Three Problems Involving Permutations

by

Justin H.C. Chan

M.Sc., University of Victoria, 2010
B.Sc., University of Victoria, 2007

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Mathematics
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2016

Approval

Name: **Justin H.C. Chan**

Degree: **Doctor of Philosophy (Mathematics)**

Title: ***Three Problems Involving Permutations***

Examining Committee: **Chair:** Matthew DeVos
Associate Professor

Jonathan Jedwab
Senior Supervisor
Professor

Petr Lisoněk
Supervisor
Professor

Marni Mishna
Internal Examiner
Associate Professor

Ronald Graham
External Examiner
Professor
Department of Computer Science and Engineering
University of California, San Diego

Date Defended: 5 December 2016

Abstract

We study three problems involving permutations: the n-card problem, inv-Wilf-equivalence, and suitable sets of permutations.

The n-card problem is to determine the minimal intervals $[u, v]$ such that for every $n \times n$ stochastic matrix A there is an $n \times n$ permutation matrix P (depending on A) such that $\mathrm{tr}(PA) \in [u, v]$. The minimal intervals for the n-card problem are known only for $n \leq 4$. We answer a question posed by Sands, by showing that $[1, 2]$ is a solution to the n-card problem for all $n \geq 2$. We also show that each closed interval of length $\frac{n}{n-1}$ contained in $[0, 2)$ is a solution to the n-card problem for all $n \geq 2$.

Wilf-equivalence is one of the central concepts of pattern-avoiding permutations. The two known infinite families of Wilf-equivalent permutation pairs both satisfy the stronger condition of shape-Wilf-equivalence. Dokos et al. studied a different strengthening of Wilf-equivalence called inv-Wilf-equivalence. They conjectured that all inv-Wilf-equivalent permutation pairs arise from trivial symmetries. We disprove this conjecture with an infinite family of counterexamples, obtained by generalizing simultaneously the concepts of shape-Wilf-equivalence and inv-Wilf-equivalence. We also prove the Baxter-Jaggard conjecture on even-shape-Wilf-equivalent permutation pairs.

A set of N permutations of $\{1, 2, \ldots, v\}$ is (N, v, t)-suitable if each symbol precedes each subset of $t-1$ others in at least one permutation. We give examples of suitable sets of permutations for new parameter triples (N, v, t). We relate certain suitable sets of permutations with parameter t to others with parameter $t+1$, thereby showing that one of the two infinite families presented by Colbourn can be constructed directly from the other. We prove an exact nonexistence result for suitable sets of permutations using elementary combinatorial arguments. We then establish an asymptotic nonexistence result using Ramsey's theorem.

Keywords: permutation matrix, n-card problem, pattern-avoiding permutation, Wilf-equivalence, suitable sets of permutations, Ramsey's theorem

Table of Contents

List of Tables

List of Figures

Chapter 1

Overview

A *permutation* σ of length n is a sequence of n numbers, denoted $\sigma_1\sigma_2\ldots\sigma_n$, which contains each element of $[n] = \{1, 2, \ldots, n\}$ exactly once. For example, 1423 is a permutation of length 4. A permutation σ may be regarded as a bijective function that maps $[n]$ to $[n]$, defined by $\sigma(i) = \sigma_i$. The set of all permutations of length n is denoted S_n.

There are two further representations of the permutation σ:

- An $n \times n$ permutation matrix P, where the entry in the ith row and jth column of P is 1 if $\sigma_j = i$ and 0 otherwise.

- An $n \times n$ square diagram of cells, such that each cell either contains a dot or is blank, and the cell in the ith row from the top and jth column from the left has a dot if and only if $\sigma_j = i$.

Figure 1.1 shows these two representations for the permutation 1423. Note that, in this thesis, the row labels in the square diagram increase from top to bottom; some of the literature (for example, [23]) uses the opposite labelling convention.

Permutations occur in many areas of mathematics, including group theory (every group is isomorphic to a group of permutations) and Galois theory (which uses permutation groups to describe relationships between roots of a polynomial equation). Permutations are important in

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

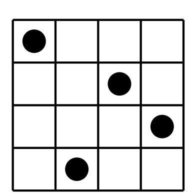

Figure 1.1: Permutation matrix and square diagram representing permutation 1423

information science, for example in the design of sorting algorithms, hash functions, and error correction code interleavers. Permutations also underlie the solution of puzzles such as Rubik's cube and the 15-puzzle, and the design of sequences of permutations to be applied to tuned bells in change ringing.

We study three problems involving permutations: the n-card problem (Chapter 2), inv-Wilf-equivalence of permutations (Chapter 3), and suitable sets of permutations (Chapter 4). The main results of these chapters are contained in the published papers [17] and [16], and in the accepted paper [18].

The n-card problem is to determine the minimal intervals $[u, v]$ such that for every $n \times n$ stochastic matrix A there is an $n \times n$ permutation matrix P (depending on A) such that $\operatorname{tr}(PA) \in [u, v]$. This problem is closely related to classical mathematical problems from industry and management involving the possible values of $\operatorname{tr}(PA)$ for a permutation matrix P and a fixed square matrix A, including the linear assignment problem and the travelling salesman problem. The linear assignment problem, "one of the most famous problems in linear programming and in combinatorial optimization [15]," is to minimize $\operatorname{tr}(PA)$ over all permutation matrices P (see [15, Chapter 4] for a detailed historical account of the development of algorithms for its solution, and an equivalent formulation in terms of weighted bipartite matchings); the travelling salesman problem is the special case in which the permutation corresponding to P is cyclic [32].

The minimal intervals for the n-card problem are known only for $n \leq 4$. We introduce a new method of analysis for the n-card problem that makes repeated use of the Extreme Principle. We use this method to answer a question posed by Sands [54], by showing that $[1, 2]$ is a solution to the n-card problem for all $n \geq 2$. We also show that each closed interval of length $\frac{n}{n-1}$ contained in $[0, 2)$ is a solution to the n-card problem for all $n \geq 2$. Brualdi and Fritscher [14] reported that these results motivated their study of the possible sets of traces of a $(0, 1)$-matrix, and linked that study to the problem of determining the possible values of the "quality assignments" of n people to n jobs.

Wilf-equivalence is one of the central concepts of pattern-avoiding permutations, and has been studied for more than thirty years. The two known infinite families of Wilf-equivalent permutation pairs, due to Stankova and West [59] and Backelin, West, and Xin [3], both satisfy the stronger condition of shape-Wilf-equivalence. In 2012, Dokos et al. [23] studied a different strengthening of Wilf-equivalence called inv-Wilf-equivalence, which takes account of the inversion number of a permutation. They conjectured that all inv-Wilf-equivalent permutation pairs arise from trivial symmetries. We disprove this conjecture by constructing an infinite family of counterexamples derived from the permutation pair (231) and (312). The key to this construction is to generalize simultaneously the concepts of shape-Wilf-equivalence and inv-Wilf-equivalence. A further consequence is a proof of the 2011 conjecture by Baxter and Jaggard [4] on even-shape-Wilf-equivalent permutation pairs.

A set of N permutations of $[v] = \{1, 2, \ldots, v\}$ is (N, v, t)-suitable if each symbol precedes each subset of $t-1$ others in at least one permutation. The central problems are to determine the smallest N for which such a set exists for given v and t, and to determine the largest v for which such a set exists for given N and t. These extremal problems were the subject of classical studies by Dushnik [25] in 1950 and Spencer [56] in 1971, as well as by Colbourn [22] who found connections to many other combinatorial objects, including binary covering arrays, directed packings, Golomb rulers, and Paley matrices. We give examples of suitable sets of permutations for new parameter triples (N, v, t), found both by hand and by interactive computer search. We relate certain suitable sets of permutations with parameter t to others with parameter $t+1$, thereby showing that one of the two infinite families presented by Colbourn [22] in 2015 can be constructed directly from the other. We give small examples of suitable sets of permutations which suggest the existence of an infinite family. However, we show that this apparent existence pattern is misleading, by proving an exact nonexistence result for sufficiently large suitable sets of permutations using only elementary combinatorial arguments. We then establish an asymptotic nonexistence result by introducing a new tool for the study of suitable sets of permutations, namely Ramsey's theorem [49].

Chapter 2

The n-card problem

2.1 Introduction

Let $n \geq 2$ be an integer. An $n \times n$ *stochastic matrix* is an $n \times n$ matrix (a_{ij}) of non-negative real numbers, each of whose row sums is 1. A *transversal sum* of (a_{ij}) is a sum of the form $\sum_{i=1}^{n} a_{\sigma(i),i}$, for some permutation σ of $\{1, 2, \ldots, n\}$, or in other words, the sum of all the entries contained in a particular transversal of the matrix. The matrix (a_{ij}) has $n!$ transversal sums, some of which may be repeated.

For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

is a 3×3 stochastic matrix with two transversal sums of 0 and four transversal sums of $\frac{3}{2}$.

The "n-card problem" is given in [45] in the form of numbers written in a column on cards:

> Let $n \geq 4$ and suppose you are given n cards, each containing n nonnegative real numbers written in a column, for which the sum of the numbers on each card is 1. You may place the cards in any order, and take the first number of the first card, the second number of the second card, ..., the nth number of the nth card, and add them together. What are the minimal intervals $[a, b]$ such that, no matter which cards are given, there is an ordering such that the sum will lie in $[a, b]$?

We recast the problem (expressed using rows instead of columns) in terms of stochastic matrices and transversal sums. For a fixed $n \geq 2$, a *solution to the n-card problem* is a closed interval $[u, v]$ (of nonzero length) such that every $n \times n$ stochastic matrix contains at least one transversal sum in $[u, v]$. Equivalently, we may say that a solution to the n-card problem is a closed interval $[u, v]$ such that for every $n \times n$ stochastic matrix A there is an $n \times n$ permutation matrix P such that

$\text{tr}(PA) \in [u, v]$. A *minimal solution of the n-card problem* is a solution $[u, v]$ to the n-card problem for which no proper closed subinterval of $[u, v]$ is a solution to the n-card problem.

Solutions to the n-card problem, not necessarily minimal, trivially exist. For example, $[0, n]$ is a solution to the n-card problem for all $n \geq 2$; every transversal sum of a stochastic $n \times n$ matrix lies in $[0, n]$, since every entry of the matrix lies in $[0, 1]$. However, the intervals that we show are solutions to the n-card problem have a significantly smaller interval length. We will also show that minimal solutions to the n-card problem exist and have length as small as 1 (for example, $[1, 2]$ as shown in Theorem 2.2.1), but no smaller (as shown by Lemma 2.1.2).

The problem of finding all solutions to the n-card problem can be reduced to finding all minimal solutions to the n-card problem, as follows from the next proposition combined with the fact that every interval that contains a minimal solution to the n-card problem is clearly a solution to the n-card problem.

Proposition 2.1.1. *Every solution $[a, b]$ to the n-card problem contains as a subinterval (or is equal to) a minimal solution to the n-card problem.*

Proof. Firstly, note that, of the matrices

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix},$$

the left matrix has transversal sums only of 1, whereas the right matrix has transversal sums only of $\frac{1}{n}$ and $1 + \frac{1}{n}$.

Let $[a, b]$ be a solution to the n-card problem. For an $n \times n$ stochastic matrix A, we may then define $m(A)$ to be the smallest transversal sum of A that lies in $[a, b]$. Let S be the set of all $m(A)$ as A ranges over all $n \times n$ stochastic matrices. Now all elements of S lie in $[a, b]$ and there are at least two distinct elements of S because of the above matrices. So then $a < \sup S \leq b$, and so $[a, \sup S]$ is a solution to the n-card problem whose right endpoint cannot be made smaller.

We then define $m'(A)$ to be the largest transversal sum of A that lies in $[a, \sup S]$. Let S' be the set of all $m'(A)$ as A ranges over all $n \times n$ stochastic matrices. Now all elements of S' lie in $[a, \sup S]$ and there are at least two distinct elements of S' because of the above matrices. So then $a \leq \inf S' < \sup S$ and so $[\inf S', \sup S]$ is a solution to the n-card problem whose left endpoint cannot be made larger. Therefore $[\inf S', \sup S]$ is a minimal solution to the n-card problem which is contained as a subinterval in $[a, b]$ (or is equal to $[a, b]$). □

The following lemma, which first appeared in [54], is a necessary condition that will be used frequently to show that solutions to the n-card problem are minimal.

Lemma 2.1.2 ([54]). *Every solution to the n-card problem must contain as a subinterval at least one interval in the set*

$$\left\{\left[\frac{k}{n-1}, 1+\frac{k}{n-1}\right] : k = 0, 1, \dots, n-1\right\}.$$

Proof. Let $[a, b]$ be a solution to the n-card problem.

Firstly, $[a, b]$ must contain 1 because the matrix

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix}$$

contains transversal sums only of 1.

If $a = 0$, then $[a, b]$ contains $[0, 1]$ which gives the result for $k = 0$. Otherwise $0 < a \leq 1$. Let i be the integer between 1 and $n-1$ such that

$$\frac{i-1}{n-1} < a \leq \frac{i}{n-1}.$$

Then the matrix consisting of i rows of

$$\begin{bmatrix} 0 & \frac{1}{n-1} & \frac{1}{n-1} & \dots & \frac{1}{n-1} \end{bmatrix}$$

and $n-i$ rows of

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix}$$

has transversal sums only of $\frac{i-1}{n-1}$ and $1 + \frac{i}{n-1}$. Then since $\frac{i-1}{n-1}$ does not lie in $[a, b]$, it follows that $1 + \frac{i}{n-1}$ does, and so $[a, b]$ contains $[\frac{i}{n-1}, 1 + \frac{i}{n-1}]$ which gives the result for $k = i$. ☐

We now turn to showing that intervals are solutions to the n-card problem. The following lemma is useful in this regard, and along with Lemma 2.1.2 immediately shows that $[0, 1]$ is a minimal solution to the n-card problem for all $n \geq 2$ (although it is not the only minimal solution).

Lemma 2.1.3. *In an $n \times n$ stochastic matrix A, there exists a transversal sum which is at most 1, and a transversal sum which is at least 1.*

Proof. Consider the n cyclic shifts of the diagonal transversal $\{a_{11}, a_{22} \dots, a_{nn}\}$. The sum of all the corresponding transversal sums is n, and so their mean is 1. Therefore at least one of these transversal sums is at most 1, and at least one is at least 1. ☐

Corollary 2.1.4 ([54] Lemma 2.1)**.** $[0,1]$ *is a minimal solution to the* n*-card problem for all* $n \geq 2$.

Proof. By Lemma 2.1.3, there exists a transversal sum which is at most 1. Therefore $[0,1]$ is a solution to the n-card problem for $n \geq 2$. This interval cannot be made smaller because of Lemma 2.1.2. □

2.2 History of the n-card problem

A specific version of the n-card problem first appeared as Problem 2620 in *Crux Mathematicorum with Mathematical Mayhem* [53], which asks for a proof that $[\frac{1}{2}, \frac{3}{2}]$ is a solution to the 3-card problem. In 2003, Lenza found a new proof of Problem 2620, whose core concept was extended to establish:

- $[\frac{1}{3}, \frac{4}{3}]$ is a solution to the 4-card problem [45].
- $[1, 2]$ is a solution to the 3-card problem [45] and the 4-card problem [54].
- $[\frac{2}{3}, \frac{5}{3}]$ is a solution to the 4-card problem [54].

These results, along with Corollary 2.1.4 and Lemma 2.1.2, show that the minimal solutions to the 3-card problem are

$$[0,1], \quad [\tfrac{1}{2}, \tfrac{3}{2}], \quad [1,2], \tag{2.1}$$

and the minimal solutions to the 4-card problem are

$$[0,1], \quad [\tfrac{1}{3}, \tfrac{4}{3}], \quad [\tfrac{2}{3}, \tfrac{5}{3}], \quad [1,2]. \tag{2.2}$$

We note that the 2-card problem, left as an exercise in [45], has the minimal solutions

$$[0,1], \quad [1,2]. \tag{2.3}$$

This is because of Lemma 2.1.2 and the fact that every 2×2 stochastic matrix is of the form

$$\begin{bmatrix} a & 1-a \\ 1-b & b \end{bmatrix},$$

with transversal sums of $a + b$ and $(1-a) + (1-b) = 2 - (a+b)$.

The method of [45] and [54] is to establish a particular interval $[u, v]$ as a solution to the n-card problem (for $n = 3$ or 4), by using intersecting permutations and careful counting to show that the maximum number of transversal sums greater than v plus the maximum number of transversal sums less than u is less than $n!$. For $n = 3$ we give a short proof that $[1, 2]$ is a solution, based on these lines, in Section 2.3.

However, the method of [45] and [54] relies on a laborious case analysis for each of the intervals in (2.2) ($n = 4$) other than $[0, 1]$. More importantly, it does not extend to $n \geq 5$ [45, p.6]. For example, regarding the interval $[1, 2]$ in the 5-card problem, the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

has 96 transversal sums of $\frac{3}{4}$, whereas the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{bmatrix}$$

has 24 transversal sums of $2 + \frac{1}{5}$, for a total of $120 = 5!$ transversal sums lying outside the interval $[1, 2]$; therefore the method of intersecting permutations fails to show that $[1, 2]$ is a solution to the 5-card problem.

In this chapter we introduce a new method for analyzing the n-card problem that makes repeated use of the Extreme Principle [64]. We believe that this method could shed light on other problems involving $\mathrm{tr}(PA)$, where P is a permutation matrix and A is a fixed square matrix. The Extreme Principle directs attention to the "largest" and "smallest" elements of a problem. In the present context, we assume for a contradiction that no transversal sum of an $n \times n$ stochastic matrix lies in some interval $[u, v]$, and then consider the smallest transversal sum d greater than v. Then, if a transversal sum is less than d, it must be less than u. We seek such transversal sums, involving exactly $n - 2$ of the original summands of d, in order to reach a contradiction. We thereby obtain strong new restrictions for all $n \geq 5$. In particular, we solve Problem 5 of [54] as follows.

Theorem 2.2.1. *For all $n \geq 2$, the interval $[1, 2]$ is a solution to the n-card problem.*

A proof of this theorem is given in Section 2.4.

Another problem of interest is to find the maximum possible length of a minimal solution. The maximum possible length is at most 2, as shown by the following proposition which is a simple generalization of Lemma 5.1 of [54]:

Proposition 2.2.2. *For all $n \geq 2$, each closed interval of length 2 containing 1 is a solution to the n-card problem.*

Proof. Let $[a, a+2]$ be a closed interval containing 1 and suppose for a contradiction that a stochastic matrix A exists with no transversal sum in $[a, a+2]$. From Lemma 2.1.3, A has a transversal T_1 whose sum is less than a and a transversal T_2 whose sum is greater than $a+2$. Now transposing two rows of a transversal changes the sum by at most 2. So in a sequence of row transpositions that takes T_1 to T_2, there is an intermediate transversal whose sum is in $[a, a+2]$, which is a contradiction. □

However, the following result significantly strengthens Proposition 2.2.2, showing that minimal solutions are no longer than $\frac{n}{n-1}$.

Theorem 2.2.3. *For all $n \geq 4$, each closed interval of length $\frac{n}{n-1}$ contained in $[0, 2)$ is a solution to the n-card problem.*

A proof of this theorem is given in Section 2.5.

The following corollary is a consequence of the above theorem as well as Theorem 2.2.1 and previous results for the 2- and 3-card problems.

Corollary 2.2.4. *For all $n \geq 2$, each closed interval of length $\frac{n}{n-1}$ containing 1 is a solution to the n-card problem.*

By Lemma 2.1.2, the length $\frac{n}{n-1}$ in Corollary 2.2.4 is the smallest possible ℓ for which every closed interval of length ℓ containing 1 is a solution to the n-card problem. On the other hand, the known complete set of minimal solutions (2.1) for $n = 3$ and (2.2) for $n = 4$ show that such intervals of length $\frac{n}{n-1}$ are not necessarily minimal solutions.

2.3 Proof that $[1, 2]$ is a solution to the 3-card problem

The following proposition was stated without proof in [45]. Here we give a short proof along similar lines to the proofs used in [45] and [54].

Proposition 2.3.1. $[1, 2]$ *is a solution to the* 3*-card problem.*

Proof. Suppose, for a contradiction, that $A = (a_{ij})$ is a 3×3 stochastic matrix having no transversal sum in $[1, 2]$. Consider the set of transversals T_1 consisting of the three cyclic shifts of $\{a_{11}, a_{22}, a_{33}\}$, as well as the set of transversals T_2 consisting of the three cyclic shifts of $\{a_{11}, a_{23}, a_{32}\}$. Since T_1 comprises all entries of A, the sum of all its transversal sums is 3. So at most two of the transversal sums in T_1 can be less than 1. Likewise, at most two of the transversal sums in T_2 can be less than 1. In total, at most four of the six transversal sums of A can be less than 1.

So by assumption there exist at least two transversal sums of A which are greater than 2; let T be a set of two such corresponding transversals. The transversals of T are not disjoint; otherwise the sum of all entries in A, which include all the entries of the transversals of T, is greater than 4 (contradicting that the sum of the entries of A is 3). Therefore the transversals of T coincide at exactly one entry. Rearrange the rows and columns so that the transversals of T are $\{a_{11}, a_{22}, a_{33}\}$ and $\{a_{11}, a_{23}, a_{32}\}$. Since $0 \leq a_{11} \leq 1$, we have $a_{23} + a_{32} > 1$, and so the sum of the entries of A is at least $(a_{11} + a_{22} + a_{33}) + (a_{23} + a_{32}) > 2 + 1 = 3$, which is a contradiction. □

2.4 Proof of Theorem 2.2.1

We firstly establish some preliminary lemmas.

Lemma 2.4.1. *Let (a_{ij}) be an $n \times n$ stochastic matrix, all of whose transversal sums lie outside an interval $[u, v]$ containing 1. Then (a_{ij}) has at least one transversal sum less than u, and at least one transversal sum greater than v.*

Proof. Follows from Lemma 2.1.3. □

The rows and columns of an $n \times n$ stochastic matrix can be permuted without changing the set of its $n!$ transversal sums. Our method relies on examining the effect of transposing two rows of an $n \times n$ stochastic matrix, and thereby bounding the matrix entries. We now show that if an $n \times n$ stochastic matrix has a sufficiently large diagonal sum then there must be a transposition of two rows that decreases this diagonal sum. We prove this result for the following slightly more general case of an $n \times n$ substochastic matrix (each of whose row sums is at most 1).

Lemma 2.4.2. *Let (a_{ij}) be an $n \times n$ substochastic matrix. Suppose (a_{ij}) has diagonal sum greater than 1. Then*

$$a_{kk} + a_{\ell\ell} > a_{k\ell} + a_{\ell k} \quad \textit{for some } k, \ell.$$

Proof. Suppose, for a contradiction, that $a_{ii} + a_{jj} \leq a_{ij} + a_{ji}$ for all i, j. Sum this inequality over all i, j to obtain $2n \sum_i a_{ii} \leq 2 \sum_{i,j} a_{ij} \leq 2n$, since by assumption the row sums of (a_{ij}) are each at most 1. This implies that the diagonal sum satisfies $\sum_i a_{ii} \leq 1$, giving the required contradiction. □

We next give conditions under which the sum of two diagonal entries of an $n \times n$ stochastic matrix can be bounded from below.

Lemma 2.4.3. *Let (a_{ij}) be an $n \times n$ stochastic matrix with diagonal sum d, and suppose all transversal sums of (a_{ij}) lie outside the interval $[u, d)$. Then, for all i, j,*

$$a_{ii} + a_{jj} > a_{ij} + a_{ji} \quad \textit{implies} \quad a_{ii} + a_{jj} > d - u.$$

Proof. Suppose $a_{ii} + a_{jj} > a_{ij} + a_{ji}$. Then the positive quantity $a_{ii} + a_{jj} - a_{ij} - a_{ji}$ is the decrease in the diagonal sum caused by transposing rows i and j of the matrix, and so by assumption is greater than $d - u$. We therefore have $a_{ii} + a_{jj} \geq a_{ii} + a_{jj} - a_{ij} - a_{ji} > d - u$. □

Define an $n \times n$ stochastic matrix (a_{ij}) to be *diagonally ordered* if its diagonal entries are in non-increasing order:

$$a_{11} \geq a_{22} \geq \cdots \geq a_{nn}.$$

We are now ready to prove Theorem 2.2.1.

