-a-random-permutation-from-probabilist, 2011. [Accessed 2020-06-16.].

[46] A. Frieze. On the length of the longest monotone subsequence in a random permutation. *The Annals of Applied Probability*, **1**(2):301–305, 1991.

[47] W. Gao and A. Geroldinger. Zero-sum problems in finite abelian groups: a survey. *Expositiones Mathematicae*, **24**(4):337–369, 2006.

[48] M. Gawron. Izomorficzne podstruktury w słowach i permutacjach. Master's thesis, Uniwersytet Jagielloński, 2014.

[49] I. Gent, S. Kitaev, A. Konovalov, S. Linton and P. Nightingale. *S*-crucial and bicrucial permutations with respect to squares. *Journal of Integer Sequences*, **18**(6), 2015.

[50] R. Graham. Reflections on a theme of Ulam. In R. Gera, T. W. Haynes and S. T. Hedetniemi, editors, *Graph Theory: Favorite Conjectures and Open Problems. 2*, Problem Books in Mathematics, pages 55–61. Springer, 2018.

[51] C. Groenland. *Structure for algorithms, graph reconstruction and hypercube intersections*. DPhil thesis, University of Oxford, 2020.

[52] C. Groenland and T. Johnston. Intersection sizes of linear subspaces with the hypercube. *arXiv preprint arXiv:810.02729*, 2018.

[53] C. Groenland and T. Johnston. Intersection sizes of linear subspaces with the hypercube. *Journal of Combinatorial Theory, Series A*, **170**:105–142, 2020.

[54] C. Groenland and T. Johnston. The lengths for which bicrucial square-free permutations exist. *arXiv preprint arXiv:2109.00502*, 2021.

[55] J. Grytczuk, H. Kordulewski and A. Niewiadomski. Extremal square-free words. *Electronic Journal of Combinatorics*, **27**(1), 2020.

[56] F. Harary and R. W. Robinson. Isomorphic factorizations VIII: Bisectable trees. *Combinatorica*, **4**(2-3):169–179, 1984.

[57] F. Harary and R. W. Robinson. Isomorphic factorizations X: Unsolved problems. *Journal of Graph Theory*, **9**(1):67–86, 1985.

[58] J. Håstad, T. Leighton and M. Newman. Reconfiguring a hypercube in the presence of faults. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, pages 274–284. ACM, 1987.

[59] L. Hong and S. Zhang. No extremal square-free words over large alphabets. *arXiv preprint arXiv:2107.13123*, 2021.

[60] J. E. Hopcroft and R. M. Karp. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing*, **2**(4):225–231, 1973.

[61] S. Hui and C. Park. The representation of hypergeometric random variables using independent Bernoulli random variables. *Communications in Statistics-Theory and Methods*, **43**(19):4103–4108, 2014.

[62] A. Ignatiev, A. Morgado and J. Marques-Silva. PySAT: A Python toolkit for prototyping with SAT oracles. In *SAT*, pages 428–437, 2018.

[63] T. Johnston. Zero-sum squares in $\{-1, 1\}$-matrices with low discrepancy. *arXiv preprint arXiv:2010.10310*, 2020.

[64] T. Johnston and A. Scott. Lipschitz bijections between boolean functions. *Combinatorics, Probability and Computing*, **30**(4):513–525, 2021.

[65] J. Kahn, G. Kalai and N. Linial. The influence of variables on boolean functions. In *Proceedings of the Twenty-Ninth Annual Symposium on Foundations of Computer Science*. IEEE, 1988.

[66] G. Kindler, N. Kirshner and R. O'Donnell. Gaussian noise sensitivity and Fourier tails. *Israel Journal of Mathematics*, **225**(1):71–109, 2018.

[67] S. Kitaev. Personal communication at *Permutation Patterns 2021*.

[68] D. E. Knuth. *The Art of Computer Programming*, volume 1. Pearson Education, 1997.

[69] D. E. Knuth. *The Art of Computer Programming*, volume 3. Pearson Education, 1997.

[70] C. Lee, P.-S. Loh and B. Sudakov. Self-similarity of graphs. *SIAM Journal on Discrete Mathematics*, **27**(2):959–972, 2013.

[71] T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica*, **9**(2):161–187, 1989.

[72] S.-Y. R. Li. Annihilators in nonrepetitive semigroups. *Studies in Applied Mathematics*, **55**(1):83–85, 1976.

[73] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. *Journal of the London Mathematical Society*, **s1-13**(4):288–295, 1938.

[74] S. Lovett and E. Viola. Bounded-depth circuits cannot sample good codes. In *Computational Complexity*, volume 21, pages 245–266, 2012.

[75] P. A. MacMahon. *Combinatory Analysis*, volume 1. Cambridge University Press, 1915.

[76] A. Marcus and G. Tardos. Excluded permutation matrices and the Stanley-Wilf conjecture. *Journal of Combinatorial Theory, Series A*, **107**(1):153–160, 2004.

[77] N. Melo and A. Winter. Intersection patterns of linear subspaces with the hypercube. *Journal of Combinatorial Theory, Series A*, **164**:60–71, 2019.

[78] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[79] L. Mol and N. Rampersad. Lengths of extremal square-free ternary words. *Contributions to Discrete Mathematics*, to appear.

[80] E. Mossel, R. O'Donnell and K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, **171**:295–341, 2010.

[81] Y. Nakata, C. Hirche, M. Koashi and A. Winter. Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics. *Physical Review X*, **7**(2):021006, 2017.

[82] P. S. Novikov and S. I. Adjan. Infinite periodic groups. I. *Mathematics of the USSR-Izvestiya*, **2**(1):209–236, 1968.

[83] P. S. Novikov and S. I. Adjan. Infinite periodic groups. II. *Mathematics of the USSR-Izvestiya*, **2**(2):241–479, 1968.

[84] P. S. Novikov and S. I. Adjan. Infinite periodic groups. III. *Mathematics of the USSR-Izvestiya*, **2**(3):665–685, 1968.

[85] S. Rao and I. Shinkar. On Lipschitz bijections between boolean functions. *Combinatorics, Probability and Computing*, **27**:411–426, 2018.

[86] A. Schrijver. *Combinatorial optimization: Polyhedra and Efficiency*. Number 24 in Algorithms and Combinatorics. Springer, 2003.

[87] Z. Stankova and J. West. A new class of Wilf-equivalent permutations. *Journal of Algebraic Combinatorics*, **15**(3):271–290, 2002.

[88] Z. E. Stankova. Forbidden subsequences. *Discrete Mathematics*, **132**(1-3):291–316, 1994.

[89] J. M. Steele. Variations on the monotone subsequence theme of Erdős and Szekeres. In D. Aldous, P. Diaconis, J. Spencer and J. M. Steele, editors, *Discrete Probability and Algorithms*, pages 111–131. Springer, 1995.

[90] The OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences, 2021.

[91] A. Thue. Uber unendliche zeichenreihen. *Skrifter udgivne af Videnskabsselskabet i Christiania. I Mathematisk-naturvidenskabelig klasse.*, **7**:1–22, 1906.

[92] A. Waksman. A permutation network. *Journal of the ACM (JACM)*, **15**(1):159–163, 1968.

[93] A. C.-C. Yao. Bounds on selection networks. *SIAM Journal on Computing*, **9**(3):566–582, 1980.