Title
Two topics in combinatorics: Generalized coinvariant algebras and Catalan-pair graphs

Author
Kroes, Daniël

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Two topics in combinatorics:
Generalized coinvariant algebras and Catalan-pair graphs

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

Daniël Kroes

Committee in charge:

Professor Brendon Rhoades, Chair
Professor Fan Chung Graham
Professor Russell Impagliazzo
Professor Jonathan Novak
Professor Alexander Vardy

2021

The dissertation of Daniël Kroes is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the endless support, encouragement, motivation and sometimes necessary distraction during the last five years here in San Diego.

First and foremost, I would like to thank my advisor Brendon Rhoades. You were always able to provide me with interesting problems to work on, to share your knowledge on concepts and backgrounds I was unfamiliar with, to discuss my ideas and progress, to encourage me and give me suggestions on how to proceed with those ideas, and to proofread my work and help me become both a better mathematician and better writer.

Secondly, I am very grateful to Fan Chung Graham. You and Ron always had a large collection of interesting and suitable problems to work on, and you never stopped motivating me to expand on my ideas and get even more out of the current progress. The first and last version of the paper on Catalan-pair graphs feel like worlds apart, both in content and style, which is due to your commitment to keep asking questions to Sam and me, to keep pushing us to get more out of our methods, and to keep pushing us to improve the exposure of the material.

I would also like to thank Russell Impagliazzo, Jonathan Novak and Alexander Vardy for taking the time to be on my committee.

Moving 9000 kilometers across the Atlantic Ocean was a huge step, both literally and figuratively, and would not have been possible without the support from my family. I am incredibly grateful for the Skype calls, care packages with some great Dutch food,

visits to San Diego, birthday cards, and the time we would spend together when I would fly back home during summer or for Christmas. I can't express how lucky I am for all the support for and interest in my time here in San Diego. Dankjewel papa, mama, Cindy en Almar, opa, oma, ooms, tantes, neven en nichten!

I cannot be more thankful for everyone in San Diego that has made moving here so much easier. From the moment I arrived here I am happy to say that everyone here managed to make San Diego really feel like home. A particular thank you goes to Angelika, Benni, Daniel, Emma, Eva, François, Haley, Hooman, Jacqueline, Jake, Janina, Janine, Jason, Jesus, Kuang, Mark, Marvin, Michelle, Mike, Nikki, Peter, Philipp, Pieter, Sam, Stephan, Thea, Vaki and Zach for all the board games, movie nights, frisbee, trips, trivia nights, tea breaks at the department, and so much more.

Lastly, I would like to show my gratitude for everyone back home that always makes the time to meet up and catch up whenever I'm in the Netherlands. Thank you Eelke, Floris, Franziska, Jetze, Julian, Maarten, Max, Merlijn, Rick, Roel and Yuri for staying in touch.

Daniël

May 27th, 2021

Chapters 1, 2 and 3 contain material from: D. Kroes, *"Generalized coinvariant algebras for $G(r,1,n)$ in the Stanley-Reisner setting"*, Electronic Journal of Combinatorics, vol. 26 (3), P.3.11, 2018. The dissertation author was the primary investigator and author of this paper.

Chapters 1, 2 and 4 contain material from: D. Kroes and B. Rhoades, *"Packed words and quotient rings"*, submitted (2021). The dissertation author was one of the primary investigators and authors of this paper.

Chapters 1, 2 and 5 contain material from: D. Kroes and S. Spiro, *"Random Graphs Induced by Catalan Pairs"*, Journal of Combinatorics, Accepted (2020). The dissertation author was one of the primary investigators and authors of this paper.

VITA

2014	B.Sc. in Mathematics, Utrecht University
2014	B.Sc. in Physics and Astronomy, Utrecht University
2016	M.Sc. in Mathematical Sciences, Utrecht University
2018	C.Phil. in Mathematics, University of California San Diego
2021	Ph.D. in Mathematics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Two topics in combinatorics:
Generalized coinvariant algebras and Catalan-pair graphs

by

Daniël Kroes

Doctor of Philosophy in Mathematics

University of California San Diego, 2021

Professor Brendon Rhoades, Chair

In this dissertation we study two combinatorial problems. The starting point of the first problem are coinvariant algebras, quotients of the polynomial ring in n variables that serve as a remarkable connection between symmetric functions, representation theory and permutation statistics. Recently, inspired by the Delta Conjecture, generalized quotients were introduced, whose combinatorics are controlled by set partitions of a set of size n into a given number of blocks. We exhibit quotients of the Stanley-Reisner ring of the Boolean

algebra isomorphic to the given generalizations, extending an isomorphism known in the classical setting. Additionally, we introduce a quotient whose combinatorics are related to all set partitions of a given set, without any restrictions on the number of blocks.

Secondly, we look at Catalan numbers, a well-known combinatorial sequence with a variety of interpretations and applications. We study the interaction between two objects chosen from one of these interpretations, and represent this interaction in terms of a graph, also known as a bipartite circle graph. We introduce a random model to generate such graphs, and describe the asymptotic behaviour of various properties, including the number of edges, the number of isolated vertices, and its subgraphs.

Chapter 1

Introduction

1.1 Coinvariant algebras

The polynomial ring $\mathbb{Q}[\mathbf{x}_n] := \mathbb{Q}[x_1, \ldots, x_n]$ carries an action of the symmetric group $\mathfrak{S}_n$ by variable permutation. Elements of the corresponding invariant subring $\mathbb{Q}[\mathbf{x}_n]^{\mathfrak{S}_n}$ are known as *symmetric functions*, and the subring has algebraically independent homogeneous generators given by the elementary symmetric functions $e_1(\mathbf{x}_n), \ldots, e_n(\mathbf{x}_n)$. The *invariant ideal* $I_n = \langle e_1(\mathbf{x}_n), \ldots, e_n(\mathbf{x}_n) \rangle$ leads to a quotient $R_n = \mathbb{Q}[\mathbf{x}_n]/I_n$, called the *coinvariant algebra*.

As I_n is $\mathfrak{S}_n$-stable, R_n is a $\mathfrak{S}_n$-module and it is known [Che55] that as ungraded $\mathfrak{S}_n$-module its structure concides with that of the regular representation $\mathbb{Q}[\mathfrak{S}_n]$. In particular, R_n has dimension $n!$ and R_n has many interesting ties with permutations. For example, the Hilbert series of R_n agrees with the generating function of many important permutation statistics, and R_n has vector space bases naturally indexed by permutations. Moreover, the graded $\mathfrak{S}_n$ structure of the has been studied extensively and can (for example) be described in terms of dual Hall-Littlewood symmetric functions.

Inspired by the Delta Conjecture [HRW18] of Haglund, Remmel and Wilson, a generalization of these coinvariant algebras was introduced by Haglund, Rhoades and Shimozono [HRS18]. The Delta Conjecture asserts an equality between three quasisymmetric functions, two of which are $\mathrm{Rise}_{n,k}(\mathbf{x};q,t)$ and $\mathrm{Val}_{n,k}(\mathbf{x};q,t)$ for positive integers $k \leq n$. These functions are defined in terms of combinatorial objects and following work of Rhoades [Rho18] and Wilson [Wil16] it is known that

$$\mathrm{Rise}_{n,k}(\mathbf{x};q,0) = \mathrm{Rise}_{n,k}(\mathbf{x};0,q) = \mathrm{Val}_{n,k}(\mathbf{x};q,0) = \mathrm{Val}_{n,k}(\mathbf{x};0,q) \tag{1.1}$$

and denoting this common function by $C_{n,k}(\mathbf{x};q)$ we can write

$$C_{n,k}(\mathbf{x};q) = \sum_{T \in \mathrm{SYT}(n)} q^{\mathrm{maj}(T)+\binom{n-k}{2}-(n-k)\cdot\mathrm{des}(T)} \begin{bmatrix} \mathrm{des}(T) \\ n-k \end{bmatrix}_q s_{\mathrm{shape}(T)}(\mathbf{x}). \tag{1.2}$$

Haglund, Rhoades and Shimozono introduced quotients $R_{n,k}$ of $\mathbb{Q}[\mathbf{x}_n]$ whose graded $\mathfrak{S}_n$-structure, up to some minor twists, coincides with $C_{n,k}(\mathbf{x};q)$. Additionally, as ungraded modules the structure is equal to $\mathbb{Q}[\mathcal{OP}_{n,k}]$ where $\mathcal{OP}_{n,k}$ is the set of set partitions of $[n] := \{1, 2, \ldots, n\}$ into exactly k blocks. Finally, the Hilbert series of $R_{n,k}$ is governed by various statistics on ordered set partitions similar to those on permutations in the classical case.

The third quasisymmetric function in the Delta Conjecture is defined in terms of symmetric functions and Macdonald eigenoperators. Recently, D'Adderio and Mellit

[DM20] and Blasiak, Haiman, Morse, Pun and Seelinger [BHM$^+$21] announced proofs that show the equality between this third quasisymmetric function and the function $\mathrm{Rise}_{n,k}(\mathbf{x};q,t)$ from above.

The expression of $C_{n,k}$ in Equation (1.2) also appears in the setting of the *superspace ring* Ω_n. Originally studied in physics, recently this ring has received attention in coinvariant theory as well, see [RW20] and [Zab19]. Here, superspace of rank n is the tensor product

$$\Omega_n := \mathbb{Q}[x_1, \dots, x_n] \otimes \wedge\{\theta_1, \dots, \theta_n\}$$

of a rank n polynomial ring with a rank n exterior algebra. We let $\mathfrak{S}_n$ act diagonally on Ω_n and let $(\Omega_n)_+^{\mathfrak{S}_n}$ be the $\mathfrak{S}_n$-invariants with zero constant term. The *superspace coinvariant ring* $\Omega_n/\langle(\Omega_n)_+^{\mathfrak{S}_n}\rangle$ is a bigraded $\mathfrak{S}_n$-module and the Combinatorics Group and Fields Institute conjectured [Zab19] that

$$\mathrm{grFrob}(\Omega_n/\langle(\Omega_n)_+^{\mathfrak{S}_n}\rangle; q, z) = \sum_{k=1}^{n} z^{n-k} \cdot C_{n,k}(\mathbf{x}; q). \tag{1.3}$$

We will build upon the work of Haglund, Rhoades and Shimozono in two directions. Firstly, it has been shown that the coinvariant algebra can also be obtained as a quotient of the *Stanley-Reisner ring of the Boolean algebra*

$$\mathbb{C}[\mathcal{B}_n^*] := \frac{\mathbb{C}[\mathbf{y}_S]}{\langle y_S \cdot y_T \rangle}$$

where $\mathbb{C}[\mathbf{y}_S]$ is the polynomial ring in $2^n - 1$ variables indexed by non-empty subsets $S \subseteq [n]$, and the generators of the ideal range over all pairs (S,T) such that $S \not\subseteq T$ and $T \not\subseteq S$. We will define a quotient of $\mathbb{C}[\mathcal{B}_n^*]$ that is isomorphic to $R_{n,k}$ and using methods inspired by those used by Braun and Olsen in the classical case $k = n$ [BO18] we describe a basis of our quotient that intimately relates to a basis found for $R_{n,k}$. Moreover, our quotient carries a multigraded Frobenius series that can be viewed as a refinement of the graded Frobenius series of $R_{n,k}$. Finally, our methods carry over to a similar family of quotients for certain reflection groups, which were introduced by Chan and Rhoades [CR20].

Secondly, we develop a quotient of $\mathbb{Q}[\mathbf{x}_n]$ whose combinatorics are controlled by *all* set partitions of $[n]$, or equivalently by the set of all *packed words* of length n. Packed words have appeared in various other settings, including Hopf algebras [NT06] and polytopes [CL20]. Let $\mathcal{W}_n$ be the set of packed words of length n, then we introduce a $\mathfrak{S}_n$-module S_n with ungraded structure

$$S_n \cong \mathbb{Q}[\mathcal{W}_n].$$

Moreover, the graded Frobenius series of S_n is equal to

$$\operatorname{grFrob}(S_n; q) = \sum_{k=1}^{n} q^{n-k} \cdot (\operatorname{rev}_q \circ \omega) C_{n,k}(\mathbf{x}; q), \tag{1.4}$$

which has a striking similarity with the conjectured superspace Frobenius series in Equation (1.3). Therefore, studying S_n in more detail and exploring similarities between S_n and $\Omega_n / \langle (\Omega_n)_+^{\mathfrak{S}_n} \rangle$ might inspire a proof of Equation (1.3).

1.2 Catalan-Pair graphs

The sequence $1, 1, 2, 5, 14, 42, \ldots$ of the *Catalan numbers* $\frac{1}{n+1}\binom{2n}{n}$ is a famous and abundant sequence in combinatorics. A large body of work has been devoted to studying this sequence and the objects it enumerates. An overview of many of these objects and properties can be found in [Pak14] and [Sta15]. Among these objects are polygon triangulations, binary trees, plane trees, and Dyck paths. One more object counted by the Catalan numbers is the set of all ways to draw n non-intersecting chords given $2n$ points on a circle.

While there are many generalizations and many unanswered questions about Catalan numbers, we will be concerned with studying the interaction between two sets of non-intersecting chords on a circle. By interpreting the chords as vertices of a graph, and connecting two vertices if their corresponding chords intersect, we obtain a bipartite graph. These graphs appeared [BDD$^+$] as a result of studying and generalizing a magic trick involving a dollar bill and some paperclips, and the graphs were named paperclip graphs as a result.

In general, graphs that represent the intersection pattern of a set of chords on a circle are known as *circle graphs*, and hence paperclip graphs are precisely the bipartite circle graphs. Circle graphs have been extensively studied, especially from an algorithmic point of view. For example, Spinrad [Spi94] produced an $O(n^2)$-time algorithm for identifying whether a given graph is a circle graph. Many problems that are know to be NP-complete for general graphs turn out to have polynomial time algorithms when restricted to circle graphs. Recently Tiskin showed that a maximum clique of a circle graph can be found in $O(n(\log n)^2)$ time [Tis15], and Gregg and Nash have shown that a maximum independent set can be found in time $O(\alpha n)$, where α denotes the independence number of the circle graph [NG10].

We will introduce a random model for these graphs, and study its various properties. The idea of studying graphs through various random models is well known, see [Bol84], [ER60] and [Gil59] for a few famous examples. Our model for these graphs allows us to study the behaviour of these graphs as the number of vertices n increases. In particular, when $n \to \infty$ we show that the expected number of edges is asymptotically equal to $\frac{1}{\pi} n \log n$ and by studying the variance we show that we have a strong concentration around this mean.

Due to the frequent occurence of short chords, our graphs will have many isolated vertices and small components. In particular, we will show that the number of isolated vertices is linear in the total number of vertices, determine the coefficient, and show strong concentration around the mean in this case as well. Additionally, we show that our random graph is expected to have many components of various small sizes, with components of sizes of at least order $\log(n)$ to be expected.

1.3 Structure of this dissertation

The remainder of this dissertation is organized as follows. In Chapter 2 we will cover the necessary background material for the remainder of the dissertation. The majority of this chapter is devoted to the details of the (generalized) coinvariant algebras and the necessary algebraic definitions and techniques. The final part of this chapter is used to discuss Catalan-arc matchings, and a probability result needed to study the random graph model.

In Chapter 3 we will introduce our quotient of the Stanley-Reisner ring of the Boolean algebra and prove the desired isomorphisms with the rings introduced by Haglund, Rhoades and Shimozono, and Chan and Rhoades. In Chapter 4 we introduce the quotient S_n, study its combinatorics and algebraic structure. Finally, in Chapter 5 we introduce the random model for the bipartite circle graphs and study various interesting properties of these random graphs.

This chapter contains material from: D. Kroes, *"Generalized coinvariant algebras for $G(r, 1, n)$ in the Stanley-Reisner setting"*, Electronic Journal of Combinatorics, vol. 26 (3), P.3.11, 2018. The dissertation author was the primary investigator and author of this paper.

This chapter also contains material from: D. Kroes and B. Rhoades, *"Packed words and quotient rings"*, submitted (2021). The dissertation author was one of the primary investigators and authors of this paper.

This chapter also contains material from: D. Kroes and S. Spiro, *"Random Graphs Induced by Catalan Pairs"*, Journal of Combinatorics, Accepted (2020). The dissertation author was one of the primary investigators and authors of this paper.

Chapter 2

Preliminaries

In this chapter we will cover the background material needed to state and prove the main results of our thesis. We will also prove some initial results that will be of use in the later chapters. With an eye towards our results about generalized coinvariant algebras we will discuss permutations, ordered set partitions, and statistics thereon. Additionally, we will cover symmetric functions, representation theory, Gröbner theory, and some of the previous results on (generalized) coinvariant algebras.

In a different direction, looking ahead to the Catalan-pair graphs we will cover Catalan numbers, as well as a concentration result on the sum of independent identically distributed random variables.

Notation

First we fix some notation. We denote by $\mathbf{x} = (x_1, x_2, x_3, \ldots)$ an infinite set of variables and by $\mathbf{x}_n = (x_1, x_2, \ldots, x_n)$ a set of n variables. Similarly, for any field K, $K[\mathbf{x}_n]$ denotes the polynomial ring $K[x_1, x_2, \ldots, x_n]$ in n variables.

For any positive integer we write $[n] = \{1, 2, \ldots, n\}$. For integers $1 \leq k \leq n$ we let $\mathfrak{S}_n$ be the symmetric group of degree n and $\mathcal{OP}_{n,k}$ the set of ordered set partitions of $[n]$ into k blocks.

The expression C_n will denote the n^{th} Catalan number $\frac{1}{n+1}\binom{2n}{n}$.

2.1 Permutations and ordered set partitions

We start this section by describing some statistics on permutations. Let $\sigma = \sigma_1 \sigma_2 \cdots \sigma_n$ be a permutation written in one-line notation. The ascent set $\mathrm{Asc}(\sigma)$ and descent set $\mathrm{Des}(\sigma)$ are defined by

$$\mathrm{Asc}(\sigma) = \{1 \leq i \leq n-1 \ : \ \sigma_i < \sigma_{i+1}\} \quad \text{and} \quad \mathrm{Des}(\sigma) = \{1 \leq i \leq n-1 \ : \ \sigma_i > \sigma_{i+1}\},$$

and $\mathrm{asc}(\sigma) = |\mathrm{Asc}(\sigma)|$ and $\mathrm{des}(\sigma) = |\mathrm{Des}(\sigma)|$ denote the number of ascents and descents of σ respectively. The *major index* and *comajor index* of σ are defined via

$$\mathrm{maj}(\sigma) = \sum_{i \in \mathrm{Des}(\sigma)} i \qquad \text{and} \qquad \mathrm{comaj}(\sigma) = \sum_{i \in \mathrm{Asc}(\sigma)} i, \tag{2.1}$$

which are complimentary in the sense that $\mathrm{maj}(\sigma) + \mathrm{comaj}(\sigma) = \binom{n}{2}$.

We define the *inversion* and *coinversion* statistics via

$$\operatorname{inv}(\sigma) = |\{1 \le i < j \le n \;:\; \sigma_i > \sigma_j\}| \quad \text{and} \quad \operatorname{coinv}(\sigma) = |\{1 \le i < j \le n \;:\; \sigma_i < \sigma_j\}|, \tag{2.2}$$

which once again satisfy $\operatorname{inv}(\sigma) + \operatorname{coinv}(\sigma) = \binom{n}{2}$.

The q-binomials are given by

$$[n]_q := 1 + q + \cdots + q^{n-1}, \quad [n]!_q := [n]_q[n-1]_q \cdots [1]_q, \quad \begin{bmatrix} n \\ k \end{bmatrix}_q := \frac{[n]!_q}{[k]!_q \cdot [n-k]!_q}.$$

Theorem 2.1.1. *The statistics* maj*,* comaj*,* inv *and* coinv *are equidistributed on* $\mathfrak{S}_n$*.*

Proof. By [Mac15] (for maj) and [Net01] (for inv) we know that

$$\sum_{\sigma \in \mathfrak{S}_n} q^{\operatorname{inv}(\sigma)} = [n]_q! = \sum_{\sigma \in \mathfrak{S}_n} q^{\operatorname{maj}(\sigma)}$$

showing that inv and maj are equidistributed. As by definition $[n]_q!$ is invariant under reversal of its coefficients, both coinv and comaj have the same generating function as well. □

An *ordered set partition* is a set partition of $[n]$ with a total order on the blocks. For example, $\{1,2\} \prec \{3\}$ and $\{3\} \prec \{1,2\}$ are two different ordered set partitions of $[3]$. We will have two alternative ways to represent ordered set partitions, where we will use $\sigma = \{1,3,6\} \prec \{2,4\} \prec \{5,7\} \in \mathcal{OP}_{7,3}$ as the running example. We let $B_1 = \{1,3,6\}$, $B_2 = \{2,4\}$ and $B_3 = \{5,7\}$

1. We can write the blocks in list form, seperated by vertical bars. In this notation, $\sigma = (B_1 \mid B_2 \mid B_3)$ or even more succintly $\sigma = (136 \mid 24 \mid 57)$.

2. We can write σ using the *ascent starred model*. In this model, write the numbers in the blocks in increasing order, write them in one-line notation, and put stars between numbers in the same block. This turns our above example into $1_*3_*6\ 2_*4\ 5_*7$. Alternatively we can represent this by $\sigma = (\tau, S)$ where $\tau = 1362457 \in \mathfrak{S}_7$ and $S = \{1, 2, 4, 6\} \subseteq \mathrm{Asc}(\tau)$ indicates the positions of the stars.

Upon close examination the ascent starred model identifies $\mathcal{OP}_{n,k}$ with

$$\{(\tau, S) \mid \tau \in \mathfrak{S}_n, S \subseteq \mathrm{Asc}(\tau), |S| = n - k\}.$$

We now define generalizations of maj and inv to the set $\mathcal{OP}_{n,k}$. Consider $\sigma = (B_1 \mid B_2 \mid \cdots \mid B_k) \in \mathcal{OP}_{n,k}$ an inversion is a pair $1 \leq i < j \leq n$ such that i is the minimal element of B_m, $j \in B_\ell$ and $\ell < m$. Let $\mathrm{inv}(\sigma)$ denote the number of inversions. We let $\mathrm{coinv}(\sigma) = (n-k)(k-1) + \binom{k}{2} - \mathrm{inv}(\sigma)$ where $(n-k)(k-1) + \binom{k}{2}$ is the maximal value of $\mathrm{inv}(\sigma)$ on $\mathcal{OP}_{n,k}$, (uniquely) achieved by

$$\sigma = (k\ (k+1)\ \cdots\ n \mid k-1 \mid \cdots \mid 1).$$

To define $\mathrm{maj}(\sigma)$ we use the ascent-starred representation $\sigma = (\tau, S)$. Write $i^c = n + 1 - i$ and for $\tau = \tau_1\tau_2\cdots\tau_n$ let $\tau^c = \tau_1^c\tau_2^c\cdots\tau_n^c$. We define

$$\mathrm{maj}(\sigma) = \mathrm{maj}(\tau^c) - \sum_{i \in S} |\mathrm{Asc}(\tau) \cap \{i, i+1, \ldots, n-1\}|. \tag{2.3}$$

Again, $\mathrm{maj}(\sigma)$ has maximal value $(n-k)(k-1) + \binom{k}{2}$ with unique maximizer

$$\sigma = (1 \mid \cdots \mid k-1 \mid k\ (k+1)\ \cdots\ n),$$

so we define $\mathrm{comaj}(\sigma) = (n-1)(k-1) + \binom{k}{2} - \mathrm{maj}(\sigma)$.

Remark. Note that in the case $n = k$, so when $\mathcal{OP}_{n,k} = \mathfrak{S}_n$, this actually reverses the roles of maj and comaj on the symmetric group. ◀

Just as in the case of the symmetric group, one can compute the generating functions of these statistics to show the following result, see for example [Ste19] and [RW15].

Theorem 2.1.2. *The statistics* inv *and* maj *are equidistributed on* $\mathcal{OP}_{n,k}$.

Lastly, we introduce the notion of a *partition*. Let n be a positive integer, then a partition of (size) n is a non-increasing sequence of positive integers $\lambda = (\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 + \ldots + \lambda_k = n$. We denote $\lambda \vdash n$ to indicate that λ is a partition of n, and write $|\lambda| = n$ and $\ell(\lambda) = k$ for the size and number of parts of λ respectively.

2.2 r-colored ordered set partitions

Let $\mathcal{OP}_{n,k}^{(r)}$ be the set of *r-colored ordered set partitions*, which are the ordered set partitions of $[n]$ where each number $1 \leq i \leq n$ is assigned a color from $\{0, 1, \ldots, r-1\}$. Similar to before we can represent these ordered set partitions as ascent-starred words on the alphabet $\{i^c \mid 1 \leq i \leq n, 0 \leq c \leq r-1\}$ where we order the alphabet in the following way

$$1^{r-1} < 2^{r-1} < \cdots < n^{r-1} < 1^{r-2} < \cdots < n^{r-2} < \cdots < 1^0 < \cdots < n^0.$$

Given *any* word $w = w_1^{c_1} \cdots w_n^{c_n}$ on the above alphabet we define the descent set $\mathrm{Des}(w) = \{1 \leq i \leq n-1 \mid w_i^{c_i} > w_i^{c_{i+1}}\}$ and the major index

$$\mathrm{maj}(w) = c(w) + r \cdot \sum_{i \in \mathrm{Des}(w)} i, \tag{2.4}$$

where $c(w) = c_1 + \cdots + c_n$ denotes the sum of the colors of w.

When $n = k$ we get the set of r-colored permutations, which can also be interpreted as the reflection group $G(r, 1, n) \subseteq \mathrm{GL}_n(\mathbb{C})$ consisting of all monomial matrices with entries in $\{1, \zeta, \zeta^2, \ldots, \zeta^{r-1}\}$ where $\zeta = \exp(2\pi i/r)$ is a primitive r-th root of unity. The correspondence extends the usual connection between permutation matrices and permutations by choosing the color according to the exponent of ζ in the respective column. For example, the monomial matrix

$$\begin{pmatrix} 0 & i & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ -i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix} \in G(4, 1, 5)$$

will be interpreted as $3^3 1^1 5^2 2^2 4^0$.

We can now represent the ascent-starred representations of elements of $\mathcal{OP}_{n,k}^{(r)}$ as pairs (g, λ) where $g \in G(r, 1, n)$ satisfies $\mathrm{des}(g) < k$ and λ is a partition with $k - \mathrm{des}(g) - 1$ parts that are all at most $n - k$. For example, consider

$$\{4^3, 2^2, 3^2\} \prec \{9^1\} \prec \{6^1, 1^0\} \prec \{5^2\} \prec \{7^2, 8^1\} \in \mathcal{OP}_{9,5}^{(4)}$$

which we can represent as $4^3_*2^2_*3^2\ 9^1\ 6^1_*1^0\ 5^2\ 7^2_*8^1$. Here $g = 4^3 2^2 3^2 9^1 6^1 1^0 5^2 7^2 8^1$ and of the ascents at positions $\{1, 2, 3, 5, 7, 8\}$, only the third and fifth are unstarred. Representing each star by a step left and each non-star by a step down, this traces out the following path from $(0, 2)$ to $(4, 0)$

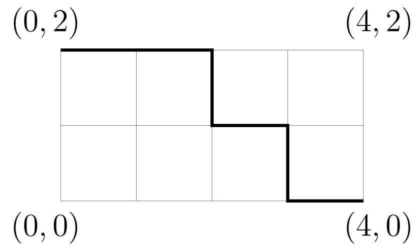

Figure 2.1: A partition traced out by the ascent-starred representation.

which represents the corresponding partition $\lambda = (3, 2)$. We define the comajor index on $\mathcal{OP}^{(r)}_{n,k}$ via

$$\mathrm{comaj}((g, \lambda)) = \mathrm{maj}(g) + r \cdot |\lambda|. \tag{2.5}$$

Lemma 2.2.1. *For* $r = 1$ *the statisic* comaj *in* (2.5) *coincides with the statistic* comaj *introduced after* (2.3).

Proof. Consider $\sigma \in \mathcal{OP}_{n,k}$, which is represented as $\sigma = (g, \lambda)$ where in this case $g \in \mathfrak{S}_n$.