Proof of Theorem 2.2.1. We know from (2.3) and (2.1) that the result holds for $n = 2$ and 3, so we may take $n \geq 4$. Suppose, for a contradiction, that (a_{ij}) is an $n \times n$ stochastic matrix whose transversal sums all lie outside the interval $[1, 2]$. Then by Lemma 2.4.1, (a_{ij}) has a transversal sum greater than 2 and a transversal sum less than 1. Let $2 + \epsilon$ be the smallest transversal sum greater than 2, and reorder the rows and columns of (a_{ij}) so that the summands of this transversal sum occur on the matrix diagonal and so that the matrix is diagonally ordered. By Lemma 2.4.3 with $d = 2 + \epsilon$ and $u = 1$,

$$a_{ii} + a_{jj} > a_{ij} + a_{ji} \quad \text{implies} \quad a_{ii} + a_{jj} > 1 + \epsilon. \tag{2.4}$$

Now the $(n-1) \times (n-1)$ submatrix of (a_{ij}) formed by deleting the first row and column has diagonal sum $2 + \epsilon - a_{11} > 1$. Apply Lemma 2.4.2 to this submatrix to show that, for some distinct $k > 1$ and $\ell > 1$,

$$a_{kk} + a_{\ell\ell} > a_{k\ell} + a_{\ell k}.$$

We therefore have $a_{kk} + a_{\ell\ell} > 1 + \epsilon$ by (2.4), and so

$$a_{22} + a_{33} > 1 + \epsilon \tag{2.5}$$

since the matrix is diagonally ordered and k, ℓ are distinct. Since the diagonal sum of (a_{ij}) is $2 + \epsilon$, we have

$$a_{ii} + a_{11} \leq 2 + \epsilon - a_{22} - a_{33} \quad \text{for all } i > 3,$$

and therefore $a_{ii} + a_{11} < 1$ for all $i > 3$, by (2.5). Then, since the matrix is diagonally ordered,

$$a_{ii} + a_{jj} < 1 \quad \text{for all } i, j \text{ with } i > 3,$$

which in turn implies by (2.4) that

$$a_{ii} + a_{jj} \leq a_{ij} + a_{ji} \quad \text{for all } i, j \text{ with } i > 3. \tag{2.6}$$

We complete the proof by showing that (2.5) and (2.6) force the sum of the entries of (a_{ij}) to be too large. We have

$$\begin{aligned}\sum_{i,j} a_{ij} &\geq \sum_{i\leq 3} a_{ii} + \sum_{i>3} a_{ii} + \sum_{i>3,\, j\leq 3} (a_{ij}+a_{ji}) \\ &\geq (n-2)\sum_{i\leq 3} a_{ii} + 4\sum_{i>3} a_{ii}\end{aligned}$$

by substitution from (2.6). Therefore

$$\begin{aligned}\sum_{i,j} a_{ij} &\geq (n-2)\sum_{i\leq 3} a_{ii} + 2\sum_{i>3} a_{ii} \\ &= (n-4)\sum_{i\leq 3} a_{ii} + 2\sum_{i} a_{ii} \\ &\geq (n-4)(1+\epsilon) + 2(2+\epsilon)\end{aligned}$$

by (2.5), using $n \geq 4$. Therefore $\sum_{i,j} a_{ij} > n$, which is a contradiction because each row sum of (a_{ij}) is 1. □

2.5 Proof of Theorem 2.2.3

Proposition 2.5.1. *Let $n \geq 4$ and let (a_{ij}) be a diagonally ordered $n \times n$ stochastic matrix. Suppose the diagonal sum d of (a_{ij}) satisfies $d \in (1, 2]$. Then (a_{ij}) has a transversal sum lying in the interval $[d - \frac{n}{n-1}, d)$.*

Proof. Suppose, for a contradiction, that no transversal sum of (a_{ij}) lies in the interval $[d - \frac{n}{n-1}, d)$. Then by Lemma 2.4.3 with $u = d - \frac{n}{n-1}$,

$$a_{ii} + a_{jj} > a_{ij} + a_{ji} \quad \text{implies} \quad a_{ii} + a_{jj} > \frac{n}{n-1}. \tag{2.7}$$

Since $d > 1$, Lemma 2.4.2 gives

$$a_{kk} + a_{\ell\ell} > a_{k\ell} + a_{\ell k} \quad \text{for some distinct } k, \ell,$$

and it follows from (2.7) that $a_{kk} + a_{\ell\ell} > \frac{n}{n-1}$. Since the matrix is diagonally ordered and k, ℓ are distinct, this implies

$$a_{11} + a_{22} > \frac{n}{n-1} \tag{2.8}$$

and so

$$a_{11} > \frac{1}{2} \cdot \frac{n}{n-1}. \tag{2.9}$$

We now claim that

$$a_{ii} + a_{jj} \leq \frac{n}{n-1} \quad \text{for all distinct } i > 1 \text{ and } j > 1. \tag{2.10}$$

Suppose otherwise, for a contradiction, so that $a_{rr} + a_{ss} > \frac{n}{n-1}$ for some distinct $r > 1$ and $s > 1$. Since the matrix is diagonally ordered, this gives

$$a_{22} + a_{33} > \frac{n}{n-1}. \tag{2.11}$$

Therefore, for all $i > 3$, we have $a_{ii} + a_{11} \leq d - a_{22} - a_{33} < \frac{n-2}{n-1}$ because $d \leq 2$. Since the matrix is diagonally ordered, we then have

$$a_{ii} + a_{jj} < \frac{n-2}{n-1} \quad \text{for all } i, j \text{ with } i > 3,$$

which by (2.7) implies

$$a_{ii} + a_{jj} \leq a_{ij} + a_{ji} \quad \text{for all } i, j \text{ with } i > 3. \tag{2.12}$$

Now

$$\begin{aligned} \sum_{i,j} a_{ij} &\geq \sum_{i \leq 3} a_{ii} + \sum_{i>3, j \leq 3} (a_{ij} + a_{ji}) \\ &\geq (n-2) \sum_{i \leq 3} a_{ii} + 3 \sum_{i>3} a_{ii} \end{aligned}$$

by substitution from (2.12), so that

$$\begin{aligned} \sum_{i,j} a_{ij} &\geq (n-2) \sum_{i \leq 3} a_{ii} \\ &> (n-2) \cdot \frac{3}{2} \cdot \frac{n}{n-1} \end{aligned}$$

from (2.9) and (2.11). Since $\sum_{ij} a_{ij} = n$ and $n \geq 4$, this is a contradiction and proves the claim (2.10).

It then follows from (2.7) that

$$a_{ii} + a_{jj} \leq a_{ij} + a_{ji} \quad \text{for all distinct } i > 1 \text{ and } j > 1.$$

Summing over all i, j satisfying $1 < i < j$, we find that

$$(n-2) \sum_{i>1} a_{ii} \leq \sum_{1<i<j} (a_{ij} + a_{ji}). \tag{2.13}$$

Now let m be the largest integer i such that $a_{11} + a_{ii} > \frac{n}{n-1}$. Note that $m \geq 2$, by (2.8). By (2.7) we have $a_{11} + a_{ii} \leq a_{1i} + a_{i1}$ for $i > m$, so that

$$a_{11} \leq a_{1i} + a_{i1} \quad \text{for } i > m. \tag{2.14}$$

We now show that (2.13) and (2.14) force the entries of (a_{ij}) to be too large. We have

$$\begin{aligned}\sum_{i,j} a_{ij} &\geq a_{11} + \sum_{i>1} a_{ii} + \sum_{i>m}(a_{1i} + a_{i1}) + \sum_{1<i<j}(a_{ij} + a_{ji}) \\ &\geq (n-m+1)a_{11} + (n-1)\sum_{i>1} a_{ii}\end{aligned}$$

by substitution from (2.13) and (2.14). Therefore

$$\begin{aligned}\sum_{i,j} a_{ij} &\geq (n-m+1)a_{11} + (n-1)\sum_{1<i\leq m} a_{ii} \\ &> (n-m+1)a_{11} + (n-1)(m-1)\left(\frac{n}{n-1} - a_{11}\right)\end{aligned}$$

by definition of m and the diagonal ordering of (a_{ij}), and so

$$\sum_{i,j} a_{ij} > n(m-1-(m-2)a_{11}).$$

Since $m \geq 2$ and $a_{11} \leq 1$, this implies the contradiction $\sum_{i,j} a_{ij} > n$ and so completes the proof. □

We now combine Proposition 2.5.1 with Theorem 2.2.1 to prove Theorem 2.2.3.

Proof of Theorem 2.2.3. Suppose, for a contradiction, that (a_{ij}) is an $n \times n$ stochastic matrix whose transversal sums all lie outside the interval $\left[u, u + \frac{n}{n-1}\right]$ for some $u \in \left[0, \frac{n-2}{n-1}\right)$. Since this interval contains 1, by Theorem 2.2.1 the matrix (a_{ij}) therefore has a transversal sum in the interval $\left(u + \frac{n}{n-1}, 2\right]$. Let d be the smallest such transversal sum. Reorder the rows and columns of (a_{ij}) so that the summands of this transversal sum occur on the matrix diagonal and so that the matrix is diagonally ordered. Then by Proposition 2.5.1, (a_{ij}) has a transversal sum lying in the interval $\left[d - \frac{n}{n-1}, d\right)$. By choice of d, this gives the required contradiction. □

2.6 Open problems

1. Lemma 2.1.2 says that every solution to the n-card problem must contain at least one interval in the set
$$\left\{\left[\frac{k}{n-1}, 1 + \frac{k}{n-1}\right] : k = 0, 1, \ldots, n-1\right\}.$$

Problem 3 of [54] asks whether each of the intervals given in the statement of Lemma 2.1.2 is a solution to the n-card problem, which would imply that the complete set of minimal solutions to the n-card problem comprises these n intervals. This question remains open for $n > 4$. In particular, is $[\frac{1}{2}, \frac{3}{2}]$ a solution for all odd n?

2. Is there a meaningful generalization of the n-card problem to a union of intervals, or more generally to a subset of the real numbers?

Chapter 3

Inv-Wilf equivalence

3.1 Introduction

A permutation $\sigma = \sigma_1\sigma_2\ldots\sigma_n$ *avoids* a permutation $\pi = \pi_1\pi_2\ldots\pi_k$ if there is no subsequence $\sigma_{a_1}\sigma_{a_2}\ldots\sigma_{a_k}$ of σ (having $a_1 < a_2 < \ldots < a_k$) such that

$$\pi_i < \pi_j \quad \text{if and only if} \quad \sigma_{a_i} < \sigma_{a_j}.$$

If σ does not avoid π, we say that σ *contains* π (alternatively, that π is *contained* in σ). For example, $\sigma = 314256$ avoids $\pi = 1243$, but $\sigma = 214365$ contains π because of its subsequence 1465. Let $S_n(\pi)$ be the set of all permutations of size n (acting on $\{1, 2, \ldots, n\}$) that avoid the permutation π.

There are two main problems in the study of pattern avoidance in permutations. The first problem is to determine the asymptotic behavior of $|S_n(\pi)|$ for a given π. The second problem is to determine whether two permutations α and β are Wilf-equivalent. The problem of classifying all permutations up to Wilf-equivalence is "the basic problem in the theory of forbidden subsequences" [59], and has attracted a great deal of attention since it was posed by H. Wilf in the early 1980s. Permutations α and β are *Wilf-equivalent*, denoted $\alpha \sim \beta$, if

$$|S_n(\alpha)| = |S_n(\beta)| \quad \text{for all positive integers } n.$$

Wilf-equivalence is an equivalence relation, and so partitions the set of permutations of a given size into Wilf-equivalence classes.

As previously discussed in Chapter 1, we can represent a permutation $\sigma = \sigma_1\sigma_2\ldots\sigma_n$ with an $n \times n$ square diagram of cells as follows: each cell either contains a dot or is blank, and the cell in the ith row from the top and jth column from the left has a dot if and only if $\sigma_j = i$. For example, Figure 3.1 represents the permutation 1423. An equivalent formulation of pattern avoidance in permutations is that σ avoids π exactly when the square diagram associated with σ does not contain as a subdiagram the square diagram associated with π.

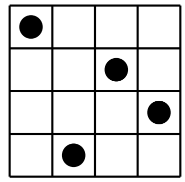

Figure 3.1: Square diagram representing permutation 1423

For a permutation σ of length n, the complement, denoted $c(\sigma)$, is defined to be the permutation $(n+1-\sigma_1)(n+1-\sigma_2)\dots(n+1-\sigma_n)$. The reverse, denoted $r(\sigma)$, is defined to be $\sigma_n\sigma_{n-1}\dots\sigma_1$. We denote the inverse of σ (that is, γ where $\gamma_i = j$ whenever $\sigma_j = i$) to be $t(\sigma)$. We let D denote the group generated by the functions c, r, t; that is, $\{i, c, r, t, c\circ r, c\circ t, r\circ t, c\circ r\circ t\}$, where i is the identity function.

The group D corresponds to the group of rotations and reflections of the square diagram associated with σ, with c being a reflection across a horizontal line, r a reflection across a vertical line, and t a reflection across a diagonal line from top-left to bottom-right. Therefore D is isomorphic to D_4, the symmetry group of the square.

The square diagram of a permutation will be used heavily from Section 3.4 onward, when we discuss transversals in Young diagrams.

We have the following preliminary observations.

Proposition 3.1.1.

(i) If π is contained in τ, then $|S_n(\pi)| \leq |S_n(\tau)|$ for all $n \geq 1$.

(ii) If $\alpha \sim \beta$, then α and β have the same length.

(iii) $|S_n(12)| = |S_n(21)| = 1$ for all $n \geq 1$, so that $12 \sim 21$.

(iv) For all permutations σ, and $d \in D$, we have $\sigma \sim d(\sigma)$.

Proof.

(i) If σ avoids π, then σ must avoid τ.

(ii) Otherwise, if α is of length m and β of length n with $m < n$, then $S_m(\alpha) = S_m \setminus \{\alpha\}$ and $S_m(\beta) = S_m$, showing that $|S_m(\alpha)| \neq |S_m(\beta)|$.

(iii) For all $n \geq 1$, we have $S_n(12) = \{n(n-1)\dots 1\}$ and $S_n(21) = \{12\dots n\}$.

(iv) By symmetry, σ avoids π if and only if $d(\sigma)$ avoids $d(\pi)$.

□

The next corollary follows by applying Proposition 3.1.1 (iv).

Corollary 3.1.2.

(i) $123 \sim 321$

(ii) $132 \sim 231 \sim 213 \sim 312$.

Thus, to settle the problem of determining the Wilf-equivalence classes of permutations in S_3, it suffices to determine whether or not $123 \sim 132$. We give a proof that in fact $123 \sim 132$ using the bijection of Lemma 4.3 of [11], itself motivated by a bijection due to Simion and Schmidt [55].

Proposition 3.1.3 ([11] Lemma 4.3)**.**

(i) $123 \sim 132$.

(ii) *If* π *is a permutation of length* 3*, then* $|S_n(\pi)| = C_n$ *for all* $n \geq 1$*, where* $C_0, C_1, C_2, \ldots$ *are the Catalan numbers, defined by the recurrence*

$$C_n = \sum_{i=1}^{n} C_{i-1}C_{n-i}, \quad C_0 = 1.$$

Proof.

(i) We say that an entry of a permutation is a *left-to-right minimum* if it is smaller than all entries preceding it. For example, in the permutation 67351284, the left-to-right minima are 6, 3, and 1.

Now, given a permutation σ of length n, let S be the set of permutations of length n that have the same set of left-to-right minima in the same positions as σ. Consider σ' in S obtained by rearranging all the entries that are not left-to-right minima (if any) of σ in decreasing order from left to right. For example, if $\sigma = 67351284$, then $\sigma' = 68371542$. Now σ' avoids 123 since, if there were a 123-pattern xyz in σ', then y and z would be an increasing pair of non-minima, contradicting how σ' was formed. On the other hand, a permutation of $S \setminus \{\sigma'\}$ (if it exists) has an increasing pair of non-minima yz, and so taking the left-to-right minimum x nearest y on the left gives a 123-pattern xyz. Therefore σ' is the only permutation of S that avoids 123.

Now consider σ'' in S obtained by rearranging all the non-minima (if any) of σ as follows: from left to right, place in each position the smallest non-minimum not yet placed that is greater than the closest left-to-right minima to its left. For example, if $\sigma = 67351284$, then $\sigma'' = 67341258$. Now σ'' avoids 132 since, if there were a 132-pattern xyz in σ'', then y and z would be a decreasing pair of non-minima both greater than x, contradicting how σ'' was formed. On the other hand, a permutation of $S \setminus \{\sigma''\}$ (if it exists) has a decreasing pair of non-minima yz, both of which are greater than some left-to-right minimum x to the left of y, giving a 132-pattern xyz. Therefore σ'' is the only permutation of S that avoids 132.

Therefore, in each possible class of permutations with the same set of left-to-right minima in the same positions, there is exactly one permutation that avoids 123 and exactly one permutation (possibly the same) that avoids 132. Therefore $|S_n(123)| = |S_n(132)|$ for all $n \geq 1$.

(ii) From (i) and Corollary 3.1.2, it suffices to find $|S_n(132)|$. We define $a_n = |S_n(132)|$ for $n \geq 1$ and $a_0 = 1$. Let σ be a permutation of length n avoiding 132. Then σ can be written as $\alpha n \beta$, where α and β are (possibly empty) subsequences avoiding 132. Note that all the entries of α are necessarily larger than all those of β; otherwise if x in α is smaller than y in β, then xny is a 132-pattern. So α is composed of $\{n-i+1, n-i+2, \ldots, n-1\}$ and β is composed of $\{1, 2, \ldots, n-i\}$ for some $1 \leq i \leq n$. So we have the recurrence relation

$$a_n = \sum_{i=1}^{n} a_{i-1} a_{n-i}, \quad a_0 = 1,$$

which is the recurrence relation for the Catalan numbers.

□

Up to equivalence, the only permutations π of size greater than 3 for which an exact formula for $|S_n(\pi)|$ is known are 1234 and 1342 [11, p161].

In 1980, Stanley and Wilf conjectured that $S_n(\pi)$ grows at most exponentially in n for all permutations π. The conjecture is true for π up to length 3, since $|S_n(\pi)| = 1$ for π of length 2 by Proposition 3.1.1 (iii), and

$$|S_n(\pi)| = C_n = \frac{1}{n+1}\binom{2n}{n} \approx \frac{4^n}{n^{1.5}} \tag{3.1}$$

for all π of length 3 by Proposition 3.1.3 (ii), where the approximation is by Stirling's formula for the factorial function. It turns out that the conjecture is true for all π, as proved by Marcus and Tardos [47] in 2003.

Theorem 3.1.4 (Stanley-Wilf conjecture, proved in [47])**.** *For all permutations π, there exists a constant c, depending on π, such that $|S_n(\pi)| \leq c^n$.*

The following result at first appears to be a stronger form of Theorem 3.1.4, but is actually equivalent.

Corollary 3.1.5 (Alternative version of Theorem 3.1.4 [11], [1])**.** *For all permutations π, the limit*

$$L(\pi) = \lim_{n\to\infty} \sqrt[n]{|S_n(\pi)|}$$

exists.

Note that Corollary 3.1.5 immediately implies Theorem 3.1.4. The implication in the other direction, proved by Arratia in 1999 [1], requires the observation that $|S_{n+m}(\pi)| \geq |S_n(\pi)||S_m(\pi)|$ for all $m, n \geq 1$ as well as Fekete's lemma on subadditive sequences [31] applied to $-\log(|S_n(\pi)|)$.

The value of $L(\pi)$ is known for some π. For example:

- $L(123) = L(132) = 4$, from (3.1).
- $L(12 \dots k) = (k-1)^2$ for $k \geq 2$ [11].
- $L(1342) = 8$ [11].
- $L(12453) = 9 + 4\sqrt{2}$ [10].

The shortest permutation π for which $L(\pi)$ is not known is $\pi = 1324$ and the members of its equivalence class under the action of D. It is known that $L(1324) \geq 9.81$ [5] and $L(1324) \leq 13.74$ [12].

One area of particular interest is determining the asymptotic growth rate in k of both $\max_{\pi \in S_k} L(\pi)$ and $\min_{\pi \in S_k} L(\pi)$. It is known that there exists a positive constant c for which $L(\pi) \geq ck^2$ for all permutations π of size k [38]. We define a *layered permutation* to be a permutation of the form

$$a_1(a_1-1)\dots(a_0+1)\ a_2(a_2-1)\dots(a_1+1)\ \dots\ a_i(a_i-1)\dots(a_{i-1}+1),$$

for some sequence $0 = a_0 < a_1 < a_2 < \dots < a_{i-1} < a_i = n$. For example, 1324 is a layered permutation with $a_0 = 0, a_1 = 1, a_2 = 3, a_3 = 4$. In 2012, Claesson, Jelínek, and Steingrímsson showed that for a layered permutation π of length k, we have $L(\pi) \leq 4k^2$ [20]. Bóna [11, p182] conjectured that the maximum $L(\pi)$ over all permutations π of length k is achieved when π is a layered permutation, which together with known results would imply that $\max_{\pi \in S_k} L(\pi)$ is bounded above, and $\min_{\pi \in S_k} L(\pi)$ bounded below, by functions in $\Theta(k^2)$. However, in 2013, Fox [33] disproved the conjecture by showing that

$$\max_{\pi \in S_k} L(\pi) = 2^{\Omega(k^{1/4})}.$$

3.2 Wilf-equivalence

We now focus on the problem of Wilf-equivalence: which permutations π_1 and π_2 satisfy $|S_n(\pi_1)| = |S_n(\pi_2)|$ for all $n \geq 1$?

Proposition 3.1.3 (i) already shows that $123 \sim 132$. In 1985, Simion and Schmidt [55] found the first explicit proof that $123 \sim 132$ via the bijection described in the proof of Proposition 3.1.3 (i). Many other combinatorial correspondences have been studied in the literature. For example, bijections have been found between:

- $S_n(132)$ and Dyck paths, based on a similar idea to the proof of Proposition 3.1.3 (ii) [40], [42].
- $S_n(321)$ and Dyck paths, using standard Young tableaux [41].
- $S_n(321)$ and binary trees [52].
- $S_n(123)$ and ballot sequences (equivalent to Dyck paths) [51].
- $S_n(123)$ and $S_n(132)$ using generating trees [62].
- $S_n(123)$ and Dyck paths [42].
- $S_n(132)$ and $S_n(321)$ [50].

For permutations of length 4, West [62] used generating trees to show that $|S_n(1234)| = |S_n(1243)| = |S_n(2143)|$, while Stankova [60] used generating trees to show that $|S_n(4132)| = |S_n(3142)|$. Stankova also showed [58] that $|S_n(1234)| = |S_n(4123)|$. These results, along with Proposition 3.1.1 (iv) as well as the fact that $|S_7(1234)|$, $|S_7(1342)|$, and $|S_7(1324)|$ are all distinct, complete the classification of Wilf-equivalence for S_4 as follows: Wilf-equivalence partitions S_4 into the equivalence class containing 1234, the equivalence class containing 1342, and the equivalence class containing 1324.

For permutations α of length i and β of length j, let $\alpha \oplus \beta$ denote the permutation $\alpha_1\alpha_2 \ldots \alpha_i(\beta_1 + i)(\beta_2 + i) \ldots (\beta_j + i)$. For example, $312 \oplus 231 = 312564$. All known Wilf-equivalences, up to symmetry, can be summarized as two infinite families and one sporadic example, as follows.

Proposition 3.2.1.

(i) $123 \ldots n \oplus \gamma \sim n(n-1)(n-2) \ldots 1 \oplus \gamma$ *for all* n *and all permutations* γ *[3].*

(ii) $231 \oplus \gamma \sim 312 \oplus \gamma$ *for all permutations* γ *[59].*

(iii) $4132 \sim 3142$ *[60].*

The proof of Proposition 3.2.1 (i) and (ii) relies on shape-Wilf-equivalence, which we will describe in Section 3.4.

3.3 Permutation statistics

Wilf-equivalence is concerned with the size of the sets $S_n(\pi_1)$ and $S_n(\pi_2)$. Stronger forms of Wilf-equivalence involve partitioning each of the sets $S_n(\pi_1)$ and $S_n(\pi_2)$ into smaller sets, and comparing the sizes of the smaller sets of $S_n(\pi_1)$ with those of $S_n(\pi_2)$. This is done using *permutation statistics*, namely functions whose domain is the set of all permutations.

Given a permutation statistic f, we say that two permutations π_1 and π_2 are *f-Wilf-equivalent*, denoted $\pi_1 \overset{f}{\sim} \pi_2$, if

$$|\{\sigma \in S_n(\pi_1) : f(\sigma) = y\}| = |\{\sigma \in S_n(\pi_2) : f(\sigma) = y\}|$$

for all $n \geq 1$ and all y in the range of f restricted to S_n. For example, if f is a function that maps σ to $\{(j, \sigma_j) : \sigma_i > \sigma_j \text{ for all } i < j\}$ (in other words, the positions and values of left-to-right minima of σ), then the proof of Proposition 3.1.3 (i) shows that 123 is f-Wilf-equivalent to 132; specifically, $|\{\sigma \in S_n(123) : f(\sigma) = y\}| = 1 = |\{\sigma \in S_n(132) : f(\sigma) = y\}|$ for all $n \geq 1$ and all y in the range of f restricted to S_n.

Let $\sigma = \sigma_1\sigma_2 \ldots \sigma_n$ be a permutation. Some of the main permutation statistics of interest are:

- Inversion number, $\text{inv}(\sigma) := |\{(i, j) : i < j \text{ and } \sigma_i > \sigma_j\}|$.
- Descent set, $\text{Des}(\sigma) := \{i : \sigma_i > \sigma_{i+1}\}$.
- Major index, $\text{maj}(\sigma) := \sum_{i \in \text{Des}(\sigma)} i$.
- Number of fixed points, $\text{fp}(\sigma) := |\{i : \sigma_i = i\}|$.
- Number of excedances, $\text{exc}(\sigma) := |\{i : \sigma_i > i\}|$.

Permutation statistics have been studied in the context of Wilf-equivalence by Dokos et al. [23], Elizalde [26], and Claesson and Kitaev [21]. We now examine various properties of f-Wilf-equivalence. Among the results of Proposition 3.3.1, parts (i), (iii), and (iv) are generalizations of results in [23]. Recall the complement $c(\sigma)$, reverse $r(\sigma)$, and inverse $t(\sigma)$ of a permutation σ, and the group $D = \{i, c, r, t, c \circ r, c \circ t, r \circ t, c \circ r \circ t\}$.

Proposition 3.3.1.

(i) If $\pi_1 \overset{f}{\sim} \pi_2$, then $\pi_1 \sim \pi_2$.

(ii) If $\pi_1 \overset{f}{\sim} \pi_2$, then π_1 and π_2 have the same length. Furthermore, $f(\pi_1) = f(\pi_2)$.

(iii) Let f be a permutation statistic and $d \in D$ such that $f(\sigma) = f(d(\sigma))$ for all permutations σ. Then $\pi \overset{f}{\sim} d(\pi)$ for all permutations π.

(iv) Let f be a permutation statistic and $d \in D$ such that $f(\sigma_1) = f(\sigma_2)$ if and only if $f(d(\sigma_1)) = f(d(\sigma_2))$ for all $\sigma_1, \sigma_2 \in S_n$. Then $\pi_1 \overset{f}{\sim} \pi_2$ if and only if $d(\pi_1) \overset{f}{\sim} d(\pi_2)$ for all $\pi_1, \pi_2 \in S_n$.

(v) Let f, g be permutation statistics such that $f(\sigma_1) = f(\sigma_2)$ implies $g(\sigma_1) = g(\sigma_2)$ for all $\sigma_1, \sigma_2 \in S_n$. Then $\pi_1 \overset{f}{\sim} \pi_2$ implies $\pi_1 \overset{g}{\sim} \pi_2$ for all $\pi_1, \pi_2 \in S_n$.

Proof.

(i) If π_1 and π_2 are f-Wilf-equivalent, then $|S_n(\pi_1)| = \sum_y |\{\sigma \in S_n(\pi_1) : f(\sigma) = y\}| = \sum_y |\{\sigma \in S_n(\pi_2) : f(\sigma) = y\}| = |S_n(\pi_2)|$ and so π_1 and π_2 are Wilf-equivalent.

(ii) By (i) and Proposition 3.1.1 (ii), π_1 and π_2 have the same length. Now if π_1 and π_2 are of length n, then $S_n(\pi_1) = S_n \setminus \{\pi_1\}$ and $S_n(\pi_2) = S_n \setminus \{\pi_2\}$. Putting $y = f(\pi_1)$ in the definition of f-Wilf-equivalence then gives

$$|\{\sigma \in S_n(\pi_2) : f(\sigma) = y\}| = |\{\sigma \in S_n(\pi_1) : f(\sigma) = y\}| = |\{\sigma \in S_n : f(\sigma) = y\}| - 1,$$

which forces $y = f(\pi_2)$.

(iii) By symmetry, σ avoids π if and only if $d(\sigma)$ avoids $d(\pi)$. Since $f(\sigma) = f(d(\sigma))$ for all permutations σ,

$$|\{\sigma \in S_n(\pi) : f(\sigma) = y\}| = |\{\sigma' \in S_n(d(\pi)) : f(\sigma') = y\}|$$

for all $n \geq 1$ and all y in the range of f restricted to S_n.

(iv) Let $\pi_1, \pi_2 \in S_n$. Let y be a value in the range of f restricted to S_n and let τ be such that $f(d(\tau)) = y$ (which exists because d is a bijection). Then for all $n \geq 1$, since σ avoids π if and only if $d(\sigma)$ avoids $d(\pi)$,

$$|\{\sigma' \in S_n(d(\pi_1)) : f(\sigma') = y = f(d(\tau))\}| = |\{\sigma \in S_n(\pi_1) : f(\sigma) = f(\tau)\}|$$

and

$$|\{\sigma' \in S_n(d(\pi_2)) : f(\sigma') = y = f(d(\tau))\}| = |\{\sigma \in S_n(\pi_2) : f(\sigma) = f(\tau)\}|.$$

Therefore $d(\pi_1) \overset{f}{\sim} d(\pi_2)$ if and only if $\pi_1 \overset{f}{\sim} \pi_2$.

(v) Let $\pi_1, \pi_2 \in S_n$, where $\pi_1 \overset{f}{\sim} \pi_2$. Let y be a value in the range of g restricted to S_n, and let T be the set $\{\tau \in S_n : g(\tau) = y\}$ and Z be the set $\{f(\tau) : \tau \in T\}$. Then for all $n \geq 1$, since $f(\tau_1) = f(\tau_2)$ implies $g(\tau_1) = y = g(\tau_2)$ for all $\tau_1, \tau_2 \in T$,

$$|\{\sigma \in S_n(\pi_1)) : g(\sigma) = y\}| = \sum_{z \in Z} |\{\sigma \in S_n(\pi_1)) : f(\sigma) = z\}|$$

$$= \sum_{z \in Z} |\{\sigma \in S_n(\pi_2)) : f(\sigma) = z\}| = |\{\sigma \in S_n(\pi_2)) : g(\sigma) = y\}|.$$

Therefore $\pi_1 \overset{g}{\sim} \pi_2$.

□

In the case that f is a non-negative integer-valued function for which the range of f restricted to S_n is finite for all $n \geq 1$ (such as the inversion number), we may equivalently express f-Wilf-

π	$I_2(\pi,q)$	$I_3(\pi,q)$	$I_4(\pi,q)$	$I_5(\pi,q)$	$I_6(\pi,q)$
21	1	1	1	1	1
231		$[1,2,1,1]$	$[1,3,3,3,2,1,1]$	$[1,4,6,7,7,5,5,3,2,1,1]$	$[1,5,10,14,17,16,16,14,11,9,7,5,3,2,1,1]$
321		$[1,2,2]$	$[1,3,5,4,1]$	$[1,4,9,12,10,4,2]$	$[1,5,14,25,31,26,16,9,4,1]$
3142			$[1,3,5,5,5,3,1]$	$[1,4,9,13,16,17,16,13,9,4,1]$	$[1,5,14,26,39,50,57,64,64,57,50,39,26,14,5,1]$
3412			$[1,3,5,6,4,3,1]$	$[1,4,9,15,18,18,16,11,6,4,1]$	$[1,5,14,29,46,61,70,72,64,53,41,27,16,8,5,1]$
4123			$[1,3,5,5,5,3,1]$	$[1,4,9,13,17,17,15,13,9,4,1]$	$[1,5,14,26,41,53,61,65,63,56,45,37,26,14,5,1]$
4213			$[1,3,5,6,4,3,1]$	$[1,4,9,15,18,18,15,11,7,4,1]$	$[1,5,14,29,46,61,68,68,62,53,40,29,19,11,5,1]$
4231			$[1,3,5,6,5,2,1]$	$[1,4,9,15,20,20,16,10,5,2,1]$	$[1,5,14,29,49,68,80,79,67,51,32,20,10,5,2,1]$
4312			$[1,3,5,6,5,2,1]$	$[1,4,9,15,20,20,16,10,5,2,1]$	$[1,5,14,29,49,68,80,79,67,50,33,20,10,5,2,1]$
4321			$[1,3,5,6,5,3]$	$[1,4,9,15,20,22,18,11,3]$	$[1,5,14,29,49,71,87,91,77,52,26,10,1]$

Table 3.1: Examples of $I_n(\pi,q)$ for $2 \leq n \leq 6$, where $[a_0, a_1, \ldots, a_n]$ denotes the polynomial $\sum_{i=0}^{n} a_i q^i$.

equivalence in terms of polynomials. In the case of the inversion number, we define the generating function

$$I_n(\pi,q) = \sum_{\sigma \in S_n(\pi)} q^{\mathrm{inv}(\sigma)}. \tag{3.2}$$

Applying Proposition 3.3.1 (iii) to $f = \mathrm{inv}$ gives that, for $d \in \{i, t, c\circ r, c\circ r\circ t\}$, we have $\pi \overset{\mathrm{inv}}{\sim} d(\pi)$ for all permutations π. Furthermore, $\mathrm{inv}(\sigma) = \binom{n}{2} - \mathrm{inv}(d(\sigma))$ for all $d \in D \setminus \{i, t, c\circ r, c\circ r\circ t\}$; for such a d, we have $I_n(\pi,q) = q^{\binom{n}{2}} I_n(d(\pi), q^{-1})$. Thus, knowing $I_n(\pi,q)$ immediately gives $I_n(d(\pi),q)$ for all $d \in D$. As an example, Table 3.1 lists $I_n(\pi,q)$ $(n \leq 6)$ for all distinct π of size $s \leq n$ with $2 \leq s \leq 4$, up to symmetry under the action of D on π.

Define the q-analogue of the factorial $[n]_q! = \prod_{i=1}^{n}(1 + q + \ldots + q^{i-1})$. The major index is notable because it is "equidistributed" with inversion number, in the sense that

$$\sum_{\sigma \in S_n} q^{\mathrm{maj}(\sigma)} = \sum_{\sigma \in S_n} q^{\mathrm{inv}(\sigma)} = [n]_q! \tag{3.3}$$

for all $n \geq 1$. This result was proved in 1913 by MacMahon [46]. An integer-valued permutation statistic f which satisfies

$$\sum_{\sigma \in S_n} q^{f(\sigma)} = [n]_q!$$

for all $n \geq 1$ is called *Mahonian.* Babson and Steingrímsson [2] classified the Mahonian statistics which can be written as linear combinations of "generalized permutation patterns"; inv and maj are two such statistics.

For permutation statistics f, g we use the notation $f \times g$ to denote the permutation statistic given by $(f \times g)(\sigma) = (f(\sigma), g(\sigma))$ for all σ. By Proposition 3.3.1 (v), $f \times g$-Wilf-equivalence implies both f-Wilf-equivalence and g-Wilf-equivalence.

Elizalde and Pak [27] used a bijection to show that $321 \overset{\text{fp}\times\text{exc}}{\sim} 132$. In particular, $321 \overset{\text{fp}}{\sim} 132$ and $321 \overset{\text{exc}}{\sim} 132$. We note that these results, along with Proposition 3.3.1 (ii) and (iii), show that the fp-Wilf-equivalence classes of S_3 are

$$\{123\}, \{132, 213, 321\}, \{231, 312\},$$

and that, also using the result $|\{\sigma \in S_4(132) : \text{exc}(\sigma) = 1\}| = 6 \neq 5 = |\{\sigma \in S_4(312) : \text{exc}(\sigma) = 1\}|$, the exc-Wilf-equivalence classes of S_3 are

$$\{123\}, \{132, 213, 321\}, \{231\}, \{312\}.$$

Dokos et al. [23] focused on the permutation statistics of inversion number and major index. They determined the inv-Wilf-equivalence classes of S_3 to be

$$\{123\}, \{321\}, \{132, 213\}, \{231, 312\},$$

which can readily be shown by Proposition 3.3.1 (ii) and (iii). They also determined the maj-Wilf-equivalence classes of S_3 to be

$$\{123\}, \{321\}, \{132, 231\}, \{213, 312\},$$

by proving $132 \overset{\text{maj}}{\sim} 231$ inductively; then $213 \overset{\text{maj}}{\sim} 312$ follows by Proposition 3.3.1 (iv) and the rest of the classification by Proposition 3.3.1 (ii). Note by Proposition 3.3.1 (v) that $\pi_1 \overset{\text{Des}}{\sim} \pi_2$ implies $\pi_1 \overset{\text{maj}}{\sim} \pi_2$ (since major index depends only on the values of the descent set). We shall give a short proof that, in fact, $132 \overset{\text{Des}}{\sim} 231$, which strengthens the result of Dokos et al. that $132 \overset{\text{maj}}{\sim} 231$.

Proposition 3.3.2 (Special case of [23, Theorem 2.6])**.** $132 \overset{\text{Des}}{\sim} 231$.

Proof. For permutations α, β, γ of length i, j, k, respectively, define the following notation:

$$231[\alpha, \beta, \gamma] := (\alpha_1 + k)(\alpha_2 + k) \ldots (\alpha_i + k)(\beta_1 + k + i)(\beta_2 + k + i) \ldots (\beta_j + k + i)\gamma_1\gamma_2 \ldots \gamma_k,$$

and

$$132[\alpha,\beta,\gamma] := \alpha_1\alpha_2\dots\alpha_i(\beta_1+k+i)(\beta_2+k+i)\dots(\beta_j+k+i)(\gamma_1+i)(\gamma_2+i)\dots(\gamma_k+i).$$

By induction on $n \geq 1$, we prove that there exists a bijection h_n which maps $S_n(132)$ to $S_n(231)$ and preserves Des. (For an empty permutation ϵ we define $h_0(\epsilon) = \epsilon$.)

The statement is clearly true for $n = 1$, since $S_1(132) = \{1\} = S_1(231)$.

Now assume that the statement is true for all $n < N$. Let $\sigma \in S_N(132)$. Since σ avoids 132, all the entries of σ to the left of N are necessarily larger than all those to the right of N. Therefore σ can be written $231[\alpha, 1, \beta]$, where α and β are permutations of length i and $N-i-1$ ($0 \leq i \leq N-1$), respectively, each avoiding 132.

We then define $h_N(\sigma)$ to be $132[h_i(\alpha), 1, h_{N-i-1}(\beta)]$. Then $h_N(\sigma)$ avoids 231, since $h_i(\alpha)$ and $h_{N-i-1}(\beta)$ each avoid 231, and all the entries of $h_N(\sigma)$ to the left of N are smaller than all those to the right of N. Also, h_N is clearly injective since $h_i(\alpha)$ and $h_{N-i-1}(\beta)$ uniquely determine α and β; since $|S_N(132)| = |S_N(231)|$ by Corollary 3.1.2 (ii), it follows that h is bijective. Finally, h_N preserves Des: if $\beta \neq \epsilon$, then we have

$$\mathrm{Des}(h_N(\sigma)) = \mathrm{Des}(132[h_i(\alpha), 1, h_{N-i-1}(\beta)]) = \mathrm{Des}(h_i(\alpha)) \cup \{i+1\} \cup (\mathrm{Des}(h_{N-i-1}(\beta)) + i + 1)$$

$$= \mathrm{Des}(\alpha) \cup \{i+1\} \cup (\mathrm{Des}(\beta) + i + 1) = \mathrm{Des}(231[\alpha, 1, \beta]) = \mathrm{Des}(\sigma);$$

otherwise,

$$\mathrm{Des}(h_N(\sigma)) = \mathrm{Des}(h_{N-1}(\alpha)N) = \mathrm{Des}(h_{N-1}(\alpha)) = \mathrm{Des}(\alpha) = \mathrm{Des}(\alpha N) = \mathrm{Des}(\sigma).$$

□

Dokos et al. [23] also made further conjectures involving the major index.

Theorem 3.3.3 ([23, Conjecture 2.8], proved in [6])**.**

- $1423 \overset{\mathrm{maj}}{\sim} 2314 \overset{\mathrm{maj}}{\sim} 2413$, *and equivalently*
- $3142 \overset{\mathrm{maj}}{\sim} 3241 \overset{\mathrm{maj}}{\sim} 4132$.

Theorem 3.3.3 was proved by Bloom [6], by exhibiting a bijection between $S_n(1423)$ and $S_n(2413)$ that preserves various permutation statistics. In particular, Bloom shows that $1423 \overset{\mathrm{Des}}{\sim} 2413$ (and so $1423 \overset{\mathrm{maj}}{\sim} 2413$). All other maj-Wilf-equivalences stated in Theorem 3.3.3 follow by applying Proposition 3.3.1 (iv) to the result $1423 \overset{\mathrm{maj}}{\sim} 2413$.

Dokos et al. also made a conjecture which generalizes the result $132 \overset{\mathrm{maj}}{\sim} 231$ as follows:

Conjecture 3.3.4 ([23, Conjecture 2.7])**.** *For all* k, n *satisfying* $1 \leq k \leq n-2$, *we have*

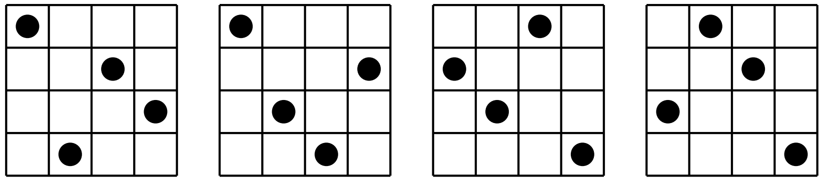

Figure 3.2: Square diagrams representing the trivial symmetries of 1423 under the action of the group $\{i, t, c \circ r, c \circ r \circ t\}$

- $12 \ldots kn(n-1) \ldots (k+1) \overset{\text{maj}}{\sim} (n-k)(n-k+1) \ldots n(n-k-1)(n-k-2) \ldots 1$, *and equivalently*
- $n(n-1) \ldots (n-k+1)12 \ldots (n-k) \overset{\text{maj}}{\sim} (k+1)k \ldots 1(k+2)(k+3) \ldots n.$

Note that Conjecture 3.3.4 was proved for all n and $k = n-2$ in 2015 by Ge, Yan, and Zhang [63] by showing that $123 \ldots (n-2)n(n-1) \overset{\text{Des}}{\sim} 234 \ldots n1$ (and so $123 \ldots (n-2)n(n-1) \overset{\text{maj}}{\sim} 234 \ldots n1$). Applying Proposition 3.3.1 (iv) also shows that the conjecture is true for all n and $k = 1$. The conjecture remains open for $1 < k < n-2$.

In the case of inv-Wilf-equivalence, a bold conjecture of [23] states that two permutations are inv-Wilf-equivalent if and only if they are related by trivial symmetry (under the action of the group $\{i, t, c \circ r, c \circ r \circ t\}$). Figure 3.2 illustrates trivial symmetry for the permutation 1423.

Conjecture 3.3.5 (Dokos et al., special case of [23, Conj. 2.4])**.** *Let α and β be permutations of size n. Then $\alpha \overset{\text{inv}}{\sim} \beta$ if and only if $f(\alpha) = \beta$ for some $f \in \{i, t, c \circ r, c \circ r \circ t\}$.*

Dokos et al. verified Conjecture 3.3.5 by computer for $n \leq 5$: two permutations α and β which are not related by trivial symmetry under the action of the group $\{i, t, c \circ r, c \circ r \circ t\}$ can be shown not to be inv-Wilf-equivalent by finding a value m for which $I_m(\alpha, q) \neq I_m(\beta, q)$. They also proposed a stronger version of Conjecture 3.3.5, involving "multiple-permutation avoidance" [23, Conj. 2.4], which was disproved by Trongsiriwat [61]. However, we shall show that even the single-permutation version stated in Conjecture 3.3.5 fails, for $n = 6$. In fact, we shall construct an infinite family of counterexamples to Conjecture 3.3.5, as part of Theorem 3.3.6. Theorem 3.3.6 strengthens Proposition 3.2.1 (ii), replacing Wilf-equivalence by inv-Wilf-equivalence; this theorem is a consequence of a more general result, which we shall state as Theorem 3.5.1 and prove in Sections 3.5 to 3.8.

Theorem 3.3.6. *For all permutations γ, we have $231 \oplus \gamma \overset{\text{inv}}{\sim} 312 \oplus \gamma$.*

Thus, choosing a permutation γ for which $\gamma \neq t(\gamma)$ gives a counterexample to Conjecture 3.3.5. In particular, taking $\gamma = 231$ in Corollary 3.5.3 shows that $231 \oplus 231 = 231564$ and $312 \oplus 231 = 312564$ are inv-Wilf-equivalent, even though they are not trivially related by one of the four symmetries of D. Table 3.2 gives all permutations γ of size at most 4 such that $231 \oplus \gamma$ and $312 \oplus \gamma$

γ	$231 \oplus \gamma$	$312 \oplus \gamma$	$\overline{I_7(231 \oplus \gamma, q)}$	$\overline{I_8(231 \oplus \gamma, q)}$	$\overline{I_9(231 \oplus \gamma, q)}$
231	231564	312564	$q^4[3, 6, 8, 8, 6, 4, 2]$	$q^4[6, 21, 45, 76, 101, 123, 132, 126, 104, 75, 46, 24, 10, 3]$	$q^4[10, 48, 133, 289, 511, 796, 1126, 1463, 1762, 1968, 2038, 1961, 1750, 1446, 1098, 761, 476, 266, 130, 54, 18, 4]$
1423	2314756	3124756	q^4	$q^4[3, 7, 10, 10, 8, 6, 4, 2]$	$q^4[6, 24, 57, 103, 147, 186, 208, 211, 194, 162, 120, 79, 46, 24, 10, 3]$
2314	2315647	3125647	q^4	$q^4[3, 7, 10, 10, 8, 6, 4, 2]$	$q^4[6, 24, 57, 103, 147, 186, 208, 211, 194, 162, 120, 79, 46, 24, 10, 3]$
2341	2315674	3125674	q^5	$q^5[3, 6, 10, 11, 8, 6, 4, 2]$	$q^5[6, 21, 52, 97, 143, 184, 209, 215, 198, 166, 124, 81, 46, 24, 10, 3]$
2413	2315746	3125746	q^5	$q^5[3, 7, 10, 10, 8, 6, 4, 2]$	$q^5[6, 24, 58, 105, 147, 184, 208, 210, 194, 162, 120, 79, 46, 24, 10, 3]$
2431	2315764	3125764	q^6	$q^6[3, 7, 10, 10, 8, 6, 4, 2]$	$q^6[6, 24, 59, 105, 149, 184, 206, 209, 194, 162, 120, 79, 46, 24, 10, 3]$
3241	2316574	3126574	q^6	$q^6[3, 7, 10, 10, 8, 6, 4, 2]$	$q^6[6, 24, 59, 105, 149, 184, 206, 209, 194, 162, 120, 79, 46, 24, 10, 3]$
3421	2316754	3126754	q^7	$q^7[3, 7, 10, 10, 8, 6, 4, 2]$	$q^7[6, 24, 59, 106, 150, 185, 204, 208, 193, 162, 120, 79, 46, 24, 10, 3]$

Table 3.2: Examples of nontrivial inv-Wilf-equivalent pairs, where $[a_0, a_1, \ldots, a_n]$ denotes the polynomial $\sum_{i=0}^{n} a_i q^i$.

form a nontrivial inv-Wilf-equivalent pair, up to symmetry under the action of D on the pair as well as the symmetric property of inv-Wilf-equivalence.

Furthermore, define $\overline{S_n(\pi)}$ to be the set of permutations of size n containing (rather than avoiding) π, and define $\overline{I_n(\pi, q)} = \sum_{\sigma \in \overline{S_n(\pi)}} q^{\mathrm{inv}(\sigma)}$. By (3.3) we note that $I_n(\pi, q)$ and $\overline{I_n(\pi, q)}$ are related by

$$I_n(\pi, q) + \overline{I_n(\pi, q)} = [n]_q!$$

In Table 3.2, we list $\overline{I_n(231 \oplus \gamma, q)}$ rather than $I_n(231 \oplus \gamma, q)$ for $7 \leq n \leq 9$, since $\overline{I_n(231 \oplus \gamma, q)}$ is a shorter polynomial. Note that the listed polynomials $\overline{I_n(231 \oplus \gamma, q)}$ are identical for some permutation pairs (e.g. those in rows 2 and 3). However, none of the permutations pairs in distinct rows of Table 3.2 are inv-Wilf-equivalent, although a value of n greater than 9 is needed to demonstrate this.

The proof of our main result (Theorem 3.5.1) requires the concept of shape-Wilf-equivalence, as described in Section 3.4, and a new generalization of this concept, as described in Section 3.5.

3.4 Shape-Wilf-equivalence

As seen in Proposition 3.2.1, two infinite families of Wilf-equivalent permutation pairs and one sporadic example are known [3], [59], [60]. Both of the infinite families were constructed by means

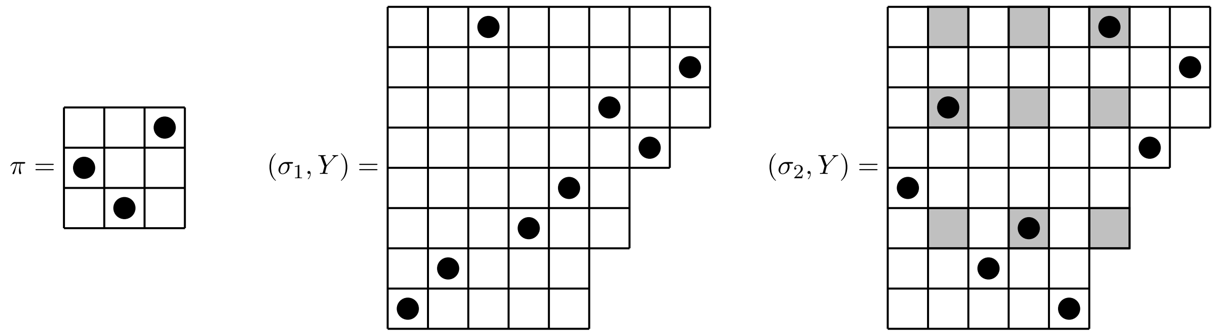

Figure 3.3: Permutation avoidance, demonstrating that σ_1 avoids $\pi = 231$ and σ_2 contains π

of the stronger property of *shape-Wilf-equivalence*, as we now describe. (Note that shape-Wilf-equivalence is not the same thing as f-Wilf-equivalence for a permutation statistic f.)

We first generalize permutations by introducing the concept of transversals in Young diagrams. A *Young diagram* is a finite collection of left-justified cells arranged in rows and columns, where the number of cells in each row is nonincreasing from top to bottom. For example, the cells of the middle diagram of Figure 3.3 is an example of a Young diagram.

As discussed previously in Chapter 1, each permutation σ of length n is associated with a square diagram (which is a Young diagram) of order n, which we denote SQ_n. For example, the permutation 1423 is associated with the square Young diagram SQ_4 as shown in Figure 3.4.

Now we extend the idea of permutation pattern avoidance to Young diagrams Y that are not necessarily square, but which admit at least one *transversal* (namely, a set of dots such that each row and each column of Y contains exactly one dot, as in the middle diagram of Figure 3.3 for example). In particular, a Young diagram Y admits a transversal if and only if the number of columns of Y and the number of rows of Y are the same number n, and the ith row from the top of Y contains at least $(n+1)-i$ cells, $1 \leq i \leq n$.

We use (σ, Y) to denote a transversal σ contained in a Young diagram Y (abbreviating (σ, Y) to σ when the context is clear). We now say that (σ, Y) *avoids* the permutation π if Y does not contain as a subdiagram the square Young diagram associated with π. For example, in Figure 3.3, (σ_1, Y) avoids π, even though the rightmost three dots of σ_1 form a pattern coinciding with the dots of π, because the column of the rightmost of these dots and the row of the lowermost of these dots intersect outside Y. On the other hand, (σ_2, Y) contains (does not avoid) π, as illustrated by the shaded cells.

Let $S_Y(\pi)$ denote the set of all transversals (σ, Y) that avoid the permutation π. Permutations α and β are *shape-Wilf-equivalent*, written $\alpha \overset{s}{\sim} \beta$ (not to be confused with the symbol for f-Wilf-

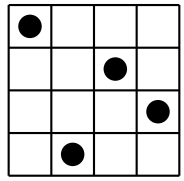

Figure 3.4: Young diagram associated with permutation 1423

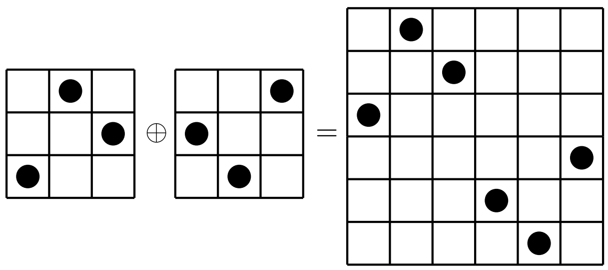

Figure 3.5: The $\oplus$ operation on two permutations, demonstrating $312 \oplus 231 = 312564$

equivalence, $\overset{f}{\sim}$), if

$$|S_Y(\alpha)| = |S_Y(\beta)| \quad \text{for all Young diagrams } Y.$$

Shape-Wilf-equivalence implies Wilf-equivalence, by restricting the range of Y to the set $\{SQ_n\}$.