We note that $\mathrm{maj}(\sigma)$ as in (2.3) equals the sum over all the *ascents* of σ with respect to the weight sequence $(w_1, \ldots, w_n)$, where w_i is the number of completed blocks when reaching the i^{th} element of g. Also, $\mathrm{comaj}(g)$ is the sum over all *ascents* of σ with respect to the weight sequence $(1, 2, \ldots, n)$.

Now observe that every starred ascent of (g, λ) results in the weights of that ascent and all ascents after it to be decreased by 1. As λ is a partition with parts at most $n-k$, and the stars correspond to horizontal segments in the bottom left justified Ferrers diegram, the number of affected ascents is equal to

$$\begin{aligned}&(1+\text{the height of the last column}) + (2 + \text{the height of the second to last column}) + \dots \\ &\quad + ((n-k) + \text{the height of the first column}) = (1 + 2 + \dots + (n-k)) + |\lambda|.\end{aligned}$$

Therefore, we have

$$\mathrm{maj}((g, \lambda)) = \mathrm{comaj}(g) - (1 + \dots + (n-k)) - |\lambda|,$$

so we can compute

$$\begin{aligned}\mathrm{comaj}((g,\lambda)) &= (1 + \dots + (k-1)) + (n-k)(k-1) - \mathrm{maj}((g, \lambda)) \\ &= (1 + \dots + (k-1)) + (n-k)(k-1) + (1 + \dots + (n-k)) - \mathrm{comaj}(g) + |\lambda| \\ &= (1 + \dots + (k-1)) + (k + \dots + (n-1)) - \mathrm{comaj}(g) + |\lambda| \\ &= 1 + 2 + \dots + (n-1) - \mathrm{comaj}(g) + |\lambda| = \mathrm{maj}(g) + |\lambda|,\end{aligned}$$

where we expanded $\binom{k}{2} = 1 + \dots + (k-1)$. □

We need one small generalization of r-colored ordered set partitions.

Definition 2.2.2. A $G(r, 1, n)$-face is a set partition of $[n]$ where the letters in every block, with the possible exception of the first block, are colored with one of r colors from $\{0, 1, \dots, r-1\}$. ◀

If the letters of the first block are uncolored we call this block the *zero block*. In particular, $G(r,1,n)$-faces come in two types: (1) r-colored ordered set partitions and (2) a subset of $[n]$ together with an r-colored set partition on the remaining elements of $[n]$. We write $\mathcal{F}_{n,k}$ for the set of $G(r,1,n)$-faces with exactly k nonzero blocks.

Remark. The use of the word *face* originates from the fact that $\mathcal{F}_{n,k}$ can be identified with the set of k-dimensional faces in the Coxeter complex of $G(r,1,n)$. As such, a $G(r,1,n)$-face in $\mathcal{F}_{n,k}$ is also said to have dimension k. ◀

2.3 Symmetric functions and representation theory of $\mathfrak{S}_n$

Let $\Lambda \subseteq \mathbb{Q}[\mathbf{x}]$ be the set of *symmetric functions*. This ring has many bases, but the most useful from the perspective of representation theory is the set of Schur functions $s_\lambda(\mathbf{x})$ where λ is a partition. In particular, the degree n graded piece Λ_n has a basis $\{s_\lambda(\mathbf{x}) \mid \lambda \vdash n\}$. For a thorough definition and properties of Λ, the Schur functions, and other bases we refer to [MR15] and [Sag01].

We now focus on the representation theory of $\mathfrak{S}_n$. It is well known that the irreducible representations of $\mathfrak{S}_n$ biject with partitions $\lambda \vdash n$. Consequently, every $\mathfrak{S}_n$-module V decomposes as

$$V = \bigoplus_{\lambda \vdash n} \left(S^\lambda\right)^{c_\lambda} \tag{2.6}$$

for some integers $c_\lambda \geq 0$. The *Frobenius character* of V is the symmetric function

$$\mathrm{Frob}(V) = \sum_{\lambda \vdash n} c_\lambda \cdot s_\lambda(\mathbf{x}). \tag{2.7}$$

Lastly, let V be a graded vector space such that for every $d \geq 0$ the degree d homogeneous component V_d is finite dimensional. The *Hilbert series* of V is the power series in q given by

$$\mathrm{Hilb}(V; q) = \sum_{d \geq 0} \dim(V_d) \cdot q^d. \tag{2.8}$$

If further V carries a graded $\mathfrak{S}_n$-action, we define the *graded Frobenius character* by

$$\mathrm{grFrob}(V; q) = \sum_{d \geq 0} \mathrm{Frob}(V_d) \cdot q^d. \tag{2.9}$$

A thorough description of the representation theory of $\mathfrak{S}_n$ can be found in [Sag01].

2.4 Generalized coinvariant algebras

The classical coinvariant algebra is defined as follows. The symmetric functions in n variables have algebraically independent generators $e_1(\mathbf{x}_n), \ldots, e_n(\mathbf{x}_n)$ where

$$e_d(\mathbf{x}_n) = \sum_{1 \leq i_1 < \ldots < i_d \leq n} x_{i_1} \cdots x_{i_d} \tag{2.10}$$

is the *elementary symmetric function* of degree d. Consider the ideal in $\mathbb{Q}[\mathbf{x}_n]$ generated by all symmetric functions with zero constant term:

$$I_n = \langle \mathbb{Q}[\mathbf{x}_n]_+^{\mathfrak{S}_n} \rangle = \langle e_1(\mathbf{x}_n), \ldots, e_n(\mathbf{x}_n) \rangle. \tag{2.11}$$

The *coinvariant algebra* $R_n = \mathbb{Q}[\mathbf{x}_n]/I_n$ has long been studied and this algebraic object turns out to interact with permutations and their statistics in the following way.

- $\dim(R_n) = n!$ and $\mathrm{Hilb}(R_n) = [n]_q!$, which we have seen to be the generating function of any of inv, coinv, maj and comaj on $\mathfrak{S}_n$.

- As ungraded $\mathfrak{S}_n$-module $R_n \cong \mathbb{Q}[\mathfrak{S}_n]$ is isomorphic to the regular representation of $\mathfrak{S}_n$.

- As graded $\mathfrak{S}_n$-module we have

$$\mathrm{grFrob}(R_n; q) = \sum_{T \in \mathrm{SYT}(n)} q^{\mathrm{maj}(T)} s_{\mathrm{shape}(T)} = \sum_{w = w_1 \cdots w_n} q^{\mathrm{maj}(w)} x_{w_1} \cdots x_{w_n},$$

  where $w \in \mathcal{W}_n$ is a word of length n on the alphabet of positive integers.

Inspired by the Delta Conjecture, Haglund, Rhoades and Shimozono [HRS18] generalized this and provided a family of algebraic quotients whose combinatorics is controlled by $\mathcal{OP}_{n,k}$ rather than $\mathfrak{S}_n$.

Definition 2.4.1. [HRS18, Def. 1.1] Given two positive integers $k \leq n$, let $I_{n,k} \subseteq \mathbb{Q}[\mathbf{x}_n]$ be the ideal

$$I_{n,k} := \langle x_1^k, \ldots, x_n^k, e_n(\mathbf{x}_n), e_{n-1}(\mathbf{x}_n), \ldots, e_{n-k+1}(\mathbf{x}_n) \rangle. \tag{2.12}$$

Let $R_{n,k}$ be the corresponding quotient ring:

$$R_{n,k} := \frac{\mathbb{Q}[\mathbf{x}_n]}{I_{n,k}}. \qquad ◀ \tag{2.13}$$

Remark. It can be shown that when $k = n$, x_i^n belongs to I_n, hence we have $I_{n,n} = I_n$ and $R_{n,n} = R_n$. ◀

They show various similar properties to above, namely

- $\dim(R_n) = k! \cdot \mathrm{Stir}(n,k) = |\mathcal{OP}_{n,k}|$ and $\mathrm{Hilb}(R_{n,k};q) = \mathrm{rev}_q([k]_q! \cdot \mathrm{Stir}_q(n,k))$, which is the generating function of coinv and comaj on $\mathcal{OP}_{n,k}$.

- As ungraded $\mathfrak{S}_n$-module we have $R_{n,k} \cong \mathbb{Q}[\mathcal{OP}_{n,k}]$.

- As graded $\mathfrak{S}_n$-module we have

$$\mathrm{grFrob}(R_{n,k};q) = \sum_{T \in \mathrm{SYT}(n)} q^{\mathrm{maj}(T)} \begin{bmatrix} n-\mathrm{des}(T)-1 \\ n-k \end{bmatrix}_q s_{\mathrm{shape}(T)}(\mathbf{x}).$$

  This equation can also be written as $\mathrm{grFrob}(R_{n,k};q) = (\mathrm{rev}_q \circ \omega) C_{n,k}(q)$, where $C_{n,k}(\mathbf{x};q)$ is as in Equation (1.2) and rev_q is the operator that reverses a polynomial with respect to the variable q. Finally, ω is the involution on symmetric functions that sends $s_\lambda(\mathbf{x})$ to $s_{\lambda'}(\mathbf{x})$, or equivalently trades $e_n(\mathbf{x})$ for $h_n(\mathbf{x})$.

Chan and Rhoades [CR20] further generalized this to r-colored ordered set partitions. We use $\mathbf{x}_n^r$ to denote $(x_1^r, \ldots, x_n^r)$.

Definition 2.4.2. [CR20, Def. 1.1] Let n, k, and r be nonnegative integers which satisfy $n \geq k$, $n \geq 1$, and $r \geq 2$. We define two quotients of the polynomial ring $\mathbb{C}[\mathbf{x}_n]$ as follows.

(1) Let $I_{n,k} \subseteq \mathbb{C}[\mathbf{x}_n]$ be the ideal

$$I_{n,k} := \langle x_1^{kr+1}, x_2^{kr+1}, \ldots, x_n^{kr+1}, e_n(\mathbf{x}_n^r), \ldots, e_{n-k+1}(\mathbf{x}_n^r) \rangle \tag{2.14}$$

and let $R_{n,k}$ be the corresponding quotient:

$$R_{n,k} := \mathbb{C}[\mathbf{x}_n]/I_{n,k}. \tag{2.15}$$

(2) Let $J_{n,k} \subseteq \mathbb{C}[\mathbf{x}_n]$ be the ideal

$$I_{n,k} := \langle x_1^{kr}, x_2^{kr}, \ldots, x_n^{kr}, e_n(\mathbf{x}_n^r), \ldots, e_{n-k+1}(\mathbf{x}_n^r) \rangle \tag{2.16}$$

and let $R_{n,k}$ be the corresponding quotient:

$$R_{n,k} := \mathbb{C}[\mathbf{x}_n]/I_{n,k}. \quad \blacktriangleleft \tag{2.17}$$

Remark. Note that there is a slight conflict in notation here: the quotient $R_{n,k}$ by Haglund, Rhoades and Shimozono is obtained by taking the $r = 1$ specialization of $S_{n,k}$ introduced by Chan and Rhoades. ◀

Once again, these quotients have nice combinatorial properties: Chan and Rhoades show that as ungraded $G(r, 1, n)$-modules we have

$$R_{n,k} \cong \mathbb{C}[\mathcal{F}_{n,k}] \quad \text{and} \quad S_{n,k} \cong \mathbb{C}[\mathcal{OP}_{n,k}^{(r)}].$$

For the Hilbert series and graded Frobenius series of $R_{n,k}$ and $S_{n,k}$ we refer to [CR20].

2.5 Gröbner theory

Let K be any field and let $I \subseteq K[\mathbf{x}_n]$ be an ideal. On various occasions we will be interested in the dimension of $K[\mathbf{x}_n]/I$. One useful tool in such calculations is Gröbner theory. Following [CLO15], a *monomial order* $<$ on $K[\mathbf{x}_n]$ is a total order on the monomials satisfying

1. $1 \leq m$ for any monomial m;

2. for any monomials m, m_1, m_2 with $m_1 < m_2$ we have $m \cdot m_1 < m \cdot m_2$.

Given such an order and a nonzero $f \in K[\mathbf{x}_n]$, the *leading monomial* $\mathrm{LM}(f)$ is the largest monomial (with respect to the monomial order) that has nonzero coefficient in f. For an ideal $I \subseteq K[\mathbf{x}_n]$ we let $\mathrm{LM}(I)$ be the ideal generated by the leading monomials of all the nonzero elements of I.

Gröbner theory [CLO15, Ch.5, §3, Prop. 1] tells us that one basis for the vector space $K[\mathbf{x}_n]/I$ is given by all the monomials that do not belong to $\mathrm{LM}(I)$, or equivalently all monomials that are not divisible by any $\mathrm{LM}(f)$ for $f \in I$. We will refer to this basis as the *standard monomial basis* of $K[\mathbf{x}_n]/I$ with respect to $<$.

The following monomial orders are of interest to us.

1. The *lexicographical order*: Here, for two monomials $m_1 \neq m_2$ write $m_1 = x_1^{a_1} \cdots x_n^{a_n}$ and $m_2 = x_1^{b_1} \cdots x_n^{b_n}$. Let $j \in \{1, 2, \ldots, n\}$ be minimal such that $a_j \neq b_j$, then $m_1 <_{lex} m_2$ if and only if $a_j <_{lex} b_j$.

2. The *graded lexicographical order*: Here, for we have $m_1 <_{grlex} m_2$ if and only if either $\deg(m_1) < \deg(m_2)$ or $\deg(m_1) = \deg(m_2)$ and $m_1 <_{lex} m_2$

2.6 Catalan numbers

The *Catalan numbers* $C_n = \frac{1}{n+1}\binom{2n}{n}$ form a sequence of numbers that is ubiquitous in combinatorics. They enumerate a wide variety of objects, including polygon triangulations, binary trees, plane trees, and Dyck paths. For a thorough background of Catalan numbers and their properties we refer to [Pak14] and [Sta15]. In our case we will be interested in yet another set of objects counted by the Catalan numbers, which we will refer to as the *Catalan-arc matchings (of size n)*, which are placements of n non-intersecting, semi-circular arcs on $2n$ given collinear points.

For example, below one can see the $C_3 = 5$ Catalan-arc matchings of size 3.

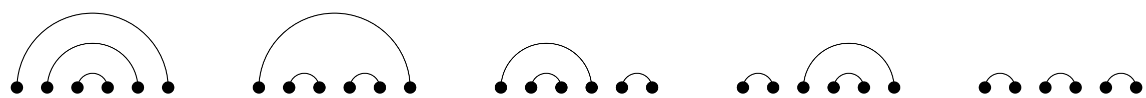

Figure 2.2: Catalan-arc matchings of size 3.

2.7 Results in probability

In our probabilistic approach to the Catalan-pair graphs we will use the following concentration result.

Lemma 2.7.1. *Let n be a positive integer and let $X_1, X_2, \ldots, X_n$ be mutually independent random variables with $\mathbb{P}[X_i = 0] = \mathbb{P}[X_i = 1] = \frac{1}{2}$. Define $S_n = X_1 + X_2 + \ldots + X_n$ and let $a > 0$. Then*

$$\mathbb{P}[|S_n - n/2| > a] < 2e^{-2a^2/n}. \tag{2.18}$$

Proof. Let $Y_i = 2X_i - 1$, and $T_n = Y_1 + Y_2 + \ldots + Y_n = 2S_n - n$. Using [AS08, Cor A.1.2] we see that

$$\mathbb{P}[|S_n - n/2| > a] = \mathbb{P}[|T_n| > 2a] < 2e^{-(2a)^2/2n} = 2e^{-2a^2/n}. \qquad \square$$

Additionally, we will make use of Chebyshev's inequality [AS08, Thm. 4.1.1].

Lemma 2.7.2. *Let X be a random variable with finite expected value $\mathbb{E}[X]$ and finite nonzero variance* $\mathrm{Var}[X]$*. Then, for any positive k we have*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq k] \leq \frac{\mathrm{Var}[X]}{k^2}. \tag{2.19}$$

This chapter contains material from: D. Kroes, *"Generalized coinvariant algebras for $G(r,1,n)$ in the Stanley-Reisner setting"*, Electronic Journal of Combinatorics, vol. 26 (3), P.3.11, 2018. The dissertation author was the primary investigator and author of this paper.

This chapter also contains material from: D. Kroes and B. Rhoades, *"Packed words and quotient rings"*, submitted (2021). The dissertation author was one of the primary investigators and authors of this paper.

This chapter also contains material from: D. Kroes and S. Spiro, *"Random Graphs Induced by Catalan Pairs"*, Journal of Combinatorics, Accepted (2020). The dissertation author was one of the primary investigators and authors of this paper.

Chapter 3

Generalized coinvariant algebras in the Stanley–Reisner setting

3.1 The Stanley-Reisner ring

Consider the polynomial ring $\mathbb{C}[\mathbf{y}_S]$ where $\mathbf{y}_S = \{y_{\{1\}}, \ldots, , y_{\{n\}}, \ldots, y_{\{1,2,\ldots,n\}}\}$ is a list of variables indexed by the non-empty subsets $S \subseteq [n]$. The *Stanley–Reisner ring* of the Boolean algebra is given by $\mathbb{C}[\mathcal{B}_n^*] := \frac{\mathbb{C}[\mathbf{y}_S]}{\langle y_S \cdot y_T \rangle}$, where the generators of the ideal satisify $S \not\subseteq T$ and $T \not\subseteq S$. It is easy to see that $\mathbb{C}[\mathcal{B}_n^*]$ has a $\mathbb{C}$-basis given by multichain monomials, which are monomials of the form $y = y_{S_1} \cdots y_{S_t}$ with $\emptyset \neq S_1 \subseteq S_2 \subseteq \ldots \subseteq S_t \subseteq [n]$.

One important tool will be the following ring homomorphism $\mathbb{C}[\mathbf{y}_S] \to \mathbb{C}[\mathbf{x}_n]$.

Definition 3.1.1. Let n be a positive integer. Let $\varphi : \mathbb{C}[\mathbf{y}_S] \to \mathbb{C}[\mathbf{x}_n]$ be the ring homomorphism defined by

$$\varphi(y_S) = \prod_{i \in S} x_i. \qquad \blacktriangleleft$$

Note that φ does not vanish on $\langle y_S \cdot y_T \rangle$, so φ does not induce a ring isomorphism between $\mathbb{C}[\mathcal{B}_n^*]$ and $\mathbb{C}[\mathbf{x}_n]$. However, it is well known that defining φ on the multichain basis of $\mathbb{C}[\mathcal{B}_n^*]$ yields a $\mathfrak{S}_n$-module isomorphism between $\mathbb{C}[\mathcal{B}_n^*]$ and $\mathbb{C}[\mathbf{x}_n]$. Furthermore, this even gives us a G_n-module isomorphism if we equip $\mathbb{C}[\mathcal{B}_n^*]$ with a G_n-structure in the following way. For $g \in G_n$ and y_S we set $g \cdot y_S = \alpha y_T$ where α and T are selected according to $g \cdot \prod_{i \in S} x_i = \alpha \prod_{j \in T} x_j$, and extend this multiplicatively to $\mathbb{C}[\mathbf{y}_S]$.

Garsia and Stanton [GS84] show that one can obtain the coinvariant algebra as a quotient of $\mathbb{C}[\mathcal{B}_n^*]$. For $1 \leq i \leq n$, denote

$$\theta_i = \sum_{S \subseteq [n], |S| = i} y_S. \tag{3.1}$$

Note that applying our homomorphism from above we have $\varphi(\theta_i) = e_i(\mathbf{x}_n)$. It is shown by Garsia and Stanton that there is an isomorphism between the coinvariant algebra R_n and the quotient

$$\mathcal{R}_n = \frac{\mathbb{C}[\mathcal{B}_n^*]}{\langle \theta_1, \ldots, \theta_n \rangle}.$$

3.2 The main result

For the remainder of this chapter we fix $r \geq 1$, and we denote $G_n = G(r, 1, n)$ as usual. From now on, we also write $\mathcal{OP}_{n,k}$ to denote $\mathcal{OP}_{n,k}^{(r)}$. We first define the quotients of $\mathbb{C}[\mathcal{B}_n^*]$ that will be the analogues of $R_{n,k}$ introduced in Definition 2.4.1 and $R_{n,k}$ and $S_{n,k}$ from Definition 2.4.2.

Definition 3.2.1. Let $0 \le k \le n$ be integers with $n \ge 1$. In $\mathbb{C}[\mathbf{y}_S]$ we define the following ideals:

$$\mathcal{I}_{n,k} = \langle y_S \cdot y_T, \theta_{n-k+1}, \ldots, \theta_n, y_{S_1} \cdots y_{S_{kr+1}} \rangle; \tag{3.2}$$

$$\mathcal{J}_{n,k} = \langle y_S \cdot y_T, \theta_{n-k+1}, \ldots, \theta_n, y_{S_1} \cdots y_{S_{kr}} \rangle, \tag{3.3}$$

where S and T range over all pairs of nonempty subsets $S, T \subseteq [n]$ with $S \not\subseteq T$ and $T \not\subseteq S$,

$$\theta_i = \sum_{S \subseteq [n], |S| = i} y_S^r$$

and $(S_1, \ldots, S_{kr+1})$ and $(S_1, \ldots, S_{kr})$ range over all multichains $S_1 \subseteq \ldots \subseteq S_{kr+1}$ and $S_1 \subseteq \ldots \subseteq S_{kr}$ of nonempty subsets of $[n]$ of length $kr+1$ and kr respectively.

Lastly, define $\mathcal{R}_{n,k} = \mathbb{C}[\mathbf{y}_S]/\mathcal{I}_{n,k}$ and $\mathcal{S}_{n,k} = \mathbb{C}[\mathbf{y}_S]/\mathcal{J}_{n,k}$. ◀

Remark. Even though we will show that there exist bases for $\mathcal{R}_{n,k}$ and $R_{n,k}$ (and $\mathcal{S}_{n,k}$ and $S_{n,k}$) such that the image of the $\mathbf{y}$-variable basis under φ is exactly the $\mathbf{x}$-variable basis, the map φ will *not* define a G_n-module isomorphism on these bases, not even in the case of the classical coinvariant algebra. Instead, φ is often referred to as the *transfer map*, indicating the analogy between the Stanley–Reisner quotients and the traditional $\mathbf{x}$-variable quotients. ◀

The main result of this chapter is the following.

Theorem 3.2.2. *For $r \geq 2$, we have G_n-module isomorphisms $\mathcal{R}_{n,k} \cong R_{n,k}$ and $\mathcal{S}_{n,k} \cong S_{n,k}$.*

Remark. Note that the assumption $r \geq 2$ is introduced to have notational consistency with [CR20] as in Definition 2.4.2. However, all the proofs transfer over to the case $r = 1$ to show an isomorphism between $\mathcal{S}_{n,k}$ and $R_{n,k}$ from [HRS18] as in Definition 2.4.1. ◀

3.3 Preliminary results

Before proving Theorem 3.2.2 we will need some auxiliary results. In this section, we will work towards bases for $\mathcal{R}_{n,k}$ and $\mathcal{S}_{n,k}$ that resemble bases introduced in [CR20]. Often we will show a result for the $\mathcal{S}_{n,k}$ quotient and then most of the arguments will directly transfer over to the case of $\mathcal{R}_{n,k}$.

In order to prove some results we need a monomial order on $\mathbb{C}[\mathbf{y}_S]$. In our case, we will equip $\mathbb{C}[\mathbf{y}_S]$ with the graded lexicographical monomial order with respect to the ordering of the variables by $y_S > y_T$ if $|S| > |T|$ or $|S| = |T|$ and $\min(S \backslash T) < \min(T \backslash S)$. For example, for $n = 3$, this order is given by

$$y_{\{1,2,3\}} > y_{\{1,2\}} > y_{\{1,3\}} > y_{\{2,3\}} > y_{\{1\}} > y_{\{2\}} > y_{\{3\}}.$$

We remark that only this ordering on the variables is essential, because we will mainly work in homogeneous components of $\mathbb{C}[\mathbf{y}_S]$. Therefore, one could use any monomial order with this ordering on the variables instead.

3.3.1 Garsia–Stanton type bases

Let us recall the Garsia–Stanton type bases for $R_{n,k}$ and $S_{n,k}$, as introduced by Chan and Rhoades [CR20, Def. 5.7 & Def. 5.9]. In order to do so we need the classical Garsia–Stanton basis for R_n, indexed by elements $g \in G_n$. When $g = \pi_1^{c_1} \cdots \pi_n^{c_n}$, set $d_i(g) = \#\{j \geq i \ : \ j \in \mathrm{Des}(g)\}$ for the number of descents at or after position i. The *descent monomial* b_g is defined by

$$b_g = \prod_{i=1}^{n} x_{\pi_i}^{rd_i(g)+c_i}. \tag{3.4}$$

Now, the following set descends to a $\mathbb{C}$-vector space bases for $S_{n,k}$ [CR20, Def. 5.7 & Thm. 5.8]:

$$\mathcal{D}_{n,k} = \{b_g \cdot x_{\pi_1}^{ri_1} \cdots x_{\pi_{n-k}}^{ri_{n-k}} \ : \ g \in G_n, \mathrm{des}(g) < k, k - \mathrm{des}(g) > i_1 \geq \ldots \geq i_{n-k} \geq 0\}. \tag{3.5}$$

Furthermore, $R_{n,k}$ has a similar basis $\mathcal{E}_{n,k}$ [CR20, Def. 5.9 & Thm. 5.10] given by all elements of the form

$$\prod_{j \in Z} x_j^{kr} \cdot b_{\pi_{z+1}^{c_{z+1}} \cdots \pi_n^{c_n}} \cdot x_{\pi_{z+1}}^{ri_{z+1}} \cdots x_{\pi_{n-k}}^{ri_{n-k}}, \tag{3.6}$$

where $Z \subseteq [n]$ satisfies $0 \leq |Z| = z \leq n-k$, $\pi_{z+1}^{c_{z+1}} \ldots \pi_n^{c_n}$ is a word on $[n] - Z$ with $\mathrm{des}(\pi_{z+1}^{c_{z+1}} \ldots \pi_n^{c_n}) < k$ and $k - \mathrm{des}(\pi_{z+1}^{c_{z+1}} \ldots \pi_n^{c_n}) > i_{z+1} \geq \ldots \geq i_{n-k} \geq 0$.

Since, $|\mathcal{D}_{n,k}| = |\mathcal{OP}_{n,k}|$ one might wonder whether there is a natural way to index those basis elements by elements of $\mathcal{OP}_{n,k}$. One way to do so is using our ascent starred model for $\mathcal{OP}_{n,k}$.

Definition 3.3.1. Given an element in $\mathcal{OP}_{n,k}$ represented by (g, λ) we define

$$b_{(g,\lambda)} := b_g \cdot x_{\pi_1}^{ri_1} \cdots x_{\pi_{n-k}}^{ri_{n-k}}, \tag{3.7}$$

where $i_j = \#\{m \ : \ \lambda_m \geq j\}$. ◀

Similarly, since $|\mathcal{E}_{n,k}| = |\mathcal{F}_{n,k}|$ we would like to index elements of $\mathcal{E}_{n,k}$ by elements of $\mathcal{F}_{n,k}$. Again, we will use our model for elements of $\mathcal{F}_{n,k}$. To this end, note that the definitions of b_g and $b_{(g,\lambda)}$ make sense even if g is just a word on the alphabet $\{i^j \ : \ 1 \leq i \leq n, 0 \leq j \leq r-1\}$. Therefore, we have the following definition.

Definition 3.3.2. Let (Z, g, λ) represent an element in $\mathcal{F}_{n,k}$. Set

$$b_{(Z,g,\lambda)} = \prod_{i \in Z} x_i^{kr} \cdot b_{(g,\lambda)}. \qquad \blacktriangleleft \qquad (3.8)$$

It is an easy check that $\mathcal{D}_{n,k} = \{b_{(g,\lambda)} \ : \ (g,\lambda) \in \mathcal{OP}_{n,k}\}$ and $\mathcal{E}_{n,k} = \{b_{(Z,g,\lambda)} \ : \ (Z, g, \lambda) \in \mathcal{F}_{n,k}\}$.

3.3.2 An intermediate quotient

We will first consider the quotient $\mathbb{C}[\mathbf{y}_S]$ by the ideal $\langle y_S \cdot y_T, \theta_{n-k+1}, \ldots, \theta_n \rangle$.