The power of shape-Wilf-equivalence lies in the following proposition, which gives rise to an infinite family of shape-Wilf-equivalent pairs of permutations from a single example. Recall the definition of $\oplus$, which in terms of transversals in Young diagrams is shown in Figure 3.5.

Proposition 3.4.1 ([3])**.** *Let α and β be permutations and suppose $\alpha \overset{s}{\sim} \beta$. Then $\alpha \oplus \gamma \overset{s}{\sim} \beta \oplus \gamma$ for all permutations γ.*

The following two theorems, along with Proposition 3.4.1, imply the existence of the two infinite families of Proposition 3.2.1.

Theorem 3.4.2 (Stankova and West [59])**.** *$231 \overset{s}{\sim} 312$.*

Theorem 3.4.3 (Backelin, West and Xin [3])**.** *$123 \ldots n \overset{s}{\sim} n(n-1)(n-2)\ldots 1$ for all n.*

Alternative proofs of Theorem 3.4.2 have been given by Jelínek [37] and by Bloom and Saracino [8]. Jelínek [37, p.204] points out that Theorem 3.4.3 is implied by a result on pattern avoidance in matchings due to Chen et al. [19, Thm. 1.1]. An alternative proof of Theorem 3.4.3 has been given by Krattenthaler [43], and alternative proofs of special cases of Theorem 3.4.3 have been given by Jelínek [37] and by Bloom and Elizalde [7].

3.5 Shape-inv-Wilf-equivalence

First extend the definition of *inversion number* to a transversal (σ, Y): we define $\mathrm{inv}(\sigma)$ to be the number of unordered pairs of distinct dots u, v of σ such that the line through u and v runs from bottom-left to top-right. Note that part of the line between u and v may extend outside of Y. For example, in Figure 3.3 we have $\mathrm{inv}(\sigma_1) = 22$. This definition of inversion number for the square Young diagram associated with a permutation π is consistent with the definition of inversion number of π; indeed, $\mathrm{inv}(1423) = 2$ as shown in Figure 3.4.

Then, for a permutation π, define the generating function

$$I_Y(\pi, q) = \sum_{\sigma \in S_Y(\pi)} q^{\mathrm{inv}(\sigma)}$$

and call permutations α and β *shape-inv-Wilf-equivalent*, denoted $\alpha \overset{si}{\sim} \beta$, if

$$I_Y(\alpha, q) = I_Y(\beta, q) \quad \text{for all Young diagrams } Y.$$

Shape-inv-Wilf-equivalence implies shape-Wilf-equivalence by setting $q = 1$, and implies inv-Wilf-equivalence by restricting the range of Y to the set $\{SQ_n\}$. Note that $I_n(\pi, q) = I_{SQ_n}(\pi, q)$ (where $I_n(\pi, q)$ is defined in (3.2)).

The hierarchy of these relationships is shown in Figure 3.6. (Note that even-Wilf-equivalence will be defined later.)

Our main theorem generalizes the shape-Wilf-equivalence of 231 and 312 stated in Theorem 3.4.2 to shape-inv-Wilf-equivalence.

Theorem 3.5.1 (Main Theorem)**.** $231 \overset{si}{\sim} 312$.

We shall also prove the following extension of Proposition 3.4.1, in Section 3.7.

Proposition 3.5.2. *Let α and β be permutations and suppose $\alpha \overset{si}{\sim} \beta$. Then $\alpha \oplus \gamma \overset{si}{\sim} \beta \oplus \gamma$ for all permutations γ.*

Combination of Theorem 3.5.1 with Proposition 3.5.2 constructs an infinite family of shape-inv-Wilf-equivalent pairs of permutations.

Corollary 3.5.3. $231 \oplus \gamma \overset{si}{\sim} 312 \oplus \gamma$ *for all permutations γ.*

Since shape-inv-Wilf-equivalence implies inv-Wilf-equivalence, Corollary 3.5.3 implies Theorem 3.3.6.

We now describe another corollary of Theorem 3.5.1. Let (σ, Y) be a transversal. Let $A_Y(\pi)$ be the set of all transversals (σ, Y) with $\mathrm{inv}(\sigma)$ even, that avoid the permutation π, and write

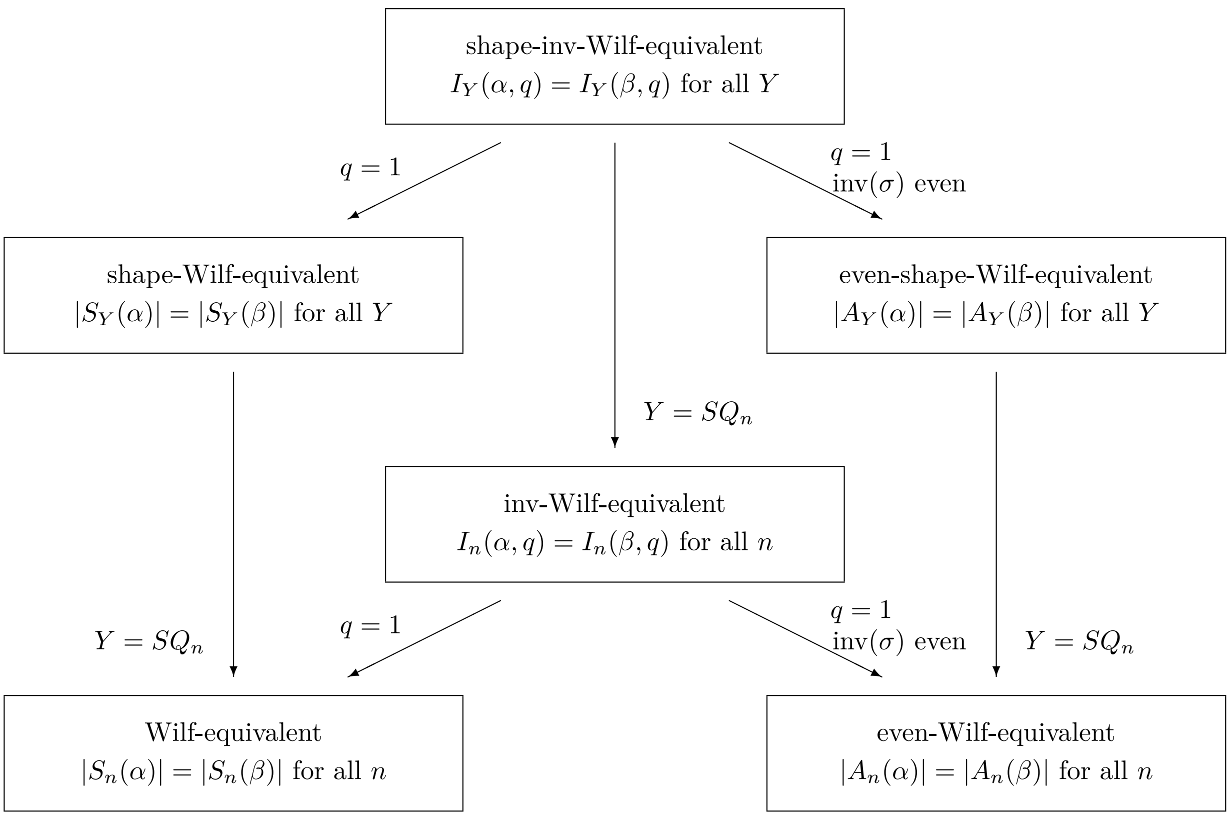

Figure 3.6: Hierarchy of Wilf-equivalence relationships for permutations α, β, where Y is a Young diagram and σ is a transversal

$A_n(\alpha) = A_{SQ_n}(\alpha)$. Baxter and Jaggard [4] define permutations α and β to be *even-shape-Wilf-equivalent* if

$$|A_Y(\alpha)| = |A_Y(\beta)| \quad \text{for all Young diagrams } Y,$$

and *even-Wilf-equivalent* if

$$|A_n(\alpha)| = |A_n(\beta)| \quad \text{for all } n.$$

See Figure 3.6 for the relationship of these definitions to the other Wilf-equivalent definitions.

Conjecture 13 of [4] states that 231 is even-shape-Wilf-equivalent to 312. Since $|A_Y(\pi)| = (I_Y(\pi, -1) + I_Y(\pi, 1))/2$ for all permutations π, shape-inv-Wilf-equivalence implies even-shape-Wilf-equivalence. Thus, we prove this conjecture as an immediate consequence of Theorem 3.5.1. Combination with [4, Lemma 7] and [59] proves that 231564 and 312564 are even-Wilf-equivalent, as are 465132 and 465213, thus completing the classification of pairs of even-Wilf-equivalent permutations of size 6, as laid out in Table 3 of [4, Sect. 5.3]. Note that this new result does not contribute to the classification of even-Wilf-equivalent permutations of size 5, as laid out in Table 2 of [4, Sect. 5.2], which therefore remains an open problem.

We mention, without proof, an extension to another result appearing in [59] that can be obtained using our methods. Stankova and West [59, Cor. 2] describe specific pairs of Young diagrams Y, Y' for which $|S_Y(231)| = |S_{Y'}(231)|$. Using essentially the same argument as in [59, Section 5], the stronger result $I_Y(231, q) = I_{Y'}(231, q)$ holds for these pairs.

3.6 Notation and overview of the proof of Theorem 3.5.1

From this point on, all Young diagrams Y admit at least one transversal; that is, as described previously, Y must have n rows and n columns for some n, and must contain the staircase Young diagram of size n, namely the Young diagram whose row sizes are $n, n-1, \ldots, 1$ when reading down the rows. In this case, all transversals σ of Y have size $|\sigma| = n$ and we define the *size* of Y to be $|Y| = n$. We use the convention that the rows of a Young diagram of size $n+1$ are labelled $0, 1, 2, \ldots, n$ from top to bottom, the columns are labelled $0, 1, 2, \ldots, n$ from left to right, and (i, j) denotes the cell at row i and column j.

Let $\overline{Y}$ denote the transpose of a Young diagram (reflected across its main diagonal which runs from top-left to bottom-right).

Figure 3.3 uses shaded cells to illustrate that the transversal (σ_2, Y) contains the permutation π. A subdiagram of (σ, Y) equal to the square Young diagram associated with π is called an *instance* of π in (σ, Y).

We say that a Young diagram Z is *smaller* than a Young diagram Y if Z is a proper subdiagram of Y.

We now summarize the idea of the proof method for Proposition 3.4.1 given by Backelin, West, and Xin [3]. Suppose that σ is a transversal in a Young diagram Y that avoids the permutation $\alpha \oplus \gamma$.

We shade the cells of (σ, Y) in such a way that the shaded portion itself forms a transversal in a Young diagram and avoids α. Then we can establish a bijection between $\alpha \oplus \gamma$-avoiding transversals and $\beta \oplus \gamma$-avoiding transversals by simply replacing the shaded portion avoiding α with a portion having the same shape that avoids β, using the assumed bijection between α-avoiding transversals and β-avoiding transversals.

We shall prove Proposition 3.5.2 in Section 3.7 by adapting the Backelin-West-Xin proof to take account of inversion number. Although the shading of cells is the same, we must now account for the positions of the dots. We shall consider four cases, according to whether each of the two dots of a pair lies in a shaded or an unshaded cell.

We next summarize the proof method for Theorem 3.4.2 given by Stankova and West [59].

1. Define a function $T(Y)$ which counts the number of 231-avoiding transversals in Young diagram Y (in other words, $T(Y) = |S_Y(231)|$). Then Theorem 3.4.2 is equivalent to the statement that $T(Y) = T(\overline{Y})$ for all Young diagrams Y.

2. Establish a functional relation called "row decomposition" [59, Thm. 2 and Cor. 1], which computes $T(Y)$ from values $T(Z)$ for Young diagrams Z smaller than Y. Row decomposition has a relatively straightforward combinatorial proof. The form of row decomposition given in [59, Thm. 2] is a summation of multiplied pairs of values of $T(Z)$, and the consequence given in [59, Cor. 1] is of the form $T(Y) = T(Z_1) + T(Z_2)T(Z_3)$. This latter result [59, Cor. 1] forms a central part of the proof.

3. Establish another functional relation called "column decomposition" [59, Lem. 3], which is a reflected version of row decomposition. Whereas row decomposition has a combinatorial proof, column decomposition is proved instead by induction using the commutativity of row and column decompositions shown in the left side of Figure 3.7. That is, if we apply row decomposition to $T(Y)$ and then (using the inductive hypothesis) column decomposition to the smaller resulting parts, we obtain the same formula as if we applied column decomposition to $T(Y)$ and then row decomposition to the smaller resulting parts.

4. Use row decomposition and column decomposition together to prove by induction that $T(Y) = T(\overline{Y})$ for all Young diagrams Y.

We shall prove Theorem 3.5.1 in Section 3.8 by adapting the Stankova-West proof to take account of inversion number, as follows.

1. Define a function $I(Y)$ which maps a Young diagram Y to a generating function in q, and which counts the number of 231-avoiding transversals σ in Y in terms of the inversion number $\mathrm{inv}(\sigma)$. Then Theorem 3.5.1 is equivalent to the statement that $I(Y) = I(\overline{Y})$ for all Young diagrams Y (Lemma 3.8.3).

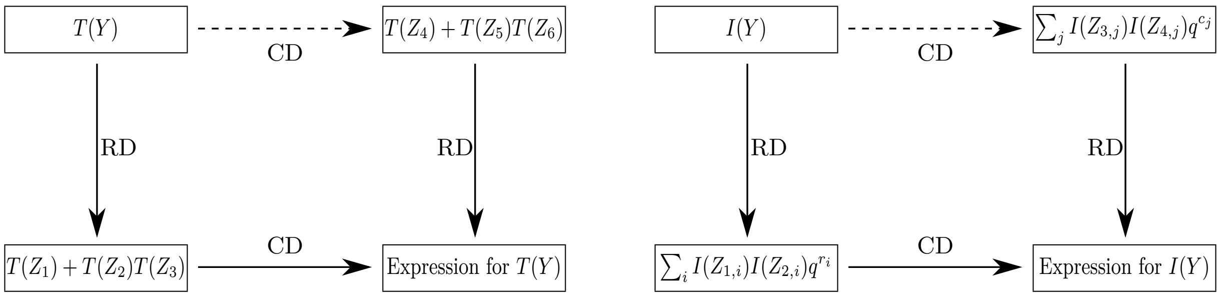

Figure 3.7: Commutativity of row decomposition (RD) and column decomposition (CD), where the Z are Young diagrams smaller than Y.

2. Establish an inversion-number version of row decomposition which computes $I(Y)$ from values $I(Z)$ for Young diagrams Z smaller than Y (Proposition 3.8.5). We use a similar argument to that used in the combinatorial proof of [59, Thm. 2], but we must incorporate inversion number carefully. The form of this row decomposition is the same as the form given by [59, Thm. 2] except that each term is multiplied by a power of q which depends on the index of summation. This proof, which depends on the shape of Y and which requires additional notation, will be motivated in Section 3.8. We also give in passing the inversion-number version of [59, Cor. 1] as Corollary 3.8.6, although it is not used as a central part of the proof as it is in Stankova-West.

3. Establish column decomposition as a reflected version of row decomposition (Proposition 3.8.8). Column decomposition is likewise proved by induction using the commutativity of row and column decompositions as shown in the right side of Figure 3.7. Various modifications have been applied to [59, Lem. 3] to form the proof, and we leave further details for Section 3.8.

4. Use row decomposition and column decomposition together to prove by induction that $I(Y) = I(\overline{Y})$ for all Young diagrams Y (Proposition 3.8.4).

3.7 Proof of Proposition 3.5.2

For convenience, we recall the statement of Proposition 3.5.2.

Proposition 3.5.2. *Let α and β be permutations and suppose $\alpha \overset{si}{\sim} \beta$. Then $\alpha \oplus \gamma \overset{si}{\sim} \beta \oplus \gamma$ for all permutations γ.*

Proof of Proposition 3.5.2. (See Figure 3.8.) Let (σ, Y) be a transversal avoiding $\alpha \oplus \gamma$. Color a cell of Y light grey if it lies to the upper-left of all dots of some instance of γ contained in Y. Let σ' be the set of dots of σ that are on light grey cells, and let δ be the set of dots of σ that are not in σ'. Recolor a light grey cell dark grey if there exist dots $u, v \in \sigma'$, not necessarily distinct,

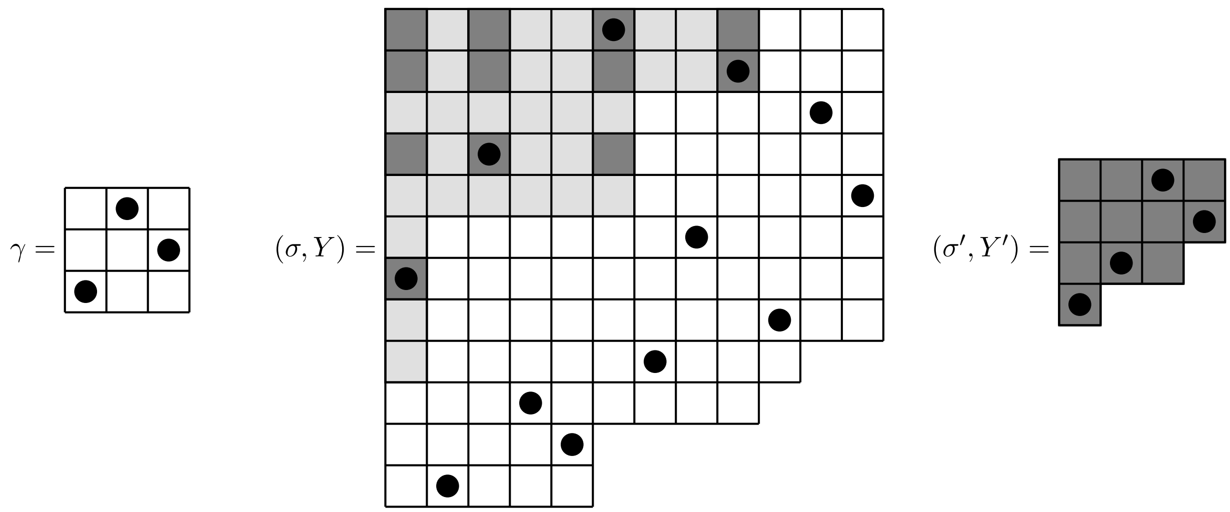

Figure 3.8: Illustration used in the proof of Proposition 3.5.2

such that the cell is in the same row as u and in the same column as v. Let the dark grey cells be Young diagram Y' as a subdiagram of Y. Then σ' is a transversal of Y'. The transversal (σ', Y') avoids α; otherwise the original transversal (σ, Y) contains $\alpha \oplus \gamma$.

Figure 3.8 illustrates this shading for a particular transversal σ and Young diagram Y, with $\gamma = 312$.

Use the implied bijection $\alpha \overset{si}{\sim} \beta$ to map σ' to transversal τ' of Y', which avoids β and has the same inversion number as σ'. Form a new transversal $\tau = \tau' \cup \delta$ of Y that avoids $\beta \oplus \gamma$. We claim that $\text{inv}(\tau) = \text{inv}(\sigma)$. If so, then by mapping $\sigma = \sigma' \cup \delta$ to $\tau = \tau' \cup \delta$, we preserve the inversion number and obtain a $(\beta \oplus \gamma)$-avoiding permutation in Y. By an analogous argument, we can reverse this process to obtain a map from τ to σ. So we have a bijection between the set of $(\alpha \oplus \gamma)$-avoiding permutations on Y and the set of $(\beta \oplus \gamma)$-avoiding permutations on Y that preserves inversion number. It remains to prove that $\text{inv}(\tau) = \text{inv}(\sigma)$.

Let $Q(u, v) = 1$ if dot v is to the upper-right of dot u and $Q(u, v) = 0$ otherwise. Then

$$\text{inv}(\pi) = \sum_{u,v \in \pi} Q(u, v)$$

for all transversals π, and so

$$\begin{aligned} \text{inv}(\tau) &= \sum_{u,v \in \tau} Q(u, v) = \sum_{u,v \in \tau' \cup \delta} Q(u, v) \\ &= \sum_{u,v \in \tau'} Q(u, v) + \sum_{u,v \in \delta} Q(u, v) + \sum_{u \in \delta, v \in \tau'} Q(u, v) + \sum_{u \in \tau', v \in \delta} Q(u, v). \end{aligned} \tag{3.4}$$

From the assumed inversion-preserving bijection, we have $\sum_{u,v\in\tau'} Q(u,v) = \sum_{u,v\in\sigma'} Q(u,v)$. We next show that $\sum_{v\in\tau'} Q(u,v) = \sum_{v\in\sigma'} Q(u,v)$ for a given $u \in \delta$. Let $u \in \delta$ and consider the set of cells of Y' (possibly empty) which are to the right of u, which we call Z. Now all cells of Z must be above u; otherwise, a cell of Z (and therefore Y') to the right or lower-right of u would mean u is in σ' instead of δ. Therefore, for some integer z, the cells of Z compose the z rightmost columns of Y'. Since every transversal of Y' must have exactly one dot in each of its columns, then $\sum_{v\in\tau'} Q(u,v) = z = \sum_{v\in\sigma'} Q(u,v)$. It follows that $\sum_{u\in\delta,v\in\tau'} Q(u,v) = \sum_{u\in\delta,v\in\sigma'} Q(u,v)$. By a similar argument involving rows, $\sum_{u\in\tau',v\in\delta} Q(u,v) = \sum_{u\in\sigma',v\in\delta} Q(u,v)$. Thus we switch τ' with σ' in (3.4), completing the proof.

□

3.8 Proof of Theorem 3.5.1

We define the "non-inversion number" $\text{inv}^*(\sigma)$ to be the number of unordered pairs of distinct dots u, v of σ such that the line through u and v in the Young diagram Y runs from top-left to bottom-right. It is easily shown that $\text{inv}(\sigma) + \text{inv}^*(\sigma) = \binom{|\sigma|}{2}$, and so for transversals σ_1, σ_2 with equally many dots, $\text{inv}(\sigma_1) = \text{inv}(\sigma_2)$ if and only if $\text{inv}^*(\sigma_1) = \text{inv}^*(\sigma_2)$. For example, the transversal σ_1 of Figure 3.3 satisfies $\text{inv}(\sigma_1) = 22$ and $\text{inv}^*(\sigma_1) = 6$, with $\text{inv}(\sigma_1) + \text{inv}^*(\sigma_1) = 28 = \binom{8}{2}$.

We also define, as an analogue of $I_Y(\pi, q)$,

$$I^*_Y(\pi, q) = \sum_{\sigma\in S_Y(\pi)} q^{\text{inv}^*(\sigma)}.$$

Since $\text{inv}(\sigma) + \text{inv}^*(\sigma) = \binom{|\sigma|}{2}$ for all transversals σ, then

$$I^*_Y(\pi, q) = I_Y(\pi, q^{-1})q^{\binom{|Y|}{2}} \quad \text{for all permutations } \pi.$$

Therefore, $\alpha \overset{si}{\sim} \beta$ if and only if

$$I^*_Y(\alpha, q) = I^*_Y(\beta, q) \quad \text{for all Young diagrams } Y.$$

We use the shorthand $I(Y)$ to represent $I^*_Y(231, q)$. We consider the size 0 (empty) Young diagram to be trivial, and all others to be nontrivial. We use the convention that $I(Y) = 1$ when Y is trivial. Let $a_0a_1 \ldots a_n$ denote the Young diagram of size $n + 1$ with a_i cells in row i for $0 \le i \le n$, and $\emptyset$ denote the trivial Young diagram.

Table 3.3 gives $I(Y) = I^*_Y(231, q)$ for all Young diagrams up to size 4. The table suggests some functional relations between generating functions $I(Y)$ for certain related Y. In particular, we note the following simple functional relation:

Proposition 3.8.1. *Let* $Y_1 = a_0 a_1 \dots a_{n-1}1$ *and* $Y_2 = a_0 a_1 \dots a_{n-1}2$ *be Young diagrams. Then*

$$I(Y_2) = I(Y_1)(1+q).$$

Proof. Let $Y_0 = (a_0-1)(a_1-1)\dots(a_{n-1}-1)$ be the Young diagram formed by deleting the leftmost column of Y_1. We calculate $I(Y_2)$ by summing the contributions when the dot in the bottom row of Y_2 lies in column 0 and when it lies in column 1. Since that dot can never play the role of 3 in an instance of 231 in Y_2, we obtain

$$I(Y_2) = I(Y_0) + I(Y_0)q,$$

where the additional factor of q when the dot lies in column 1 arises from an additional contribution of 1 to $\mathrm{inv}^*(\sigma)$. Similarly, $I(Y_1) = I(Y_0)$. □

Example 3.8.2. *As can be seen in Table 3.3,* $I(4442) = 1 + 2q + 3q^2 + 3q^3 + q^4 = (1 + q + 2q^2 + q^3)(1 + q) = I(4441)(1 + q)$.

Other relations, however, are not so obvious. The table suggests that $I(Y) = I(\overline{Y})$, a statement linked to the shape-Wilf-equivalence of 231 and 312.

Lemma 3.8.3. $231 \overset{si}{\sim} 312$ *if and only if*

$$I(Y) = I(\overline{Y}) \quad \text{for all Young diagrams } Y.$$

Proof. Since the square Young diagrams associated with 231 and 312 are transposes of each other, and transposition does not change the inversion number of a transversal,

$$\begin{aligned} 231 \overset{si}{\sim} 312 &\iff I^*_Y(231,q) = I^*_Y(312,q) \text{ for all } Y \\ &\iff I^*_Y(231,q) = I^*_{\overline{Y}}(231,q) \text{ for all } Y \\ &\iff I(Y) = I(\overline{Y}) \text{ for all } Y. \end{aligned}$$

□

Thus the proof of Theorem 3.5.1 only requires the proof of the following statement.