Definition 3.3.3. Let $g \in G_n$ with $g = \sigma_1^{c_1} \cdots \sigma_n^{c_n}$, and let $d \in \mathbb{Z}_{\geq 0}^n$.

1. Define $\tilde{b}_g = \prod_{i=1}^n y_{T_i}^{m_i}$, where $T_i = \{\sigma_j \ : \ 1 \leq j \leq i\}$ and

$$m_i = \begin{cases} c_i - c_{i+1} + r & \text{if } i < n \text{ and } i \in \mathrm{Des}(g); \\ c_i - c_{i+1} & \text{if } i < n \text{ and } i \notin \mathrm{Des}(g); \\ c_n & \text{if } i = n. \end{cases}$$

2. Set $\tilde{b}_{(g,d)} = \tilde{b}_g \cdot \prod_{i=1}^n y_{T_i}^{rd_i}$. ◀

As an example let $n = 5$, $r = 3$ and $g = 4^0 2^2 5^2 3^2 1^1$. Then we have descents at positions 1 and 3, so $\tilde{b}_g = y_{\{4\}} y^3_{\{2,4,5\}} y_{\{2,3,4,5\}} y_{\{1,2,3,4,5\}}$.

Lemma 3.3.4. *Every multichain monomial* $y = y_{S_1} \cdots y_{S_t}$ *is equal to* $\tilde{b}_{(g,d)}$ *for a unique* $(g, d) \in G_n \times \mathbb{Z}^n_{\geq 0}$.

Before we give the proof, let us illustrate the idea of the proof. Let $n = 6$, $r = 3$ and consider $y = y^5_{\{4\}} y^7_{\{1,3,4\}} y_{\{1,2,3,4,6\}} y^4_{\{1,2,3,4,5,6\}}$. If we want to write this in the form $\tilde{b}_{(g,d)}$ the underlying permutation of g has to be $4abcd5$ with $\{a, b\} = \{1, 3\}$ and $\{c, d\} = \{2, 6\}$. Now, note that if $ab = 31$, then $y_{\{4,3\}}$ will have exponent at least 1 in b_g, either because 3 and 1 have different colors, or because 3 and 1 have the same color, which implies that g has a descent at the second position. Similarly, we have $cd = 26$ and hence the underlying permutation is 413265. Now, let $c_1, \ldots, c_6$ be the colors (of 4, 1, 3, 2, 6 and 5). By the above, we have $c_2 = c_3$ and $c_4 = c_5$. Note that $y_{\{1,2,3,4,5,6\}}$ has exponent c_6 in b_g, hence exponent equivalent to c_6 modulo 3 in $b_{(g,d)}$. Therefore, since $0 \leq c_6 \leq 2$, we need $c_6 = 1$. Equivalently, $y_{\{1,2,3,4,6\}}$ has exponent equivalent to $c_5 - c_6$ modulo 3 in b_g (it is either $c_5 - c_6$ or $c_5 - c_6 + 3$) hence we have $c_5 - c_6 \equiv 1 \mod 3$ in $b_{(g,d)}$ as well. We conclude that $c_4 = c_5 = 2$. Similarly, $c_2 = c_3 = 0$ and $c_1 = 2$, hence the only option for g is $4^2 1^0 3^0 2^2 6^2 5^1$. Note that in this case $b_g = y^2_{\{4\}} y_{\{1,3,4\}} y_{\{1,2,3,4,6\}} y_{\{1,2,3,4,5,6\}}$, so we can uniquely write $y = b_{(g,d)}$ for $d = (1, 0, 2, 0, 0, 1)$.

Proof. Suppose that our multichain monomial is of the form $y = y^{a_1}_{S_{i_1}} \cdots y^{a_j}_{S_{i_j}}$, where $1 \leq i_1 < \ldots, i_j \leq n$, $|S_{i_k}| = i_k$ for $1 \leq k \leq j$ and $a_1, \ldots, a_j > 0$. Let $S_{i_1} = \{g_1 < \ldots < g_{i_1}\}$,

$S_{i_m}\backslash S_{i_{m-1}} = \{g_{i_{m-1}+1} < \ldots < g_{i_m}\}$ for $2 \le m \le j$ and $[n]\backslash S_{i_j} = \{g_{i_j+1} < \ldots < g_n\}$ (if this set is non-empty). Note that if $y = \tilde{b}_{(g,d)}$ then the one-line notation of the underlying permutation of g has to be $g_1 g_2 \ldots g_n$ and all elements that are in the same set (from $\{g_1 < \ldots < g_{i_1}\}$, $\{g_{i_{m-1}+1} < \ldots < g_{i_m}\}$ and $\{g_{i_j+1} < \ldots < g_n\}$) need to have the same color. Let these colors be $c_1, \ldots, c_{i_j}$ and c_{i_j+1} (the last one appearing only if necessary). Indeed, from the definition, if $h \in G_n$ and h_i and h_{i+1} have different colors then $y_{\{h_1,\ldots,h_i\}}$ has exponent $c_i - c_{i+1}$ or $c_i - c_{i+1} + r$ (depending on whether there is a descent or not) and in both cases this exponent is nonzero, so $\tilde{b}_{(h,d)}$ does not equal y. And if h_i and h_{i+1} have the same color, but $h_i > h_{i+1}$, then $y_{\{h_1,\ldots,h_i\}}$ would appear with exponent $r > 0$, so again this cannot happen. Therefore, the underlying permutation of g is uniquely determined (if it exists). On the other hand, if such c_1 up to c_j (and possibly c_{j+1}) exist, they are also uniquely determined, by a backwards inductive argument. Indeed, if $S_{i_j} \neq [n]$, then $Y_{[n]}$ has coefficient 0 modulo r, hence we need $c_{j+1} = 0$, and else $S_{i_j} = [n]$ and c_j has to be the exponent of S_{i_j} taken modulo n. Now, suppose c_k has been determined, then we will determine c_{k-1}. It is clear that we need $c_{k-1} - c_k \equiv a_{k-1} \mod r$, and since c_{k-1} has to be taken from $\{0, \ldots, r-1\}$ this gives a unique choice. Now, for this choice of the colors, and the corresponding g, we show that there is a suitable $d \in \mathbb{Z}^n_{\geq 0}$. Note that by construction, $\tilde{b}_g = y^{b_1}_{S_{i_1}} \cdots y^{b_j}_{S_{i_j}}$, where $b_i \equiv a_i \mod r$. Furthermore, $b_i \in \{0, 1, \ldots, r\}$. It is clear that we can get d by taking $d_m = 0$ when $m \neq i_t$ and taking $d_{i_m} = (b_m - a_m)/r$ when $m \in \{1, \ldots, j\}$. Note that this is an integer by $b_m \equiv a_m \mod r$. Furthermore, it is nonnegative, since $a_m > 0$, $b_m \geq 0$, $a_m \equiv b_m \mod r$ and $b_m \in \{0, 1, \ldots, r\}$ implies that

$b_m \geq a_m$. □

Using this we can find a different basis for $\mathbb{C}[\mathcal{B}_n^*]$.

Definition 3.3.5. Let $g \in G_n$ and $d \in \mathbb{Z}_{\geq 0}$. Define

$$\tilde{b}'_{(g,d)} = \theta_{n-k+1}^{d_{n-k+1}} \cdots \theta_n^{d_n} \tilde{b}_{(g,(d_1,\dots,d_{n-k},0,\dots,0))}.$$ ◀

Lemma 3.3.6. *The set* $\{\tilde{b}'_{(g,d)} \ : \ g \in G_n, d \in \mathbb{Z}_{\geq 0}\}$ *is a* $\mathbb{C}$*-basis for* $\mathbb{C}[\mathcal{B}_n^*]$.

Proof. Order the basis $\tilde{b}_{(g,d)}$ according to the monomial order from above. Note that for each monomial y, the set of monomials y' with $y' \leq y$ is finite, since any such monomial y' must have $\deg(y') \leq \deg(y)$ and there are finitely many such monomials.

Now, if we expand $\tilde{b}'_{(g,d)}$ in terms of the basis $\{\tilde{b}_{(g,d)}\}$ we find that

$$\tilde{b}'_{(g,d)} = \tilde{b}_{(g,d)} + \text{lower terms with respect to } <.$$

Indeed, suppose g has underlying permutation $g_1 \cdots g_n$. Set $S_i = \{g_1, \dots, g_i\}$. Note that if $g_i > g_{i+1}$, or $c_i \neq c_{i+1}$ then necessarily we have that y_{S_i} occurs in $\tilde{b}_g$ with a positive exponent. Note that (since we only allow multichains), we have $\theta_a^b = \sum_{|S|=a} y_S^{rb}$. Now, terms in $\tilde{b}'_{(g,d)}$ correspond to picking one of the terms from each of the θ_a^b with positive b, in such a way that the result is still a multichain. Because of our monomial order, we should pick from larger a first. Suppose we are picking a subset of size i and suppose $i_t < i < i_{t+1}$ (set $i_{j+1} = n$) (we can exclude $i = i_t$, because of the multichain condition we must pick S_i). Then, we are asking for the largest y_S with $|S| = i$ and $S_{i_t} \subseteq S \subseteq S_{i_{t+1}}$, which is $S = \{g_1, \dots, g_i\}$, due to the fact that $g_{i_t+1}, \dots, g_{i_{t+1}}$ all have the same color and are in increasing order, by the proof of the lemma above. Therefore, the largest possible monomial that could possibly appear is obtained by picking y_{S_i} for every $i > n-k$ with $d_i > 0$. Now note that if we take this choice for all i simultaneously we indeed get a multichain monomial, and this monomial is equal to $\tilde{b}_{(g,d)}$, as desired.

Therefore, $\tilde{b}'_{(g,d)}$ expands in a unitriangular way in terms of the basis $\{\tilde{b}_{(g,d)}\}$ and because of the initial observation in this proof, it follows that $\{\tilde{b}'_{(g,d)} \ : \ g \in G_n, d \in \mathbb{Z}_{\geq 0}\}$ is a basis for $\mathbb{C}[\mathcal{B}_n^*]$. □

For $(d_1, \ldots, d_{n-k}) = d \in \mathbb{Z}_{\geq 0}^{n-k}$ set $\tilde{b}_{(g,d)} = \tilde{b}_{(g,(d_1,\ldots,d_{n-k},0,\ldots,0))}$. Then the following is immediate.

Corollary 3.3.7. *$\mathbb{C}[\mathcal{B}_n^*]$ is a free $\mathbb{C}[\theta_{n-k+1}, \ldots, \theta_n]$-module with basis given by*

$$\{\tilde{b}_{(g,d)} \ : \ g \in G_n, d \in \mathbb{Z}_{\geq 0}^{n-k}\}. \tag{3.9}$$

Furthermore, this set descends to a vector space basis for $\mathbb{C}[\mathcal{B}_n^*]/\langle\theta_{n-k+1}, \ldots, \theta_n\rangle =$ $\mathbb{C}[\mathbf{y}_S]/\langle y_S \cdot y_T, \theta_{n-k+1}, \ldots, \theta_n\rangle$.

Additionally, this allows us to quickly determine a vector space basis for the quotient $\mathbb{C}[\mathbf{y}_S]/\langle y_S \cdot y_T, \theta_{n-k+1}, \ldots, \theta_n, y_{S_1,\ldots,S_m}\rangle$, of which we will be interested in the cases $m = kr$ and $m = kr + 1$. Again, the result is immediate, so the proof is omitted.

Corollary 3.3.8. *Let $m \in \mathbb{Z}_{>0}$ and consider $\mathbb{C}[\mathbf{y}_S]/\langle y_S \cdot y_T, \theta_{n-k+1}, \ldots, \theta_n, y_{S_1} \cdots y_{S_m}\rangle$, where (S,T) runs over all pairs with $S \not\subseteq T$ and $T \not\subseteq S$, and $(S_1, \ldots, S_m)$ runs over all $\emptyset \neq S_1 \subseteq \ldots \subseteq S_m \subseteq [n]$. This is a finite-dimensional $\mathbb{C}$-vector space with basis given by all elements $\tilde{b}_{(g,d)}$ with $g \in G_n$, $d \in \mathbb{Z}_{\geq 0}^{n-k}$ and $\deg(\tilde{b}_{(g,d)}) < m$.*

3.3.3 Bases for the rings $\mathcal{R}_{n,k}$ and $\mathcal{S}_{n,k}$

Note that Corollary 3.3.8 yields bases for $\mathcal{R}_{n,k}$ and $\mathcal{S}_{n,k}$. In this section we will show that these bases can be indexed by elements of $\mathcal{F}_{n,k}$ and $\mathcal{OP}_{n,k}$ respectively. We will use the models introduced before.

Definition 3.3.9.

1. For $(g,\lambda) \in \mathcal{OP}_{n,k}$, let $\tilde{b}_{(g,\lambda)} = \tilde{b}_g \cdot y_{S_1}^r \cdots y_{S_t}^r$, where $S_i = \{g_i : 1 \leq j \leq \lambda_i\}$.

2. Let $(Z,g,\lambda) \in \mathcal{F}_{n,k}$. If (loosely extending the definition above) we have $\tilde{b}_{(g,\lambda)} = y_{S_1} \cdots y_{S_j}$, then set $\tilde{b}'_{(g,\lambda)} = y_{S_1 \cup Z} \cdots y_{S_j \cup Z}$. Now, set $\tilde{b}_{(Z,g,\lambda)} = y_Z^{kr - \deg(\tilde{b}_{(g,\lambda)})} \cdot \tilde{b}'_{(g,\lambda)}$. ◀

It is an easy check that $\varphi(\tilde{b}_{(g,\lambda)}) = b_{(g,\lambda)}$ and $\varphi(\tilde{b}_{(Z,g,\lambda)}) = b_{(Z,g,\lambda)}$. The main result is now the following.

Theorem 3.3.10. *The sets $\{\tilde{b}_{(g,\lambda)} \ : \ (g,\lambda) \in \mathcal{OP}_{n,k}\}$ and $\{\tilde{b}_{(Z,g,\lambda)} \ : \ (Z,g,\lambda)\}$ are bases for $\mathcal{S}_{n,k}$ and $\mathcal{R}_{n,k}$ respectively.*

Proof. Let us first show that there is a bijection between elements of the form $\tilde{b}_{(g,\lambda)}$ and $\tilde{b}_{(g,d)}$ with $\deg(\tilde{b}_{(g,d)}) < kr$. Note that for any partition λ with parts at most $n-k$, we have (after extending the above definition to allow for any partition) $\tilde{b}_{(g,\lambda)} = \tilde{b}_{(g,d)}$, where $d = (d_1, \ldots, d_{n-k})$ with $d_i = \#\{j \ : \ \lambda_j = i\}$. Therefore, it suffices to show that λ has at most $k - \operatorname{des}(g) - 1$ parts if and only if $\deg(\tilde{b}_{(g,\lambda)}) < kr$. Now, note that if λ has m parts, we have

$$\deg(\tilde{b}_{(g,\lambda)}) = \deg(\tilde{b}_g) + mr = \sum_{i=1}^{n-1}(c_i - c_{i+1} + r \cdot \chi(i \text{ is a descent})) + c_n + mr$$

$$= c_1 + r\text{des}(g) + mr = c_1 + (m + \text{des}(g))r,$$

where χ is the indicator function given by $\chi(S) = 1$ if statement S is true and $\chi(S) = 0$ otherwise. Now, since $c_1 \in \{0, 1, \ldots, r-1\}$ we have $\deg(\tilde{b}_{(g,\lambda)}) < kr$ if and only if $m + \text{des}(g) \leq k - 1$, that is if and only if λ has at most $k - \text{des}(g) - 1$ parts.

Similarly, we have to show that there is a bijection between elements of the form $\tilde{b}_{(Z,g,\lambda)}$ and $\tilde{b}_{(g,d)}$ with $\deg(\tilde{b}_{(g,d)}) \leq kr$. A similar calculation to above shows that $\deg(\tilde{b}_{(Z,g,\lambda)}) < kr$ if $Z = \emptyset$ and clearly $\deg(\tilde{b}_{(Z,g,\lambda)} = kr$ when $Z \neq \emptyset$, so it suffices to show that there is a bijection between elements of the form $\tilde{b}_{(Z,g,\lambda)}$ with $Z \neq \emptyset$ and $\tilde{b}_{(g,d)}$ with $\deg(\tilde{b}_{(g,d)}) = kr$. Note that $\deg(\tilde{b}_{(g,d)}) = c_1 + r(\text{des}(g) + d_1 + \ldots + d_{n-k})$, so $\deg(\tilde{b}_{(g,d)}) = kr$ if and only if $c_1 = 0$ and $\text{des}(g) + d_1 + \ldots + d_{n-k} = k$.

Now, given $\tilde{b}_{(g,d)}$ with $\deg(\tilde{b}_{(g,d)}) = kr$, we show that there is a unique (Z, h, λ) such that $\tilde{b}_{(Z,h,\lambda)} = b_{(g,d)}$. Let S be the smallest subset (in size) such that y_S has positive exponent in $\tilde{b}_{(g,\lambda)}$. It is clear that if (Z, h, λ) exists we must have $Z = S$. Now, suppose that $|S| > n-k$. Then in particular we have $d_1 = \ldots = d_{n-k} = 0$, and g has no descents at positions $1, \ldots, n-k$. But then, using $c_1 = 0$, we have $\deg(\tilde{b}_{g,d}) = \deg(\tilde{b}_g) = c_1 + r\text{des}(g) = r\text{des}(g) < r(k-1)$, a contradiction. Therefore, let $z = |S|$, so that $1 \leq z \leq n-k$. Using $c_1 = 0$ and minimality of S, we see that $g = g_1^0 \cdots g_z^0 \cdots$ with $g_1 < \ldots < g_z$. Additionally,

$d_1 = \ldots = d_{z-1} = 0$. Set $b = \tilde{b}_{(g,d)}/y_S^e$, where e is the exponent of y_S, and write

$$b = \prod_{i=1}^{m} y_{S \cup S_i},$$

where $\emptyset \neq S_1 \subseteq \ldots \subseteq S_m \subseteq [n] \backslash S$. Note that $\prod_{i=1}^{m} y_{S_i} = \tilde{b}_{(h,d)}$ for $h = g_{z+1}^{c_{z+1}} \cdots g_n^{c_n}$ and $d = (d_{z+1}, \ldots, d_{n-k})$. We now want to show that there is a unique (h, λ) such that $(Z, h, \lambda) \in \mathcal{F}_{n,k}$ and $\tilde{b}_{(h,\lambda)} = \tilde{b}_{(h,d)}$. However, since $\deg(\tilde{b}_{(h,d)}) < kr$ the first part of the proof shows that indeed we can find such a (h, λ).

Conversely, we show that $\tilde{b}_{(Z,h,\lambda)}$ is of the form $\tilde{b}_{(g,d)}$ for a unique (g, d). Write $Z = \{g_1 < \ldots < g_z\}$ and let $h = h_1^{c_1} \cdots h_{n-z}^{c_{n-z}}$. It is clear that we must have $g = g_1^c \cdots g_z^c h_1^{c_1} \cdots h_{n-z}^{c_{n-z}}$ for a suitable c. Furthermore, since we need $\deg(\tilde{b}_g) \equiv 0 \mod r$, we in fact have to pick $c = 0$. Therefore, g is uniquely determined, and hence d (if it exists) is also uniquely determined. By construction, if $S_t = \{g_1, \ldots, g_t\}$, the exponent of y_{S_t} in $\tilde{b}_g$ and in $\tilde{b}_{(Z,h,\lambda)}$ agree modulo r. Indeed, this is obvious for $t > z$, and for $t \leq z$ the choice of $c = 0$ guarantees this. Furthermore, for $t > n - k$ we still have that the exponents agree as integers (so not only modulo r). Therefore, it only suffices to show that for any $1 \leq t \leq n-k$ the exponent of y_{S_t} in $\tilde{b}_{(Z,h,\lambda)}$ is at least the exponent of y_{S_t} in $\tilde{b}_g$. Again, this is obvious for $z < t \leq n - k$. Additionally, it is clear for $1 \leq t < z$, since by construction y_{S_t} has exponent 0 in $\tilde{b}_g$. Now, for $t = z$, we are immediately okay if $y_{S_z} = y_Z$ occurs with exponent less $\{0, 1, \ldots, r-1\}$ in $\tilde{b}_g$. Therefore, the only thing that might fail is that y_Z occurs with exponent r in $\tilde{b}_g$ but exponent 0 in $\tilde{b}_{(Z,h,\lambda)}$. However, since $\deg(\tilde{b}_{(h,\lambda)}) < kr$, we know that y_Z occurs with exponent at least 1 in $\tilde{b}_{(Z,h,\lambda)}$ and therefore, with exponent at least r, as desired. □

3.3.4 A Gröbner theory result

In this section we will show that the above bases are actually the standard monomial bases with respect to the monomial order used. Our proof methods are inspired by Braun and Olsen [BO18], who obtain similar results in the case that $n = k$.

Theorem 3.3.11. *Let $0 \leq k \leq n$ be integers with $n \geq 1$. Then the set $\{\tilde{b}_{(g,\lambda)} \ : \ (g, \lambda) \in \mathcal{OP}_{n,k}\}$ is precisely the standard monomial basis for $\mathcal{S}_{n,k}$.*

Proof. Since we know that the given set is a basis, it suffices to show that the standard monomial basis of $\mathcal{S}_{n,k}$ is contained in $\{\tilde{b}_{(g,\lambda)} \ : \ (g, \lambda) \in \mathcal{OP}_{n,k}\}$.

Similar to Braun and Olsen [BO18] we show that $y \in \{\tilde{b}_{(g,\lambda)} \ : \ (g, \lambda) \in \mathcal{OP}_{n,k}\}$ if and only if y is not divisible by any of the monomials in the list below. The proof will then be completed by showing that each of these monomials occurs as the leading term of some element of $\mathcal{J}_{n,k}$. The list of monomials is given by

1. $y_S \cdot y_T$ for $S \not\subseteq T$ and $T \not\subseteq S$;

2. $y_{[m]}^r$ for $m \geq n - k + 1$;

3. y_S^{r+1} for $|S| \geq n - k + 1$;

4. $y_S^r \cdot y_T$ for $S \subsetneq T$, $|S| \geq n - k + 1$ and $\min(T \backslash S) > \max(S)$;

5. $y_S \cdot y_T^r$ for $S \subsetneq T$, $|T| \geq n - k + 1$ and $T = S \cup [\ell]$ for some ℓ;

6. $y_{S_1} \cdot y_{S_2}^r \cdot y_{S_3}$ for $S_1 \subsetneq S_2 \subsetneq S_3$, $|S_2| \geq n - k + 1$ and $\max(S_2 \backslash S_1) < \min(S_3 \backslash S_2)$;

7. $y_{S_1} \cdots y_{S_{kr}}$ where $S_1 \subseteq \ldots \subseteq S_{kr}$.

We will first show necessity of these conditions, then sufficiency and lastly will exhibit these monomials as leading terms in $\mathcal{J}_{n,k}$.

Necessity: we will assume that g is of the form $\pi_1^{c_1} \cdots \pi_n^{c_n}$. Note that if y_S with $|S| \geq n-k+1$ occurs in some $\tilde{b}_{(g,\lambda)}$ then its contribution completely comes from $\tilde{b}_g$. Now, if there is no descent at position $|S|$, y_S will have exponent $c_{|S|+1} - c_{|S|} \leq (r-1) - 0 < r$. Furthermore, if there is a descent at position $|S|$, we have $c_{|S|+1} \geq c_{|S|}$, so y_S will have exponent $r + c_{|S|+1} - c_{|S|} \leq r$. Therefore, if y_S occurs with exponent at least r, it occurs with exponent exactly r, we have a descent at position $|S|$ and $c_{|S|+1} = c_{|S|}$.

1. Each variable occurring in $\tilde{b}_{(g,\lambda)}$ is of the form y_{S_i} for $1 \leq i \leq n$, where $S_i = \{\pi_1, \ldots, \pi_i\}$. Since $S_1 \subseteq S_2 \subseteq \ldots \subseteq S_n$, every two variables in $\tilde{b}_{(g,\lambda)}$ will automatically be indexed by subsets one of which is contained in the other.

2. Since $m \geq n-k+1$, $y_{[m]}^r$ would have to come from a descent of g at position m with $c_m = c_{m+1}$. In order to have a descent we need $\pi_{m+1} < \pi_m$. However, $\pi_m \in [m]$, hence $\pi_m \leq m$, whereas $\pi_{m+1} \in [n] \backslash [m]$, so $\pi_{m+1} \geq m+1$.

3. This was observed above.

4. Suppose such a product $y_S^r \cdot y_T$ actually occurs. Since $|S| \geq n-k+1$, y_S^r comes from a descent at position $|S|$ with $c_{|S|+1} = c_{|S|}$, so $\pi_{|S|} > \pi_{|S|+1}$. Since $\{\pi_1, \ldots, \pi_{|S|}\} = S$ and $\{\pi_1, \ldots, \pi_{|T|}\} = T$, we have $\min(T \backslash S) \leq \pi_{|S|+1} < \pi_{|S|} \leq \max(S)$, which is an obvious contradiction.

5. Suppose that such a product occurs. Again, y_T^r has to come from a descent at position $|T|$ with $c_{|T|} = c_{|T|+1}$, hence $\pi_{|T|} > \pi_{|T|+1}$. Note that $\pi_{|T|} \in T \backslash S \subseteq [\ell]$, so $\pi_{|T|} \leq \ell$. Furthermore, $\pi_{|T|+1} \notin T$, hence in particular $\pi_{|T|+1} > \ell$, which is a contradiction.

6. Suppose such a triple product occurs. Since $|S_2| \geq n - k + 1$, $y_{S_2}^r$ comes from a descent at position $|S_2|$ with $c_{|S_2|} = c_{|S_2|+1}$, so we must have $\pi_{|S_2|} > \pi_{|S_2|+1}$. However, $\pi_{|S_2|} \in S_2 \backslash S_1$ and $\pi_{|S_2|+1} \in S_3 \backslash S_2$, so by assumption we have $\pi_{|S_2|} \leq \max(S_2 \backslash S_1) < \min(S_3 \backslash S_2) \leq \pi_{|S_2|+1}$.

7. We note that

$$\begin{aligned}
\deg(\tilde{b}_{(g,\lambda)}) &\leq \deg(\tilde{b}_g) + (k - \mathrm{des}(\sigma) - 1)r = \sum_{i=1}^{n} m_i + (k - \mathrm{des}(\sigma) - 1)r \\
&= \sum_{i=1}^{n} (c_i - c_{i+1} + r\chi(i \text{ is a descent})) + (k - \mathrm{des}(\sigma) - 1)r \\
&= c_1 + r\mathrm{des}(\sigma) + (k - \mathrm{des}(\sigma) - 1)r = kr + c_1 - r \leq kr - 1,
\end{aligned}$$

where $c_{n+1} = 0$, and χ is the indicator function of the indicated event.

Sufficiency: Let $m = y_{S'_1} \cdots y_{S'_t}$ be a monomial not divisible by any of the above mentioned monomials. Then combining properties 1. and 7. we may assume $S'_1 \subseteq S'_2 \subseteq \ldots \subseteq S'_t$ and $t < kr$. However, we will rewrite this as $m = y_{S_1}^{t_1} \cdots y_{S_u}^{t_u}$, where $S_1 \subsetneq S_2 \subsetneq \ldots \subsetneq S_u$.

We will first construct the corresponding $g \in G_n$, after which the augmentation λ will follow automatically. Firstly, the underlying permutation of σ will be given by putting the elements of S_1 in ascending order, then the elements of $S_2\backslash S_1$, ..., the elements of $S_u\backslash S_{u-1}$ in ascending order and finally the elements of $[n]\backslash S_u$ in ascending order. Now, we have to assign colors to each of the elements. We will give all elements of S_1 the same color, all elements of $S_2\backslash S_1$ the same color, ..., all elements of $S_u\backslash S_{u-1}$ the same color and finally all elements of $[n]\backslash S_u$ the same color. We will assign these colors in reverse order. Firstly, assign color 0 to everything in $[n]\backslash S_u$, then assign color t_u to $S_u\backslash S_{u-1}$, then color $t_u + t_{u-1}$ to $S_{u-1}\backslash S_{u-2}$, ... and finally color $t_u + t_{u-1} + \ldots + t_1$ to S_1. Here, everything should be interpreted modulo n. It is an easy check that $m = \tilde{b}_g \cdot y_{S_1}^{rv_1} \cdots y_{S_u}^{rv_u}$, where $v_1, v_2, \ldots, v_u \geq 0$.

Now, let us check that g together with some appropriate λ satisfies the condition that $m = b_{(g,\lambda)}$. Firstly, using a similar computation to above, $r\text{des}(g) \leq \deg(\tilde{b}_g) \leq \deg(m) < kr$, hence $\text{des}(g) < k$, as desired. So, to see that the augmented part corresponds to an appropriate λ we have to check two things, namely that $v_j = 0$ if $|S_j| \geq n-k+1$ and that $v_1 + \ldots + v_u \leq (k - \text{des}(g) - 1)$. For the latter, note that

$$r(v_1 + \ldots + v_u) = \deg(m) - \deg(b_g) < kr - \deg(b_g) \leq kr - \text{des}(g)r = (k - \text{des}(g))r,$$

so $v_1 + \ldots + v_u < k - \text{des}(g)$, as desired. For the first part, note that if $m = |S_j| \geq n-k+1$, then y_{S_j} has exponent at most r by condition 3. Therefore, if $v_j > 0$, we need $v_j = 1$, and the exponent of y_{S_j} in b_g equals 0. In particular, σ_m and σ_{m+1} have the same color and $\sigma_m < \sigma_{m+1}$. Furthermore, note that y_{S_j} now has exponent exactly r, so in particular we have $S_j \neq [m]$ by condition 2. Now, we distinguish four cases.

- $j = u = 1$: In this case, $\sigma_m = \max(S_1) > m$ and $\sigma_{m+1} = \min([n]\backslash S_1) \leq m$, a contradiction.

- $j = 1, u > 1$: In this case, $\sigma_m = \max(S_1)$ and $\sigma_{m+1} = \min(S_2\backslash S_1)$. By condition 4. this implies $\sigma_m > \sigma_{m+1}$, a contradiction.

- $1 < j < u$: Now, $\sigma_m = \max(S_j\backslash S_{j-1})$ and $\sigma_{m+1} = \min(S_{j+1}\backslash S_j)$, but then $\sigma_m < \sigma_{m+1}$ contradicts condition 6.

- $1 < j = u$: Now $\min([n]\backslash S_u) = \sigma_{m+1} > \sigma_m$, so $[\sigma_m] \subseteq S_u$. Furthermore, $\max(S_u\backslash S_{u-1}) = \sigma_m]$ hence $S_u\backslash S_{u-1} \subseteq [\sigma_m]$ so by $[\sigma_m] \subseteq S_u$ this implies $S_u = S_{u-1} \cup [\sigma_m]$. However, this contradicts condition 5.

Therefore, we need to have $v_j = 0$ if $|S_j| \geq n - k + 1$, completing this part of the proof.

Leading monomials:

1. These monomials are among the generators of $\mathcal{J}_{n,k}$.

2. These monomials are the leading monomials of $\theta_m \in \mathcal{J}_{n,k}$.

3. Write $m = |S|$ and consider $y_S\theta_m \in \mathcal{J}_{n,k}$. All monomials in this polynomial are of the form $y_S \cdot y_T^r$ where $|T| = m = |S|$. Note that all such products have S and T incomparable, except for when $T = S$. Therefore, modulo $\mathcal{J}_{n,k}$ this equals y_S^{r+1}, showing that y_S^{r+1} in fact occurs in $\mathcal{J}_{n,k}$.

4. Write $m = |S|$ and consider $\theta_m \cdot y_T$. All monomials in this polynomial are of the form $y_R^r \cdot y_T$ where $|R| = m = |S|$. Modulo $J_{n,k}$ this is equal to $\sum_R y_R^r \cdot y_T$ where

R runs over all such subsets with $R \subseteq T$. By assumption, S is the smallest such set with respect to the monomial order, hence $y_S^r \cdot y_T$ is the leading term of this monomial.

5. Let $m = |T|$ and note that $\mathcal{J}_{n,k}$ contains $y_S \cdot \theta_T$ which modulo $\mathcal{J}_{n,k}$ reduces to $\sum_R y_S \cdot y_R^r$ where $S \subseteq R$ and $|R| = m$. Since $T = S \cup [\ell]$ it is clear that T is the lexicographically smallest such set, so this polynomial has leading monomial $y_S \cdot y_T^r$.

6. Let $m = |S_2|$ and consider $y_{S_1} y_{S_3} \cdot \theta_m$. Similarly, this equals $\sum_T y_{S_1} y_T^r y_{S_3}$ modulo $\mathcal{J}_{n,k}$ where T runs over all m-element subsets $S_1 \subseteq T \subseteq S_3$. By assumption, S_2 is the lexicographically smallest such set, hence $y_{S_1} y_{S_2} y_{S_3}$ can be obtained as a leading monomial.

7. These monomials are among the generators of $\mathcal{J}_{n,k}$.

This completes the proof. □

Similarly, we have the following result.

Theorem 3.3.12. *Let* $0 \leq k \leq n$ *be integers with* $n \geq 1$. *Then the set* $\{\tilde{b}_{(Z,g,\lambda)} : (Z, g, \lambda) \in \mathcal{F}_{n,k}\}$ *is precisely the standard monomial basis for* $\mathcal{R}_{n,k}$.

Proof. Again it suffices to show that the standard monomial basis of $\mathcal{R}_{n,k}$ is contained in $\{\tilde{b}_{(Z,g,\lambda)} : (Z, g, \lambda) \in \mathcal{F}_{n,k}\}$. We will show that a monomial y belongs to this set if and only if it is not divisible by any of the monomials in the exact same list as before, except that we need to change the 7th condition into

7'. $y_{S_1} \cdots y_{S_{kr+1}}$ where $S_1 \subseteq \ldots \subseteq S_{kr+1}$.

Again we will go through the steps necessity, sufficiency and show that they occur as leading monomials in $\mathcal{J}_{n,k}$.

Necessity: Condition 1 is clearly still satisfied, and conditions 2-6 follow by the exact same argument, since the appropriate monomials y_S with $|S| \geq n-k+1$ still have to come from the contribution of g to $b_{(Z,g,\lambda)}$, since neither Z, nor λ will affect the exponent of these. For condition 7', we note that $b_{(Z,g,\lambda)}$ might now have degree kr (when $Z \neq \emptyset$), but will never have degree $kr + 1$ or more.

Sufficiency: There are two cases to consider. Let y be a monomial not divisible by any of the monomials specified in the list. We will show that y is of the form $b_{(Z,g,\lambda)}$. If $\deg(y) < kr$, then we set $Z = \emptyset$ and use the same procedure as in Theorem 3.3.11 to find the appropriate (g, λ). If $\deg(y) = kr$, set Z to be the smallest subset S of $[n]$ (in size) such that y_S has positive exponent in y. Let e_S be the exponent of S and set $y' = y/y_S^{e_S}$. Now, set y'' to be the same monomial as y' where each y_T is replaced by $y_{T \setminus S}$. Since $\deg(y'') < kr$, the same procedure as in Theorem 3.3.11 can be used to find appropriate (g, λ) to complete the triple (Z, g, λ).

Leading monomials: For conditions 1-6 the reasoning is exactly the same, since none of them use the multichain generators of $\mathcal{I}_{n,k}$. For condition 7', it again follows immediately since these multichain monomials belong to the generators of $\mathcal{J}_{n,k}$. □

3.4 A filtration of $\mathcal{R}_{n,k}$ and $\mathcal{S}_{n,k}$.

We are now ready to prove Theorem 3.2.2. We first need two definitions.

Definition 3.4.1. 1. For $y = y_{S_1} \cdots y_{S_m} \in \mathbb{C}[\mathbf{y}_S]$ a multichain monomial with $S_1 \subseteq \ldots \subseteq S_m$, we let $\mu(y)$ be the partition $(|S_m|, \ldots, |S_1|)$.

2. For $m \in \mathbb{C}[\mathbf{x}_n]$ a monomial, we let $\mu(m)$ be the partition given by $\mu(y)$, where y is the unique multichain monomial with $\varphi(y) = m$. ◀

Let $\trianglerighteq$ be the dominance order on partitions and let $(\mathcal{A}, A)$ be $(\mathcal{R}_{n,k}, R_{n,k})$ or $(\mathcal{S}_{n,k}, S_{n,k})$. Now fix $d \geq 0$ and let $\mu \vdash d$. Set

$$U_{\trianglerighteq\mu} = \text{span}\{m \ : \ \mu(m) \vdash d, \mu(m) \trianglerighteq \mu\} \qquad \text{and} \qquad \mathcal{U}_{\trianglerighteq\mu} = \text{span}\{y \ : \ \mu(y) \vdash d, \mu(y) \trianglerighteq \mu\}. \tag{3.10}$$

and define $U_{\triangleright\mu}$ and $\mathcal{U}_{\triangleright\mu}$ in a similar fashion. Let $V_{\trianglerighteq\mu}$ be the image of $U_{\trianglerighteq\mu}$ in A, $\mathcal{V}_{\trianglerighteq\mu}$ be the image of $V_{\trianglerighteq\mu}$ in $\mathcal{A}$ and similarly for the other 2. Now, A and $\mathcal{A}$ decompose as G_n-modules as

$$\bigoplus_{d\geq 0}\bigoplus_{\mu\vdash d} V_{\trianglerighteq\mu}/V_{\triangleright\mu} \qquad \text{and} \qquad \bigoplus_{d\geq 0}\bigoplus_{\mu\vdash d} \mathcal{V}_{\trianglerighteq\mu}/\mathcal{V}_{\triangleright\mu},$$

respectively. The proof of Theorem 3.2.2 now follows from the lemma below.

Lemma 3.4.2. *For each μ, $V_{\trianglerighteq\mu}/V_{\triangleright\mu}$ and $\mathcal{V}_{\trianglerighteq\mu}/\mathcal{V}_{\triangleright\mu}$ have bases $\{b \ : \ \mu(b) = \mu\}$ and $\{\tilde{b} \ : \ \mu(\tilde{b}) = \mu\}$ respectively, where b and $\tilde{b}$ belong to the Garsia-Stanton type bases mentioned before. Furthermore, the map $\tilde{b} \to b = \varphi(b)$ induces a G_n-module isomorphism $\mathcal{V}_{\trianglerighteq\mu}/\mathcal{V}_{\triangleright\mu} \to V_{\trianglerighteq\mu}/V_{\triangleright\mu}$.*

In turn, this lemma follows from two other lemmas, for which we need another definition.

Definition 3.4.3. Let $(\mathcal{A}, A) = (\mathcal{S}_{n,k}, S_{n,k})$ (resp. $(\mathcal{A}, A) = (\mathcal{R}_{n,k}, R_{n,k})$). Given a partition $\mu \vdash d$ with parts that are at most n we say that μ is

1. *admissible* if μ has less than kr (resp. $kr + 1$) parts, $n - k + 1 \leq i \leq n - 1$ occurs at most r times and n occurs at most $r - 1$ times.

2. *semi-admissible* if λ has less than kr (resp. $kr + 1$) parts, has at most $r - 1$ parts equal to n, but some $n - k + 1 \leq i \leq n - 1$ occurs at least $r + 1$ times.

3. *non-admissible* if λ has at least kr (resp. $kr + 1$) parts or has at least r parts equal to n. ◀

For example, when $n = 6$, $k = 3$ and $r = 2$, the partitions $(5, 5, 2, 2, 2)$, $(6, 5, 5, 5, 1)$, $(6, 5, 4, 4, 2, 2, 2, 1)$ and $(6, 6, 2)$ are admissable, semi-admissable, non-admissable and non-admissable respectively, both when $(\mathcal{A}, A) = (\mathcal{S}_{n,k}, S_{n,k})$ and when $(\mathcal{A}, A) = (\mathcal{R}_{n,k}, R_{n,k})$. However, the partition $(6, 5, 5, 2, 2, 2)$ is non-admissable if $(\mathcal{A}, A) = (\mathcal{S}_{6,3}, S_{6,3})$, but admissable for $(\mathcal{R}_{6,3}, R_{6,3})$.

Note that μ is admissible if and only if there exists a basis element $\tilde{b}$ with $\mu(\tilde{b}) = \mu$. A *move* is replacing y_S^r by $y_S^r - \theta_{|S|}$ and cancelling out all non-multichain terms or replacing $x_{i_1}^r \cdots x_{i_j}^r$ by $x_{i_1}^r \cdots x_{i_j}^r - e_j(\mathbf{x}_n^r)$ (for $i_1 < \ldots < i_j$), depending on what setting one is working in.

The two main lemmas are now as follows:

Lemma 3.4.4. *Let y be a multichain monomial in $\mathbb{C}[\mathbf{y}_S]$ with $\mu(y) = \mu$. Then*

1. *if μ is semi-admissible or non-admissible, $y = 0$ in $\mathcal{A}$.*

2. *if μ is admissible, one can perform a finite number of moves to find the expansion of y in $\mathcal{A}$ in terms of the Garsia-Stanton type basis. Additionally, any multichain monomial Y that ever appears in this process has $\mu(Y) = \mu$.*

Proof. For part 1, note that since μ is semi-admissible or non-admissible, y is divisible by $y_{[n]}^r$, y_S^{r+1} for $n-k+1 \leq |S| \leq n-1$ or a multichain monomial of length kr (resp. $kr+1$). Since the ideal we quotient out by to get $\mathcal{A}$ contains $y_S^{r+1} \equiv y_S \cdot \theta_{|S|}$ for $n-k+1 \leq |S| \leq n-1$, we see that all of $y_{[n]}^r$, y_S^{r+1} and the multichain monomials belong to the ideal, hence $y = 0$ in $\mathcal{A}$.

For the second part, recall the monomial order on $\mathbb{C}[\mathbf{y}_S]$ from before. Now, consider a monomial y with $\mu(y) = \mu$. We claim that if y is not a Garsia-Stanton type monomial we can perform a move and rewrite y as a $\mathbb{C}$-linear combination of smaller monomials y' with $\mu(y') = \mu(y)$. Indeed, since μ is admissible, a monomial y that is not a basis monomial is this for one of four reasons (by the classification of monomials that are of this form given in Theorems 3.3.11 and 3.3.12):

1. y is divisible by $y_{[t]}^r$ for some $n-k+1 \leq t \leq n-1$.

2. y is divisible by $y_S^r y_T$ for $S \subsetneq T$, $|S| \geq n-k+1$ and $\min(T \backslash S) > \max(S)$.

3. y is divisible by $y_S y_T^r$ for $S \subsetneq T$, $|T| \geq n-k+1$ and $T = S \cup [\ell]$ for some ℓ.

4. y is divisible by $y_{S_1}y_{S_2}^r y_{S_3}$ for $S_1 \subsetneq S_2 \subseteq S_3$, $|S_2| \geq n-k+1$ and $\max(S_2 \backslash S_1) < \min(S_3 \backslash S_2)$.

In these cases, apply the move, replacing $y_{[t]}^r$, y_S^r, y_T^r and $y_{S_2}^r$ respectively. Any monomial that remains after crossing out non-multichain monomials is obtained by replacing this specific variable by some y_R^r, so it suffices to show that y_R is smaller than the replaced monomial. In the first case, $|R| = t$, so $y_R < y_{[t]}$. In the second case, R is a subset of T of size $|S|$ and since $\min(T \backslash S) > \max(S)$, S was the subset corresponding to the largest possible monomial over all R, and similarly in the other 2 cases. Therefore, we can rewrite each non-basis monomial in terms of smaller monomials, so at some point we will be left with only basis-monomials as desired. $\square$

For the proof of the second lemma we need the following observation. Note that if y is any monomial in $\mathbb{C}[\mathbf{y}_S]$ we can still define $\mu(y)$, even if y is not a multichain monomial. Now, if y is a non-multichain monomial, let y' be the unique multichain monomial with $\varphi(y) = \varphi(y')$. We claim that $\mu(y') \triangleright \mu(y)$. Indeed, starting from y we can repeatedly replace $y_A y_B$ (for A and B incomparable) by $y_{A\cup B}y_{A\cap B}$. Note that on the μ-level this corresponds to replacing $(\ldots, |A|, \ldots, |B|, \ldots)$ by $(\ldots, |A \cup B|, \ldots, |A \cap B|, \ldots)$ which strictly increases the corresponding partition in dominance order (since A and B are incomparable). Note that this local replacement does not change the image under φ and since μ increases every time we can only do this finitely many times. So we will end up with some multichain monomial and by uniqueness this is y'. Also, we have done at least one replacement, so indeed $\mu(y')$ is strictly larger than $\mu(y)$.

The $\mathbf{x}_n$-variable analogue of the above lemma is the following.

Lemma 3.4.5. *Let m be a monomial in $\mathbb{C}[\mathbf{x}_n]$ with $\mu(m) = \mu$. Then*

1. *if μ is non-admissible, $m = 0$ in A.*

2. *if μ is semi-admissible, then in A we can rewrite m as a sum of monomials m_α with $\mu(m_\alpha) \triangleright \mu$.*

3. *if μ is admissible, then a finite number of moves can be used to rewrite m as a $\mathbb{C}$-linear combination of Garsia-Stanton monomials m' with $\mu(m') = \mu$, together with monomials m_α with $\mu(m_\alpha) \triangleright \mu$. Moreover, if the moves in part 2 of Lemma 3.4.4 are replacing $y^r_{S_1}$, $y^r_{S_2}$, $\ldots$, $y^r_{S_m}$ respectively, then the moves in this case are replacing $\prod_{i \in S_1} x_i^r$, $\prod_{i \in S_2} x_i^r$, $\ldots$, $\prod_{i \in S_m} x_i^r$ respectively.*

Proof. Let y be the multichain monomial associated to m.

For the first case, since λ is non-admissible, y is divisible by either $y^r_{[n]}$ or a multichain of length kr (resp. $kr + 1$). In the first case, $e_n(\mathbf{x}_n^r) = x_1^r \cdots x_n^r$ divides m, hence $m = 0$ in $R_{n,k}$. In the second case, let j be an element that is in the smallest S such that y_S occurs in the multichain. Then x_j^{kr} (resp. x_j^{kr+1}) divides m and consequently $m = 0$ in A.

In the second case, we have that y is divisible by y_S^{r+1} for some S with $n - k + 1 \leq S \leq n - 1$. Suppose $S = \{i_1, \ldots, i_j\}$. Then apply the move by replacing $x^r_{i_1} \cdots x^r_{i_j}$. We can "pull back" the move to $\mathbb{C}[\mathbf{y}_S]$, where we replace y^r_S by $y^r_S - \theta_{|S|}$, but we *do not* get rid of non-multichains. Now, any monomial that occurs will contain $y_S y_T$ for $|S| = |T|$

but $S \neq T$, hence would have been removed in the y-setting, but in the x-setting these monomials remain. However, by the above observation, all of these monomials have strictly smaller μ-partition, as desired.

In the third case, again "pull back" to the y-setting and do the exact same sequence of moves as in part 2 of the above lemma, but again we do not get rid of non-multichains. Instead, we replace them by multichains with the same image under φ and again this will strictly increase the μ-partition. $\square$

As an example of this phenomenon, consider $r = 2$ and $\mathcal{S}_{5,4}$. In the $\mathbf{y}$-variable setting, consider $y = y_{\{5\}}^3 y_{\{2,5\}}^2 y_{\{1,2,3,5\}}^2$. Note that this is not yet of the form $\tilde{b}_{(g,\lambda)}$, for example since the appearance of $y_{\{2,5\}} y_{\{1,2,3,5\}}^2$ violates condition 5 in the proof of Theorem 3.3.11. Therefore, we apply a step and replace $y_{\{1,2,3,5\}}^2$ by $y_{\{1,2,3,5\}}^2 - \theta_4$ and after getting rid of any monomial that is not a multichain monomial we find that

$$y \equiv -y_{\{5\}}^3 y_{\{2,5\}}^2 y_{\{1,2,4,5\}}^2 - y_{\{5\}}^3 y_{\{2,5\}}^2 y_{\{2,3,4,5\}}^2.$$

Here, the first monomial is $\tilde{b}_{(5^1 2^0 1^0 4^0 3^0),(1)}$ is of the desired form. However, the second monomial contains $y_{\{5\}} y_{\{2,5\}}^2 y_{\{2,3,4,5\}}$, which violates condition 6 in the proof of Theorem 3.3.11. Therefore, we perform a step on $y_{\{2,5\}}^2$ and get that

$$y \equiv -y_{\{5\}}^3 y_{\{2,5\}}^2 y_{\{1,2,4,5\}}^2 + y_{\{5\}}^3 y_{\{3,5\}}^2 y_{\{2,3,4,5\}}^2 + y_{\{5\}}^3 y_{\{4,5\}}^2 y_{\{2,3,4,5\}}^2,$$

and one can check that all monomials appearing here are indeed of the form $\tilde{b}_{(g,\lambda)}$. Also, note that we started with a monomial with μ-partition $(4, 4, 2, 2, 1, 1, 1)$ and each monomial kept that form.

Now, in the x-variable setting we have to consider $x_5^7x_2^4x_1^2x_3^2$. Replacing the monomial $(x_1x_2x_3x_5)^2$ by $(x_1x_2x_3x_5)^2 - e_4(x_1^2, x_2^3, x_3^2, x_4^2, x_5^2)$ we get

$$x_5^7x_2^4x_1^2x_3^2 \equiv -x_5^7x_2^4x_1^2x_4^2 - x_5^7x_2^4x_3^2x_4^2 - x_5^7x_1^2x_2^2x_3^2x_4^2 - x_5^5x_2^4x_1^2x_3^2x_4^2.$$

Here, the first monomial is a generalized Garsia-Stanton monomial, the second monomial is one we have to perform another step on. Now, the last two monomials have y-monomial $y_{\{5\}}^5y_{\{1,2,3,4,5\}}^2$ and $y_{\{5\}}y_{\{2,5\}}^2y_{\{1,2,3,4,5\}}^2$ respectively, hence they have μ-partitions $(5,5,1,1,1,1,1)$ and $(5,5,2,2,1)$. Now, it holds that $(5,5,1,1,1,1,1) \triangleright (4,4,2,2,1,1,1)$ and $(5,5,2,2,1) \triangleright (4,4,2,2,1,1,1)$. Therefore,

$$x_5^7x_2^4x_1^2x_3^2 \equiv -x_5^7x_2^4x_1^2x_4^2 - x_5^7x_2^4x_3^2x_4^2 + \text{monomials with larger } \mu\text{-partition},$$

and hence the first step of the algorithm carries out in the a way similar to the first step in the $\mathbf{y}$-variable setting. Now, applying an analogous step to $x_5^7x_2^4x_3^2x_4^2$ we find that

$$x_5^7x_2^4x_1^2x_3^2 \equiv -x_5^7x_2^4x_1^2x_4^2 + x_5^7x_3^4x_2^2x_4^2 + x_5^7x_4^4x_2^2x_3^2 + \text{monomials with larger } \mu\text{-partition},$$

which indeed show that even though the expansion of y and $x_5^7x_2^4x_1^2x_3^2$ in the Garsia-Stanton bases are not identical, the monomials that appear and have the same μ-partition as the original monomial *do* coincide, and their coefficients agree.

Lemma 3.4.2 is now an easy application of the Lemmas 3.4.4 and 3.4.5.

Proof of Lemma 3.4.2. If μ is non-admissible or semi-admissible the above lemmas show that $V_{\trianglerighteq\mu}/V_{\triangleright\mu}$ and $\mathcal{V}_{\trianglerighteq\mu}/\mathcal{V}_{\triangleright\mu}$ are both trivial G_n-modules.

Now, suppose μ is admissible. We need to show that sending $\tilde{b}$ to $b = \varphi(\tilde{b})$ (for $\tilde{b}$ a Garsia-Stanton monomial with $\mu(\tilde{b}) = \mu$) induces a G_n-module isomorphism. For $g \in G_n$ we can rewrite $g \cdot \tilde{b}$ in the Garsia-Stanton basis using the moves from part 2 of Lemma 3.4.4. Since the multichain monomial corresponding to πb is given by $\pi \tilde{b}$ we can use part 3 of Lemma 3.4.5 to rewrite πb in the same way in this given basis (viewed as basis for $V_{\trianglerighteq \mu}/V_{\triangleright \mu}$), since all the additional monomials that appear belong to $V_{\triangleright \mu}$ and hence are 0 in the quotient. □

3.5 Multi-graded Frobenius series

We now specialize to the case $r = 1$ and determine the multi-graded Frobenius character of $\mathcal{S}_{n,k}$. We show that the appropriate specialization of this Frobenius character agrees with the graded Frobenius character of $R_{n,k}$, as seen in [HRS18, Corollary 6.13] and [HR18, Corollary 6.3].

Note that if μ is a partition with parts at most n and if d is a nonnegative integer, the subspaces $U_d = \mathrm{span}\{m : \deg(m) = d\}$ and $\mathcal{U}_\mu = \mathrm{span}\{m : \mu(m) = \mu\}$ are $\mathfrak{S}_n$-stable subspaces of $\mathbb{C}[\mathbf{x}_n]$ and $\mathbb{C}[\mathbf{y}_S]$ respectively, and hence so are their images V_d and $\mathcal{V}_\mu$ in $S_{n,k}$ and $\mathcal{S}_{n,k}$ (for $r = 1$) respectively. The graded Frobenius character and multi-graded Frobenius character of $S_{n,k}$ and $\mathcal{S}_{n,k}$ are

$$\operatorname{grFrob}(S_{n,k};q) = \sum_{d=0}^{\infty} q^d \operatorname{Frob}(V_d);$$

$$\operatorname{grFrob}(\mathcal{S}_{n,k};t_1,\ldots,t_n) = \sum_{\mu} t_1^{m_1(\mu)} \cdots t_n^{m_n(\mu)} \operatorname{Frob}(\mathcal{V}_\mu),$$

where the sum is over all partitions with parts at most n, and $m_i(\mu)$ is the number of parts of μ equal to i. We can determine the graded Frobenius image $\operatorname{grFrob}(\mathcal{S}_{n,k};t_1,\ldots,t_n)$.

Theorem 3.5.1. *Suppose that $r=1$. Then*

$$\operatorname{grFrob}(\mathcal{S}_{n,k};t_1,\ldots,t_n) = \sum_{\substack{\alpha\models n\\ \ell(\alpha)\le k}} \left(\prod_{i\in D(\alpha)} t_i\right)\left(\sum_{j_1+\cdots+j_{n-k}\le k-\ell(\alpha)} \prod_{i=1}^{n-k} t_i^{j_i}\right) s_\alpha,$$

where the sum runs over compositions $\alpha=(\alpha_1,\ldots,\alpha_m)$ of n and $D(\alpha)=\{\alpha_1,\ldots,\alpha_1+\ldots+\alpha_{m-1}\}$. Furthermore, by setting $t_i=q^i$ we recover the graded Frobenius character of $S_{n,k}$, in accordance with [HRS18, Corollary 6.13] and [HR18, Corollary 6.3].