Proposition 3.8.4. $I(Y) = I(\overline{Y})$ *for all Young diagrams* Y*.*

The proof of Proposition 3.8.4 relies on equating the expressions obtained from Propositions 3.8.5 and 3.8.8, which we call *row decomposition* and *column decomposition*, respectively. The motivation for row decomposition, a vital part of the proof, is based on the bijection in Stankova and West's proof of [59, Thm. 2]. The bijection here is the same, but the proof is adapted to include inversion number in this row decomposition.

Y	$I(Y)$
$\emptyset$	1
$SQ_1 = 1$	1
21	1
$SQ_2 = 22$	$1+q$
321	1
322	$1+q$
331	$1+q$
332	$1+2q+q^2$
$SQ_3 = 333$	$1+q+2q^2+q^3$
4321	1
4322	$1+q$
4331	$1+q$
4332	$1+2q+q^2$
4333	$1+q+2q^2+q^3$
4421	$1+q$
4422	$1+2q+q^2$
4431	$1+2q+q^2$
4432	$1+3q+3q^2+q^3$
4433	$1+2q+3q^2+3q^3+q^4$
4441	$1+q+2q^2+q^3$
4442	$1+2q+3q^2+3q^3+q^4$
4443	$1+q+3q^2+3q^3+3q^4+q^5$
$SQ_4 = 4444$	$1+q+2q^2+3q^3+3q^4+3q^5+q^6$

Table 3.3: The generating function $I(Y) = I^*_Y(231, q)$ for various Y

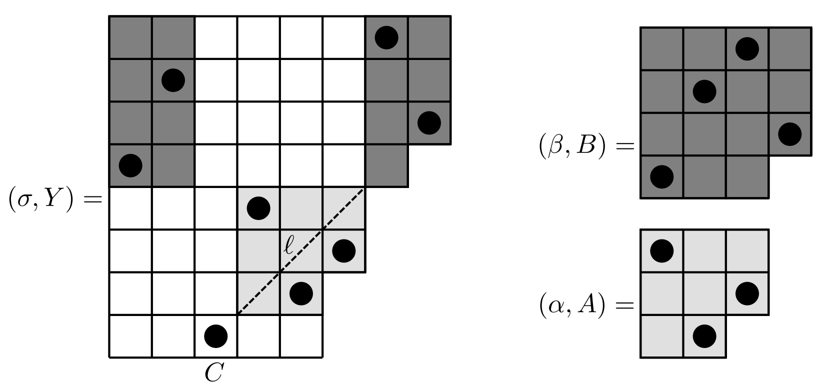

Figure 3.9: Row decomposition

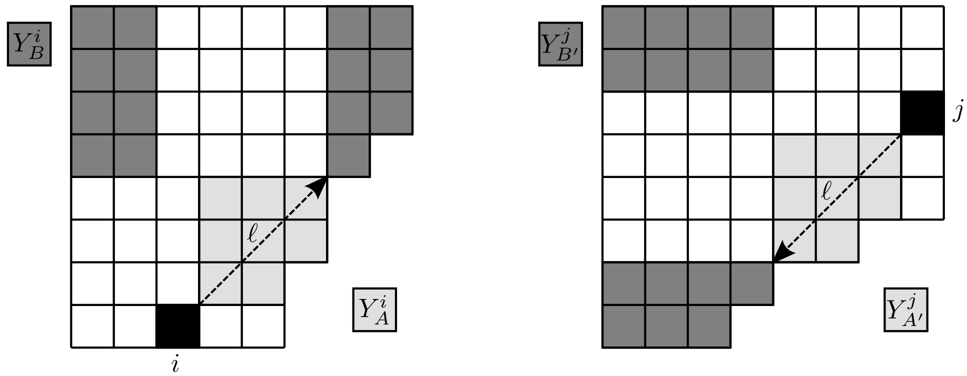

Figure 3.10: Row decomposition (left) and column decomposition (right)

We use Figure 3.9 as an example for this motivation. The transversal (σ, Y) in Figure 3.9 avoids 231, and the cell C marks the location of the bottom dot. Draw a line ℓ from the top right corner of C diagonally up and right until it touches the border of Y, and form A as the light grey region in the figure with ℓ as its long diagonal.

We shall see that A is a Young diagram, and that the part of σ that is in A forms a transversal α of A. Excluding α and (n, i), the rest of σ lies in B (the dark grey region formed by taking "the rest of Y", joining parts together), and we shall see also that B is a Young diagram, and that the part of σ that is in B forms a transversal β of B.

In this manner, we break up σ into two smaller transversals α and β, one which is 231-avoiding in A, and one which is 231-avoiding in B. Furthermore, the converse holds, since if transversals are chosen for subregions A and B of Y, an instance of 231 must be contained entirely in A or in B as enforced by the locations of A and B as well as the shape of Y. This method implicitly forms a bijection between the set of 231-avoiding transversals σ of Y with a given dot in C, and the set of pairs of 231-avoiding transversals α of A and β of B.

Furthermore, we must account for the inversion number. We shall see that $\mathrm{inv}^*(\sigma)$ can be calculated by adding $\mathrm{inv}^*(\alpha)$ and $\mathrm{inv}^*(\beta)$, and adding $i(|A|+1)$, where i is the number of columns of Y to the left of C. As illustrated in Figure 3.9, $\mathrm{inv}^*(\alpha) = 2$, $\mathrm{inv}^*(\beta) = 2$, $i = 2$, and $|A| = 3$, and so $\mathrm{inv}^*(\alpha) + \mathrm{inv}^*(\beta) + i(|A|+1) = 2 + 2 + 2 \cdot 4 = 12 = \mathrm{inv}^*(\sigma)$.

This motivates the definitions used for row decomposition, as illustrated in Figure 3.10. Let Y be a nontrivial Young diagram of size $n+1$, whose bottom row has $r+1$ cells. Let i be an integer satisfying $0 \le i \le r$. Draw a (possibly zero-length) line segment ℓ from the top-right corner of the cell (n, i) upward and right at a 45-degree angle until it first touches the border of Y. Define Y_A^i to be the (possibly trivial) subdiagram obtained by deleting all rows and columns which ℓ does not cross, and define Y_B^i to be the (possibly trivial) subdiagram obtained by deleting all rows and columns which ℓ crosses as well as the row and column of (n, i).

From this point on, $\sum_i$ denotes summation over all i for which the summation term is defined.

Proposition 3.8.5 (Row decomposition)**.**

$$I(Y) = \sum_i I(Y_A^i)I(Y_B^i)q^{i(|Y_A^i|+1)} \text{ for all nontrivial Young diagrams } Y.$$

We will delay the proof of Proposition 3.8.5 for now.

Proposition 3.8.5 gives rise to another functional relation between generating functions $I(Y)$, as given in Corollary 3.8.6. This generalizes [59, Cor. 1], which was central to that paper's use of row and column decompositions. In contrast, Corollary 3.8.6 is not required for our proofs, and its proof is omitted.

Corollary 3.8.6. *Let $Y_0 = a_0a_1 \dots a_{n-1}r$ and $Y = a_0a_1 \dots a_{n-1}(r+1)$ be Young diagrams, where $r \geq 1$. Then*

$$I(Y) = I(Y_0) + I(Y_B^r)(q^r - q^{r-1}) + I(Y_A^{r-1})I(Y_B^{r-1})q^{(r-1)(|Y_A^{r-1}|+1)}.$$

Although Table 3.3 is small enough to be generated by hand, Proposition 3.8.5 and Corollary 3.8.6 can be used to compute $I(Y)$ recursively. For computing the value of $I(Y)$ where the values of $I(Z)$ are known for all Young diagrams Z smaller than Y, Corollary 3.8.6 usually has an advantage over Proposition 3.8.5, since there are only three terms to compute in Corollary 3.8.6, whereas there is a summation of potentially more terms in Proposition 3.8.5.

Example 3.8.7. *Let $Y = 55544$. Then Y_A^i and Y_B^i are defined only for $0 \leq i \leq 3$. Then*

$$Y_A^0 = 4443, Y_B^0 = \emptyset, Y_A^1 = 332, Y_B^1 = SQ_1, Y_A^2 = SQ_1, Y_B^2 = SQ_3, Y_A^3 = \emptyset, Y_B^3 = 4443.$$

Referring to $I(Y)$ when $Y = \emptyset, SQ_1, 332, SQ_3, 4443$ in Table 3.3,

$$\begin{aligned} I(Y) &= I(4443)I(\emptyset) + I(332)I(SQ_1)q^{1\cdot 4} + I(SQ_1)I(SQ_3)q^{2\cdot 2} + I(\emptyset)I(4443)q^{3\cdot 1} \\ &= (1+q^3)(1+q+3q^2+3q^3+3q^4+q^5) + q^4(1+2q+q^2) + q^4(1+q+2q^2+q^3) \\ &= 1+q+3q^2+4q^3+6q^4+7q^5+6q^6+4q^7+q^8. \end{aligned}$$

Alternatively, if we know that $I(55543) = 1+q+4q^2+4q^3+7q^4+6q^5+4q^6+q^7$, we can use Corollary 3.8.6 directly with $r = 3$:

$$\begin{aligned} I(Y) &= I(55543) + I(4443)(q^3 - q^2) + I(SQ_1)I(SQ_3)q^{2\cdot 2} \\ &= 1+q+4q^2+4q^3+7q^4+6q^5+4q^6+q^7 \\ &\quad + (1+q+3q^2+3q^3+3q^4+q^5)(q^3-q^2) + q^4(1+q+2q^2+q^3) \\ &= 1+q+3q^2+4q^3+6q^4+7q^5+6q^6+4q^7+q^8. \end{aligned}$$

We use a similar concept for column decomposition, which is the reflected version of row decomposition as shown in Figure 3.10. The subdiagrams $Y^j_{A'}$ and $Y^j_{B'}$, illustrated in Figure 3.10, are given by $Y^j_{A'} = \overline{\overline{Y}^j_A}$ and $Y^j_{B'} = \overline{\overline{Y}^j_B}$.

Unlike the proof for row decomposition, the proof for column decomposition relies on induction and use of row and column decomposition to equate the two sides. The idea of the proof, a simplified version of the proof of [59, Lem. 3], is based on "commutativity of row and column decompositions", as shown in the right side of Figure 3.7.

Proposition 3.8.8 (Column decomposition)**.**

$$I(Y) = \sum_j I(Y^j_{A'})I(Y^j_{B'})q^{j(|Y^j_{A'}|+1)} \text{ for all nontrivial Young diagrams } Y.$$

We will also delay the proof of Proposition 3.8.8.

Example 3.8.9. *Let $Y = 55544$. Then $Y^j_{A'}$ and $Y^j_{B'}$ are defined only for $0 \le j \le 2$. Then*

$$Y^0_{A'} = SQ_4, Y^0_{B'} = \emptyset, Y^1_{A'} = SQ_3, Y^1_{B'} = SQ_1, Y^2_{A'} = \emptyset, Y^2_{B'} = SQ_4.$$

Then, referring to $I(Y)$ when $Y = \emptyset, SQ_1, SQ_3, SQ_4$ in Table 3.3,

$$\begin{aligned} I(Y) &= I(SQ_4)I(\emptyset) + I(SQ_3)I(SQ_1)q^{1\cdot 4} + I(\emptyset)I(SQ_4)q^{2\cdot 1} \\ &= (1+q^2)(1+q+2q^2+3q^3+3q^4+3q^5+q^6) + q^4(1+q+2q^2+q^3) \\ &= 1+q+3q^2+4q^3+6q^4+7q^5+6q^6+4q^7+q^8, \end{aligned}$$

as previously calculated in Example 3.8.7 using row decomposition.

Propositions 3.8.5 and 3.8.8 imply Proposition 3.8.4 as follows, which gives us Theorem 3.5.1.

Proof of Proposition 3.8.4. We need to prove $I(Y) = I(\overline{Y})$ for all Young diagrams Y. The statement is true when Y is trivial. Suppose it is true for all Young diagrams whose size is less than that of Y. By the inductive hypothesis and the fact that $|Z| = |\overline{Z}|$ for all Young diagrams Z,

$$\begin{aligned} I(Y) &= \sum_j I(Y^j_{A'})I(Y^j_{B'})q^{j(|Y^j_{A'}|+1)} \\ &= \sum_j I(\overline{Y^j_{A'}})I(\overline{Y^j_{B'}})q^{j(|\overline{Y^j_{A'}}|+1)} \\ &= \sum_j I(\overline{Y}^j_A)I(\overline{Y}^j_B)q^{j(|\overline{Y}^j_A|+1)} \\ &= I(\overline{Y}). \end{aligned}$$

∎

To complete the proof of Theorem 3.5.1, it remains to prove Proposition 3.8.5 and 3.8.8. We first prove Proposition 3.8.5, based on the motivation given by Figure 3.9 and the text after the statement of Proposition 3.8.4.

Proof of Proposition 3.8.5. We need to prove

$$I(Y) = \sum_i I(Y_A^i)I(Y_B^i)q^{i(|Y_A^i|+1)} \text{ for all nontrivial Young diagrams } Y.$$

Let Y be a Young diagram of size $n+1$. We use the shorthand $S(Y)$ to mean $S_Y(231)$, the set of all transversals in Y that avoid 231, and $S_i(Y)$ for the set of all $\sigma \in S(Y)$ that have a dot in (n, i). Let $\sigma \in S_i(Y)$ and let $a = |Y_A^i|$.

We claim that all dots of σ in columns $0, 1, 2, \dots, i-1$ lie in rows above all those in columns $i, i+1, i+2, \dots, i+a$. If either Y_A^i or Y_B^i is trivial, then this holds true. Otherwise, suppose for a contradiction that the claim does not hold. Then there exists a dot t in column $j < i$ which is lower than a dot v in column $i+k$ with $0 < k \le a$ but higher than all dots in columns $i, i+1, \dots, i+k-1$. Then because there are k dots in columns $i, \dots, i+k-1$, then at least one of these dots, which we call u, must be in row $n-k+1$ or above, and below t. Since the column $i+k$ containing v extends as far down as row $n-k+1$ by definition of the Young diagram Y_A^i, then dots t, u, v form an instance of 231 in Y, contradicting that σ avoids 231 in Y. Thus, the above claim holds.

Let α be the set of dots in columns $i+1, i+2, \dots, i+a$. It follows immediately that the dots in α span the rows $n-a, n-a+1, \dots, n-1$, and all dots above the row $n-a$, whose set we call β, are either in columns to the left of i or to the right of $i+a$. Thus, if $\sigma \in S_i(Y)$, then $\alpha \in S(Y_A^i)$ and $\beta \in S(Y_B^i)$, and α and β are uniquely determined from σ.

Conversely, for a fixed i, if we choose $\alpha \in S(Y_A^i)$ and $\beta \in S(Y_B^i)$ in advance, then there is only one possible σ which results in the given α and β from the above process. We will show that σ avoids 231. Suppose for a contradiction that σ contains an instance of 231 formed from dots t, u, v from left to right. We show that t, u, v are all in Y_A^i or all in Y_B^i, contradicting $\alpha \in S(Y_A^i)$ and $\beta \in S(Y_B^i)$. We divide by case depending on the location of v. Clearly, v cannot be (n, i). If v is in Y_B^i and left of column i, then t and u must also be in Y_B^i. If v is in Y_B^i and right of column i, then u cannot be (n, i) or in Y_A^i; otherwise the intersection of the row of u with the column of v would be outside Y, contradicting containment. So u, and therefore t, are in Y_B^i. If v is in Y_A^i, then u cannot be (n, i); otherwise, t would be above v. So u, and therefore t, are in Y_A^i. All these cases result in a contradiction. Therefore, $\sigma \in S_i(Y)$.

Thus,

$$\sigma \in S_i(Y) \text{ if and only if } \alpha \in S(Y_A^i) \text{ and } \beta \in S(Y_B^i). \tag{3.5}$$

We claim that

$$\text{inv}^*(\sigma) = \text{inv}^*(\alpha) + \text{inv}^*(\beta) + i(a+1), \tag{3.6}$$

where $a = |Y_A^i|$. Let $Q^*(u,v) = 1$ if dot v is to the lower-right of dot u and $Q^*(u,v) = 0$ otherwise. Then

$$\mathrm{inv}^*(\pi) = \sum_{u,v\in\pi} Q^*(u,v)$$

for all transversals π, and so

$$\begin{aligned}
\mathrm{inv}^*(\sigma) &= \sum_{u,v\in\sigma} Q^*(u,v)\\
&= \sum_{u,v\in\alpha\cup\{(n,i)\}} Q^*(u,v) + \sum_{u,v\in\beta} Q^*(u,v) + \sum_{u\in\alpha\cup\{(n,i)\},v\in\beta} Q^*(u,v) + \sum_{u\in\beta,v\in\alpha\cup\{(n,i)\}} Q^*(u,v)\\
&= \mathrm{inv}^*(\alpha\cup\{(n,i)\}) + \mathrm{inv}^*(\beta) + \sum_{u\in\beta,v\in\alpha\cup\{(n,i)\}} Q^*(u,v)\\
&= \mathrm{inv}^*(\alpha) + \mathrm{inv}^*(\beta) + i(a+1),
\end{aligned}$$

since all the dots in columns $0,1,2,\ldots,i-1$ lie in rows above all those in columns $i,i+1,i+2,\ldots,i+a$. Thus, the above claim holds.

By combining (3.5) and (3.6), we find that

$$\begin{aligned}
I(Y) &= \sum_i \sum_k |\{\sigma\in S_i(Y) : \mathrm{inv}^*(\sigma) = k\}|\, q^k\\
&= \sum_i \sum_k \left|\{\alpha\in S(Y_A^i), \beta\in S(Y_B^i) : \mathrm{inv}^*(\alpha) + \mathrm{inv}^*(\beta) = k - i(|Y_A^i|+1)\}\right| q^k\\
&= \sum_i \left(\sum_c \left|\{\alpha\in S(Y_A^i) : \mathrm{inv}^*(\alpha) = c\}\right| q^c\right)\left(\sum_d \left|\{\beta\in S(Y_B^i) : \mathrm{inv}^*(\beta) = d\}\right| q^d\right) q^{i(|Y_A^i|+1)}\\
&= \sum_i I(Y_A^i) I(Y_B^i) q^{i(|Y_A^i|+1)}.
\end{aligned}$$

□

We now give a brief outline for the proof of Proposition 3.8.8; a full proof will be given later.

Let Y be a Young diagram of size $n+1$. We say that Y_A^i is *inner* if Y_A^i has no cell in the rightmost column of Y; otherwise Y_A^i is *outer*. Similarly, $Y_{A'}^j$ is *inner* if $Y_{A'}^j$ has no cell in the bottom row of Y; otherwise $Y_{A'}^j$ is *outer*. We note that there exists an integer i for which Y_A^i is inner; in particular, if the bottom row of Y has $r+1$ cells, then Y_A^r is trivial. Furthermore, if k is the smallest integer such that Y_A^k is inner, then Y_A^i is inner if and only if $i \geq k$, and $Y_{A'}^j$ is inner if and only if $j \geq k$.

This distinction of inner and outer is used in our motivation for column decomposition. We also require Y to be a non-square Young diagram for the following method; the case where $Y = SQ_{n+1}$ has a simple separate proof.

In Y, mark cell (n,i) on the bottom row, and cell (j,n) on the right column, as shown in the top-left diagram of Figure 3.11. We focus only on the particular term in the row and column decomposition indicated by the marked cells and we want to show that applying row and column decomposition in that order ("row-column decomposition") give the same particular term as applying column and row decomposition in that order ("column-row decomposition"). Now, based on whether each of Y_A^i and $Y_{A'}^j$ is inner or outer, we have four important cases.

Case 1: Suppose that both Y_A^i and $Y_{A'}^j$ are inner, as in Figure 3.11. If we first apply row decomposition, then (j,n) is in Y_B^i and so we apply column decomposition to $I(Y_B^i)$. On the other hand, if we first apply column decomposition, then (n,i) is in $Y_{B'}^j$ and so we apply row decomposition to $I(Y_{B'}^j)$. Figure 3.11 illustrates that the resulting terms are the same.

Case 2: Suppose that Y_A^i is outer and $Y_{A'}^j$ is inner, as in Figure 3.12. Unlike Case 1 above, if we first apply row decomposition, then (j,n) is in Y_A^i instead of Y_B^i. So we apply column decomposition to $I(Y_A^i)$ using summation index $j^* = j-i$. On the other hand, if we first apply column decomposition, then (n,i) is in $Y_{B'}^j$ and so we apply row decomposition to $I(Y_{B'}^j)$, like the above case. Figure 3.12 illustrates that the resulting terms are the same.

Case 3: The case where Y_A^i is inner and $Y_{A'}^j$ is outer is just a mirrored version of Case 2 above.

Case 4: Suppose that both Y_A^i and $Y_{A'}^j$ are outer, as in Figure 3.13. This case requires more care since (j,n) need not be in Y_A^i or Y_B^i, nor does (n,i) need to be in $Y_{A'}^j$ or $Y_{B'}^j$. Let $i < j$. If we apply row decomposition, then (j,n) is in Y_A^i so we apply column decomposition to $I(Y_A^i)$ using summation index $j^* = j - i$.

Now since (n,i) is neither in $Y_{A'}^j$ nor in $Y_{B'}^j$, then the particular term in column-row decomposition does not exist. However, if we switch i and j so that the marked cells are (n,j) and (i,n), then although the particular term in row-column decomposition does not exist, the particular term in column-row decomposition now exists, since (n,j) is in $Y_{A'}^i$. So we apply row decomposition to $I(Y_{A'}^i)$ using summation index $j^* = j - i$. Figure 3.13 illustrates that the resulting terms are the same.

The above switch implicitly covers the case where the cells are $(n,i),(j,n)$ with $i > j$. The above argument also deals with $i = j$.

Putting all these cases together establishes the commutativity of row and column decompositions.

Example 3.8.10. *Let* $Y = \mathrm{TTTTTT9855}$ *(where* $\mathrm{T} = 10$*)*, $i = 1, j = 4$, *as shown in Figure 3.12. Starting from row decomposition, the row decomposition term of index* $i = 1$ *is* $I(Y_A^i)I(Y_B^i)q^{i(|Y_A^i|+1)} = I(Z)I(1)q^9 = I(Z)q^9$, *where* $Z = 88888763$. *Since* $(4,9)$ *of* Y *is in the subdiagram of* Z *in* Y, *apply column decomposition to* $I(Z)$ *and take the term of index* $j-i = 3$: $I(Z_{A'}^{j-i})I(Z_{B'}^{j-i})q^{(j-i)(|Y_{A'}^{j-i}|+1)} =$

$I(332)I(4443)q^{12}$. So the particular term in row-column decomposition is

$$I(332)I(4443)q^{12}q^9 = I(332)I(4443)q^{21}.$$

On the other hand, starting from column decomposition, the column decomposition term of index $j = 4$ is

$$I(Y^j_{A'})I(Y^j_{B'})q^{j(|Y^j_{A'}|+1)} = I(332)I(W)q^{16},$$

where $W = 666655$. Since $(9,1)$ of Y is in the subdiagram of W in Y, apply row decomposition to $I(W)$ and take the term of index $i = 1$: $I(W^i_A)I(W^i_B)q^{i(|W^i_A|+1)} = I(4443)I(1)q^5 = I(4443)q^5$. So the particular term in column-row decomposition is

$$I(332)I(4443)q^{16}q^5 = I(332)I(4443)q^{21},$$

which is the same as above.

We now give a full proof of Proposition 3.8.8.

Proof of Proposition 3.8.8. We need to prove

$$I(Y) = \sum_j I(Y^j_{A'})I(Y^j_{B'})q^{j(|Y^j_{A'}|+1)} \text{ for all nontrivial Young diagrams } Y.$$

The proof is by induction on the size of Y.

The statement holds for the size 1 Young diagram. Now assume it holds for all Young diagrams whose size is less than that of Y. Let $|Y| = n+1$ where $n > 0$.

By Proposition 3.8.5, it is sufficient to prove

$$\sum_i I(Y^i_A)I(Y^i_B)q^{i(|Y^i_A|+1)} = \sum_j I(Y^j_{A'})I(Y^j_{B'})q^{j(|Y^j_{A'}|+1)}. \tag{3.7}$$

If $Y = SQ_{n+1}$, then $Y^i_A = SQ_{n-i} = Y^i_{A'}$ and $Y^i_B = SQ_i = Y^i_{B'}$, and so (3.7) holds trivially.

Otherwise, $Y \neq SQ_{n+1}$. We will prove (3.7) by applying the assumed (under induction) column decomposition to Y^i_A or Y^i_B of the LHS depending on i, applying row decomposition to $Y^j_{A'}$ or $Y^j_{B'}$ of the RHS depending on j, and comparing the terms to show that they are equal.

Let k be the smallest integer such that Y^k_A is inner. As stated above, k exists, and Y^i_A is inner if and only if $i \geq k$, and $Y^j_{A'}$ is inner if and only if $j \geq k$. Furthermore, since $Y \neq SQ_{n+1}$, then $i, j, k < n$.

Now consider (3.7). On the LHS, if $i \geq k$, then Y^i_A is inner, so apply the assumed column decomposition to $I(Y^i_B)$ using summation index j; otherwise Y^i_A is outer, so apply it to $I(Y^i_A)$ using summation index $j^* = j - i \geq 0$. On the RHS, if $j \geq k$, then $Y^j_{A'}$ is inner, so apply row decomposition to $I(Y^j_{B'})$ using summation index i; otherwise $Y^j_{A'}$ is outer, so apply it to $I(Y^j_{A'})$ using summation index $i^* = i - j \geq 0$. For ease of presentation, $Y^{i,j}_{AB'}$ denotes $(Y^i_A)^j_{B'}$, and so on.

Rewriting the exponent of q, equation (3.7) is equivalent to

$$\sum_{i\geq k,\text{all}\, j} I(Y^i_A)I(Y^{i,j}_{BA'})I(Y^{i,j}_{BB'})q^{i|Y^i_A|+j|Y^{i,j}_{BA'}|+i+j} + \sum_{i<k,\text{all}\, j} I(Y^{i,j-i}_{AA'})I(Y^{i,j-i}_{AB'})I(Y^i_B)q^{i|Y^i_A|+(j-i)|Y^{i,j-i}_{AA'}|+j}$$
$$= \sum_{j\geq k,\text{all}\, i} I(Y^j_{A'})I(Y^{j,i}_{B'A})I(Y^{j,i}_{B'B})q^{j|Y^j_{A'}|+i|Y^{j,i}_{B'A}|+j+i} + \sum_{j<k,\text{all}\, i} I(Y^{j,i-j}_{A'A})I(Y^{j,i-j}_{A'B})I(Y^j_{B'})q^{j|Y^j_{A'}|+(i-j)|Y^{j,i-j}_{A'A}|+i}.$$

It is easily verified that this equation holds provided the following four identities hold. We label the identities according to whether Y^i_A is inner (i) or outer (o) on the left side of the label, and whether $Y^j_{A'}$ is inner or outer on the right.

$$\sum_{i\geq k,j\geq k} I(Y^i_A)I(Y^{i,j}_{BA'})I(Y^{i,j}_{BB'})q^{i|Y^i_A|+j|Y^{i,j}_{BA'}|+i+j} = \sum_{j\geq k,i\geq k} I(Y^j_{A'})I(Y^{j,i}_{B'A})I(Y^{j,i}_{B'B})q^{j|Y^j_{A'}|+i|Y^{j,i}_{B'A}|+j+i}. \tag{ii}$$

$$\sum_{i<k,j\geq k} I(Y^{i,j-i}_{AA'})I(Y^{i,j-i}_{AB'})I(Y^i_B)q^{i|Y^i_A|+(j-i)|Y^{i,j-i}_{AA'}|+j} = \sum_{j\geq k,i<k} I(Y^j_{A'})I(Y^{j,i}_{B'A})I(Y^{j,i}_{B'B})q^{j|Y^j_{A'}|+i|Y^{j,i}_{B'A}|+j+i}. \tag{oi}$$

$$\sum_{i\geq k,j<k} I(Y^i_A)I(Y^{i,j}_{BA'})I(Y^{i,j}_{BB'})q^{i|Y^i_A|+j|Y^{i,j}_{BA'}|+i+j} = \sum_{j<k,i\geq k} I(Y^{j,i-j}_{A'A})I(Y^{j,i-j}_{A'B})I(Y^j_{B'})q^{j|Y^j_{A'}|+(i-j)|Y^{j,i-j}_{A'A}|+i}. \tag{io}$$

$$\sum_{i<k,j<k} I(Y^{i,j-i}_{AA'})I(Y^{i,j-i}_{AB'})I(Y^i_B)q^{i|Y^i_A|+(j-i)|Y^{i,j-i}_{AA'}|+j} = \sum_{i<k,j<k} I(Y^{i,j-i}_{A'A})I(Y^{i,j-i}_{A'B})I(Y^i_{B'})q^{i|Y^i_{A'}|+(j-i)|Y^{i,j-i}_{A'A}|+j} \tag{oo}$$

$$\left(= \sum_{j<k,i<k} I(Y^{j,i-j}_{A'A})I(Y^{j,i-j}_{A'B})I(Y^j_{B'})q^{j|Y^j_{A'}|+(i-j)|Y^{j,i-j}_{A'A}|+i}\right).$$

(Note: RD and CD in Figures 3.11-3.13 indicate row decomposition and column decomposition, respectively.)

Proof of (ii)**:** Let $i \geq k$, $j \geq k$; see Figure 3.11.

As illustrated in Figure 3.11, every cell of Y^i_A is to the lower-left of every cell of $Y^j_{A'}$. This is because, from the definition of k, there exist integers x, y, with $0 \leq x \leq n, 0 \leq y \leq n$, such that $x + y = n + k + 1$ and (x, y) lies outside Y. Then every cell of Y^i_A lies on or below row x, and every cell of $Y^j_{A'}$ lies on or right of column y.

Therefore, Y^i_A is completely contained in $Y^j_{B'}$, and so $Y^i_A = Y^{j,i}_{B'A}$. By symmetric arguments, $Y^{i,j}_{BA'} = Y^j_{A'}$.

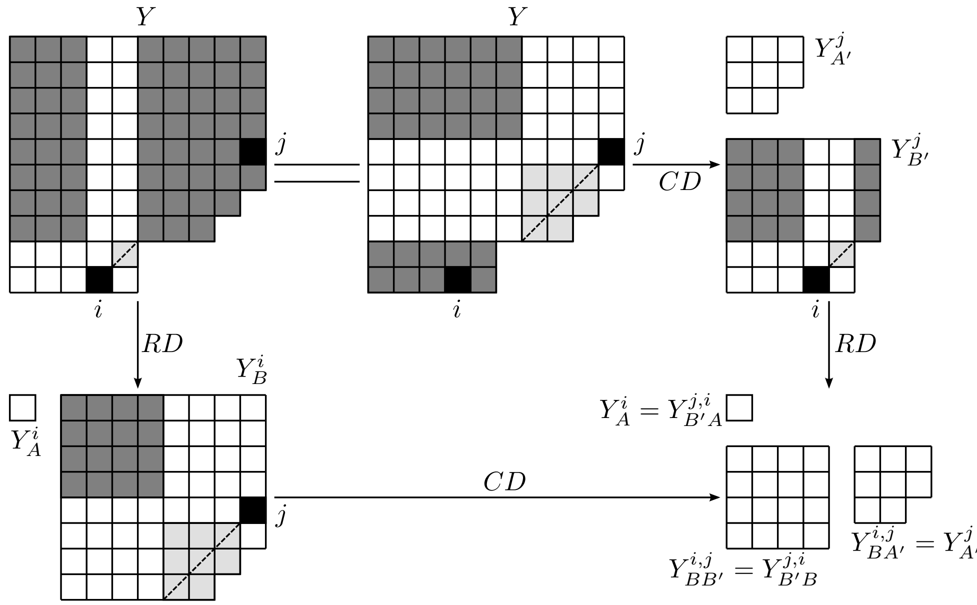

Figure 3.11: Y^i_A and $Y^j_{A'}$ are inner ($i \geq k$, $j \geq k$), corresponds to (ii)

Since Y^i_A is entirely to the lower-left of $Y^j_{A'}$, $Y^{i,j}_{BB'}$ is formed by deleting from Y all rows and columns intersecting $Y^i_A \cup (n,i)$, then deleting all rows and columns intersecting $Y^j_{A'} \cup (j,n)$. $Y^{j,i}_{B'B}$ is formed by reversing these two operations. The operations commute, and so $Y^{i,j}_{BB'} = Y^{j,i}_{B'B}$.

This concludes the proof of (ii).

Proof of (oi)**:** Let $i < k, j \geq k$; see Figure 3.12.

Since Y^i_A is outer and $Y^j_{A'}$ is inner, $Y^j_{A'}$ is completely contained in Y^i_A. Therefore, $Y^{i,j-i}_{AA'} = Y^j_{A'}$.

Now consider $Y^{i,j-i}_{AB'}$, which is just Y^i_A with all rows and columns intersecting $Y^j_{A'} \cup (j,n)$ deleted. Note that $Y^{i,j-i}_{AB'}$ lies along the columns labelled $i+1, \ldots, n - |Y^j_{A'}| - 1$. Now $Y^{j,i}_{B'A}$ is itself outer in $Y^j_{B'}$, since Y^i_A is outer. Therefore, $Y^{j,i}_{B'A}$ also lies along the columns $i+1, \ldots, n - |Y^j_{A'}| - 1$. Since $|Y^i_A| = n - i$, we have the important identity

$$|Y^i_A| - |Y^j_{A'}| - 1 = |Y^{j,i}_{B'A}|. \tag{3.8}$$

Furthermore, $Y^{i,j-i}_{AB'}$ lies along the first $|Y^i_A| - |Y^j_{A'}| - 1$ rows above (n,i) that do not intersect $Y^j_{A'} \cup (j,n)$, and likewise for $Y^{j,i}_{B'A}$. Therefore $Y^{i,j-i}_{AB'} = Y^{j,i}_{B'A}$.

Since Y^i_A is outer, then $Y^i_B = SQ_i$. Now, $Y^{j,i}_{B'B}$ consists of the i rows above $Y^{j,i}_{B'A}$ (and thus above Y^i_A) intersected with the i columns to the left of (n,i), and so $Y^{j,i}_{B'B} = SQ_i$. Therefore $Y^i_B = Y^{j,i}_{B'B}$.

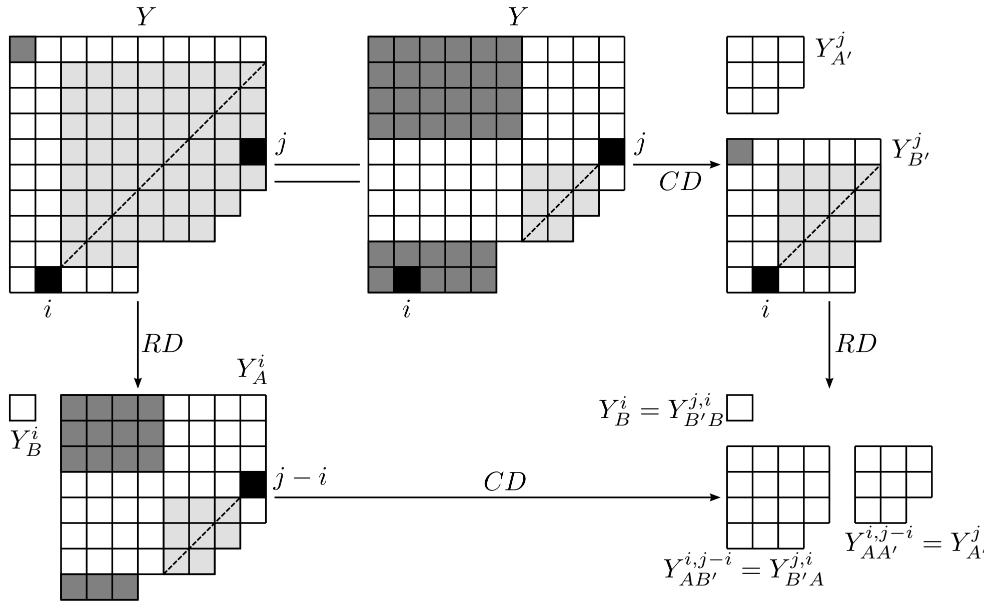

Figure 3.12: Y^i_A is outer and $Y^j_{A'}$ is inner ($i < k, j \geq k$), corresponds to (oi)

The above observations settle (oi), since by (3.8),

$$\begin{aligned} & i|Y^i_A| + (j-i)|Y^{i,j-i}_{AA'}| + j \\ = \; & i|Y^i_A| + (j-i)|Y^j_{A'}| + j \\ = \; & j|Y^j_{A'}| + i(|Y^i_A| - |Y^j_{A'}|) + j \\ = \; & j|Y^j_{A'}| + i(|Y^{j,i}_{B'A}| + 1) + j \\ = \; & j|Y^j_{A'}| + i|Y^{j,i}_{B'A}| + j + i. \end{aligned}$$

Proof of (io)**:** Let $i \geq k, j < k$.

Follows from (oi) by symmetry.

Proof of (oo)**:** Let $i < k, j < k$; see Figure 3.13.

Since Y^i_A and $Y^i_{A'}$ are outer, then $Y^i_B = Y^i_{B'}$, since they are both SQ_i. It is easy to check that $|Y^i_A| = |Y^i_{A'}|$. Likewise, $Y^{i,j-i}_{AB'} = Y^{i,j-i}_{A'B}$, since they are both SQ_{j-i}.

Now $Y^{i,j-i}_{AA'}$ is just $Y^j_{A'}$ with the bottom row and left column deleted. Likewise, $Y^{i,j-i}_{A'A}$ is just Y^j_A with the right column and top row deleted. So $Y^{i,j-i}_{AA'} = Y^{i,j-i}_{A'A}$, since they are both formed from Y by deleting all columns and rows other than those labelled $j+1, \ldots, n-1$.

This concludes the proof of (oo).

□

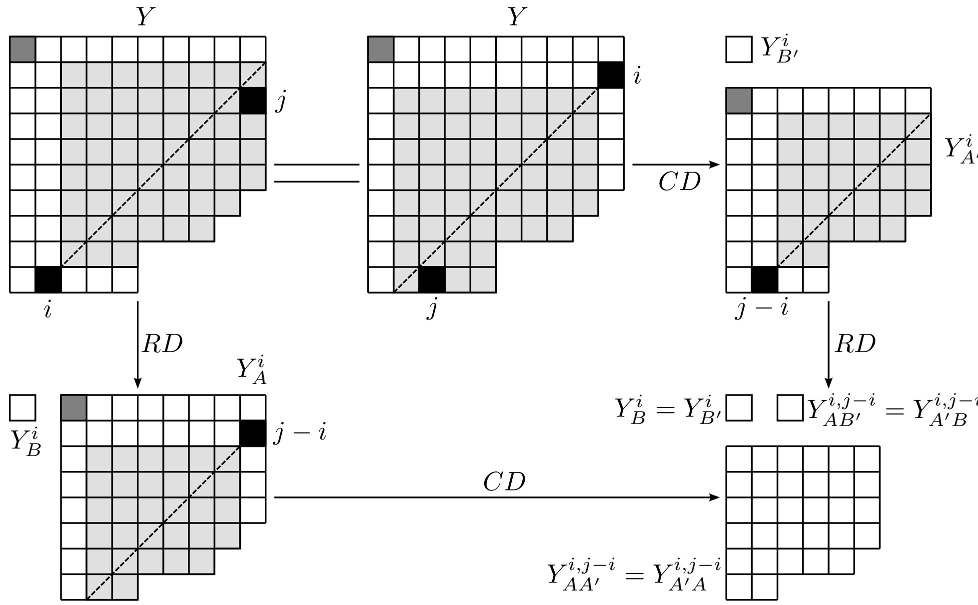

Figure 3.13: Y^i_A and $Y^j_{A'}$ are outer $(i < k, j < k)$, corresponds to (oo)

3.9 Conclusion

Corollary 3.5.3 provides the first known infinite family of nontrivial inv-Wilf-equivalent permutation pairs. This disproves Conjecture 3.3.5, due to Dokos et al. [23]. A key aspect of our method is to combine the definitions of inv-Wilf-equivalence and shape-Wilf-equivalence into the stronger definition of shape-inv-Wilf-equivalence.

It would be interesting to find explicitly the bivariate generating function $\sum_{n\geq 0} I_n(231 \oplus \gamma, q)t^n$ associated with the inv-Wilf-equivalent permutation pairs of Corollary 3.5.3 for specific permutations γ. However, this appears to be a difficult problem. In particular, it would give an explicit form for the generating function of the sequence $\{|S_n(231 \oplus \gamma)| : n \geq 0\}$, by setting $q = 1$. Even this special case would be significant: up to symmetry, the only permutations α of size greater than 3 for which the generating function of the sequence $\{|S_n(\alpha)| : n \geq 0\}$ is currently known are $\alpha = 1342$ (see Bóna [9], and a simpler proof provided by Bloom and Elizalde [7]) and $\alpha = 12\ldots k$ for all k (see Gessel [35] for $k = 4$ and Bousquet-Mélou [13] for general k).

The methods and results of this chapter suggest the following questions.

1. Is there an inv-Wilf-equivalent permutation pair α and β of size n that does not arise from Corollary 3.5.3 or trivial symmetries of D? Such a pair α, β could exist only for $n \geq 8$, since for each permutation pair α, β of size $n \leq 7$ we have found a suitable value of $m \leq 11$ for which $I_m(\alpha, q) \neq I_m(\beta, q)$, by computer search; a single such value of m demonstrates that α is not inv-Wilf-equivalent to β.

2. Can other variants of Wilf-equivalent definitions be fruitfully combined in order to address the many open questions and conjectures in the theory of pattern avoidance in permutations? In particular, can similar methods be used to understand when two permutations are maj-Wilf-equivalent?

Chapter 4

Suitable sets of permutations

4.1 Introduction

A set of N permutations of $[v] = \{1, 2, \ldots, v\}$ is (N, v, t)*-suitable* if each symbol precedes each subset of $t-1$ others in at least one permutation, where t is less than or equal to v. We represent such a set as an $N \times v$ array called an (N, v, t)*-suitable array.* For example, $\{2413, 3421, 1423\}$ is a $(3, 4, 3)$-suitable set of permutations and its corresponding array is

$$\begin{bmatrix} 2 & 4 & 1 & 3 \\ 3 & 4 & 2 & 1 \\ 1 & 4 & 2 & 3 \end{bmatrix}.$$

We give preliminary observations about (N, v, t)-suitable arrays.

Lemma 4.1.1.

(i) If an (N, v, t)-suitable array exists, then an $(N+1, v, t)$-suitable array exists, an $(N, v-1, t)$-suitable array exists, and, for $t \geq 2$, an $(N, v, t-1)$-suitable array exists.

(ii) An $(N, v, 1)$-suitable array exists for all N and v, and an $(N, v, 2)$-suitable array exists for all $N \geq 2$ and all v.

(iii) If an (N, v, t)-suitable array exists, then $N \geq t$.

(iv) An (N, v, t)-suitable array exists for all $t \leq v \leq N$.

Proof.

(i) Given an (N, v, t)-suitable array, we can readily form an $(N+1, v, t)$-suitable array by adding an arbitrary extra row. We can also form an $(N, v-1, t)$-suitable array by removing all occurrences of a single symbol (and left-justifying the remaining symbols), provided $v \geq t+1$. For $t \geq 2$, an (N, v, t)-suitable array is also an $(N, v, t-1)$-suitable array.

(ii) An $(N, v, 1)$-suitable array trivially exists, and an $(N, v, 2)$-suitable array exists when there are at least two rows because of the following array:

$$\begin{bmatrix} 1 & 2 & \dots & v-1 & v \\ v & v-1 & \dots & 2 & 1 \end{bmatrix}$$

(iii) Given a set of t symbols of $[v]$, each symbol appears before the others in at least one row of the array, so there are at least t rows in the array.

(iv) Place each symbol of $[v]$ first in a row. Then each symbol σ of $[v]$ precedes each subset of $t-1$ others in the row where σ appears first.

□

We consider two problems (P1) and (P2) on suitable permutations in this chapter, as described below.

(P1) Given v and t, what is the smallest N, denoted $N(v,t)$, for which an (N, v, t)-suitable array exists?

(P2) Given N and t, what is the largest v, denoted $\mathrm{SUN}(t, N)$ in [22], for which an (N, v, t)-suitable array exists?

Problem (P1) was proposed in 1950 by Dushnik [25], who introduced suitable permutations (originally called suitable arrangements). Problem (P2) was introduced in 2015 by Colbourn [22].

4.2 Dushnik's problem (P1)

The value $N(v,t)$ exists for all $v \geq t \geq 1$ by Lemma 4.1.1 (iv). Preliminary observations given in [25], which follow immediately from Lemma 4.1.1, include:

Lemma 4.2.1 (Dushnik [25])**.**

- *For $v_1 \leq v_2$ and $t_1 \leq t_2$, we have $N(v_1, t_1) \leq N(v_1, t_2) \leq N(v_2, t_2)$.*
- *$t \leq N(v,t) \leq v$.*
- *$N(v, 1) = 1$, $N(v, 2) = 2$, $N(v, v) = v$.*

One of the main results of Dushnik's paper [25] is that:

Theorem 4.2.2 (Dushnik [25])**.** *For each j satisfying $2 \leq j \leq \sqrt{v}$ and for each t satisfying*

$$\left\lfloor \frac{v}{j} \right\rfloor + j - 1 \leq t < \left\lfloor \frac{v}{j-1} \right\rfloor + j - 2,$$

we have $N(v,t) = v - j + 1$.

The above theorem determines [25] the value of $N(v,t)$ for all t in the range

$$\left\lfloor \frac{v}{\lfloor\sqrt{v}\rfloor} \right\rfloor + \lfloor\sqrt{v}\rfloor - 1 \leq t < v.$$

In particular, when the lower bound is attained (arising by taking $j = \lfloor\sqrt{v}\rfloor$), both v and $N(v,t)$ grow as $\Theta(t^2)$.

Spencer [56] in 1971 showed that $N(v,t) \geq \log_2 \log_2 v$ for all $t \geq 3$, using a theorem due to Erdős and Szekeres [29]. Furthermore, Spencer established an upper bound as follows: for $t \geq 3$ fixed, $N(v,t) = O(\log\log v)$, using Sperner's lemma [57] and the Erdős-Ko-Rado theorem. The two results show that $N(v,t) = \Theta(\log\log v)$ for fixed $t \geq 3$.

Füredi and Kahn [34] in 1986 showed using probabilistic methods that $N(v,t) \leq t^2(1+\log(v/t))$. Kierstead [39] in 1996 refined this result when t is approximately $\log v$.

4.3 Colbourn's problem (P2)

From Lemma 4.1.1 (iv), there exist (N,v,t)-suitable arrays for all v in the range $t \leq v \leq N$ at least; however, it is not immediately clear that there is a largest such v. For example, $\mathrm{SUN}(1,N)$ and $\mathrm{SUN}(2,N)$ do not exist by Lemma 4.1.1 (ii). However, Spencer's result that $N(v,t) \geq \log_2 \log_2 v$ when $t \geq 3$ immediately implies that $\mathrm{SUN}(t,N) \leq 2^{2^N}$ for $t \geq 3$, showing that $\mathrm{SUN}(t,N)$ exists for $t \geq 3$. Furthermore, Colbourn [22] extended Spencer's result by showing that $\mathrm{SUN}(t,N) = \Theta(2^{2^N})$ for fixed $t \geq 3$, using a connection with binary covering arrays [44].

Colbourn examined the case when v and N both grow as t^2, by making a connection with Golomb rulers and their variants [24], [28], [30]. He also found results for the case when t is $O(\log N)$ by making a connection with Hadamard matrices [36] and Paley matrices [48]. Our principal interest in this chapter is the study of problem (P2) when t has a growth rate between $\log(N)$ and $\sqrt{N}$.

Colbourn recast the problem of finding suitable arrays as the equivalent problem of finding "suitable cores", whose motivation is based on the proof of [25, Lemma 4].

4.4 Suitable cores

Since (N,v,t)-suitable arrays with $v \leq N$ exist according to Lemma 4.1.1 (iv), we may restrict attention to (N,v,t)-suitable arrays having $v > N$. We begin with a straightforward lemma.

Lemma 4.4.1 (Colbourn [22])**.** *Let A be an (N,v,t)-suitable array, and let α occur in the leftmost position of some row of A. Then the array B obtained by moving an occurrence of α in a different row of A to the rightmost position of its row results in another (N,v,t)-suitable array.*

Proof. The symbol α precedes all other symbols in some row of B, and therefore precedes each set of $t-1$ other symbols in this row. Each symbol β other than α precedes each set of $t-1$ other symbols in at least one row of A, and it still precedes the same $t-1$ symbols when some occurrence of α is moved rightwards to form B. □

For example, we can transform the following $(5,7,3)$-suitable array on the left to the $(5,7,3)$-suitable array on the right by applying Lemma 4.4.1 repeatedly (moving $\alpha = 3$ in row 4, then $\alpha = 4$ in row 4, then $\alpha = 5$ in row 4, then $\alpha = 6$ in row 5, then $\alpha = 7$ in rows 1 through 4).

$$\begin{bmatrix} 3 & 1 & 2 & 7 & 4 & 5 & 6 \\ 4 & 1 & 2 & 7 & 3 & 5 & 6 \\ 5 & 7 & 2 & 1 & 3 & 4 & 6 \\ 6 & 3 & 4 & 5 & 2 & 1 & 7 \\ 6 & 7 & 2 & 1 & 3 & 4 & 5 \end{bmatrix} \to \left[\begin{array}{c|cc|cccc} 3 & 1 & 2 & 4 & 5 & 6 & 7 \\ 4 & 1 & 2 & 3 & 5 & 6 & 7 \\ 5 & 2 & 1 & 3 & 4 & 6 & 7 \\ 6 & 2 & 1 & 3 & 4 & 5 & 7 \\ 7 & 2 & 1 & 3 & 4 & 5 & 6 \end{array}\right]$$

Conversely, the boxed 5×2 subarray on the right can be transformed back into a $(5,7,3)$-suitable array by choosing 5 new symbols, prepending a different one to each row, and in each row appending the remaining 4 new symbols in arbitrary order.

In general, by applying Lemma 4.4.1 repeatedly we can transform an (N,v,t)-suitable array A into another (N,v,t)-suitable array B having the following properties:

- The leftmost column of B consists of N distinct symbols; call these the *first symbols* of B.
- Columns 2 to $v-N+1$ of B consist only of the $v-N$ symbols which are not first symbols.
- Columns $v-N+2$ to v of B consist only of first symbols.

We can also transform the $N \times (v-N)$ array formed from columns 2 to $v-N+1$ of B back into an (N,v,t)-suitable array by choosing N new symbols, prepending a different one to each row, and in each row appending the remaining $N-1$ new symbols in arbitrary order. We call an $N \times (v-N)$ array that can be transformed into an (N,v,t)-suitable array by this procedure an $(N, v-N, t)$*-suitable core*. In the example above, the boxed 5×2 subarray is a $(5,2,3)$-suitable core.

We see in this way that the existence of an $(N, v+N, t)$-suitable array is equivalent to the existence of an (N,v,t)-suitable core. Given N and t, define $\mathrm{SCN}(t,N)$ as in [22] to be the largest v for which an (N,v,t)-suitable core exists. From the above, we obtain $\mathrm{SCN}(t,N) = \mathrm{SUN}(t,N) - N$, and so determining $\mathrm{SUN}(t,N)$ is equivalent to determining $\mathrm{SCN}(t,N)$.

Colbourn [22, Section 1] derives the value of $\mathrm{SCN}(2s,N)$ for all $N < s(s+1)$, and the value of $\mathrm{SCN}(2s+1,N)$ for all $N < (s+1)^2$, and gives the following lower bounds for the next largest value of N.

Theorem 4.4.2 (Colbourn [22]).

(i) [22, Lemma 1.3] $\mathrm{SCN}(2s, s(s+1)) \geq s+2$ *for all* $s \geq 2$.

(ii) [22, Lemma 1.4] $\mathrm{SCN}(2s+1, (s+1)^2) \geq s+2$ *for all* $s \geq 1$.