Proof. We write $\beta\preceq\alpha$ if $D(\beta)\subseteq D(\alpha)$. It is well known that

$$h_\alpha = \sum_{\beta\preceq\alpha} s_\beta.$$

It follows from work of Garsia and Stanton [GS84] that

$$\begin{aligned}
\mathrm{grFrob}(\mathbb{C}[\mathcal{B}_n^*];t_1,\ldots,t_n) &= \sum_{\gamma\models n}\left(\prod_{i\in D(\gamma)}(t_i+t_i^2+\cdots)\right)h_\gamma \\
&= \sum_{\alpha\models n}\sum_{\alpha\preceq\gamma}\left(\prod_{i\in D(\gamma)}(t_i+t_i^2+\cdots)\right)s_\alpha \\
&= \sum_{\alpha\models n}\left(\prod_{i\in D(\alpha)}(t_i+t_i^2+\cdots)\right)\sum_{\alpha\preceq\gamma}\left(\prod_{i\in D(\gamma)\setminus D(\alpha)}(1+t_i+t_i^2+\cdots)\right)s_\alpha \\
&= \sum_{\alpha\models n}\left(\prod_{i\in D(\alpha)}(t_i+t_i^2+\cdots)\right)\left(\prod_{i\in D(\alpha^c)}(1+t_i+t_i^2+\cdots)\right)s_\alpha \\
&= \left(\prod_{i=1}^{n}(1+t_i+t_i^2+\cdots)\right)\sum_{\alpha\models n}\left(\prod_{i\in D(\alpha)}t_i\right)s_\alpha
\end{aligned}$$

The Hilbert series of the polynomial algebra $\mathbb{C}[\theta_{n-k+1},\ldots,\theta_n]$ is

$$\mathrm{Hilb}(\mathbb{C}[\theta_{n-k+1},\ldots,\theta_n];t_1,\ldots,t_n) = \prod_{i=n-k+1}^{n}(1+t_i+t_i^2+\cdots). \tag{3.11}$$

Since $\mathbb{C}[\mathcal{B}_n^*]$ is a free module over $\mathbb{C}[\theta_1,\ldots,\theta_n]$ (Corollary 3.3.7) and the action of $\mathfrak{S}_n$ on $\mathbb{C}[\mathcal{B}_n^*]$ is linear over $\mathbb{C}[\theta_1,\ldots,\theta_n]$, we can rewrite $\mathrm{grFrob}(\mathbb{C}[\mathcal{B}_n^*];t_1,\ldots,t_n)$ as

$$\mathrm{Hilb}(\mathbb{C}[\theta_{n-k+1},\ldots,\theta_n];t_1,\ldots,t_n)\cdot\mathrm{grFrob}(\mathbb{C}[\mathcal{B}_n^*]/\langle\theta_{n-k+1},\ldots,\theta_n\rangle;t_1,\ldots,t_n),$$

hence we have

$$\mathrm{grFrob}(\mathbb{C}[\mathcal{B}_n^*]/\langle\theta_{n-k+1},\ldots,\theta_n\rangle;t_1,\ldots,t_n) = \left(\prod_{i=1}^{n-k}(1+t_i+t_i^2+\cdots)\right)\sum_{\alpha\models n}\left(\prod_{i\in D(\alpha)}t_i\right)s_\alpha.$$

Modulo all length k multichains, which results in retaining everything in degree less than k and removing everything in degree k and above, we have

$$\mathrm{grFrob}(\mathcal{S}_{n,k};t_1,\ldots,t_n) = \sum_{\substack{\alpha\models n\\ \ell(\alpha)\le k}}\left(\prod_{i\in D(\alpha)}t_i\right)\left(\sum_{j_1+\cdots+j_{n-k}\le k-\ell(\alpha)}\prod_{i=1}^{n-k}t_i^{j_i}\right)s_\alpha.$$

We can interpret each $(j_1, \ldots, j_{n-k})$ as a partition fitting inside a $(n-k)\times(k-\ell(\alpha))$ box, by letting j_i be the number of rows of length j_i (and by letting $(j_1, \ldots, j_{n-k})$ run we obtain all such partitions). Now, the size of the corresponding partition is equal to $j_1 + 2j_2 + \ldots + (n-k)j_{n-k}$ so setting $t_i = q^i$ yields

$$\sum_{\alpha \models n} q^{\mathrm{maj}(\alpha)} \sum_{\lambda \in (n-k)\times(k-\ell(\alpha))} q^{|\lambda|} s_\alpha = \sum_{\alpha \models n} q^{\mathrm{maj}(\alpha)} \binom{n-\ell(\alpha)}{k-\ell(\alpha)}_q s_\alpha$$

using the fact that $\sum_{\lambda \in a\times b} q^{|\lambda|} = \binom{a+b}{b}_q$. Note that this is indeed the expression for $\mathrm{grFrob}(S_{n,k}; q)$. □

Remark. The Frobenius character map has an analogue for G_n as well [CR20, Section 2.4]. The proofs in Section 3.4 show that $\mathcal{R}_{n,k}$ and $\mathcal{S}_{n,k}$ are a refined version of the graded G_n-modules $R_{n,k}$ and $S_{n,k}$, in the sense that

$$\mathrm{grFrob}(\mathcal{R}_{n,k}; q, q^2, \ldots, q^n) = \mathrm{grFrob}(R_{n,k}; q);$$

$$\mathrm{grFrob}(\mathcal{S}_{n,k}; q, q^2, \ldots, q^n) = \mathrm{grFrob}(S_{n,k}; q),$$

of which we just explicitly handled the case $(\mathcal{S}_{n,k}, S_{n,k})$ for $r = 1$. By finding the graded Frobenius image of $\mathbb{C}[\mathcal{B}_n^*]$ as a G_n-module and factoring out $\prod_{i=1}^n (1 + x_i^r + x_i^{2r} + \ldots)$ one can obtain a similar result for general r. ◀

This chapter contains material from: D. Kroes, *"Generalized coinvariant algebras for $G(r,1,n)$ in the Stanley-Reisner setting"*, Electronic Journal of Combinatorics, vol. 26 (3), P.3.11, 2018. The dissertation author was the primary investigator and author of this paper.

Chapter 4

A quotient ring for packed words

4.1 Packed words

In this chapter we consider a quotient ring of $\mathbb{Q}[\mathbf{x}_n]$ whose combinatorics is controlled by the set of *packed words*. A word $w = w_1 \dots w_n$ over the alphabet $\{1, 2, \dots\}$ of positive integers is packed if, for all $i > 0$, whenever $i + 1$ appears as a letter in w, so does i. Let $\mathcal{W}_n$ be the family of packed words of length n. For example, we have

$$\mathcal{W}_3 = \{123, 213, 132, 231, 312, 321, 112, 121, 211, 122, 212, 221, 111\}.$$

By interpreting w_i as the index of the block the number i belongs to, packed words in $\mathcal{W}_n$ are in natural bijection with the family $\mathcal{OP}_n$ of all ordered set partitions of $[n]$.

We will consider the following ideal inside $\mathbb{Q}[\mathbf{x}_n]$, where we use the notation $e_d^{(i)} := e_d(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ for the degree d elementary symmetric polynomial with the variable x_i omitted.

Definition 4.1.1. Let $J_n \subseteq \mathbb{Q}[\mathbf{x}_n]$ be the ideal

$$J_n = \langle x_i^d \cdot e_{n-r}^{(i)} \ : \ 1 \leq i \leq n, 1 \leq r \leq d \rangle \tag{4.1}$$

and let $S_n := \mathbb{Q}[\mathbf{x}_n]/J_n$ be the corresponding quotient ring. ◀

By convention, the degree 0 elementary symmetric polynomial is 1, so that J_n contains the variable powers x_i^n. Additionally, we use the convention that $e_d \equiv 0$ for $d < 0$.

Although the generators of the ideal J_n may appear unusual, they will arise naturally from the perspective of orbit harmonics as follows. More precisely, suppose $X \subseteq \mathbb{Q}^n$ is a finite locus of points. Consider the ideal

$$\mathbf{I}(X) := \{f \in \mathbb{Q}[\mathbf{x}_n] \ : \ f(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in X\} \tag{4.2}$$

of polynomials in $\mathbb{Q}[\mathbf{x}_n]$ which vanish on X and let

$$\mathbf{T}(X) := \langle \tau(f) \ : \ f \in \mathbf{I}(X) - \{0\} \rangle, \tag{4.3}$$

where $\tau(f)$ denotes the highest degree component of a nonzero polynomial $f \in \mathbb{Q}[\mathbf{x}_n]$. The homogeneous ideal $\mathbf{T}(X)$ is the *associated graded* ideal of $\mathbf{I}(X)$ and we have isomorphisms of $\mathbb{Q}$-vector spaces

$$\mathbb{Q}[X] \cong \mathbb{Q}[\mathbf{x}_n]/\mathbf{I}(X) \cong \mathbb{Q}[\mathbf{x}_n]/\mathbf{T}(X) \tag{4.4}$$

which are isomorphisms of ungraded $\mathfrak{S}_n$-modules when X is closed under the natural action of $\mathfrak{S}_n$ on $\mathbb{Q}^n$; the quotient $\mathbb{Q}[\mathbf{x}_n]/\mathbf{T}(X)$ has the additional structure of a graded $\mathfrak{S}_n$-module.

Given n distinct rational parameters $\alpha_1, \ldots, \alpha_n$, we have a natural point locus $X_n \subseteq \mathbb{Q}^n$ in bijection with $\mathcal{W}_n$, namely

$$X_n = \{(\beta_1, \ldots, \beta_n) \in \mathbb{Q}^n \,:\, \{\beta_1, \ldots, \beta_n\} = \{\alpha_1, \ldots, \alpha_k\} \text{ for some } k\}. \tag{4.5}$$

It will develop that

$$\mathbf{T}(X_n) = J_n. \tag{4.6}$$

In other words, the quotient $S_n = \mathbb{Q}[\mathbf{x}_n]/\mathbf{T}(X_n)$ is the graded quotient of $\mathbb{Q}[\mathbf{x}_n]$ arising from the packed word locus X_n. Equation (4.6) may be viewed as a more natural, but less computationally useful, alternative to Definition 4.1.1. We prove the following facts regarding the module S_n.

- The ungraded $\mathfrak{S}_n$-structure of S_n coincides with the natural $\mathfrak{S}_n$-action on $\mathcal{W}_n$ (without sign twist)

$$S_n \cong \mathbb{Q}[\mathcal{W}_n]. \tag{4.7}$$

- The graded $\mathfrak{S}_n$-structure is described by

$$\mathrm{grFrob}(S_n; q) = \sum_{k=1}^{n} q^{n-k} \cdot (\mathrm{rev}_q \circ \omega) C_{n,k}(\mathbf{x}; q). \tag{4.8}$$

  Here $C_{n,k}$ is as in Equation (1.2) and rev_q is the operator on polynomials in q which reverses their coefficient sequences and ω is the symmetric function involution which trades $e_n(\mathbf{x})$ for $h_n(\mathbf{x})$.

4.2 A combinatorial bijection

We first establish a bijection between ordered set partitions and coinversion codes. The starting point will be a bijection established by Rhoades and Wilson [RW19, Thm. 2.2]. Given an ordered set partition $\sigma = (B_1 \mid \cdots \mid B_k) \in \mathcal{OP}_{n,k}$, define a sequence $\mathbf{code}(\sigma) = (c_1, \ldots, c_n)$ as follows. If $1 \leq i \leq n$ and $i \in B_j$, then

$$c_i = \begin{cases} |\{\ell > j \ : \ \min(B_\ell) > i\}| & \text{if } i = \min(B_j); \\ \\ |\{\ell > j \ : \ \min(B_\ell) > i\}| + (j-1) & \text{if } i \neq \min(B_j). \end{cases} \tag{4.9}$$

The sequence $\mathbf{code}(\sigma)$ was called the *coinversion code* of σ in [RW19] and is a variant of the classical *Lehmer code* on permutations in the case $k = n$.

The coinversion $\mathbf{code}(\sigma)$ of ordered set partitions $\sigma \in \mathcal{OP}_{n,k}$ were characterized in [RW19] as follows. Given a subset $S = \{s_1 < \cdots < s_d\} \subseteq [n]$, define the *skip sequence* by $\gamma(S) = (\gamma_1, \ldots, \gamma_n)$ where

$$\gamma_i = \begin{cases} i - j + 1 & \text{if } i = s_j \in S \\ \\ 0 & \text{if } i \notin S. \end{cases} \tag{4.10}$$

Also let $\gamma(S)^* = (\gamma_n, \ldots, \gamma_1)$ be the *reverse skip sequence*. For example, if $n = 7$ and $S = \{2, 3, 6\}$ we have $\gamma(S) = (0, 2, 2, 0, 0, 4, 0)$ and $\gamma(S)^* = (0, 4, 0, 0, 2, 2, 0)$.

Theorem 4.2.1. *([RW19, Thm. 2.2]) Let $1 \leq k \leq n$. The map $\sigma \mapsto \mathbf{code}(\sigma)$ is a bijection from ordered set partitions of $[n]$ with k blocks to the family of nonnegative integer sequences $(c_1, \ldots, c_n)$ such that*

- *for all* $1 \leq i \leq n$ *we have* $c_i < k$,

- *for any subset* $S \subseteq [n]$ *with* $|S| = n - k + 1$, *the componentwise inequality* $\gamma(S)^* \leq (c_1, \ldots, c_n)$ *fails to hold.*

For future reference we recall the inverse map introduced in the proof of the above theorem. This inverse map uses the following insertion procedure.

For $(B_1 \mid \cdots \mid B_k)$ a sequence of k (possibly empty) sets of positive integers we define the *coinversion labels* as follows. First, label the empty sets $0, 1, \ldots, j$ from right to left, and then label the nonempty sets $j+1, \ldots, j+k-1$ from left to right.

For a sequence $(c_1, \ldots, c_n)$ satisfying the conditions in Theorem 4.2.1, we construct an ordered set partition as follows. Start with a sequence $(\emptyset \mid \cdots \mid \emptyset)$ of k copies of the empty set, and for $i = 1, 2, \ldots, n$ insert the number i in the block with label c_i under the coinversion labeling.

For example, let $n = 7$, $k = 4$ and consider the sequence $c = (2, 1, 2, 0, 2, 0, 2)$. The resulting ordered set partition will be $(6 \mid 13 \mid 257 \mid 4)$, as shown by the following process, starting with the labeled sequence of blocks $(\emptyset^3 \mid \emptyset^2 \mid \emptyset^1 \mid \emptyset^0)$.

Table 4.1: Construction of the ordered set partition in $\mathcal{OP}_{7,4}$ with coinversion code $(2, 1, 2, 0, 2, 0, 2)$.

i	c_i	updated labeled sequence of blocks
1	2	$(\emptyset^2 \mid 1^3 \mid \emptyset^1 \mid \emptyset^0)$
2	1	$(\emptyset^1 \mid 1^2 \mid 2^3 \mid \emptyset^0)$
3	2	$(\emptyset^1 \mid 13^2 \mid 2^3 \mid \emptyset^0)$
4	0	$(\emptyset^0 \mid 13^1 \mid 2^2 \mid 4^3)$
5	2	$(\emptyset^0 \mid 13^1 \mid 25^2 \mid 4^3)$
6	0	$(6^0 \mid 13^1 \mid 25^2 \mid 4^3)$
7	2	$(6^0 \mid 13^1 \mid 257^2 \mid 4^3)$

In our algebraic analysis of S_n we will need a version of this insertion which maps the family of ordered set partitions of $[n]$ with *at least* k blocks bijectively onto a certain collection $(c_1, \ldots, c_n)$ of length n 'code words' over the nonnegative integers. In the bijection **code** of Theorem 4.2.1, the ordered set partition $(1|2| \cdots |m, m+1, \ldots, n)$ has code $(0, 0, \ldots, 0)$ for any number of blocks m, so we cannot simply take the union of these maps for $m \geq k$.

We resolve the problem in the above paragraph by working with a different version of the coinversion code, which we will call the *boosted coinversion code*. For an ordered set partition $\sigma = (B_1 \mid \cdots \mid B_k)$ we define $\overline{\mathbf{code}}(\sigma) = (c_1, \ldots, c_n)$ as follows. Suppose

$1 \leq i \leq n$ and $i \in B_j$, then

$$c_i = \begin{cases} |\{\ell > j \ : \ \min(B_\ell) > i\}| & \text{if } i = \min(B_j); \\ \\ |\{\ell > j \ : \ \min(B_\ell) > i\}| + j & \text{if } i \neq \min(B_j). \end{cases} \tag{4.11}$$

Compared to the coinversion codes from before, the difference is that all the numbers corresponding to non-minimal elements of blocks are raised by one, and we say that these numbers are *boosted*.

The remainder of the section will be devoted to the proof of the following theorem.

Theorem 4.2.2. *Let $1 \leq k \leq n$. The map $\sigma \mapsto \overline{\mathbf{code}}(\sigma)$ is a bijection from the set of ordered set partitions of $[n]$ with at least k blocks to the family of nonnegative integer sequences such that*

- *for all $1 \leq i \leq n$ we have $c_i < n$.*
- *for any subset $S \subseteq [n]$ with $|S| = n - k + 1$ the componentwise inequality $\gamma(S)^* \leq (c_1, \ldots, c_n)$ fails to hold.*
- *for any $1 \leq i, d \leq n$ and any $T \subseteq [n-1]$ with $|T| = n-d$ and $\gamma(T)^* = (\gamma_{n-1}, \ldots, \gamma_1)$, the componentwise inequality $(\gamma_{n-1}, \ldots, \gamma_i, d, \gamma_{i-1}, \ldots, \gamma_1) \leq (c_1, \ldots, c_n)$ fails to hold.*

The proof of the necessity of these conditions will be similar to that of the proof of [RW19, Thm.2.2]. For the sufficiency of the conditions we use an insertion map similar to that considered above. We begin by showing that both the number of blocks of an ordered set partition of $[n]$, as well as its classical coinversion code, can be recovered from its boosted coinversion code.

Lemma 4.2.3. *Let σ be an ordered set partition of $[n]$. Given the boosted coinversion code $\overline{\mathbf{code}}(\sigma)$ one can recover the coinversion code $\mathbf{code}(\sigma)$, as well as the number of blocks of σ.*

Proof. Note that the second part is immediate once we have recovered $\mathbf{code}(\sigma)$, as the number of blocks will be equal to the number of unboosted numbers, which is easily found by comparing $\overline{\mathbf{code}}(\sigma)$ and $\mathbf{code}(\sigma)$.

Given a boosted coinversion code $(c_1, \ldots, c_n)$ corresponding to an ordered set partition with ℓ blocks (where ℓ is unknown), we can think of creating the ordered set partition by following the same procedure as described before, with the only difference that the labels of all the nonempty blocks should be raised by one.

No matter what, at some point we will fill in the last nonempty block with some number i, which necessarily has $c_i = 0$. Additionally, from the boosting, it is clear that $c_j > 0$ for all $j > i$, hence we can recover i by looking for the last entry in our sequence that equals 0.

Now, assume that we have identified that $i_1 < \ldots < i_j$ are minimal in their block and that all other numbers in $[i_1, n]$ are not minimal in their block. If $i_1 = 1$ we are done. Otherwise, it is clear that we must have at least $j + 1$ blocks (as clearly 1 will be minimal in its block). Now, let $i_0 < i_1$ be the largest number that is also minimal in its block. By the inverse map, this must correspond to some index with $c_{i_0} \leq j$, as at the time of inserting i_0 there are exactly $j + 1$ empty blocks, labeled $0, 1, \ldots, j$. Additionally, for any $i_0 < i < i_1$, at the time of insertion there will be exactly j empty blocks, hence the coinversion label of i will be at least $j+1$ (because of the boosting). Therefore, given $\overline{\mathbf{code}}(\sigma)$ we can recognize i_0 as the largest index $i_0 < i_1$ with $c_{i_0} \leq j$. By induction we are done. □

Explicitly, the procedure above is as follows. Given a sequence $(c_1, \ldots, c_n)$, trace the sequence from right to left, marking the first 0, then the first 0 or 1, then the first 0, 1 or 2, etcetera. Now, decrease all the unmarked numbers by 1 and one recovers the coinversion code. We call this procedure the *unboosting* of a sequence $(c_1, \ldots, c_n)$.

As an example, consider the boosted coinversion code $c = (2, 4, 2, 4, 0, 0, 1, 4)$. Working from right to left we mark c_6 as it is the first 0, then c_5 as it is at most 1, then c_3 as it is the next number at most 2 and finally c_1 as it is the next number that is at most 3. Therefore, the number of blocks is equal to 4 and the unboosted coinversion code is given by $(2, 3, 2, 3, 0, 0, 0, 3)$. Applying the earlier bijection this coinversion code corresponds to $(37 \mid 124 \mid 6 \mid 58)$.

We are now ready to prove the main result of this section.

Proof of Theorem 4.2.2. We first prove the necessity of the conditions. Let σ be an ordered set partition of $[n]$ with at least k blocks and let $\overline{\mathbf{code}}(\sigma) = (c_1, \ldots, c_n)$ be its boosted coinversion code.

- If i is minimal in its block, c_i will be at most the number of blocks following the block containing i, which is at most $n-1$. If i is not minimal we have at most $n-1$

blocks, and if $i \in B_j$ we have

$$c_i = j + |\{\ell > j \ : \ \min(B_\ell) > i\}| \leq j + |\{\ell > j \ : \ \text{the } \ell^{\text{th}} \text{ block exists}\}| \leq n - 1.$$

- Suppose $S = \{n+1-t_{n-k+1}, \ldots, n+1-t_1\}$ (with $t_1 < \ldots < t_{n-k+1}$) satisfies $\gamma(S)^* \leq (c_1, \ldots, c_n)$. We show that none of the numbers $\{t_1, \ldots, t_{n-k+1}\}$ is minimal in its block of σ, contradicting that σ has at least k blocks.

  If t_{n-k+1} is minimal in its block, then

$$\begin{aligned} c_{t_{n-k+1}} &= |\{\ell > t_{n-k+1} : {\scriptstyle \ell \text{ is minimal in its block and} \atop \scriptstyle \text{occurs to the right of } t_{n-k+1} \text{ in } \sigma}\}| \\ &\leq |\{t_{n-k+1}+1, \ldots, n-1, n\}| = n - t_{n-k+1}. \end{aligned}$$

  However, the term in $\gamma(S)^*$ in position t_{n-k+1} equals $n - t_{n-k+1} + 1$, hence we conclude that t_{n-k+1} is not minimal in its block.

  Now, if t_{n-k} were minimal in its block, we would have

$$\begin{aligned} c_{t_{n-k}} &= |\{\ell > t_{n-k} : \ell \text{ is minimal in its block and occurs to the right of } t_{n-k} \text{ in } \sigma\}| \\ &\leq |\{t_{n-k}+1, \ldots, n-1, n\} - \{t_{n-k+1}\}| = n - t_{n-k} - 1. \end{aligned}$$

  But again, the term in $\gamma(S)^*$ in position t_{n-k} equals $n - t_{n-k}$, which shows that t_{n-k} cannot be minimal in its block either. An inductive argument now shows that none of $\{t_1, \ldots, t_{n-k+1}\}$ is minimal in its block.

- For $d = n$ this is equivalent to the fact that $c_i < n$ for all i, so assume $1 \leq d < n$. Assume for contradiction that $(\gamma_{n-1}, \ldots, \gamma_i, d, \gamma_{i-1}, \ldots, \gamma_1) \leq (c_1, \ldots, c_n)$ where

$(\gamma_{n-1}, \ldots, \gamma_1) = \gamma(T)^*$ for some $T \subseteq [n-1]$ of size $|T| = n-d$. Since $c_{n+1-i} \geq d$, this implies that σ has at least d blocks. Let $T = \{i_1 < \ldots < i_t \leq n+1-i < i_{t+1} < \ldots < i_{t+s}\}$. By the same argument used in the previous bullet, we see that all $n+1-i_j$ with $j \leq t$ are not minimal in their block. Now we consider two cases.

- If $n+1-i$ is not minimal in its block either, we can continue the argument as in the previous case to show that none of $n+1-i_j$ is minimal in its block. In particular we have $1+(n-d)$ elements that are not minimal in their respective blocks, contradicting the fact that σ has at least d blocks.

- Now suppose that $n+1-i$ is minimal in its block. Since $c_{n+1-i} = d$, this implies that among $\{n+2-i, \ldots, n\}$ at least d numbers are also minimal in their respective blocks. In particular, there are at least d numbers that are not of the form $n+1-j$ with $j \in T$. But this implies that T has size at most $(n-1)-d < n-d$, which is a contradiction.

Now, we show that these conditions are sufficient. Given a sequence $(c_1, \ldots, c_n)$ we can first unboost the sequence (as we can apply this procedure to every sequence of nonnegative integers) to determine how many blocks our intended ordered set partition must have. Given this extra information, we can basically run the same inverse map as before, with the exception that we should increase the label of every nonempty block by 1. It now suffices to check that we don't run into any troubles by doing so. Our proof will go through the following steps.

- First we will show that the unboosting procedure concludes that there are at most $n-k$ boosted numbers, as this will ensure that the ordered set partition we aim for has at least k blocks.
- Then we will inductively show that can basically run the same inverse map as before.
  - First we show that the conclusion of the unboosting is that 1 is unboosted, ensuring we have enough blocks to insert 1 as a minimal element in its block.
  - After that we will show that if the first $j-1$ numbers have been placed, we can place j following the appropriate procedure. This argument will depend on whether j is supposed to be minimal in its block or not (something that we know from the unboosting procedure).

We will now prove each of these steps.

- Assume that we have t boosted numbers c_{n+1-i_j} (with $i_1 < \ldots < i_t$) and assume that $t \geq n-k+1$. Let $S = \{i_1, \ldots, i_{n-k+1}\}$, then we claim that $(c_1, \ldots, c_n) \geq \gamma(S)^*$. If $i \notin S$, we have $\gamma(S){*}_{n+1-i} = 0$, so $c_{n+1-i} \geq \gamma(S)^*_{n+1-i}$ indeed holds. Furthermore, for $i = i_j$ by assumption on c_{n+1-i_j} there are $(i_j - j)$ unboosted numbers to the right of $n+1-i_j$. Therefore, since c_{n+1-i_j} was boosted, we have $c_{n+1-i_j} \geq i_j - j + 1 = \gamma(S)^*_{n+1-i_j}$, as desired.
- As mentioned before, we now show that we can run the inverse map without any issues.

– If $c_1 = 0$ it is clear that we can insert 1, so assume $c_1 = d$ with $1 \leq d \leq n-1$. Our goal is to show that in the unboosting procedure we conclude that 1 has to be minimal in its block. As $c_1 = d$ this happens precisely if the procedure shows that among $\{2, 3, \ldots, n\}$ at least d numbers were not boosted. For the sake of contradiction, assume that we have at least $n-d$ boosted numbers, and let the largest $n-d$ be $n+1-i_1 > n+1-i_2 > \ldots > n+1-i_{n-d}$. Let $T = \{i_1, \ldots, i_{n-d}\}$ then by a similar argument to before we have $(c_1, c_2, \ldots, c_n) \geq (d, \gamma_{n-1}, \ldots, \gamma_1)$ where $(\gamma_{n-1}, \ldots, \gamma_1) = \gamma(T)^*$.

– Assume that the inverse map successfully inserted all the numbers in $[j-1]$ (with $j \geq 2$) and that we now try to insert j according to c_j.

  First assume that $c_j = t$ is unboosted. Since this is unboosted, there are still at least t unboosted numbers among $\{c_{j+1}, \ldots, c_n\}$. As so far only indices corresponding to unboosted numbers have been inserted in empty blocks, and the number of total blocks it the number of unboosted numbers, we have at least $t+1$ empty blocks at this point. As a result, there will be some empty block labeled with t, so we can insert j into an empty block, as desired.

  Hence, assume that c_j was boosted. Suppose that at the time we still have t nonempty blocks, then by the unboosting procedure we know that $c_j \geq t+1$, so we can insert j appropriately, unless c_j is too big. In other words, the only thing that can go wrong is that there were ℓ unboosted numbers (hence ℓ blocks in the ordered set partition), but that $c_j \geq \ell + 1$. Let $n+1-i_1 > \ldots >$

$n+1-i_a > j > n-i_{a+1} > \ldots > n-i_{n-\ell-1}$ be all the boosted numbers. But then, for $T = \{i_1, \ldots, i_{n-\ell-1}\}$ of size $n-(\ell+1)$, with $\gamma(T)^* = (\gamma_{n-1}, \ldots, \gamma_1)$, we have $(c_1, \ldots, c_n) \geq (\gamma_{n-1}, \ldots, \gamma_{n-j+1}, \ell+1, \gamma_{n-j}, \ldots, \gamma_1)$, a contradiction. □

4.3 The algebraic quotient

Recall that a word $w = w_1 w_2 \cdots w_n$ on the alphabet $\mathbb{Z}_{>0}$ is *packed* if whenever $i+1$ appears, then so does i. It will be convenient for our inductive arguments to consider packed words in which every letter in some segment $1 \leq i \leq k$ must appear. To this end, we define

$$\mathcal{W}_{n,k} := \{\text{length } n \text{ packed words } w = w_1 w_2 \ldots w_n \ : \ \text{the letters } 1, 2, \ldots, k \text{ appear in } w\}. \tag{4.12}$$

Words in $\mathcal{W}_{n,k}$ are in bijection with ordered set partitions of $[n]$ with at least k blocks. We have the further identifications $\mathcal{W}_{n,1} = \mathcal{W}_n$ and $\mathcal{W}_{n,n} = \mathfrak{S}_n$.

The symmetric group $\mathfrak{S}_n$ acts on $\mathcal{W}_{n,k}$ by the rule $\sigma \cdot (w_1 \ldots w_n) := w_{\sigma(1)} \ldots w_{\sigma(n)}$. The quotient rings $S_{n,k}$ of the following definition will give a graded refinement of this action. Their defining ideals $J_{n,k}$ contain the ideal J_n defining the ring S_n appearing in the introduction.

Definition 4.3.1. Let $J_{n,k} \subseteq \mathbb{Q}[\mathbf{x}_n]$ be the ideal

$$J_{n,k} := J_n + \langle e_n, e_{n-1}, \ldots, e_{n-k+1} \rangle \tag{4.13}$$

and let $S_{n,k} := \mathbb{Q}[\mathbf{x}_n]/J_{n,k}$ be the corresponding quotient ring. ◀

Each of the quotients $S_{n,k}$ is a graded $\mathfrak{S}_n$-module. Their defining ideals are nested according to $J_n = J_{n,1} \subseteq J_{n,2} \subseteq \cdots \subseteq J_{n,n}$. Note that $e_n \in J_n$ will follow from the equality $J_n = \mathbf{T}(X_n)$, using the fact e_n is the top degree component of $(x_1 - \alpha_1) \cdots (x_n - \alpha_1) \in \mathbf{I}(X_n)$. We study $S_{n,k}$ by making use of a point locus $X_{n,k} \subseteq \mathbb{Q}^n$ corresponding to $\mathcal{W}_{n,k}$. Fix n distinct rational numbers $\alpha_1, \ldots, \alpha_n \in \mathbb{Q}$. For any packed word $w_1 \ldots w_n \in \mathcal{W}_{n,k}$, we have a corresponding point $(\alpha_{w_1}, \ldots, \alpha_{w_n}) \in \mathbb{Q}^n$. We let $X_{n,k} \subseteq \mathbb{Q}^n$ be the family of points corresponding to all packed words in $\mathcal{W}_{n,k}$.

The set $X_{n,k} \subseteq \mathbb{Q}^n$ is closed under the coordinate-permuting action of $\mathfrak{S}_n$ and we have an identification $\mathbb{Q}[\mathcal{W}_{n,k}] \cong \mathbb{Q}[X_{n,k}]$. Recall that we have isomorphisms of ungraded $\mathfrak{S}_n$-modules

$$\mathbb{Q}[\mathcal{W}_{n,k}] \cong \mathbb{Q}[X_{n,k}] \cong \mathbb{Q}[\mathbf{x}_n]/\mathbf{I}(X_{n,k}) \cong \mathbb{Q}[\mathbf{x}_n]/\mathbf{T}(X_{n,k}).$$

It turns out that $\mathbf{T}(X_{n,k})$ coincides with $J_{n,k}$.

Theorem 4.3.2. *For any $1 \leq k \leq n$, we have the ideal equality $J_{n,k} = \mathbf{T}(X_{n,k})$. Consequently, we have an isomorphism of ungraded $\mathfrak{S}_n$-modules $\mathbb{Q}[\mathcal{W}_{n,k}] \cong S_{n,k}$.*

Proof. To show that $J_{n,k} \subseteq \mathbf{T}(X_{n,k})$, it suffices to show that every generator of $J_{n,k}$ arises as the highest degree component of some polynomial in $\mathbf{I}(X_{n,k})$. Fix $1 \leq i \leq n$ and $1 \leq r \leq d$; we begin by showing that the generator $x_i^d e_{n-r}^{(i)}$ lies in $\mathbf{T}(X_{n,k})$.

Note that if $(x_1, \ldots, x_n) \in X_{n,k}$, we either have $x_i \in \{\alpha_1, \ldots, \alpha_d\}$, or for any $1 \leq j \leq d$ the number α_d must appear among $\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$. We let t be a new variable, and define the function

$$f(x_1,\ldots,x_n,t) := (x_i-\alpha_1)\cdots(x_i-\alpha_d)\cdot\frac{(1-tx_1)\cdots(1-tx_{i-1})(1-tx_{i+1})\cdots(1-tx_n)}{(1-t\alpha_1)\cdots(1-t\alpha_d)} \tag{4.14}$$

and expanding this function in terms of the parameter t yields

$$f(x_1,\ldots,x_n,t) = \\ (x_i-\alpha_1)\cdots(x_i-\alpha_d)\cdot\sum_{r\geq 0}\left(\sum_{a+b=r}(-1)^a e_a^{(i)}\cdot h_b(\alpha_1,\ldots,\alpha_d)\right)t^r$$

Specialization of $f(x_1,\ldots,x_n,t)$ at $(x_1,\ldots,x_n)=(\beta_1,\ldots,\beta_n)$ yields an element of $\mathbb{Q}[[t]]$. We analyze this specialization when $(\beta_1,\ldots,\beta_n)\in X_{n,k}$. If $\beta_i\in\{\alpha_1,\ldots,\alpha_d\}$, then $f(\beta_1,\ldots,\beta_n,t)=0$. Otherwise, d of the terms in the numerator of f will cancel with the d terms in the denominator, so that hence $f(\beta_1,\ldots,\beta_n,t)$ is a polynomial of degree $(n-1)-d$ in t. Either way, the coefficient of t^{n-r} in $f(x_1,\ldots,x_n,t)$ vanishes on $X_{n,k}$, so that

$$(x_i-\alpha_1)\cdots(x_i-\alpha_d)\cdot\left(\sum_{a+b=n-r}(-1)^a e_a^{(i)}\cdot h_b(\alpha_1,\ldots,\alpha_d)\right)\in\mathbf{I}(X_{n,k})$$

and taking the highest degree component gives

$$x_i^d\cdot(-1)^{n-r}e_{n-r}^{(i)}\in\mathbf{T}(X_{n,k}).$$

The remaining generators e_d (for $d>n-k$) are handled by a similar argument. We consider the rational function

$$\begin{aligned} g(x_1,\ldots,x_n,t) &:= \frac{(1-tx_1)(1-tx_2)\cdots(1-tx_n)}{(1-t\alpha_1)(1-t\alpha_2)\cdots(1-t\alpha_k)} \\ &= \sum_{r\geq 0}\left(\sum_{a+b=r}(-1)^a e_a\cdot h_b(\alpha_1,\ldots,\alpha_k)\right)\cdot t^r. \end{aligned}$$

Evaluating $(x_1, \ldots, x_n)$ at a point in $X_{n,k}$ forces the k factors in the denominator to cancel with k factors in the numerator, yielding a polynomial of degree $n-k$ in t. For any $d > n-k$, we conclude that

$$\sum_{a+b=d} (-1)^a e_a \cdot h_b(\alpha_1, \ldots, \alpha_k) \in \mathbf{I}(X_{n,k}),$$

which implies

$$e_d \in \mathbf{T}(X_{n,k}).$$

This proves the containment $J_{n,k} \subseteq \mathbf{T}(X_{n,k})$, so that

$$\dim \mathbb{Q}[\mathbf{x}_n]/J_{n,k} \geq \dim \mathbb{Q}[\mathbf{x}_n]/\mathbf{T}(X_{n,k}) = |\mathcal{W}_{n,k}| \tag{4.15}$$

In light of Equation (4.15), to prove the desired equality $J_{n,k} = \mathbf{T}(X_{n,k})$ it is enough to show that $\dim(\mathbb{Q}[\mathbf{x}_n]/J_{n,k}) \leq |\mathcal{W}_{n,k}|$. This is a Gröbner theory argument.

Since the elementary symmetric polynomials $e_n, e_{n-1}, \ldots, e_{n-k+1}$ in the full variable set $\{x_1, \ldots, x_n\}$ lie in $J_{n,k}$, [HRS18, Lem. 3.4] implies that for any subset $S \subseteq [n]$ with $|S| = n-k+1$, the *Demazure character* $\kappa_{\gamma(S)}$ corresponding to the length n sequence $\gamma(S)$ also lies in $J_{n,k}$. The lexicographical leading monomial of $\kappa_{\gamma(S)}$ has exponent sequence $\gamma(S)^*$. Similarly, for $1 \leq i, d \leq n$, since

$$x_i^d \cdot e_{n-d}^{(i)}, \ldots, x_i^d \cdot e_{n-1}^{(i)} \in J_{n,k},$$

for any $T \subseteq [n-1]$ of size $|T| = n-d$, [HRS18, Lem. 3.4] again implies that

$$x_i^d \cdot \kappa_{\gamma(T)}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in J_{n,k}.$$

Writing $\gamma(T)^* = (\gamma_{n-1}, \ldots, \gamma_1)$, the exponent sequence of the lexicographical leading term of $x_i^d \cdot \kappa_{\gamma(T)}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ is $(\gamma_{n-1}, \ldots, \gamma_i, d, \gamma_{i-1}, \ldots, \gamma_1)$. It follows that

> the exponent sequence $(c_1, \ldots, c_n)$ of any member of the standard monomial basis of $\mathbb{Q}[\mathbf{x}_n]/J_{n,k}$ satisfies the conditions in the statement of Theorem 4.2.2.

Theorem 4.2.2 implies the desired dimension bound $\dim \mathbb{Q}[\mathbf{x}_n]/J_{n,k} \leq |\mathcal{W}_{n,k}|$, completing the proof. □

The standard monomial basis of $S_{n,k}$ is governed by coinversion codes.

Corollary 4.3.3. *The standard monomial basis of $S_{n,k}$ with respect to the lexicographical term ordering are the monomials $x_1^{c_1} \cdots x_n^{c_n}$ where $(c_1, \ldots, c_n) = \overline{\mathbf{code}}(\sigma)$ is the boosted coinversion code of some ordered set partition σ of $[n]$ with at least k blocks.*

Proof. This follows from Theorem 4.2.2 and the last paragraph of the above proof. □

Our next goal is to derive the graded $\mathfrak{S}_n$-module structure of the quotients $S_{n,k}$. This result is stated most cleanly in terms of the generalized coinvariant algebra's from Definiton 2.4.1.

Theorem 4.3.4. *As graded $\mathfrak{S}_n$-module we have*

$$S_{n,k} \cong R_{n,n}\langle 0 \rangle \oplus R_{n,n-1}\langle -1 \rangle \oplus \cdots \oplus R_{n,k}\langle -n+k \rangle.$$

Proof. We proceed by descending induction on k. In the case $n = k$, we claim that $J_{n,n} = I_{n,n} = \langle e_1, \ldots, e_n \rangle$ is the classical invariant ideal so that $S_{n,n} = R_{n,n}$. Indeed, each elementary symmetric polynomial e_d appears as a generator of $J_{n,n}$. On the other hand, Theorem 4.3.2 implies that $\dim S_{n,n} = n! = \dim R_{n,n}$. This finishes the proof in the case $k = n$.

Now suppose $1 \leq k \leq n-1$. We exhibit a short exact sequence of $\mathfrak{S}_n$-modules

$$0 \to R_{n,k} \xrightarrow{\varphi} S_{n,k} \xrightarrow{\pi} S_{n,k+1} \to 0, \tag{4.16}$$

where φ is homogeneous of degree $n-k$ and π is homogeneous of degree 0. The exactness of this sequence implies

$$S_{n,k} \cong S_{n,k+1} \oplus R_{n,k}\langle -n+k \rangle,$$

proving the theorem by induction.

Since every generator of $J_{n,k+1}$ is also a generator of $J_{n,k}$, we may take $\pi : S_{n,k} \twoheadrightarrow S_{n,k+1}$ to be the canonical projection. We have a map

$$\widetilde{\varphi} : \mathbb{Q}[\mathbf{x}_n] \to S_{n,k} \tag{4.17}$$

given by multiplication by e_{n-k} followed by projection onto $S_{n,k}$. We verify that $\widetilde{\varphi}$ descends to a map $\varphi : R_{n,k} \to S_{n,k}$ by showing that $\widetilde{\varphi}$ sends every generator of $I_{n,k}$ to zero. Indeed, we have $\widetilde{\varphi}(e_j(x_1, \ldots, x_n)) = 0$ for any $j > n-k$ since $e_j(x_1, \ldots, x_n)$ is a generator of $J_{n,k}$. Furthermore, for $1 \leq i \leq n$ we have

$$\widetilde{\varphi}(x_i^k) = x_i^k e_{n-k} = x_i^k e_{n-k}^{(i)} + x_i^{k+1} e_{n-k-1}^{(i)} = 0,$$

where the final equality follows because both $x_i^k e_{n-k}^{(i)}$ and $x_i^{k+1} e_{n-k-1}^{(i)}$ are generators of $J_{n,k}$. We conclude that $\widetilde{\varphi}$ descends to a map $\varphi : R_{n,k} \to S_{n,k}$ of $\mathfrak{S}_n$-modules which is homogeneous of degree $n-k$. It is clear that φ surjects onto the kernel of π. The exactness of the sequence (4.16) follows from the dimensional equality

$$\dim(S_{n,k}) = |\mathcal{W}_{n,k}| = |\mathcal{W}_{n,k+1}| + |\mathcal{OP}_{n,k}| = \dim(S_{n,k+1}) + \dim(R_{n,k}). \qquad \square$$

The graded Frobenius image of $S_{n,k}$ is most naturally stated in terms of the C-functions defined in Equation (1.2).

Corollary 4.3.5. *For any $1 \leq k \leq n$, the graded Frobenius image of $S_{n,k}$ is given by*

$$\mathrm{grFrob}(S_{n,k}; q) = \sum_{j=k}^{n} q^{n-j} \cdot (\omega \circ \mathrm{rev}_q) C_{n,j}(\mathbf{x}; q). \tag{4.18}$$

Proof. Apply [HRS18, Thm. 6.11] and Theorem 4.3.4. $\square$

This chapter contains material from: D. Kroes and B. Rhoades, *"Packed words and quotient rings"*, submitted (2021). The dissertation author was one of the primary investigators and authors of this paper.

Chapter 5

Catalan-pair graphs

5.1 Definitions and statement of results

We start by introducing the main object of study of this chapter.

Definition 5.1.1. Let n be a positive integer. A *Catalan-pair graph* on n vertices is a graph G that can be obtained by the following procedure. Start with $2n$ collinear points, of which we color $2k$ points red for some $0 \leq k \leq n$ and color the remaining points blue. Then, choose Catalan-arc matchings of sizes k and $n-k$ and place them on the red and blue points, respectively, with the latter being faced downwards rather than upwards. Finally, construct a graph G with one vertex for each of the n arcs, where two vertices are adjacent if and only if the endpoints of the corresponding arcs alternate. ◀

For example, below is a Catalan-pair graph on 9 vertices, where we colored the arcs according to the color of the points they connect. We say that the pair of Catalan-arc matchings on the left is a *representative* for the graph on the right, or alternatively that it *represents* the graph on the right.

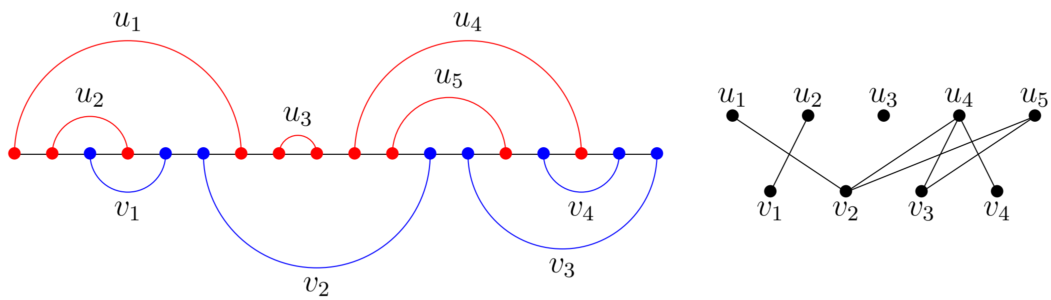

Figure 5.1: A Catalan-Pair graph on 9 vertices.

As a first observation, note that all of the arcs on the top are chosen to be non-intersecting, and similarly for all of the arcs on the bottom. Therefore, if the endpoints of two arcs alternate (and hence correspond to an edge in G) these arcs necessarily come from different sides. Thus every Catalan-pair graph is bipartite.

The main purpose of this section is to introduce a model to randomly generate a Catalan-pair graph on n vertices, which we denote by CP_n, and to establish various properties about this random graph. Before we precisely define our random graph model, we briefly summarize our main results.

Theorem 5.1.2. *The expected number of edges of the random Catalan-pair graph CP_n satisfies*

$$\mathbb{E}[e(CP_n)] \sim \frac{1}{\pi} n \log n. \tag{5.1}$$

Moreover, for any $\epsilon > 0$ we asymptotically almost surely have $|e(CP_n) - \frac{1}{\pi} n \log n| <$

$\epsilon n \log n$.

We also obtain an asymptotic formula for the expected number of isolated vertices in CP_n.

Theorem 5.1.3. *Let I_n denote the number of isolated vertices in CP_n. Then*

$$\mathbb{E}[I_n] \sim \gamma n, \tag{5.2}$$

where γ is the constant defined by

$$\gamma = 4 \sum_{m=1}^{\infty} 16^{-m} \sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b} C_b = 0.3023\ldots.$$

Moreover, for any $\epsilon > 0$ we asymptotically almost surely have $|I_n - \gamma n| < \epsilon n$.

In addition to this, we deduce the order of magnitude for the expected number of (induced) subgraphs of any connected Catalan-pair graph with at least three vertices. To this end, Let $N_H(G)$ denote the number of subgraphs of G that are isomorphic to H and let $N_H^*(G)$ denote the number of induced subgraphs of G that are isomorphic to H.

Theorem 5.1.4. *Let H be a connected Catalan-pair graph on $v \geq 3$ vertices. The expected number of (induced) subgraphs of the random Catalan-pair graph CP_n isomorphic to H satisfies*

$$\mathbb{E}[N_H(CP_n)] = \Theta(n^{v/2}).$$

$$\mathbb{E}[N_H^*(CP_n)] = \Theta(n^{v/2}).$$

Notation. Let a Catalan-pair graph on n vertices be given. For $1 \leq a < b \leq 2n$, we say that (a, b) *match* if the a^{th} and b^{th} point have the same color and if there is an arc connecting these two points. In the earlier example, the matching pairs are $(1, 7)$, $(2, 4)$, $(3, 5)$, $(6, 12)$, $(8, 9)$, $(10, 16)$, $(11, 14)$, $(13, 18)$ and $(15, 17)$. We similarly say that (a, b) match in a single Catalan-arc matching of size n if there is an arc connecting these two points. For $1 \leq a < b < c < d \leq 2n$ we say that (a, b, c, d) is *an edge* if (a, c) and (b, d) match. For example, in the the graph from before $(6, 10, 12, 16)$ is an edge, and it corresponds to the edge between u_4 and v_2. We say that an arc in a single Catalan-pair matching has length k if it covers $k - 1$ smaller arcs, or equivalently if the two points it connects have $2k - 2$ points between them. ◀

5.2 Random Catalan-pair graphs

Let us define a model to generate a random Catalan-pair graphs on n vertices. Consider the following procedure, starting with $2n$ collinear points.

1. For each of the first $2n - 1$ points, uniformly and independently color each of these points either red or blue. Then color the last point red or blue, whichever makes it so that the total number of points of each color is even.

2. Suppose that we have $2k$ red points, and consequently $2(n - k)$ blue points. Independently and uniformly pick Catalan-arc matchings of size k and $n - k$ from the set of all possible Catalan-arc matchings of that size, and place these above and

below the red and blue points respectively.

3. Create a graph according to Definition 5.1.1, and denote this (random) graph by CP_n.

One of the advantages of this model is that due to Lemma 2.7.1 with high probability roughly half of the points (or any large enough subset of the points for that matter) will be colored red. Note that because of the forced choice of the color of the last point, our setting is not completely identical to that of the above result. However, it does apply for any proper subset of the points, and the concentration result for the total number of points of a given color is almost unaffected.

5.3 Random Catalan matchings

To generate CP_n we must choose a random Catalan-arc matching from all such matchings of a given size. In this section we compute the probability of having a given set of arcs connecting a given set of points within this randomly chosen Catalan-arc matching. We note that studying the structure of a random object enumerated by the Catalan numbers is of independent interest, and other work in this direction has been done in, for example, [DFH+99] and [FS09].

Let $\mathcal{C}_n$ denote the set of Catalan-arc matchings of size n. We recall the asymptotic formula

$$C_n \sim \frac{4^n}{\sqrt{\pi} n^{3/2}}, \tag{5.3}$$

which can be derived, for example, by Stirling's formula.

Throughout this section, let C be a Catalan-arc matching chosen uniformly from $\mathcal{C}_n$. As mentioned, we are interested in the probability of having a given set of arcs connecting a given set of points within C. It is clear that in order for this to be able to happen, the points and arcs have to satisfy some conditions. First of all the endpoints of any given arc must have an even number of points between them, since any arc connecting at least one of these points must connect two of these points. Additionally, it is clear that none of the given arcs are allowed to intersect.

This leads to the following definition, where one should think of having specified arcs connecting points x_i and $x_i + 2k_i - 1$ for all i.

Definition 5.3.1. Let $\mathbf{x} = (x_1, \ldots, x_s)$ and $\mathbf{k} = (k_1, \ldots, k_s)$ be s-tuples of positive integers with $x_1 < \ldots < x_s$. We say that $(\mathbf{x}, \mathbf{k})$ is a *valid pair* if

1. For all i we have $1 \leq x_i < x_i + 2k_i - 1 \leq 2n$.

2. The integers $x_1, x_1 + 2k_1 - 1, \ldots, x_s, x_s + 2k_s - 1$ are all distinct.

3. There are no $i \neq j$ with $x_i < x_j < x_i + 2k_i - 1 < x_j + 2k_j - 1$. ◀

As an example, for $n = 8$, we have the valid pair $((2, 4), (5, 2))$ which we think of as having specified arcs connecting points 2 and 11 and 4 and 7.

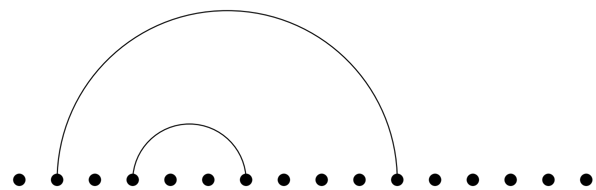

Figure 5.2: Specified arcs in a Catalan-arc matching for $n = 8$.

As mentioned before, the conditions imposed on $(\mathbf{x}, \mathbf{k})$ are necessary for there to be a Catalan-arc matching with arcs on these specified positions. In this case, it is not so hard to see that we can indeed extend this to a Catalan-arc matching, for example as follows.

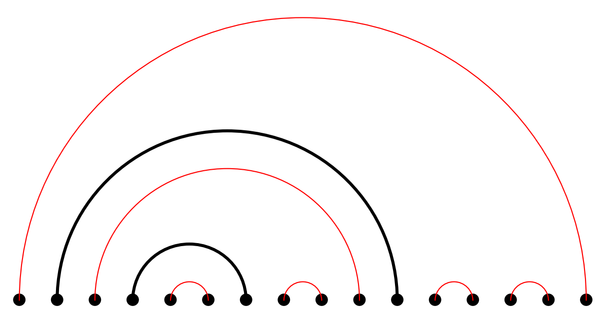

Figure 5.3: A completed Catalan-arc matching with the specified arcs.

Below we will see that the condition of $(\mathbf{x}, \mathbf{k})$ being a valid pair is also a sufficient condition to have a Catalan-arc matching with arcs connecting x_i and $x_i + 2k_i - 1$. In fact, we will determine the explicit probability of having arcs on these given positions. To this end, let $A(\mathbf{x}, \mathbf{k})$ denote the event that $(x_i, x_i + 2k_i - 1)$ match in C for all i.

Before we can determine the probability of this happening we need some notation. In the above example, we see that in order to extend to a Catalan-arc matching, we have to connect the two points within the smaller arc, we have to connect the four points within the larger arc (but outside of the smaller arc), and finally we have to connect the six points outside of the larger arc. Below we define integers that are analogues of the two, four, and six above.

For a valid pair $(\mathbf{x}, \mathbf{k})$ and $1 \leq i \leq s$, let M_i be the set of x such that $x_i < x < x_i + 2k_i - 1$ and such that there exists no $j \neq i$ with $x_j \leq x \leq x_j + 2k_j - 1$. We let M_0 be the set of x such that $1 \leq x \leq 2n$ and such that there exists no i with $x_i \leq x \leq x_i + 2k_i - 1$. Observe that every x with $1 \leq x \leq 2n$ is either of the form x_i or $x_i + 2k_i - 1$ for some i, or else belongs to a unique M_i. Furthermore, it is easy to see that each M_i has an even (possibly 0) number of elements, so the numbers $m_i = |M_i|/2$ are nonnegative integers, and from the definition it follows that these numbers sum to $n - s$.

We can now explicitly compute the probability that $(x_i, x_i + 2k_i - 1)$ match in C for all i.

Lemma 5.3.2. *If* $(\mathbf{x}, \mathbf{k})$ *is a valid pair, then*

$$\mathbb{P}[A(\mathbf{x}, \mathbf{k})] = \frac{1}{C_n} \cdot \prod_{i=0}^{s} C_{m_i}. \tag{5.4}$$

Proof. Since each Catalan-arc matching is chosen with probability $\frac{1}{C_n}$, it suffices to show that there are $\prod_{i=0}^{s} C_{m_i}$ Catalan-arc matchings for which $(x_i, x_i + 2k_i - 1)$ match for all i. We show that a Catalan-arc matching satisfies this condition if and only if points in some M_i are only connected to points in that M_i and the set of arcs on the points in M_i is a Catalan-arc matching.

First, assume for contradiction that there exists $i \neq j$ such that there is a Catalan-arc matching that connects a point x in M_i to a point y in M_j. Without loss of generality we may assume that $j \neq 0$ and that we do not have $x_j < x_i < x_i + 2k_i - 1 < x_j + 2k_j - 1$ (if the latter happens, simply switch i and j). This implies that $x_j < y < x_j + 2k_j - 1$ and $x \notin [x_j, x_j + 2k_j - 1]$, but then the arc connecting x and y would intersect the arc connecting x_j and $x_j + 2k_j - 1$, a contradiction. Furthermore, it is clear that the induced set of arcs on the points in M_i still has no intersecting arcs.

Conversely, suppose we choose Catalan-arc matchings to go on the points of each M_i. By definition, there do not exist points $a < b < c < d$ with $a, c \in M_i$ and $b, d \in M_j$ for $i \neq j$, so arcs in M_i and M_j will not intersect when $i \neq j$, and clearly also not for $i = j$. Lastly, points in M_i either lie completely inside an interval $[x_j, x_j + 2k_j - 1]$ or lie completely outside of it, so arcs on the M_i will also not intersect arcs of the form $(x_j, x_j + 2k_j - 1)$.

Therefore, since a Catalan-arc matching on the points of M_i has m_i arcs, there are C_{m_i} choices for this matching. Since these choices can be made independently, the total number of desired Catalan-arc matchings equals $\prod_{i=0}^{s} C_{m_i}$, as desired. □

By combining (5.3) and Lemma 5.3.2 we can obtain bounds for this probability.

Corollary 5.3.3. *Let* $(\mathbf{x}, \mathbf{k})$ *be a valid pair. There exist positive real numbers* α_s, β_s *such that*

$$\alpha_s \frac{n^{3/2}}{\prod' m_i^{3/2}} \leq \mathbb{P}[A(\mathbf{x}, \mathbf{k})] \leq \beta_s \frac{n^{3/2}}{\prod' m_i^{3/2}}, \tag{5.5}$$

where $\prod'$ *indicates the product over all* $0 \le i \le s$ *with* $m_i \neq 0$.