Parts (i) and (ii) of Theorem 4.4.2 are proved in [22] by two explicit constructions of families of suitable arrays. We will reprove Theorem 4.4.2 in Section 4.9, giving the construction for the proof of part (i). In Section 4.8 we will prove a general relation between certain suitable arrays with parameter t and others with parameter $t+1$. From this, part (ii) of Theorem 4.4.2 for $s > 1$ can be obtained directly from part (i).

Colbourn [22] states (without proof) that the inequality of Theorem 4.4.2 (i) is actually an equality (which we will reprove in Section 4.11). We demonstrate by example in Section 4.6 the new results that $\mathrm{SCN}(3, 4) \geq 8$ and $\mathrm{SCN}(5, 9) \geq 5$, so that equality does not hold in Theorem 4.4.2 (ii) for the cases $s = 1$ and $s = 2$. This appears to suggest that the inequality of Theorem 4.4.2 (ii) is not sharp in general, but we shall show in Section 4.10 using elementary combinatorial arguments that this is not the case:

Theorem 4.4.3. $\mathrm{SCN}(2s+1, (s+1)^2) = s+2$ *for all* $s \geq 3$.

Theorem 4.4.3 suggests a more delicate question: for $t = 2s+1$, can we increase the maximum possible value of v from $s+2$ by incrementing the value of the parameter $N = (s+1)^2$ by 1? In other words, is $\mathrm{SCN}(2s+1, (s+1)^2+1) > s+2$ for infinitely many s? Small examples suggest that this modest modification to the value of N is indeed sufficient to allow an increase in the value of v: we demonstrate by example in Section 4.6 the new results that $\mathrm{SCN}(7, 17) \geq 6$ and $\mathrm{SCN}(9, 26) \geq 7$, corresponding to the cases $s = 3$ and $s = 4$. However, such an increase is possible for only finitely many s; in particular, it can be shown that each such s is at most 14, using methods similar to the proof of Theorem 4.4.3 as given in Section 4.10. In fact, in Section 4.12 we use Ramsey's theorem [49] (to our knowledge, a new tool in the study of suitable arrays) to show the surprising result that, for the parameters of both parts of Theorem 4.4.2, the value of N can be increased any fixed amount ℓ and yet v can be increased from $s+2$ for only finitely many s:

Theorem 4.4.4.

(i) For each nonnegative integer ℓ, *there exists* s_0 *such that* $\mathrm{SCN}(2s, s(s+1)+\ell) = s+2$ *for all* $s \geq s_0$.

(ii) For each nonnegative integer ℓ, *there exists* s_0 *such that* $\mathrm{SCN}(2s+1, (s+1)^2+\ell) = s+2$ *for all* $s \geq s_0$.

The remainder of the chapter is structured in the following way.

In Section 4.5 we give necessary and sufficient conditions for suitable cores, based on results given in [22]. In Section 4.6 we give small examples of suitable cores. In Section 4.7 we establish some preliminary results to be used in later nonexistence proofs. In Section 4.8 we establish a link between suitable cores with parameters t and $t+1$. In Sections 4.10 and 4.12 we prove the central results of the chapter, Theorems 4.4.3 and 4.4.4, respectively. These sections are all new results.

In addition, we give an alternative proof of Theorem 4.4.2 in Section 4.9, and prove that the inequality of Theorem 4.4.2 (i) is actually an equality (as stated without proof by Colbourn [22]) in Section 4.11.

4.5 Necessary and sufficient conditions for suitable cores

We next give necessary and sufficient conditions for an $N \times v$ array to be an (N, v, t)-suitable core; these are essentially contained in [22, Lemma 1.1]. For an array C, symbol σ, and subset T of symbols, denote by $C_{\text{pre}}(\sigma, T)$ the set of rows of C for which σ either starts a row or is preceded only by elements of T.

Proposition 4.5.1. *Let C be an $N \times v$ array. The following statements are equivalent:*

(i) C is an (N, v, t)-suitable core.

(ii) For each s satisfying $0 \leq s \leq t-1$, each symbol of C precedes each subset of s others in at least $t-s$ rows.

(iii) For each symbol σ of C and for each subset T of other symbols, $|C_{\text{pre}}(\sigma, T)| \geq t+1-v+|T|$.

Proof. (i) $\iff$ (ii):

Construct an $N \times (v+N)$ array A from C by adding N new symbols as first symbols and completing the rows of A arbitrarily (provided that each row of A is a permutation of $[v+N]$. From the discussion following Lemma 4.4.1, statement (i) is equivalent to the statement that A is an $(N, v+N, t)$-suitable array. We now show that this is equivalent to statement (ii).

Suppose that C does not satisfy (ii), so that for some s satisfying $0 \leq s \leq t-1$ there is a symbol σ in C and a set S of s other symbols in C such that σ precedes all elements of S in at most $t-s-1$ rows of C. Combine S with the set of first symbols of the corresponding rows of A to give a set of size at most $t-1$, and extend it if necessary to a set of size $t-1$. Then there is no row of A in which σ precedes all elements of this set, and so A is not an $(N, v+N, t)$-suitable array.

On the other hand, suppose that C satisfies (ii). Let σ be a symbol in A and S be a set of $t-1$ other symbols in A. We shall show that σ precedes all elements of S in some row of A. If σ is a first symbol of A, this is immediate. Otherwise, let S' be the set of elements of S which are not first symbols. By assumption, σ precedes all elements of S' in at least $t-|S'|$ rows of A. Since there are only $t-1-|S'|$ elements of S which are first symbols, then at least one of these $t-|S'|$

rows of A does not begin with an element of S; in that row, σ precedes all of S (by construction of A from C). Thus A is an $(N, v+N, t)$-suitable array.

(ii) $\iff$ (iii):

Let σ be a symbol of C, let S be a set of s other symbols of C where $0 \leq s \leq t-1$, and let $T = [v] \setminus (S \cup \{\sigma\})$. Then σ precedes all elements of S in at least $t-s$ rows of C if and only if $|C_{\mathrm{pre}}(\sigma, T)| \geq t-s$. Note that $|T| = v-s-1$, and that $0 \leq s \leq t-1$ is equivalent to the trivial conditions $v-t \leq |T| \leq v-1$. □

We now briefly review some results due to Colbourn [22] on suitable cores, which allow the exact determination of $\mathrm{SCN}(t, N)$ for all N up to approximately $t^2/4$.

Proposition 4.5.2 ([22, Section 1]).

(i) Suppose there exists an (N, v, t)-suitable core. Then $N \geq i(t+1-i)$ for $i = 1, 2, \ldots, \min(v, t)$.

(ii) Let $v \leq (t+2)/2$. Then an (N, v, t)-suitable core exists if and only if $N \geq v(t+1-v)$.

(iii) $\mathrm{SCN}(t, N) = k$ for each $k \geq 0$ satisfying $k(t+1-k) \leq N < (k+1)(t-k)$.

Proof.

(i) Let i be an integer satisfying $1 \leq i \leq \min(v, t)$. Let S be a subset of $[v]$ of size i and let $\sigma \in S$. By Proposition 4.5.1 (ii), σ precedes all other elements of S in at least $t+1-i$ rows. As σ ranges over S we obtain $i(t+1-i)$ distinct rows of the core.

(ii) Suppose an (N, v, t)-suitable core exists. Then we may take $i = v$ in (i) to obtain $N \geq v(t+1-v)$.

Now suppose that $N \geq v(t+1-v)$. Construct an $N \times v$ array C whose first $v(t+1-v)$ rows have the following form: each symbol $i \in [v]$ starts a row $t+1-v$ times, and each symbol $j \in [v]$ that is distinct from symbol i occurs directly after i in at least one row. This is possible because $v \leq (t+2)/2$ implies $t+1-v \geq v-1$. We now use Proposition 4.5.1 (iii) to show that C is an (N, v, t)-suitable core. Let $\sigma \in [v]$ and let T be a set of symbols other than σ. Then $C_{\mathrm{pre}}(\sigma, T)$ contains the $t+1-v$ rows in which σ appears first, as well as at least $|T|$ rows in which σ appears directly after an element of T, for a total of at least $t+1-v+|T|$ rows.

(iii) Let $k \geq 0$ satisfy $k(t+1-k) \leq N < (k+1)(t-k)$. The range for N given by these inequalities is nonempty exactly when $k \leq (t-1)/2$. Therefore we may apply (ii) with $v = k$ and use the assumption $N \geq k(t+1-k)$ to show that there exists an (N, k, t)-suitable core and so $\mathrm{SCN}(t, N) \geq k$. Then apply (ii) with $v = k+1$ and use the assumption $N < (k+1)(t-k)$ to show that there does not exist an $(N, k+1, t)$-suitable core and so $\mathrm{SCN}(t, N) \leq k$. We conclude that $\mathrm{SCN}(t, N) = k$.

□

For fixed t, the smallest N for which $\mathrm{SCN}(t, N)$ is not determined by Proposition 4.5.2 (iii) is $N = (t/2)((t+2)/2)$ if t is even, and $((t+1)/2)^2$ if t is odd. These smallest undetermined cases can be written as $\mathrm{SCN}(2s, s(s+1))$ when t is even, and $\mathrm{SCN}(2s+1, (s+1)^2)$ when t is odd, and this motivates the constructions underlying Theorem 4.4.2.

4.6 New examples of suitable cores

Figures 4.1, 4.2, 4.3, 4.4 show examples of suitable cores with parameters $(4, 8, 3)$, $(9, 5, 5)$, $(17, 6, 7)$, $(26, 7, 9)$, respectively. To my knowledge, suitable cores with these parameters (and their associated suitable arrays) were not previously known. The first two were found by hand, and the second two by interactive computer search. These examples imply the bounds $\mathrm{SCN}(3, 4) \geq 8$, $\mathrm{SCN}(5, 9) \geq 5$, $\mathrm{SCN}(7, 17) \geq 6$, $\mathrm{SCN}(9, 26) \geq 7$ mentioned in Section 4.4 as motivation for the explorations leading to Theorems 4.4.3 and 4.4.4.

4.7 Preliminary results

We shall use the following two lemmas in our nonexistence results for suitable cores.

Lemma 4.7.1. *Removing all occurrences of a single symbol from an (N, v, t)-suitable core (and left-justifying the remaining symbols) results in an $(N, v-1, t)$-suitable core.*

Lemma 4.7.2. *Suppose that C is an (N, v, t)-suitable core.*

(i) Suppose $v \leq t$. Then each $k \in [v]$ starts a row at least $t+1-v$ times.

(ii) Suppose $v \leq t+1$, suppose k starts a row exactly $t+1-v$ times, and let j be another symbol. Then there is at least one row that starts with $j\,k$.

(iii) Suppose $v \leq t+2$, suppose k starts a row exactly $t+2-v$ times, and let i, j be two other distinct symbols. If neither $i\,k$ nor $j\,k$ starts a row, then there is at least one row that starts with $i\,j\,k$ or $j\,i\,k$.

Proof.

(i) For $k \in [v]$, apply Proposition 4.5.1 (iii) with $T = \emptyset$ to show that $|C_{\mathrm{pre}}(k, \emptyset)| \geq t+1-v$.

(ii) Apply Proposition 4.5.1 (iii) with $T = \{j\}$ to show that $|C_{\mathrm{pre}}(k, \{j\})| \geq t+2-v$. Since k starts a row exactly $t+1-v$ times, k must be preceded by j and by no other symbol in at least one row.

$$\begin{bmatrix} 2 & 1 & 4 & 3 & 6 & 5 & 8 & 7 \\ 3 & 4 & 1 & 2 & 7 & 8 & 5 & 6 \\ 5 & 6 & 7 & 8 & 1 & 2 & 3 & 4 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}$$

Figure 4.1: A $(4, 8, 3)$-suitable core.

$$\begin{bmatrix} 1 & 2 & 3 & 5 & 4 \\ 2 & 1 & 4 & 5 & 3 \\ 3 & 1 & 4 & 5 & 2 \\ 3 & 2 & 4 & 5 & 1 \\ 4 & 1 & 3 & 5 & 2 \\ 4 & 2 & 3 & 5 & 1 \\ 5 & 1 & 4 & 2 & 3 \\ 5 & 2 & 4 & 1 & 3 \\ 5 & 3 & 4 & 1 & 2 \end{bmatrix}$$

Figure 4.2: A $(9, 5, 5)$-suitable core.

$$\begin{bmatrix} 1 & 4 & * & * & * & * \\ 1 & 2 & 5 & * & * & * \\ 1 & 6 & 5 & * & * & * \\ 2 & 4 & * & * & * & * \\ 2 & 3 & * & * & * & * \\ 2 & 6 & 5 & * & * & * \\ 3 & 5 & * & * & * & * \\ 3 & 2 & 1 & * & * & * \\ 3 & 6 & 1 & * & * & * \\ 4 & 1 & 3 & * & * & * \\ 4 & 5 & 3 & * & * & * \\ 4 & 6 & * & * & * & * \\ 5 & 1 & 3 & * & * & * \\ 5 & 4 & 2 & * & * & * \\ 5 & 6 & * & * & * & * \\ 6 & 2 & 1 & * & * & * \\ 6 & 3 & 4 & * & * & * \end{bmatrix}$$

Figure 4.3: A $(17, 6, 7)$-suitable core (starred entries may be filled arbitrarily).

$$\begin{bmatrix} 1 & 6 & 5 & * & * & * & * \\ 1 & 7 & 5 & * & * & * & * \\ 1 & 3 & 5 & * & * & * & * \\ 1 & 4 & 2 & * & * & * & * \\ 2 & 6 & 1 & * & * & * & * \\ 2 & 7 & 1 & * & * & * & * \\ 2 & 5 & 3 & * & * & * & * \\ 2 & 4 & 1 & * & * & * & * \\ 3 & 6 & 5 & * & * & * & * \\ 3 & 7 & 5 & * & * & * & * \\ 3 & 1 & * & * & * & * & * \\ 3 & 2 & * & * & * & * & * \\ 4 & 6 & 1 & * & * & * & * \\ 4 & 7 & 1 & * & * & * & * \\ 4 & 5 & 2 & * & * & * & * \\ 4 & 3 & * & * & * & * & * \\ 5 & 6 & * & * & * & * & * \\ 5 & 7 & 3 & * & * & * & * \\ 5 & 1 & 2 & * & * & * & * \\ 5 & 4 & * & * & * & * & * \\ 6 & 7 & 5 & * & * & * & * \\ 6 & 2 & * & * & * & * & * \\ 6 & 3 & 4 & * & * & * & * \\ 7 & 6 & 1 & * & * & * & * \\ 7 & 4 & * & * & * & * & * \\ 7 & 2 & 3 & * & * & * & * \end{bmatrix}$$

Figure 4.4: A $(26, 7, 9)$-suitable core (starred entries may be filled arbitrarily).

(iii) Apply Proposition 4.5.1 (iii) with $T = \{i, j\}$ to show that $|C_{\text{pre}}(k, \{i, j\})| \geq t + 3 - v$. Since k starts a row exactly $t + 2 - v$ times, k must be preceded by one or both of i and j, and by no other symbol, in at least one row. Excluding the cases $i\,k$ and $j\,k$ for the initial symbols of this row leaves only the cases $i\,j\,k$ and $j\,i\,k$.

□

4.8 Suitable cores with parameters t and $t + 1$

The following result links suitable cores with parameters t and $t + 1$.

Theorem 4.8.1. *Suppose* $\text{SCN}(t, N) \geq v$ *and* $N > v(t + 1 - v)$*. Then* $\text{SCN}(t + 1, N + v - 1) \geq v$*.*

Proof. We suppose that C is an (N, v, t)-suitable core, where $N > v(t+1-v)$, and prove the result by constructing an $(N + v - 1, v, t + 1)$-suitable core D. Since $N > v(t + 1 - v)$, by the pigeonhole principle some symbol starts a row of C at least $t + 2 - v$ times; relabel if necessary so that this symbol is v. Form D by adding $v - 1$ rows to C, these rows starting with $1\,v, 2\,v, \ldots, (v - 1)\,v$. We now show that D is an $(N + v - 1, v, t + 1)$-suitable core using Proposition 4.5.1 (iii). Let $\sigma \in [v]$ and let T be a (possibly empty) set of symbols other than σ. We distinguish two cases.

Case 1 $\sigma \neq v$. Then $C_{\text{pre}}(\sigma, T)$ consists of at least $t + 1 - v + |T|$ rows. Combine with the extra row of D starting $\sigma\,v$ to give the necessary $t + 2 - v + |T|$ rows for $D_{\text{pre}}(\sigma, T)$.

Case 2 $\sigma = v$. When T is empty, the required condition $|D_{\text{pre}}(v, \emptyset)| \geq t + 2 - v$ is satisfied because of the $t + 2 - v$ rows that start with v. When T is nonempty, choose $c \in T$ and then the required condition $|D_{\text{pre}}(v, T)| \geq t + 2 - v + |T|$ is satisfied because there are at least $t + 1 - v + |T|$ rows in $C_{\text{pre}}(v, T)$ and the extra row of D starting $c\,v$.

□

We note two important corollaries of Theorem 4.8.1.

Corollary 4.8.2. *Theorem 4.4.2 (i) implies Theorem 4.4.2 (ii) for* $s > 1$*.*

Proof. Apply Theorem 4.8.1 with $(N, v, t) = (s(s + 1), s + 2, 2s)$ for each $s > 1$. □

Corollary 4.8.3. *Theorem 4.4.4 (ii) implies Theorem 4.4.4 (i).*

Proof. We have the general result that $\text{SCN}(t, N + 1) \geq \text{SCN}(t, N)$, by considering the addition of an arbitrary extra row to a suitable core. In view of Theorem 4.4.2 (i), we then see that Theorem 4.4.4 (i) is equivalent to: given an integer $\ell \geq 0$, we have $\text{SCN}(2s, s(s + 1) + \ell) < s + 3$ for all sufficiently large s. Likewise, in view of Theorem 4.4.2 (ii), we see that Theorem 4.4.4 (ii) is equivalent to: given an integer $\ell' \geq 1$, we have $\mathrm{SCN}(2s+1, (s+1)^2 + \ell') < s+3$ for all sufficiently large s.

Apply Theorem 4.8.1 with $(N, v, t) = (s(s+1)+\ell, s+3, 2s)$ to show that if $\mathrm{SCN}(2s, s(s+1)+\ell) \geq s+3$ then $\mathrm{SCN}(2s+1, (s+1)^2 + \ell + 1) \geq s+3$, and take $\ell' = \ell + 1$. □

4.9 Proof of Theorem 4.4.2

In this section, we prove Theorem 4.4.2. Colbourn [22] already proved this result using explicit constructions, but we will give an alternative proof here.

For Theorem 4.4.2 (i), it suffices to show that there exists an $(s(s+1), s+2, 2s)$-suitable core for all $s \geq 2$. In particular, such a suitable core is determined by the following proposition.

Proposition 4.9.1. *Let A be a $s(s+1) \times (s+2)$ array with the following properties:*

- *There is exactly one row that starts with $i\,j$ for each $1 \leq i \leq s+2$ and $1 \leq j \leq s$, $i \neq j$.*
- *For each $1 \leq i < j \leq s$, of the two rows that start with $i\,j$ and $j\,i$, one of them has a third symbol of $s+1$ and the other has a third symbol of $s+2$.*
- *For each $1 \leq i \leq s$, the row that starts with $(s+1)\,i$ has a third symbol of $s+2$, and the row that starts with $(s+2)\,i$ has a third symbol of $s+1$.*

Then A is an $(s(s+1), s+2, 2s)$-suitable core.

Proof. We use Proposition 4.5.1 (iii) with $(N, v, t) = (s(s+1), s+2, 2s)$. Consider an $s(s+1)\times(s+2)$ array A satisfying the above conditions and let $\sigma \in [s+2]$ and $T \subseteq [s+2] \setminus \{\sigma\}$. If $\sigma \in [s]$, then $A_{\text{pre}}(\sigma, T)$ contains the $s-1$ rows where σ appears first, as well as the $|T|$ rows where σ appears second after an element in T, for a total of at least $s - 1 + |T| = t - v + 1 + |T|$ rows, as desired.

On the other hand, if σ is $s+1$ or $s+2$, then $A_{\text{pre}}(\sigma, T)$ contains the s rows where σ appears first, as well as the $\binom{|T|-1}{2} + (|T|-1) = \binom{|T|}{2}$ rows where σ appears third after two elements in T, for a total of at least $s + \binom{|T|}{2}$ rows. Since $\binom{n}{2} \geq n-1$ for $n \geq 1$, this gives at least $s + (|T|-1) = t - v + 1 + |T|$ rows, as desired. □

Since the conditions of Proposition 4.9.1 are easily satisfied, this gives Theorem 4.4.2 (i). Corollary 4.8.2 then gives Theorem 4.4.2 (ii) for $s > 1$. The remaining case $s = 1$ of Theorem 4.4.2 (ii) is given by

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix},$$

for example.

4.10 Proof of Theorem 4.4.3

In this section we prove Theorem 4.4.3. As described in Section 4.4, the nonexistence result of Theorem 4.4.3 holds for all $s \geq 3$ but not for $s = 1$ and $s = 2$ (see Figures 4.1 and 4.2). We shall see where the condition $s \geq 3$ is required in the proof of Theorem 4.4.3, and why the proof does not apply to the $(9, 5, 5)$-suitable core shown in Figure 4.2.

Proof of Theorem 4.4.3. In view of Theorem 4.4.2 (ii), it is required to prove that for each $s \geq 3$ there does not exist an $((s+1)^2, s+3, 2s+1)$-suitable core. Suppose, for a contradiction, that C is such a suitable core.

Note from Lemma 4.7.2 (i) that each of the $s+1$ symbols of an $((s+1)^2, s+1, 2s+1)$-suitable core C' starts a row at least $s+1$ times, and since this accounts for all $(s+1)^2$ rows of C' we have that

each symbol of an $((s+1)^2, s+1, 2s+1)$-suitable core starts a row exactly $s+1$ times. (4.1)

It follows that

no symbol σ of C starts a row more than $s+1$ times after removing all occurrences of zero, one or two other symbols from C, (4.2)

for otherwise we could remove all occurrences of another two, one or zero other symbols, respectively, and by Lemma 4.7.1 would obtain an $((s+1)^2, s+1, 2s+1)$-suitable core in which σ starts a row more than $s+1$ times, contrary to (4.1).

Relabel if necessary so that the symbols of C are elements of $[s+3]$ and the number of rows starting with i is nondecreasing with i. By Lemma 4.7.2 (i), each of the $s+3$ symbols of C starts a row at least $s-1$ times. This accounts for $(s+3)(s-1) = (s+1)^2 - 4$ of the $(s+1)^2$ rows of C, leaving four more rows to account for. By (4.2) (with "zero"), there are three possible distributions for the symbols that start these four rows:

Case 1 Symbols 1 to $s-1$ each start a row exactly $s-1$ times, and symbols s to $s+3$ each start a row exactly s times.

By Lemma 4.7.2 (ii), each symbol that starts a row in C exactly $s-1$ times must appear second after each other symbol. Therefore C contains a row starting $j\,k$ for each $k = 1, 2, \ldots, s-1$ and for each $j \neq k$. The $s+3$ other rows of C each start with a different symbol of $[s+3]$.

Among these $s+3$ rows, no k from 1 to $s-1$ can appear second, otherwise C would contain two rows starting $j\,k$ for some j and a row starting $i\,k$ for some i distinct from j for which $i \geq s$; removing all occurrences of symbols i and j from C would then leave at least $(s-1)+3 = s+2$ rows starting with k, contradicting (4.2).

Furthermore, among these $s+3$ rows, no k from s to $s+3$ can appear second more than once, otherwise C would contain a row starting $i\,k$ and a row starting $j\,k$ for i, j not necessarily distinct; removing all occurrences of symbols i and j from C would again leave at least $s+2$ rows starting with k, contradicting (4.2).

Therefore each of the $s+3$ rows must contain a distinct symbol from s to $s+3$ in its second position, which gives the contradiction $s+3 \leq 4$.

Case 2 Symbols 1 to s each start a row exactly $s-1$ times, symbols $s+1$ and $s+2$ each start a row exactly s times, and symbol $s+3$ starts a row exactly $s+1$ times.

By Lemma 4.7.2 (ii), C contains a row starting $i\,j$ for each $j = 1, 2, \ldots, s$ and for each $i \neq j$. There is only one other row of C and it starts with $s+3$.

By Lemma 4.7.2 (iii), since $s+1$ and $s+2$ each start a row in C exactly s times, for each i, j satisfying $1 \leq i < j \leq s$ either C contains rows starting $i\,j\,(s+1)$ and $j\,i\,(s+2)$ or C contains rows starting $i\,j\,(s+2)$ and $j\,i\,(s+1)$.

It follows that $s+3$ never occurs second or third in a row of C that starts with 1, 2, or 3, and, because $s \geq 3$, no row of C starts with the symbols $1, 2, 3$ in any order.

But by Proposition 4.5.1, $C_{\text{pre}}(s+3, \{1, 2, 3\})$ contains at least $s+2$ rows. Since there are exactly $s+1$ rows of C starting with $s+3$, there is some row of C that does not start with $s+3$ in which $s+3$ is preceded only by elements of $\{1, 2, 3\}$. This gives the required contradiction.

Case 3 Symbols 1 to $s+1$ each start a row exactly $s-1$ times, and symbols $s+2$ and $s+3$ each start a row exactly $s+1$ times.

By Lemma 4.7.2 (ii) with $j = 1$ and $k = 2, 3, \ldots, s+1$, there are at least s rows starting with 1. This contradicts that symbol 1 starts a row exactly $s-1$ times.

□

Note that the proof of Theorem 4.4.3 does not apply to the $(9, 5, 5)$-suitable core shown in Figure 4.2, whose first row starts with the symbols $1, 2, s+1$ where $s = 2$.

4.11 Determination of $\mathrm{SCN}(2s, s(s+1))$

Colbourn [22] states without proof that $\mathrm{SCN}(2s, s(s+1)) = s+2$ for all $s \geq 2$, in other words that the inequality of Theorem 4.4.2 (i) is actually an equality. In this section, we recover this result using similar techniques to those in the above proof of Theorem 4.4.3.