Proof. Let us prove the lower bound, the proof for the upper bound is analogous. Because of the asymptotic formula in (5.3) there exist positive numbers $a < 1 < A$ such that

$$a\frac{4^n}{\sqrt{\pi}n^{3/2}} \le C_n \le A\frac{4^n}{\sqrt{\pi}n^{3/2}} \tag{5.6}$$

for all $n \ge 1$. Since $C_0 = 1$ we find

$$\begin{aligned}\mathbb{P}[A(\mathbf{x},\mathbf{k})] &= \frac{1}{C_n}\cdot\prod_{i=0}^{s} C_{m_i} = \frac{1}{C_n}\cdot{\prod}' C_{m_i} \\ &\ge \frac{\sqrt{\pi}n^{3/2}}{A\cdot 4^n}\cdot{\prod}'\frac{a\cdot 4^{m_i}}{\sqrt{\pi}m_i^{3/2}} \ge 4^{\sum' m_i - n}\cdot\frac{a^{s+1}}{A\cdot\pi^{s/2}}\frac{n^{3/2}}{\prod' m_i^{3/2}} = \alpha_s\frac{n^{3/2}}{\prod' m_i^{3/2}},\end{aligned}$$

where we use that $\sum' m_i = \sum_{i=0}^{s} m_i = n - s$. □

5.4 The expected number of edges

We will determine the asymptotic behavior of the expected number of edges of CP_n. To this end, we start by establishing a general upper bound on the probability that CP_n contains a given structure on a given set of points.

We consider two analogues of the valid pairs introduced in Section 5.3.

Definition 5.4.1. Let $\mathbf{x} = (x_1, \ldots, x_s)$, $\mathbf{k} = (k_1, \ldots, k_s)$, $\mathbf{y} = (y_1, \ldots, y_t)$, $\mathbf{l} = (\ell_1, \ldots, \ell_t)$ be tuples of positive integers with $x_1 < \ldots < x_s$ and $y_1 < \ldots < y_t$.

We say that this quadruple is *valid* if for all $1 \le i \le s$ and $1 \le j \le t$ we have $1 \le x_i < x_i + k_i \le 2n$ and $1 \le y_j < y_j + \ell_j \le 2n$, and if there exists at least one representative for a Catalan-pair graph on n vertices for which (x_i, x_i+k_i) and $(y_j, y_j+\ell_j)$ match for all i, j.

Similarly, we say that such a quadruple $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ is *good* if

1. $1 \leq x_i < x_i + k_i \leq 2n$ and $1 \leq y_j < y_j + \ell_j \leq 2n$ for all $1 \leq i \leq s$ and $1 \leq j \leq t$.

2. Any two numbers of the form $x_i, x_i + k_i, y_j$ or $y_j + k_j$ differ by at least 2.

3. There exists no $i \neq j$ such that $x_i < x_j < x_i + k_i < x_j + k_j$ or $y_i < y_j < y_i + \ell_i < y_j + \ell_j$. ◀

In the proof of Lemma 5.7.1 we will see that the conditions for a good quadruple imply that there exists a representative for a Catalan-pair graph G such that $(x_i, x_i + k_i)$ and $(y_i, y_i + \ell_i)$ match for all $1 \leq i \leq s$ and all $1 \leq j \leq t$. Therefore, any good quadruple is also a valid quadruple.

Given a valid quadruple $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$, we would like to have an analogue of the integers m_i defined in Section 5.3. To this end, for $1 \leq i \leq s$, set f_i to be the number of $x_i < x < x_i + k_i$ such that there is no i' with $x_{i'} \leq x \leq x_{i'} + k_{i'}$ and such that x is not of the form y_j or $y_j + \ell_j$ for any j. Set f_0 to be the number of $1 \leq x \leq 2n$ that do not belong to any interval $[x_i, x_i + k_i]$, nor are of the form y_j or $y_j + \ell_j$. Similarly define $g_0, g_1, \ldots, g_t$.

Let $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ be a valid quadruple where $\mathbf{x}$ and $\mathbf{y}$ have length s and t respectively. Let $A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ denote the intersection of the following events.

1. The points x_i and $x_i + k_i$ are colored red and the points y_j and $y_j + \ell_j$ are colored blue for all i, j.

2. For all i and j the number of red points x with $x_i < x < x_i + k_i$ and the number of blue points y with $y_j < y < y_j + \ell_j$ is even.

3. For all i and j we have that $(x_i, x_i + k_i)$ and $(y_j, y_j + \ell_j)$ match in CP_n.

We would like to point out that the second condition is necessary for $(x_i, x_i + k_i)$ and $(y_j, y_j + \ell_j)$ to match for all i and j. Therefore, we could technically omit this condition, but we have included it to improve the readability of our proofs.

We have the following upper bound for the probability that $A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ occurs.

Lemma 5.4.2. *There exists a positive real number $\beta_{s,t}$ such that for sufficiently large n, and for any valid quadruple $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ (with $\mathbf{x}$ and $\mathbf{y}$ of length s and t respectively) we have*

$$\mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] \leq \beta_{s,t} n^3 \cdot \widetilde{\prod_i} f_i^{-3/2} \cdot \widetilde{\prod_j} g_j^{-3/2}, \tag{5.7}$$

where $\widetilde{\prod}$ indicates the product over all i and j for which $f_i, g_j \geq 16(s+t)\log n$.

Proof. Let $v = (s+t)$. Note that with probability $2^{-2(s+t)} = 4^{-v}$ all of $x_i, x_i + k_i, y_j, y_j + \ell_j$ have the correct color. From now on we condition on this event happening. For each $0 \leq i \leq s$, let $2r_i$ denote the number of points counted by f_i which are colored red, where we note that r_i may not be an integer. For each i with $f_i \geq 16v \log n$, we use Lemma 2.7.1 to conclude that

$$\mathbb{P}[|2r_i - f_i/2| > \sqrt{v f_i \log n}] < 2n^{-2v}.$$

Note that if $|2r_i - f_i/2| \leq \sqrt{v f_i \log n}$, then in particular we have $2r_i \geq f_i/2 - \sqrt{v f_i \log n} \geq f_i/4$, where we used $f_i \geq 16v \log n$ in the last step. Therefore, with probability at most

$2(v+2)n^{-2v}$ we have $r_i < f_i/8$ or $b_j < g_j/8$ for some i or j for which $f_i, g_j \geq 16v\log n$.

Let B_n and R_n be the total number of blue and red points respectively. We condition on the event that $r_i \geq f_i/4$ and $b_j \geq g_j/4$ for all i and j for which $f_i, g_j \geq 16v\log n$. If any of the numbers r_i, b_j is not an integer, or equivalently if the number of red/blue points in some appropriate region is not even, the probability that all of (x_i, x_i+k_i) and $(y_j, y_j+\ell_j)$ match is 0, which definitely agrees with the proposed upper bound. If all the r_i and b_j are integers we can apply Corollary 5.3.3 to show that the probability that all of the (x_i, x_i+k_i) and $(y_j, y_j+\ell_j)$ match is at most

$$\begin{aligned}
&\beta_s R_n^{3/2} \cdot \widetilde{\prod_i} r_i^{-3/2} \cdot \beta_t B_n^{3/2} \cdot \widetilde{\prod_j} b_j^{-3/2} \\
&\leq \beta_s \cdot (2n)^{3/2} \cdot \widetilde{\prod_i} (f_i/8)^{-3/2} \cdot \beta_t (2n)^{3/2} \cdot \widetilde{\prod_j} (g_j/8)^{-3/2} \\
&= O\left(n^3 \cdot \widetilde{\prod_i} f_i^{-3/2} \cdot \widetilde{\prod_j} g_j^{-3/2}\right),
\end{aligned}$$

where in the first expression we ignored all i for which $f_i < 16v\log n$ since in the formula of Corollary 5.3.3 these terms either do not appear, or they contribute a multiplicative factor of the form $x^{-3/2}$ for some $x \geq 1$, hence leaving it out will still yield an upper bound.

Therefore, we know that

$$\mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] \leq 4^{-v} \cdot \left(2(v+2)n^{-2v} + O\left(n^3 \cdot \widetilde{\prod_i} f_i^{-3/2} \cdot \widetilde{\prod_j} g_j^{-3/2}\right)\right).$$

Since $f_i \leq 2n$, $g_i \leq 2n$, and since the above products contain at most $s+1$ and $t+1$

terms respectively, we find

$$n^3 \cdot \widetilde{\prod_i} f_i^{-3/2} \cdot \widetilde{\prod_j} g_j^{-3/2} \geq n^3(2n)^{-3/2(s+1+t+1)} = 2^{-3/2(v+2)} \cdot n^{3-3/2(v+2)} \gg n^{-2v}$$

for sufficiently large n, and hence the $n^3 \cdot \widetilde{\prod}_i f_i^{-3/2} \cdot \widetilde{\prod}_j g_j^{-3/2}$ term dominates this expression.

□

Using similar ideas, we can deduce an upper bound on the expected number of arcs in CP_n whose lengths lie in a specific range.

Lemma 5.4.3. *For any $1 \leq \alpha \leq \beta \leq 2n$, let $A_{\alpha,\beta}$ denote the number of matching arcs in CP_n of the form $(i, i+k)$ with $\alpha \leq k \leq \beta$. Then*

$$\mathbb{E}[A_{\alpha,\beta}] = O(\alpha^{-1/2}n + \beta n e^{-\alpha/16}). \tag{5.8}$$

In particular, if $32 \log n \leq \alpha$ we have

$$\mathbb{E}[A_{\alpha,\beta}] = O(\alpha^{-1/2}n). \tag{5.9}$$

Proof. We first consider some reductions of the problem. If $\alpha = O(1)$ the bound is trivial, so we will assume that $\alpha = \omega(1)$. For any $\alpha \geq n$ the proposed bound is $O(\sqrt{n})$, so in this range it suffices to prove the result for $\alpha = n$ and thus we can assume without loss of generality that $\alpha \leq n$. Also, for any $k \geq 2n - 32 \log n$, CP_n contains at most two non-intersecting arcs of length k (one for each color) since $k > n$. Thus we can assume that $\beta \leq 2n - 32 \log n$, which will cause $\mathbb{E}[A_{\alpha,\beta}]$ to decrease by at most $2 \cdot 32 \log n = O(\alpha^{-1/2}n)$ when $\alpha \leq n$.

For $\alpha \le k \le \beta$, let $A(i,k)$ denote the event that $(i, i+k)$ matches in CP_n. Let $2r_1$ denote the number of points x in $i < x < i+k$ colored red and let $2r_2$ denote the number of points x with $x < i$ or $x > i+k$ colored red, where as before we note that r_1 or r_2 may not be an integer. The probability that either $|2r_1 - (k-1)/2| > (k-1)/4$ or $|2r_2 - (2n-k-1)/2| > (2n-k-1)/4$ is at most $e^{-(k-1)/8} + e^{-(2n-k-1)/8}$. Conditional on neither of these events occurring, we can proceed as in Lemma 5.4.2 and find that the probability of $(i, i+k)$ matching is at most $cn^{3/2}(k-1)^{-3/2}(2n-(k+1))^{-3/2}$ for some absolute constant c. Combining all this we see that

$$\mathbb{P}[A(i,k)] \le cn^{3/2}(k-1)^{-3/2}(2n-(k+1))^{-3/2} + e^{-(k-1)/8} + e^{-(2n-k-1)/8}.$$

Moreover, we have that $\mathbb{P}[A(i,k)] = 0$ for $i > 2n-k$. Because

$$E[A_{\alpha,\beta}] = \sum_{k=\alpha}^{\beta} \sum_{i=1}^{2n} \mathbb{P}[A(i,k)],$$

we have that

$$\mathbb{E}[A_{\alpha,\beta}] \le \sum_{k=\alpha}^{\beta} (2n-k)(cn^{3/2}(k-1)^{-3/2}(2n-(k+1))^{-3/2} + e^{-(k-1)/8} + e^{-(2n-k-1)/8}). \quad (5.10)$$

Let $\gamma = \min(\beta, n)$. For $\alpha \le k \le \gamma$ and n sufficiently large, we have that $(2n-(k+1)) \ge \frac{1}{2}n$ and $(k-1) \ge \frac{1}{2}k$. Thus the terms in (5.10) are at most

$$2n(2^{-3}ck^{-3/2} + 2e^{-\alpha/16} + 2e^{-n/16}) \le 2^{-2}cnk^{-3/2} + 4ne^{-\alpha/16}.$$

Thus (5.10) restricted to this range is at most

$$\sum_{k=\alpha}^{\gamma} 2^{-2}cnk^{-3/2} + 4ne^{-\alpha/16} \le 2^{-2}n \int_{\alpha-1}^{\infty} cx^{-3/2}dx + 4\gamma ne^{-\alpha/16} = O(\alpha^{-1/2}n + \beta ne^{-\alpha/16}).$$

If $\beta \leq n$ then this completes the proof. Otherwise we can assume $\beta = 2n - 32\log n$. Using similar logic as before, for $n \leq k \leq 2n - 32\log n$ we have that the terms of (5.10) are at most

$$2^{-2}c(2n-k)^{-1/2} + 4ne^{-2\log n} = 2^{-2}c(2n-k)^{-1/2} + 4n^{-1}.$$

Again summing over the relevant range and bounding our sum with an integral gives an upper bound for (5.10) in this range of

$$\sum_{k=n}^{2n-32\log n} (2n-k)^{-1/2} + 4n^{-1} = O(\sqrt{n}) = O(\alpha^{-1/2}n).$$

Summing the contributions from these ranges gives the desired result. □

5.4.1 The expected number of edges

We are now ready to prove the first part of Theorem 5.1.2. We will do so by showing that for any $\epsilon > 0$ we have

$$(1-\epsilon)\frac{1}{\pi}n\log n + o(n\log n) \leq \mathbb{E}[e(CP_n)] \leq (1+\epsilon)\frac{1}{\pi}n\log n + o(n\log n). \tag{5.11}$$

It is clear that

$$\mathbb{E}[e(CP_n)] = \sum \mathbb{P}[A(x,k,y,\ell)], \tag{5.12}$$

where the sum is over all valid quadruples (x,k,y,ℓ) of positive integers such that $1 \leq x < y < x+k < y+\ell \leq 2n$ or $1 \leq y < x < y+\ell < x+k \leq 2n$.

We break up this sum into various parts, and we will show that all but one will contribute $o(n\log n)$, and that the remaining part will contribute between $(1-\epsilon)\frac{1}{\pi}n\log n$

and $(1+\epsilon)\frac{1}{\pi}n\log n$. Let $c < 1$ be a positive real number and d be a positive integer, where eventually we will pick c small and d large to get our bounds within the desired $(1 \pm \epsilon)$ region.

Proposition 5.4.4. *Consider the contribution to* (5.12) *coming from each of the following subsets of the quadruples.*

(i) Valid quadruples (x, k, y, ℓ) *with* $k < d\log n$ *or* $\ell < d\log n$.

(ii) Valid quadruples (x, k, y, ℓ) *with* $k > 2n - d\log n$ *or* $\ell > 2n - d\log n$.

(iii) Quadruples (x, k, y, ℓ) *with* $d\log n \leq k, \ell \leq 2n - d\log n$ *that are valid but not good.*

(iv) Good quadruples (x, k, y, ℓ) *with* $d\log n \leq k \leq cn < \ell \leq 2n - d\log n$ *or* $d\log n \leq \ell \leq cn < k \leq 2n - d\log n$.

(v) Good quadruples (x, k, y, ℓ) *with* $cn < k, \ell \leq 2n - d\log n$.

Each of these contributions is $o(n\log n)$.

Proof. (i) This contribution counts the expected number of edges that come from pairs of arcs with at least one arc of length at most $d\log n$. We first show that the number of such edges with at least one arc of length at most $\sqrt{\log n}$ is of order $o(n\log n)$ in any Catalan-pair graph, and therefore also in expectation. Indeed, any arc of length at most $\sqrt{\log n}$ has degree at most $\sqrt{\log n}$ since every interlacing arc must have one of its endpoints within the given arc. Since we have at most n arcs of length at most $\sqrt{\log n}$, the total number of such edges is at most $n\sqrt{\log n} = o(n\log n)$.

Now consider the edges involving an arc of length between $\sqrt{\log n}$ and $d \log n$. By Lemma 5.4.3 there are at most $O(n(\log n)^{-1/4} + \log n \cdot ne^{-\sqrt{\log n}/16}) = o(n)$ such arcs in expectation. Since each such arc can be involved in at most $d \log n$ edges, we conclude that the total expected number of edges involving vertices of this type is at most $o(n \log n)$.

(ii) This contribution counts the expected number of edges that come from a pair of arcs where at least one of the arcs has length larger than $2n - d \log n$. We show that the number of such arcs is $O((\log n)^2) = o(n \log n)$ for any Catalan-pair graph, which implies the same bound for the expected number of such edges. First, note that for n large enough and each $N > 2n - d \log n$ there is at most one arc of length N on either side. Indeed, since $2n - d \log n > n$ for n large enough, if we had two arcs of length N on one side this would contradict the condition that the arcs do not intersect. Therefore, there are at most $2d \log n$ arcs of length at least $2n - d \log n$. Furthermore, each such arc interlaces with at most $d \log n$ arcs on the opposite side. Indeed, any such interlacing arc must have one of its endpoints outside the arc in question, and there are at most $d \log n$ such points. Therefore, we have at most $2d \log n \cdot d \log n = O((\log n)^2)$ such edges, as desired.

(iii) We assume $d > 32$ in order to apply Lemma 5.4.2.

We know that for any (x, k, y, ℓ) in this range we have

$$\mathbb{P}[A(x, k, y, \ell)] = O(n^3(2n - (k + 2))^{-3/2}k^{-3/2}(2n - (\ell + 2)^{-3/2}\ell^{-3/2}).$$

Furthermore, given k and ℓ we claim that there are at most $16n$ quadruples (x, k, y, ℓ) that are valid but not good. This follows since there are at most $2n$ possibilities for x, and given x we must have that y or $y + \ell$ belongs to $\{x \pm 1, x + k \pm 1\}$.

Therefore, the total contribution is at most of the order of

$$\begin{aligned} &n^4 \sum_{k,\ell} (2n - (k+2))^{-3/2} k^{-3/2} (2n - (\ell+2))^{-3/2} \ell^{-3/2} \\ &= n^4 \left(\sum_k (2n - (k+2))^{-3/2} k^{-3/2} \right)^2 . \end{aligned}$$

We can break up $\sum_k (2n - (k+2))^{-3/2} k^{-3/2}$ in the regions $k \leq n$ and $k > n$. When $k \leq n$ we have $(2n - (k+2))^{-3/2} \leq (n-2)^{-3/2}$, hence the contribution is at most $(n-2)^{-3/2} \sum_k k^{-3/2} = O(n^{-3/2})$ since the sum of $k^{-3/2}$ is bounded. By similar reasoning the other contribution is $O(n^{-3/2})$, so

$$n^4 \left(\sum_k (2n - (k+2))^{-3/2} k^{-3/2} \right)^2 = n^4 O(n^{-3/2})^2 = O(n) = o(n \log n),$$

as was to be shown.

(iv) Again we assume $d > 32$. Also, we only consider the case $d \log n \leq k \leq cn < \ell \leq 2n - d \log n$, the other case being analogous.

We claim that for given k and ℓ there are at most $(2n - \ell) \cdot 2k$ good quadruples (x, k, y, ℓ). This holds since y has to satisfy $y + \ell \leq 2n$, and after choosing y we must have that $y - k \leq x \leq y - 1$ or $y + \ell - k \leq x \leq y + \ell - 1$, leaving at most $2k$ choices for x. Therefore, this region contributes at most

$$\sum_{k = d \log n}^{cn} \sum_{\ell = cn}^{2n - d \log n} (2n - \ell) \cdot 2k \cdot n^3 (2n - (k+2))^{-3/2} k^{-3/2} (2n - (\ell+2))^{-3/2} \ell^{-3/2}.$$

Note that this sum breaks up as

$$2n^3 \left(\sum_{k=d\log n}^{cn} k^{-1/2}(2n-(k+2))^{-3/2} \right)$$
$$\cdot \left(\sum_{\ell=cn}^{2n-d\log n} \ell^{-3/2}(2n-\ell)(2n-(\ell+2))^{-3/2} \right).$$

Using $(2n-(k+2))^{-3/2} \leq 2^{3/2}n^{-3/2}$ we find that

$$\sum_{k=d\log n}^{cn} k^{-1/2}(2n-(k+2))^{-3/2} = O(n^{-3/2}) \cdot \sum_{d\log n+2}^{cn} k^{-1/2}$$
$$= O(n^{-3/2}) \cdot O(n^{1/2}) = O(n^{-1})$$

where the second equality follows from comparison of the sum with an integral. An analogous computation shows that

$$\sum_{\ell=cn}^{2n-d\log n} \ell^{-3/2}(2n-\ell) \cdot (2n-(\ell+2))^{-3/2} = O(n^{-1}),$$

and therefore this range of k and ℓ contributes at most $2n^3 \cdot O(n^{-1}) \cdot O(n^{-1}) = O(n)$, which is in particular $o(n\log n)$ as desired.

(v) Again we estimate the number of good quadruples (x,k,y,ℓ) for given k, ℓ. Similar to above we have at most $(2n-k)$ and $(2n-\ell)$ choices for x and y respectively, and therefore we have at most $(2n-k)(2n-\ell)$ good quadruples in total. Thus this part of the sum contributes at most

$$\sum_{k,\ell=cn}^{2n-d\log n} (2n-k)(2n-\ell) \cdot n^3(2n-(k+2))^{-3/2}k^{-3/2}(2n-(\ell+2))^{-3/2}\ell^{-3/2}.$$

As in case (iv), this factors as

$$n^3 \left(\sum_{k=cn}^{2n-d\log n} k^{-3/2}(2n-k)(2n-(k+2))^{-3/2} \right)$$
$$\cdot \left(\sum_{\ell=cn}^{2n-d\log n} \ell^{-3/2}(2n-\ell)\cdot(2n-(\ell+2))^{-3/2} \right).$$

Each of the above sums will be $O(n^{-1})$ by the same argument as before. We conclude that the total contribution of these terms to the original sum is at most $n^3 \cdot O(n^{-1}) \cdot O(n^{-1}) = O(n) = o(n \log n)$, completing the proof. □

We point out that using Lemma 5.4.2 and similar arguments to the ones used in cases (iv) and (v) can be used to show that the region $d \log n \leq k, \ell \leq cn$ will contribute $O(n \log n)$ to the expected number of edges. In fact, using a later result, Lemma 5.7.1, we can also show a lower bound of $\Omega(n \log n)$ for this contribution. However, with a little bit more care it is possible to determine the exact constant. We first require a probability lemma.

Lemma. *Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{P}[X_i = 0] = \mathbb{P}[X_i = 1] = 1/2$, and set $S_j = \sum_{i=1}^{j} X_i$. For $\epsilon > 0$, $d \geq 20/\epsilon^2$ and $j > d \log n$ we have*

$$\mathbb{P}[|S_j - j/2| < \epsilon j/2] < 2n^{-10}. \tag{5.13}$$

Proof. By Lemma 2.7.1, the desired probability is at most

$$2\exp(-2(\epsilon j/2)^2/j) = 2\exp(-\epsilon^2 j/2) \leq 2\exp(-\epsilon^2 d \log n/2) = 2n^{-\epsilon^2 d/2} \leq 2n^{-10}$$

since $d \geq 20/\epsilon^2$. □

By Proposition 5.4.4, in order to show (5.11) it suffices to prove that for suitably small c and sufficiently large d the contribution from good quadruples with $d \log n \leq k, \ell \leq cn$ is between

$$(1-\epsilon)\frac{1}{\pi} n \log n \quad \text{and} \quad (1+\epsilon)\frac{1}{\pi} n \log n.$$

To this end we introduce the following notation, which intuitively means that two expression asymptotically gets arbitrarily close for $n \to \infty$, *independent of all other variables*, provided one picks a suitably small c and a suitably large d.

Definition 5.4.5. Let f and g be two functions with the same domain taking positive values, and whose inputs depend on some positive integer n and some other integer variables, some of which are restricted to the interval $[d \log n, cn]$. We say that $f \sim_{\mathrm{ac}} g$ if for any $\epsilon > 0$ there exist suitable c, d and N with

$$(1-\epsilon) f(x) \leq g(x) \leq (1+\epsilon) f(x) \tag{5.14}$$

for any input x with $n \geq N$. ◀

Here the subscript ac denotes that we do not have the exact asymptotic behavior, but that we get *arbitrary close* asymptotic behavior by choosing suitable c and d.

We now want to show that

$$\sum_{(x,k,y,\ell)} \mathbb{P}[A(x,k,y,\ell)] \sim_{\mathrm{ac}} \frac{1}{\pi} n \log n \tag{5.15}$$

where the sum is over all good quadruples (x,k,y,ℓ) with $d \log n \leq k, \ell \leq cn$. The desired result follows by the steps in the proposition below.

Proposition 5.4.6. *We have the following statements.*

(i) $\mathbb{P}[A(x,k,y,\ell)] \sim_{\mathrm{ac}} \frac{1}{16\pi}k^{-3/2}\ell^{-3/2}$.

(ii) Let $g(k,\ell)$ *be the number of pairs* (x,y) *such that* (x,k,y,ℓ) *is a good quadruple. Then* $g(k,\ell) \sim_{\mathrm{ac}} 4n \cdot \min\{k,\ell\}$.

(iii) We have

$$\frac{n}{4\pi}\sum_{d\log n \le k,\ell \le cn} k^{-3/2}\ell^{-3/2} \cdot \min\{k,\ell\} \sim_{\mathrm{ac}} \frac{1}{\pi} n \log n.$$

Before proving this proposition, we first show that this implies the asymptotic result of Theorem 5.1.2.

Corollary 5.4.7. *The expected number of edges of* CP_n *satisfies*

$$\mathbb{E}[e(CP_n)] \sim \frac{1}{\pi} n \log n. \tag{5.16}$$

Proof. Given Proposition 5.4.6, for any $\epsilon > 0$ there are some c, d and N such that for all $n \ge N$ we have

$$\mathbb{P}[A(x,k,y,\ell)] \le (1+\epsilon)\frac{1}{16\pi}k^{-3/2}\ell^{-3/2}$$

$$g(k,\ell) \le (1+\epsilon)4n\min\{k,\ell\}$$

$$\frac{n}{4\pi}\sum_{d\log n \le k,\ell \le cn} k^{-3/2}\ell^{-3/2} \cdot \min\{k,\ell\} \le (1+\epsilon)\frac{1}{\pi} n \log n.$$

This implies

$$\sum_{(x,k,y,\ell)} \mathbb{P}[A(x,k,y,\ell)] \le (1+\epsilon) \sum_{(x,k,y,\ell)} \frac{1}{16\pi} k^{-3/2}\ell^{-3/2}$$
$$= (1+\epsilon)\sum_{k,\ell} g(k,\ell) \cdot \frac{1}{16\pi} k^{-3/2}\ell^{-3/2}$$
$$\le (1+\epsilon)^2 \frac{n}{4\pi} \sum_{k,\ell} k^{-3/2}\ell^{-3/2} \min\{k,\ell\}$$
$$\le (1+\epsilon)^3 \frac{1}{\pi} n \log n,$$

and similarly for the lower bound. □

We now prove this proposition.

Proof of Proposition 5.4.6. (i) It is clear that with probability 2^{-4} all of x, $x+k$, y, $y+\ell$ have the correct color. We now claim that, conditioning on the event that this happens, with probability 2^{-2} there is an even number of red points between x and $x+k$ and an even number of blue points between y and $y+\ell$. Indeed, consider the case where $x < y < x+k < y+\ell$. Then for any possible coloring of $x+2, \dots, y-1, y+1, \dots, x+k-1, x+k+1, \dots, y+\ell-2$ there is a unique choice of colors for $x+1$ and $y+\ell-1$ that makes the number of red and blue points in the respective regions even, and with probability 2^{-2} these points will receive this color (here we used our assumption that $y \ge x+2$ and $y+\ell \ge x+k+2$).