Theorem 4.11.1 (Colbourn [22])**.** $\mathrm{SCN}(2s, s(s+1)) = s+2$ *for all* $s \geq 2$*.*

Proof. In view of Theorem 4.4.2 (i), it is required to prove that for each $s \geq 2$ there does not exist an $(s(s+1), s+3, 2s)$-suitable core. Suppose, for a contradiction, that C is such a suitable core.

Note from Lemma 4.7.2 (i) that each of the $s+1$ symbols of an $(s(s+1), s+1, 2s)$-suitable core C' starts a row at least s times, and since this accounts for all $s(s+1)$ rows of C' we have that

each symbol of an $(s(s+1), s+1, 2s)$-suitable core starts a row exactly s times. (4.3)

It follows that

no symbol σ of C starts a row more than s times after removing all occurrences of zero, one or two other symbols from C, (4.4)

for otherwise we could remove all occurrences of another two, one or zero other symbols, respectively, and by Lemma 4.7.1 would obtain an $(s(s+1), s+1, 2s)$-suitable core in which σ starts a row more than s times, contrary to (4.3).

Relabel if necessary so that the symbols of C are elements of $[s+3]$ and the number of rows starting with i is nondecreasing with i. By Lemma 4.7.2 (i), each of the $s+3$ symbols of C starts a row at least $s-2$ times. This accounts for $(s+3)(s-2) = s(s+1) - 6$ of the $s(s+1)$ rows of C, leaving six more rows to account for. By (4.4) (with "zero"), there are a few possible distributions for the first symbols of these six rows:

Case 1 Symbols 1 to $s-3$ each start a row exactly $s-2$ times, and symbols $s-2$ to $s+3$ each start a row exactly $s-1$ times.

By Lemma 4.7.2 (ii), each symbol that starts a row in C exactly $s-2$ times must appear second after each other symbol. Therefore C contains a row starting $j\,k$ for each $k = 1, 2, \ldots, s-3$ and for each $j \neq k$. Of the $2(s+3)$ rows of C not included in this description, no more than two rows start with the same symbol of $[s+3]$.

Among these $2(s+3)$ rows, no k from 1 to $s-3$ can appear second, otherwise C would contain two rows starting $j\,k$ for some j and a row starting $i\,k$ for some i distinct from j for which $i \geq s-2$; removing all occurrences of symbols i and j from C would then leave at least $(s-2)+3 = s+1$ rows starting with k, contradicting (4.4).

Furthermore, among these $2(s+3)$ rows, no k from $s-2$ to $s+3$ can appear second more than once, otherwise C would contain a row starting $i\,k$ and a row starting $j\,k$ for i, j not necessarily distinct; removing all occurrences of symbols i and j from C would again leave at least $s+1$ rows starting with k, contradicting (4.4).

Therefore each of the $2(s+3)$ rows must contain a distinct symbol from $s-2$ to $s+3$ in its second position, which gives the contradiction $2(s+3) \leq 6$.

Case 2 Symbols 1 to $s-2$ each start a row exactly $s-2$ times, symbols $s-1$ to $s+2$ each start a row exactly $s-1$ times, and symbol $s+3$ starts a row exactly s times.

By Lemma 4.7.2 (ii), each symbol that starts a row in C exactly $s-2$ times must appear second after each other symbol. Therefore C contains a row starting $j\,k$ for each $k = 1, 2, \ldots, s-2$ and for each $j \neq k$. The $s+4$ rows of C not included in this description each start with a different symbol of $[s+3]$, except for two rows which both start with $s+3$.

Among these $s+4$ rows, no k from 1 to $s-2$ can appear second, otherwise C would contain two rows starting $j\,k$ for some j and a row starting $i\,k$ for some i distinct from j; removing all occurrences of symbols i and j from C would then leave at least $(s-2)+3 = s+1$ rows starting with k, contradicting (4.4).

Furthermore, among these $s+4$ rows, no k from $s-1$ to $s+3$ can appear second more than once, otherwise C would contain a row starting $i\,k$ and a row starting $j\,k$ for i, j not necessarily distinct; removing all occurrences of symbols i and j from C would again leave at least $s+1$ rows starting with k, contradicting (4.4).

Therefore each of the $s+4$ rows must contain a distinct symbol from $s-1$ to $s+3$ in its second position, which gives the contradiction $s+4 \leq 5$.

Case 3 Symbols 1 to $s-1$ each start a row exactly $s-2$ times, symbols s and $s+1$ each start a row exactly $s-1$ times, and symbols $s+2$ and $s+3$ each start a row exactly s times.

By Lemma 4.7.2 (ii), C contains a row starting $i\,j$ for each $j = 1, 2, \ldots, s-1$ and for each $i \neq j$. There are only two rows of C not included in this description, and these rows start with $s+2$ and $s+3$, respectively. Furthermore, of these two rows:

- No symbol from 1 to $s-1$ may appear second; otherwise C would contain two rows starting $j\,k$ for $j \in \{s+2, s+3\}$ and a row starting $i\,k$ for some i distinct from j for which $i \geq s$; removing all occurrences of symbols i and j from C would then leave at least $(s-2)+3 = s+1$ rows starting with k, contradicting (4.4).
- Neither $s+2$ nor $s+3$ may appear second; otherwise if, for example, a row starts $(s+2)\,(s+3)$, removing $s+2$ leaves at least $s+1$ rows starting with $s+3$, contradicting (4.4).
- Neither s nor $s+1$ may appear second after both $s+2$ and $s+3$; otherwise, removing $s+2$ and $s+3$ leaves at least $s+1$ rows starting with some $i \in \{s, s+1\}$, contradicting (4.4).

Therefore, without loss of generality, we may take one of these two rows to begin with $(s+2)\,s$ and the other to begin with $(s+3)\,(s+1)$.

By Lemma 4.7.2 (iii), since s and $s+1$ each start a row in C exactly $s-1$ times, for each i, j satisfying $1 \leq i < j \leq s-1$ either C contains rows starting $i\,j\,s$ and $j\,i\,(s+1)$ or C contains rows starting $i\,j\,(s+1)$ and $j\,i\,s$. Furthermore, for each $1 \leq i \leq s-1$, C contains rows starting

$s\,i\,(s+1)$, $(s+1)\,i\,s$, $(s+2)\,i\,(s+1)$ and $(s+3)\,i\,s$, and rows starting $(s+2)\,s\,(s+1)$ and $(s+3)\,(s+1)\,s$. In particular, no row of C starts with the symbols $1, s+1, s+3$ in any order.

It follows that $s+2$ never occurs second or third in a row of C. However, by Proposition 4.5.1, $C_{\text{pre}}(s+2, \{1, s+1, s+3\})$ contains at least $s+1$ rows. Since there are exactly s rows of C starting with $s+2$, there is some row of C that does not start with $s+2$ in which $s+2$ is preceded only by elements of $\{1, s+1, s+3\}$. This gives the required contradiction.

Case 4 Symbols 1 to s each start a row exactly $s-2$ times, and symbols $s+1$ to $s+3$ each start a row exactly s times.

By Lemma 4.7.2 (ii) with $j = 1$ and $k = 2, 3, \ldots, s$, there are at least $s-1$ rows starting with 1. This contradicts that symbol 1 starts a row exactly $s-2$ times.

□

4.12 Proof of Theorem 4.4.4

In this section we prove Theorem 4.4.4. We first establish two auxiliary lemmas.

Lemma 4.12.1 shows that if, for a set A, we associate each element of A with a subset of A of size at most m, then some element of A appears in at most m of the subsets.

Lemma 4.12.1. *Let d and m be positive integers. Let A be a set of size d and let g be a function from A to subsets of A of size at most m. Then there exists $k \in A$ for which $\{\ell \in A : k \in g(\ell)\}$ has at most m elements.*

Proof. Let $f(k) = |\{\ell \in A : k \in g(\ell)\}|$. Then $\sum_{k \in A} f(k) \leq md$, and so the mean of $f(k)$ over $k \in A$ is at most m. So $f(k) \leq m$ for some $k \in A$. □

We next refine Lemma 4.12.1 to show that, if d is large enough, we can choose e elements of A, each of which appears in none of the subsets associated with the other $e-1$ elements.

Lemma 4.12.2. *Let e and m be positive integers and $d \geq (e-1)(2m+1)+1$. Let A be a set of size d and let g be a function from A to subsets of A of size at most m. Then there exists a subset B of A of size e such that $j \notin g(i)$ for all distinct $i, j \in B$.*

Proof. The proof is by induction on $e \geq 1$, with m and d satisfying the stated conditions.

If $e = 1$, then $|A| = d \geq 1$. Then simply choose B to comprise one element in A.

Now let $e > 1$ and assume the statement is true for all positive integers less than e and for all m and d satisfying the stated conditions. By Lemma 4.12.1, there exists $k \in A$ such that $|\{\ell \in A : k \in g(\ell)\}| \leq m$. Let $S = \{k\} \cup g(k) \cup \{\ell \in A : k \in g(\ell)\}$ and $A' = A \setminus S$. Note that $|S| \leq 2m+1$, and so $|A'| \geq d - (2m+1) \geq (e-2)(2m+1)+1$. We then define a function g'

from A' to subsets of A' of size at most m as follows: for each $b \in A'$, $g'(b) = g(b) \cap A'$. Clearly $|g'(b)| \leq m$ for each $b \in A'$.

By the inductive hypothesis applied to A' and g', there exists a subset B' of A' of size $e-1$ such that $j \notin g'(i)$ for all distinct $i, j \in B'$. Then we let $B = B' \cup \{k\}$. Since $k \notin B'$, we have $|B| = e$. Now let $j \in B$. We complete the induction by showing that $j \notin g(i)$ for all $i \in B \setminus \{j\}$.

Case 1 $j \neq k$. Then $j \notin g(k)$ by definition of S and A', since $j \in A'$ and $g(k) \subseteq S$. Also, $j \notin g(i)$ for all $i \in B' \setminus \{j\}$, since $j \notin g'(i)$ and $g'(i) = g(i) \cap A'$. Together this gives $j \notin g(i)$ for all $i \in B \setminus \{j\}$.

Case 2 $j = k$. Then $k \notin g(i)$ for all $i \in B \setminus \{k\}$, since $\{\ell \in A : k \in g(\ell)\} \subseteq S$ and $i \in A'$.

□

We are now ready to prove Theorem 4.4.4.

Proof of Theorem 4.4.4. By Corollary 4.8.3, it is sufficient to prove only part (ii) of the theorem. In view of Theorem 4.4.2 (ii), it is required to prove that for all sufficiently large s there does not exist an $((s+1)^2 + \ell, s+3, 2s+1)$-suitable core (where ℓ is a fixed nonnegative integer). Suppose, for a contradiction, that there is some arbitrarily large s for which C is such a suitable core.

Relabel if necessary so that the symbols of C are elements of $[s+3]$ and the number of rows starting with i is nondecreasing with i. By Lemma 4.7.2 (i), each of the $s+3$ symbols of C starts a row at least $s-1$ times. This accounts for $(s+3)(s-1) = (s+1)^2 - 4$ of the $(s+1)^2 + \ell$ rows of C, leaving $\ell + 4$ more rows to account for. The number of symbols that start a row more than $s-1$ times is then at most $\ell + 4$. Let c be the number of symbols that start a row exactly $s-1$ times, so that each of 1 to c starts a row exactly $s-1$ times and $c \geq s+3-(\ell+4) = s-\ell-1$.

By Lemma 4.7.2 (ii), C contains a row starting $i\,j$ for each $j = 1, 2, \ldots, c$ and for each $i \neq j$. Form C' from C by deleting the first such row for every such pair (i, j). Then in C', each of 1 to c starts a row exactly $m := s - c$ times. Since $c \geq s - \ell - 1$, we have $m \leq \ell + 1$.

The number of elements in $\{c+1, \ldots, s+3\}$ is $s - c + 3 = m + 3$. For each $i \in [c]$, since there are m rows of C' starting with i there are at least 3 elements of $\{c+1, \ldots, s+3\}$ which do not appear second after i in C', and so do not appear second after i in C. We may therefore define a function f from $[c]$ to 3-subsets of $\{c+1, \ldots, s+3\}$, such that $f(i) = \{j_1, j_2, j_3\}$ where j_1, j_2, j_3 do not appear second after $i \in [c]$.

Now choose s to be large enough to force $c \geq \binom{m+3}{3}(d-1)+1$ (via the inequality $c \geq s-\ell-1$), where $d \geq 1$ is an integer to be determined later. Then, by the pigeonhole principle, there exists a set of d numbers $A = \{a_1, a_2, \ldots, a_d\}$ in $[c]$ for which $f(a_1) = f(a_2) = \cdots = f(a_d)$. Let $\{k_1, k_2, k_3\} = f(a_1)$.

Next choose $d \geq (e-1)(2m+1)+1$, where $e \geq 1$ is an integer to be determined later. Define the function g from A to subsets of A via: for each $a \in A$, $g(a)$ is the set of elements of A appearing second in the rows of C' that start with a; so $g(a)$ has size at most m. By Lemma 4.12.2, there exists a subset $B = \{b_1, \ldots, b_e\}$ of A of size e such that $b_i \notin g(b_j)$ for all distinct i, j. It follows that no row of C' starting with an element of B has an element of B appearing second. By the construction of C' from C, we conclude that for each pair of distinct elements b_x, b_y of B there is exactly one row of C starting $b_x\, b_y$.

Now associate with C a graph G whose vertex set is $[e]$. For each $x, y \in [e]$, there is at least one element of the set $\{k_1, k_2, k_3\}$ that precedes the other two in neither the row starting $b_x\, b_y$ nor the row starting $b_y\, b_x$; choose one such element and color the edge between vertices x, y with color 1 if the choice is k_1, color 2 if it is k_2, and color 3 if it is k_3. The resulting graph G is a complete graph K_e on e vertices whose edges are colored from a set of 3 colors.

Recall that k_1, k_2, k_3 are in $\{c+1, \ldots, s+3\}$ by definition of f. Now symbols k_1, k_2, k_3 start a row $s-1+r_1, s-1+r_2, s-1+r_3$ times, respectively, for some positive integers r_1, r_2, r_3. Let T be a subset of B of size r_1+1. Then by Proposition 4.5.1, $C_{\text{pre}}(k_1, T)$ contains at least $s+r_1$ rows. Then there is some row of C, that does not start with k_1, in which k_1 is preceded only by elements of T. By the definition of f, k_1 does not appear second after $b_1, b_2, \ldots, b_e$, so this row starts $b_x\, b_y$ for some distinct elements b_x, b_y of T, and certainly in this row k_1 precedes k_2 and k_3. The edge joining vertices x and y of G is therefore not colored 1. Since this applies over all subsets T of B of size r_1+1, this means that G does not contain a K_{r_1+1} of color 1. A similar analysis holds for k_2 (T has size r_2+1) and k_3 (T has size r_3+1), and so G also contains neither a K_{r_2+1} of color 2 nor a K_{r_3+1} of color 3.

However, by Ramsey's theorem [49], for some $v \geq 1$, denoted $R(r_1+1, r_2+1, r_3+1)$, each edge coloring of a complete graph on v vertices using three colors contains either a K_{r_1+1} of color 1, or a K_{r_2+1} of color 2, or a K_{r_3+1} of color 3. Choose $e = v$ to give the required contradiction.

□

4.13 Open problems

We conclude with some open problems suggested by the results of this chapter.

1. Theorem 4.4.4 (ii) specifies the existence of s_0 for which an expression involving SCN holds for all $s \geq s_0$; but our proof, using Ramsey's theorem [49], does not determine a minimum s_0. Given a nonnegative integer ℓ, what is the smallest possible value of s_0 and how does it grow with ℓ?

2. The examples of suitable cores given in Figures 4.3 and 4.4 show that the inequality

$$\mathrm{SCN}(2s+1, (s+1)^2+\ell) > s+2 \tag{4.5}$$

holds for $\ell = 1$ and $s = 3, 4$. However, Theorem 4.4.4 (ii) shows that (4.5) holds for only finitely many s when ℓ is a fixed positive integer. But if ℓ is allowed to increase with s then (4.5) can hold for infinitely many s: substitute $s + 1$ for s in Theorem 4.4.2 (ii), and use the general result from Proposition 4.5.1 (ii) that $\mathrm{SCN}(t, N) \geq v$ implies $\mathrm{SCN}(t - 1, N) \geq v$, to show that $\ell = 2s + 3$ suffices. Does a function of s growing more slowly than $2s + 3$ suffice for (4.5) to hold for infintely many s? Does a function of s growing more slowly than linearly with s (for example, a function growing like $\log s$) suffice?

Bibliography

[1] R. Arratia. On the Stanley-Wilf conjecture for the number of permutations avoiding a given pattern. *Electron. J. Combin.*, 6:#N1, 4 pp., 1999.

[2] E. Babson and E. Steingrímsson. Generalized permutation patterns and a classification of the Mahonian statistics. *Sém. Lothar. Combin.*, 44:Art. B44b, 18 pp., 2000.

[3] J. Backelin, J. West, and G. Xin. Wilf-equivalence for singleton classes. *Adv. in Appl. Math.*, 38(2):133–148, 2007.

[4] A. Baxter and A. D. Jaggard. Pattern avoidance by even permutations. *Electron. J. Combin.*, 18(2):#P28, 15 pp., 2011.

[5] D. Bevan. Permutations avoiding 1324 and patterns in Łukasiewicz paths. *J. Lond. Math. Soc. (2)*, 92(1):105–122, 2015.

[6] J. Bloom. A refinement of Wilf-equivalence for patterns of length 4. *J. Combin. Theory Ser. A*, 124:166–177, 2014.

[7] J. Bloom and S. Elizalde. Pattern avoidance in matchings and partitions. *Electron. J. Combin.*, 20(2):#P5, 38 pp., 2013.

[8] J. Bloom and D. Saracino. A simple bijection between 231-avoiding and 312-avoiding placements. *J. Combin. Math. Combin. Comput.*, 89:23–32, 2014.

[9] M. Bóna. Exact enumeration of 1342-avoiding permutations: a close link with labeled trees and planar maps. *J. Combin. Theory Ser. A*, 80(2):257–272, 1997.

[10] M. Bóna. The limit of a Stanley-Wilf sequence is not always rational, and layered patterns beat monotone patterns. *J. Combin. Theory Ser. A*, 110(2):223–235, 2005.

[11] M. Bóna. *Combinatorics of permutations.* Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, second edition, 2012.

[12] M. Bóna. A new upper bound for 1324-avoiding permutations. *Combin. Probab. Comput.*, 23(5):717–724, 2014.

[13] M. Bousquet-Mélou. Counting permutations with no long monotone subsequence via generating trees and the kernel method. *J. Algebraic Combin.*, 33(4):571–608, 2011.

[14] R. A. Brualdi and E. Fritscher. Trace index of a $(0,1)$-matrix. *Linear Multilinear Algebra*, 63(6):1073–1085, 2015.

[15] R. Burkard, M. Dell'Amico, and S. Martello. *Assignment Problems.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.

[16] J. H. C. Chan. An infinite family of inv-Wilf-equivalent permutation pairs. *European J. Combin.*, 44(part A):57–76, 2015.

[17] J. H. C. Chan and J. Jedwab. The n-card problem, stochastic matrices, and the Extreme Principle. *Electron. J. Combin.*, 19(2):#P53, 8 pp., 2012.

[18] J. H. C. Chan and J. Jedwab. Constructions and nonexistence results for suitable sets of permutations. 2016. Accepted, *J. Combin. Theory Ser. A.*

[19] W. Y. C. Chen, E. Y. P. Deng, R. R. X. Du, R. P. Stanley, and C. H. Yan. Crossings and nestings of matchings and partitions. *Trans. Amer. Math. Soc.*, 359(4):1555–1575, 2007.

[20] A. Claesson, V. Jelínek, and E. Steingrímsson. Upper bounds for the Stanley-Wilf limit of 1324 and other layered patterns. *J. Combin. Theory Ser. A*, 119(8):1680–1691, 2012.

[21] A. Claesson and S. Kitaev. Classification of bijections between 321- and 132-avoiding permutations. In *20th Annual International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2008)*, Discrete Math. Theor. Comput. Sci. Proc., AJ, pages 495–506. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[22] C. J. Colbourn. Suitable permutations, binary covering arrays, and Paley matrices. In *Algebraic design theory and Hadamard matrices*, volume 133 of *Springer Proc. Math. Stat.*, pages 29–42. Springer, Cham, 2015.

[23] T. Dokos, T. Dwyer, B. P. Johnson, B. E. Sagan, and K. Selsor. Permutation patterns and statistics. *Discrete Math.*, 312(18):2760–2775, 2012.

[24] K. Drakakis. A review of the available construction methods for Golomb rulers. *Adv. Math. Commun.*, 3(3):235–250, 2009.

[25] B. Dushnik. Concerning a certain set of arrangements. *Proc. Amer. Math. Soc.*, 1:788–796, 1950.

[26] S. Elizalde. Fixed points and excedances in restricted permutations. *Electron. J. Combin.*, 18(2):#P29, 17 pp., 2011.

[27] S. Elizalde and I. Pak. Bijections for refined restricted permutations. *J. Combin. Theory Ser. A*, 105(2):207–219, 2004.

[28] P. Erdős. On a problem of Sidon in additive number theory. *Acta Sci. Math. Szeged*, 15:255–259, 1954.

[29] P. Erdős and G. Szekeres. A combinatorial problem in geometry. *Compositio Math.*, 2:463–470, 1935.

[30] P. Erdős and P. Turán. On a problem of Sidon in additive number theory, and on some related problems. *J. London Math. Soc.*, 16:212–215, 1941.

[31] M. Fekete. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.*, 17(1):228–249, 1923.

[32] M. M. Flood. The traveling-salesman problem. *Operations Res.*, 4:61–75, 1956.

[33] J. Fox. Stanley-Wilf limits are typically exponential, arxiv:1310.8378v1, 2013.

[34] Z. Füredi and J. Kahn. On the dimensions of ordered sets of bounded degree. *Order*, 3(1):15–20, 1986.

[35] I. M. Gessel. Symmetric functions and P-recursiveness. *J. Combin. Theory Ser. A*, 53(2):257–285, 1990.

[36] K. J. Horadam. *Hadamard matrices and their applications.* Princeton University Press, Princeton, NJ, 2007.

[37] V. Jelínek. Dyck paths and pattern-avoiding matchings. *European J. Combin.*, 28(1):202–213, 2007.

[38] T. Kaiser and M. Klazar. On growth rates of closed permutation classes. *Electron. J. Combin.*, 9(2):#R10, 20 pp., 2002/03.

[39] H. A. Kierstead. On the order dimension of 1-sets versus k-sets. *J. Combin. Theory Ser. A*, 73(2):219–228, 1996.

[40] D. E. Knuth. *The art of computer programming. Vol. 1: Fundamental algorithms.* Second printing. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1969.

[41] D. E. Knuth. *The art of computer programming. Vol. 3.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing.

[42] C. Krattenthaler. Permutations with restricted patterns and Dyck paths. *Adv. in Appl. Math.*, 27(2-3):510–530, 2001.

[43] C. Krattenthaler. Growth diagrams, and increasing and decreasing chains in fillings of Ferrers shapes. *Adv. in Appl. Math.*, 37(3):404–431, 2006.

[44] J. Lawrence, R. N. Kacker, Y. Lei, D. R. Kuhn, and M. Forbes. A survey of binary covering arrays. *Electron. J. Combin.*, 18(1):#P84, 30 pp., 2011.

[45] E. Lenza and B. Sands. A new proof and extension of problem 2620. *Crux Mathematicorum with Mathematical Mayhem*, 31:319–326, 2005.

[46] P. A. MacMahon. The indices of permutations and the derivation therefrom of functions of a single variable associated with the permutations of any assemblage of objects. *Amer. J. Math.*, 35(3):281–322, 1913.

[47] A. Marcus and G. Tardos. Excluded permutation matrices and the Stanley-Wilf conjecture. *J. Combin. Theory Ser. A*, 107(1):153–160, 2004.

[48] R. E. A. C. Paley. On orthogonal matrices. *J. Math. Phys.*, 12:311–320, 1933.

[49] F. P. Ramsey. On a problem of formal logic. *Proc. London Math. Soc.*, S2-30(1):264, 1928.

[50] A. Reifegerste. A generalization of Simion-Schmidt's bijection for restricted permutations. *Electron. J. Combin.*, 9(2):#R14, 9 pp., 2002/03.

[51] D. Richards. Ballot sequences and restricted permutations. *Ars Combin.*, 25:83–86, 1988.

[52] D. Rotem. On a correspondence between binary trees and a certain type of permutation. *Information Processing Lett.*, 4(3):58–61, 1975/76.

[53] B. Sands. Problem 2620. *Crux Mathematicorum with Mathematical Mayhem*, 27:138, 2001.

[54] B. Sands. Cards, permutations, and sums. *Contrib. Discrete Math.*, 6(2):1–19, 2011.

[55] R. Simion and F. Schmidt. Restricted permutations. *European J. Combin.*, 6:383–406, 1985.

[56] J. Spencer. Minimal scrambling sets of simple orders. *Acta Math. Acad. Sci. Hungar.*, 22:349–353, 1971/72.

[57] E. Sperner. Ein Satz über Untermengen einer endlichen Menge. *Math. Z.*, 27(1):544–548, 1928.

[58] Z. Stankova. Classification of forbidden subsequences of length 4. *European J. Combin.*, 17(5):501–517, 1996.

[59] Z. Stankova and J. West. A new class of Wilf-equivalent permutations. *J. Algebraic Combin.*, 15(3):271–290, 2002.

[60] Z. E. Stankova. Forbidden subsequences. *Discrete Math.*, 132(1-3):291–316, 1994.

[61] W. Trongsiriwat. Permutation statistics and multiple pattern avoidance. *J. Comb.*, 6(1-2):235–248, 2015.

[62] J. West. Generating trees and the Catalan and Schröder numbers. *Discrete Math.*, 146(1-3):247–262, 1995.

[63] S. H. F. Yan, H. Ge, and Y. Zhang. On a refinement of Wilf-equivalence for permutations. *Electron. J. Combin.*, 22(1):#P1.20, 26 pp., 2015.

[64] P. Zeitz. *The Art and Craft of Problem Solving.* John Wiley & Sons, Inc., New York, 1999.