Condition on the event that all of this happens. Let r_1 and r_2 be defined such that there are $2r_1$ red dots between x and $x+k$ and $2r_2$ red dots outside, and similarly define b_1 and b_2. Then, conditional on the aforementioned event, the probability of having arcs between x and $x+k$ and y and $y+\ell$ is given by

$$\frac{C_{r_1} \cdot C_{r_2}}{C_{r_1+r_2+1}} \cdot \frac{C_{b_1} \cdot C_{b_2}}{C_{b_1+b_2+1}}. \tag{5.17}$$

By Lemma , with probability at least $1-8n^{-10}$ we have $r_1 \sim_{\text{ac}} k/4$, $r_2 \sim_{\text{ac}} n/2-k/4$, $b_1 \sim_{\text{ac}} \ell/4$ and $b_2 \sim_{\text{ac}} n/2-\ell/4$. Furthermore, since $k, \ell \geq d \log n$ and $d \log n \to \infty$ we may replace all Catalan numbers by their asymptotic expressions, which yields that the probability of having arcs on the desired positions is (asymptotically arbitrary closely) given by

$$\frac{1}{16\pi} \cdot \left(\frac{r_1+r_2+1}{r_2}\right)^{3/2} r_1^{-3/2} \cdot \left(\frac{b_1+b_2+1}{b_2}\right)^{3/2} b_1^{-3/2}.$$

Since $r_1+r_2+1 \sim_{\text{ac}} n/2-k/4+k/4+1 \sim_{\text{ac}} n/2$ and $r_2 \sim_{\text{ac}} n/2-k/4 \sim_{\text{ac}} n/2$ (the latter since $n/2 \geq n/2-k/4 \geq n/2-cn/4$), we find $\frac{r_1+r_2+1}{r_1} \sim_{\text{ac}} 1$, and hence

$$\begin{aligned}\frac{1}{16\pi} \cdot \left(\frac{r_1+r_2+1}{r_2}\right)^{3/2} r_1^{-3/2} \cdot \left(\frac{b_1+b_2+1}{b_2}\right)^{3/2} b_1^{-3/2} \\ \sim_{\text{ac}} \frac{1}{16\pi}(k/4)^{-3/2}(\ell/4)^{-3/2} = 2^6 \frac{1}{16\pi} k^{-3/2}\ell^{-3/2}.\end{aligned}$$

Therefore, for any ϵ, and suitable c, d and large enough n we have

$$\begin{aligned}(1-8n^{-10})(1-\epsilon)\frac{1}{16\pi}k^{-3/2}\ell^{-3/2} &\leq \mathbb{P}[A(x,k,y,\ell)] \\ &\leq (1-8n^{-10})(1+\epsilon)\frac{1}{16\pi}k^{-3/2}\ell^{-3/2} + 8n^{-10}.\end{aligned}$$

Since $1-8n^{-10} \to 1$ for $n \to \infty$, and since $k^{-3/2}\ell^{-3/2} \geq n^{-3}$ we have $n^{-10} = o(k^{-3/2}\ell^{-3/2})$ (uniformly in n). This shows that $\mathbb{P}[A(x,k,y,\ell)] \sim_{\text{ac}} \frac{1}{16\pi}k^{-3/2}\ell^{-3/2}$.

(ii) Without loss of generality we may assume that $k \leq \ell$. We show that $(4-6c)n(k-3) \leq g(k,\ell) \leq 4nk$. Since $k \geq d\log n$ and $d \log n \to \infty$ we have $k-3 \sim_{\mathrm{ac}} k$, and the result follows.

For the upper bound, note that we have at most $2n$ choices for x. Furthermore, given x, either y or $y+\ell$ must be among $\{x+1, x+2, \ldots, x+k-1\}$, hence we have at most $2\cdot(k-1) \leq 2k$ choices for y afterwards. Therefore, $g(k,\ell) \leq 2n \cdot 2k = 4nk$.

For the lower bound, let $cn \leq x \leq (2-2c)n$. We claim that for any such x there are at least $2(k-3)$ good quadruples with that x. Indeed, let $y \in \{x+2, \ldots, x+k-2\}$ or $y \in \{x+2-\ell, \ldots, x+k-2-\ell\}$, and we claim that any such y works. Since $\ell \geq k$ these two sets are disjoint, giving us $2(k-3)$ good quadruples.

First suppose that $y = x+j$ for $2 \leq j \leq k-2$. Then we clearly have $1 \leq x < y < x+k < y+\ell$, $y \geq x+2$ and $x+k \geq y+2$. Furthermore, $y+\ell \geq x+2+\ell \geq x+2+k = (x+k)+2$. Lastly, $y+\ell \leq x+k-2+\ell \leq 2n-2cn+k+\ell \leq 2n$, since $k, \ell \leq cn$. A similar argument holds in the case $y = x+j-\ell$.

(iii) We consider the contribution to the sum coming from $k < \ell$, the analysis for the contribution coming from $k \geq \ell$ is analogous. First, note that

$$\begin{aligned}\sum_{k<\ell} k^{-1/2}\ell^{-3/2} &= \sum_{\ell=d\log n}^{cn} \ell^{-3/2} \sum_{k=d\log n}^{\ell-1} k^{-1/2} \leq \sum_{\ell} \ell^{-3/2} \int_1^{\ell} x^{-1/2}\mathrm{d}x \\ &= \sum_{\ell} \ell^{-3/2}(2\ell^{1/2}-2) \leq \sum_{\ell=d\log n}^{cn} 2\ell^{-1} \\ &\leq 2\int_{d\log n-1}^{cn} x^{-1}\mathrm{d}x \leq 2\log(cn) \leq 2\log n.\end{aligned}$$

In the other direction, note that we have a lower bound of

$$\sum_{\ell=(\log n)^2}^{cn} \ell^{-3/2} \sum_{k=d\log n}^{\ell-1} k^{-1/2} \geq \sum_{\ell=(\log n)^2}^{cn} \ell^{-3/2}\ell^{-3/2} \int_{d\log n}^{\ell} x^{-1/2}\mathrm{d}x$$
$$= \sum_{\ell=(\log n)^2}^{cn} \ell^{-3/2}(2\ell^{1/2} - 2(d\log n)^{1/2}).$$

For any ϵ we have $(d\log n) \leq \epsilon^2(\log n)^2 \leq \epsilon^2\ell^2$ for n large enough, hence $2\ell^{1/2} - 2(d\log n)^{1/2} \geq 2(1-\epsilon)\ell^{-1/2}$ for n large enough. Therefore, we get a lower bound of

$$2(1-\epsilon) \sum_{\ell=\log(n)^2}^{cn} \ell^{-1} \geq 2(1-\epsilon)\left(\log(cn+1) - \log((\log n)^2)\right)$$

by again comparing the sum with an integral. The desired result now follows from the fact that

$$\log(cn+1) - \log(\log(n)^2) \geq \log n + \log c - \log(\log(n)^2) \sim \log n,$$

hence we have $\log(cn+1) - \log(\log(n)^2) \geq (1-\epsilon)\log n$ for n large enough. □

5.5 The number of isolated vertices

In this section we will determine the asymptotic behavior of the number of isolated vertices, as stated in Theorem 5.1.3. Recall that I_n denotes the number of isolated vertices of CP_n and that we defined

$$\gamma = 4\sum_{m=1}^{\infty} 16^{-m} \sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b}C_b. \tag{5.18}$$

Before proving Theorem 5.1.3, let us first show why the sum defining γ is a convergent sum. Let $\gamma_m = 4 \cdot 16^{-m}\sum_{b=0}^{m-1}\binom{2m-2}{2b}C_{m-1-b}C_b$, then as noted in [BDD+, Section 5] we have $\gamma_m \leq \frac{1}{4(m-1)^2}$ for $m \geq 2$, from which the convergence follows since the sum of the reciprocals of the squares converges.

In fact, this gives us an error bound on how quickly the finite sums $\sum_{m=1}^{M} \gamma_m$ converge to γ. Indeed

$$\begin{aligned}\gamma = \sum_{m=1}^{\infty} \gamma_m = \sum_{m=1}^{M} \gamma_m + \sum_{m=M+1}^{\infty} \gamma_m \leq \sum_{m=1}^{M} \gamma_m + \sum_{m=M+1}^{\infty} \frac{1}{4(m-1)^2} \\ \leq \sum_{m=1}^{M} \gamma_m + \int_{x=M}^{\infty} \frac{1}{4(x-1)^2} \mathrm{d}x = \sum_{m=1}^{M} \gamma_m + \frac{1}{4(M-1)}.\end{aligned}$$

Using the trivial lower bound $\gamma \geq \sum_{m=1}^{M} \gamma_m$ and taking $M = 10^4$ one can compute that

$$0.30234 \leq \gamma \leq 0.30238.$$

We first show that $\mathbb{E}[I_n]$ is asymptotically at least γn. As a first observation we note that any arc yielding an isolated vertex must have an even number of points between its endpoints, as otherwise there would be an arc connecting a point between its endpoints with a point outside. Such an arc would necessarily be on the other side and would yield an edge involving the arc in question. Therefore, $I_n = \sum_{m=1}^{n} I_{n,m}$ where $I_{n,m}$ is the number of isolated vertices induced by an arc connecting two points with $2m-2$ points between them.

The following result will suffice to prove the lower bound for $\mathbb{E}[I_n]$.

Proposition 5.5.1. *For m a fixed positive integer we have $\mathbb{E}[I_{n,m}] \sim \gamma_m n$.*

As a result of this proposition, we can see that

$$\mathbb{E}[I_n] \geq \sum_{m=1}^{M} \mathbb{E}[I_{n,m}] \sim \sum_{m=1}^{M} \gamma_m n, \tag{5.19}$$

which gets arbitrarily close (in the multiplicative sense) to γn by picking M large enough. However, this approach does not immediately yield the upper bound, since each $\mathbb{E}[I_{n,m}]$ will converge to $\gamma_m n$ at its own rate, hence a bit more care is needed to handle the full sum $\mathbb{E}[I_n] = \sum_{m=1}^{n} \mathbb{E}[I_{n,m}]$.

Proof of Proposition 5.5.1. We count the expected number of such arcs that come from the top, and by symmetry we can multiply this quantity by two to get our final answer. As mentioned above, an arc connecting x and $x + 2m - 1$ is isolated if and only if the $2m - 2$ intermediate points are only connected to themselves. The total number of ways to connect those points is given by

$$\sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b} C_b, \tag{5.20}$$

where b is the number of arcs on the bottom, $\binom{2m-2}{2b}$ counts the number of ways to select the $2b$ points for these arcs, and C_{m-1-b} and C_b count the number of ways to choose the arcs on the top and the bottom.

Now fix one such configuration with b arcs on the bottom and a arcs on top (including the arc between x and $x + 2m - 1$). We claim that the expected number of such configurations in CP_n is given by

$$(2n - 2m + 1)2^{-2m} \sum_{r=0}^{n-m} p_r \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}},$$

where $p_r = p_r(n, a, b)$ is the probability that $2r$ of the points not among the $2m$ specified points are colored red.

This formula follows from the fact that there are $2n - 2m + 1$ possibilities for x, namely $1 \leq x \leq 2n - 2m + 1$, and that for each such x the probability of the points $x, x + 1, \ldots, x + 2m - 1$ colored exactly as in our configuration is given by 2^{-2m}. After that, given x and conditioning on these points having the correct colors and conditioning on there being $2r$ other red points, the probability that the top Catalan-arc matching (which has size $r + a$) has exactly the desired configuration on our given $2a$ red points is exactly $\frac{C_r}{C_{r+a}}$ by Lemma 5.3.2, and a similar result holds for the probability of the bottom Catalan-arc matching coinciding with our given configuration on the $2b$ points.

To complete the proof it suffices to show that

$$\sum_{r=0}^{n-m} p_r \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}} \sim 4^{-m},$$

since then

$$\mathbb{E}[I_{n,m}] \sim 2 \left(\sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b} C_b \right) (2n - 2m + 1) 2^{-2m} \cdot 4^{-m} \sim \gamma_m n.$$

Using Lemma 2.7.1 with exponential small probability we have $r \leq n/4$ or $n - m - r \leq n/4$. As a trivial lower bound we have

$$\sum_{r=0}^{n-m} p_r \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}} \geq \sum_{r=n/4}^{n-m-n/4} p_r \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}}.$$

Now in this region, since $r, r+a, n-m-r, n-m-r+b \geq n/4$ we can use the approximation for the Catalan numbers from (5.3) and find the lower bound

$$\sum_{r=n/4}^{n-m-n/4} p_r \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}}$$

$$\sim \sum_{r=n/4}^{n-m-n/4} p_r \frac{4^r}{4^{r+a}} \left(\frac{r+a}{r}\right)^{3/2} \cdot \frac{4^{n-m-r}}{4^{n-m-r+b}} \left(\frac{n-m-r}{n-m-r+b}\right)^{3/2}$$

$$\sim \sum_{r=n/4}^{n-m-n/4} p_r 4^{-(a+b)} = 4^{-m} \sum_{r=n/4}^{n-m-n/4} p_r \sim 4^{-m},$$

where the last step follows from the fact that $r < n/4$ or $r > n-m-n/4$ holds with exponentially small probability.

Similarly, we have

$$\sum_{r=0}^{n-m} p_r \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}}$$

$$\leq \sum_{r=n/4}^{n-m-n/4} p_r \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}} + \Pr(r \leq n/4 \text{ or } n-m-r \leq n/4)$$

$$\sim 4^{-m} + \Pr(r \leq n/4 \text{ or } n-m-r \leq n/4) \sim 4^{-m},$$

completing the proof. □

We now prove the desired asymptotics for the number of isolated vertices.

Proposition 5.5.2. *Let γ be the constant defined by*

$$\gamma = 4 \sum_{m=1}^{\infty} 16^{-m} \sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b} C_b = 0.3023\ldots. \tag{5.21}$$

Let I_n denote the number of isolated vertices of CP_n. Then $\mathbb{E}[I_n] \sim \gamma n$.

Proof. As mentioned after the statement of Proposition 5.5.1 we have shown an asymptotic lower bound of γn on the number of isolated vertices. For the upper bound, note that using the notation of Lemma 5.4.3 we have that $I_{n,m} \le A_{2m-1,2m-1}$, since the number of isolated vertices coming from arcs of length $2m-1$ is clearly at most the the total number of arcs of this length. By this observation, the fact that $\sum_{m=16\log n+1}^{n} A_{2m-1,2m-1} \le A_{32\log n+1,2n}$, and Lemma 5.4.3, we have

$$\sum_{m=16\log n+1}^{n} \mathbb{E}[I_{n,m}] \le \mathbb{E}[A_{32\log n+1,2n}] = o(n),$$

which shows that

$$\mathbb{E}[I_n] = \sum_{m=1}^{16\log n} \mathbb{E}[I_{n,m}] + o(n).$$

Using the argument from Proposition 5.5.1 we see that

$$\sum_{m=1}^{16\log n} \mathbb{E}[I_{n,m}] \le 4n \sum_{m=1}^{16\log n} 4^{-m} \sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b} C_b \sum_{r=0}^{n-m} p_r(n,a,b) \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}}.$$

We now see that for any m, a and b we have that there are at least n points outside of the configuration, hence $2r$ is the sum of at least n independent $0-1$ Bernoulli $p = 1/2$ variables. This means that with at most some exponentially small probability c^{-n} we have $r, n-m-r \le n/10$.

Therefore, for all cases where $r, n-m-r \ge n/10$ we can again (uniformly over all summands) replace $\frac{C_r}{C_{r+a}}$ by $4^{-a}\left(\frac{r+a}{r}\right)^{3/2}$. Since $\frac{r+a}{r} = 1 + \frac{a}{r} \le 1 + \frac{16\log n}{n/10}$ we can asymptotically replace $\frac{r+a}{r}$ by 1 over all summands. Using this and the approach as in Proposition 5.5.1 we have an asymptotic upper bound $\sum_{r=0}^{n-m} p_r(n,a,b) \frac{C_r}{C_{r+a}} \cdot \frac{C_{n-m-r}}{C_{n-m-r+b}} \le$

$4^{-m} + c^{-n}$, hence (asymptotically up to arbitrarily small multiplicative factors) we have

$$\begin{aligned}
\sum_{m=1}^{16\log n} \mathbb{E}[I_{n,m}] &\le 4n \sum_{m=1}^{16\log n} 4^{-m} \sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b} C_b \left(4^{-m} + c^{-n}\right) \\
&\le \gamma n + 4n \left(\sum_{m=1}^{16\log n} 4^{-m} \sum_{b=0}^{m-1} \binom{2m-2}{2b} C_{m-1-b} C_b \right) c^{-n} \\
&\le \gamma n + 4n \left(\sum_{m=1}^{16\log n} 4^{-m} 16^m \right) c^{-n} \le \gamma n + 4nc^{-n} \sum_{m=1}^{16\log n} 4^m \\
&\le \gamma n + 4nc^{-n} \cdot 16 \log n 4^{16\log n} \\
&= \gamma n + 64nc^{-n} \cdot \log n \cdot n^{16\log 4} = \gamma n + o(1),
\end{aligned}$$

since c^{-n} goes to zero faster than $n^{1+16\log 4}\log n$ grows to infinity. □

We can use a similar proof to bound the variance of I_n.

Proposition 5.5.3. *The variance of the number of isolated vertices in CP_n satisfies $\mathrm{Var}[I_n] = o(n^2)$.*

Before giving this proof, let us point out that using Chebyshev's inequality from Lemma 2.7.2 we can use this result to complete the proof of Theorem 5.1.3.

Proof of Theorem 5.1.3. The asymptotic result for the expected number of isolated vertices follows from Proposition 5.5.2. From this we know $|\mathbb{E}[I_n] - \gamma n| < \epsilon/2 \cdot n$ for n large enough. Hence, for sufficiently large n we have,

$$\mathbb{P}[|I_n - \gamma_n| > \epsilon n] \le \mathbb{P}[|I_n - \mathbb{E}[I_n]| > \epsilon/2 \cdot n].$$

Now, applying Chebyshev's inequality we find

$$\mathbb{P}[|I_n - \mathbb{E}[I_n]| > \epsilon/2 \cdot n] \le \frac{\mathrm{Var}[I_n]}{(\epsilon/2 \cdot n)^2} = \frac{o(n^2)}{(\epsilon/2 \cdot n)^2} = o(1), \qquad (5.22)$$

as desired. $\square$

We will now prove the result on the variance.

Proof of Proposition 5.5.3. By definition we have $\mathrm{Var}[I_n] = \mathbb{E}[I_n^2] - \mathbb{E}[I_n]^2$, where $\mathbb{E}[I_n]^2 = (\gamma n)^2 + o(n^2)$ by the first part of Theorem 5.1.3. Therefore, since variance is nonnegative, it suffices to show that

$$\mathbb{E}[I_n^2] \leq (\gamma n)^2 + o(n^2). \tag{5.23}$$

Observe that I_n^2 is the number of ordered pairs of isolated vertices.

Just as above we show that we can restrict ourselves to the isolated vertices induced by arcs of length at most $32 \log n$. Indeed, let $A_{\alpha,\beta}$ be as in Lemma 5.4.3. Then the number of pairs where at least one vertex comes from an arc of length at least $32 \log n$ is at most $2 \cdot A_{32 \log n, 2n} \cdot n$, where the factor 2 represents the choice of the vertex coming from a long arc being the first or second vertex in the pair, $A_{32 \log n, 2n}$ is the number of ways to pick this long arc, and n is the number of ways to pick the remaining vertex. Therefore, this contribution to $\mathbb{E}[I_n^2]$ is at most $\mathbb{E}[2 \cdot A_{32 \log n, 2n} \cdot n] = o(n^2)$ by Lemma 5.4.3.

Additionally, the number of pairs of isolated vertices coming from two arcs of length at most $32 \log n$, where one arc is contained in the other arc (possibly facing the other way) is deterministically at most $O(n \log n)$, since one can pick the outer arc in at most n ways and then there are at most $32 \log n$ ways to pick the smaller arc. Therefore, these pairs contribute $o(n^2)$ to $\mathbb{E}[I_n^2]$ as well. Furthermore, the number of pairs where both arcs are the same are at most n, so these will also contribute $o(n^2)$ to $\mathbb{E}[I_n^2]$.

Therefore, we can restrict our attention to pairs of isolated vertices coming from different arcs of length at most $32 \log n$ such that neither arc is contained in the other. Note that since the arcs yield isolated vertices their endpoints cannot interlace, so the sets of points covered by this arc are disjoint.

Suppose we want to calculate the probability of having a pair of isolated vertices, one of them induced by an arc connecting $(x, x + 2m - 1)$ and the other connecting an arc connecting $(y, y + 2k - 1)$, where $m, k \leq 16 \log n$. By a similar argument as in Proposition 5.5.1, after specifying configurations for $\{x + 1, \ldots, x + 2m - 2\}$ and $\{y + 1, \ldots, y + 2k - 2\}$ the probability is (asymptotically up to arbitrarily small multiplicative factors) at most

$$4^{-(m+k)} \cdot (4^{-(m+k)} + c^{-n}),$$

where $4^{-(m+k)}$ is the probability that all of $\{x, x+1, \ldots, x+2m-1\}$ and $\{y, y+1, \ldots, y+2k-1\}$ receive the correct color, and c^{-n} is once again an upper bound on the probability of *not* having at least $n/10$ more blue and red points, and the $4^{-(m+k)}$ is once again the factor that shows up by considering the asymptotic behavior of the appropriate quotient of Catalan numbers. Also, by the same argument we can do these asymptotics for all possible x, y, k, m and choice of configurations simultaneously.

Taking into account that there are at most $(2n)^2$ ways to choose x and y, and 4 ways to choose the side (top or bottom) for the arcs, and considering the possible configurations for $\{x+1, \ldots, x+2k-2\}$ and $\{y+1, \ldots, y+2k-2\}$ we find an asymptotic upper bound for the desired contribution of

$$\sum_{k,m=1}^{16\log n} 16n^2 \left(\sum_{b_1=0}^{m-1} \binom{2m-2}{2b_1} C_{m-1-b_1} C_{b_1} \right) \cdot \left(\sum_{b_2=0}^{k-1} \binom{2k-2}{2b_2} C_{k-1-b_2} C_{b_2} \right) 4^{-(m+k)} \left(4^{-(m+k)} + c^{-n}\right).$$

Using $4^{-(m+k)} + c^{-n} \leq (4^{-m} + c^{-n/2})(4^{-k} + c^{-n/2})$, we can separate the sums over k and m. Thus the contribution is at most

$$\left(\sum_{m=1}^{16\log n} 4n \cdot \sum_{b_1=0}^{m-1} \binom{2m-2}{2b_1} C_{m-1-b_1} C_{b_1} \cdot 4^{-m}(4^{-m} + c^{-n/2}) \right)^2 \leq (\gamma n)^2 + o(n^2),$$

where the last inequality once again follows from the proof of Theorem 5.1.3. □

Remark. Essentially the same proof can be used to show that $\mathbb{E}[I_n^m] \sim \gamma^m n^m$ for all $m \geq 2$. ◀

With all this we can conclude the results on the number of edges of CP_n.

Proof of Theorem 5.1.2. The asymptotic formula for the expected number of edges follows from Corollary 5.4.7. The concentration result follows from Proposition 5.6.1 and essentially the same proof used in the proof of Theorem 5.1.3. □

5.6 The variance of the number of edges

This section will be devoted to bounding the variance of the random variable $e(CP_n)$. We will prove the following result, which with a proof similar to that of Theorem 5.1.3 will imply the concentration result of Theorem 5.1.2.

Proposition 5.6.1. *The variance of the number of edges in CP_n satisfies*

$$\mathrm{Var}[e(CP_n)] = o(n^2 \log^2 n). \tag{5.24}$$

Similar to the case of isolated vertices, we will prove this statement by showing that for any $\epsilon > 0$ and n large enough we have

$$E[(e(CP_n))^2] \leq (1+\epsilon)\frac{1}{\pi^2}n^2 \log^2 n + o(n^2 \log^2 n).$$

In other words, we want to count the expected number of pairs of edges in CP_n. Just as when we determined the expected number of edges, we first have to handle some exceptional cases and show that all of these cases contribute of order $o(n^2 \log^2 n)$. The general approach of these proofs are similar to Proposition 5.4.4 and Proposition 5.5.3. However, there will be more exceptional cases to take care of, and each of the individual proofs will be slightly longer due to the involvement of more arcs and edges.

Therefore, we will first state the necessary lemmas, and show how they lead towards the proofs of Proposition 5.6.1 and consequently Theorem 5.1.2, and we will defer the proofs of the lemmas to the end of the section.

As mentioned, $e(CP_n)^2$ is the number of pairs of edges in CP_n. Typically, such a pair of edges will be induced by four arcs in the representative for CP_n. The first step will be to show that these pairs are indeed the main contribution to $E[(e(CP_n))^2]$.

Lemma 5.6.2. *The expected number of pairs of edges in CP_n induced by at most three arcs in its representative is at most $o(n^2 \log^2 n)$.*

Therefore, we can restrict to valid quadruples

$$q = (\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l}) = ((x_1, x_2), (k_1, k_2), (y_1, y_2), (\ell_1, \ell_2)),$$

where (x_i, k_i, y_i, ℓ_i) is a possible edge for $i = 1, 2$. Our goal is now to show that

$$\sum_q \mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] \leq (1+\epsilon)\frac{1}{\pi^2}n^2 \log^2 n + o(n^2 \log^2 n), \tag{5.25}$$

where the sum is over all valid quadruples $q = (\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$. We use the notation for $f_0, f_1, f_2, g_0, g_1, g_2$ as in Section 5.4. Similar to the proof for the expected number of edges, the first step will be to show that the main contribution comes from quadruples with $f_i, g_j \geq d \log n$. That is, we will show that if Q_1 is the set of quadruples for which at least one of f_i, g_j is less than $d \log n$, then

$$\sum_{q \in Q_1} \mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] = o(n^2 \log^2 n).$$

Without loss of generality we can consider the case where one of the f_i is less than $d \log n$. Then the result follows from the two lemmas below, the first one of which deals with the case that the two arcs on top are nested, and the second one deals with the unnested case.

Lemma 5.6.3. *Let $Q_{1,1}$ be the set of all valid quadruples q for which $x_1 < x_2 < x_2 + k_2 < x_1 + k_1$ and for which $k_2, k_1 - k_2$ or $2n - k_1$ is less than $d \log n$. Then*

$$\sum_{q \in Q_{1,1}} \mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] = o(n^2 \log^2 n).$$

Lemma 5.6.4. *Let $Q_{1,2}$ be the set of all valid quadruples q for which neither $x_1 < x_2 < x_2 + k_2 < x_1 + k_1$ nor $x_2 < x_1 < x_1 + k_1 < x_2 + k_2$ holds, and for which k_1, k_2 or*

$2n-(k_1+k_2)$ *is less than* $d\log n$. *Then*

$$\sum_{q\in Q_{1,2}} \mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})] = o(n^2\log^2 n).$$

In order to complete the proof of Proposition 5.6.1 we can now assume that all f_i, g_j are at least $d\log n$. The first step will be to deal with the case that some of the arcs are nested.

Lemma 5.6.5. *Let* Q_2 *be the set of quadruples with* $x_1 < x_2 < x_2+k_2 < x_1+k_1$ *and* $f_i, g_j \geq d\log n$. *Then*

$$\sum_{q\in Q_2} \mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})] = o(n^2\log^2 n).$$

For the remainder of this section on we will assume that any quadruple has no nested arcs. First we take care of the quadruples where one of the arcs is too large.

Lemma 5.6.6. *Let* Q_3 *be the set of quadruples with* $\max\{k_1,k_2,\ell_1,\ell_2\} > cn$. *Then*

$$\sum_{q\in Q_3} \mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})] = o(n^2\log^2 n)$$

We lastly rule out all of the remaining quadruples that are valid but not good.

Lemma 5.6.7. *Let* Q_4 *be the set of valid quadruples that are not good and have* $d\log n \leq k_1,k_2,\ell_1,\ell_2 \leq cn$. *Then*

$$\sum_{q\in Q_4} \mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})] = o(n^2\log^2 n).$$

Before we give the proof of Proposition 5.6.1 we recall a definition from Proposition 5.4.6. For positive integers k,ℓ, we defined $g(k,\ell)$ as the number of pairs (x,y) such that (x,k,y,ℓ) is a good quadruple. We are now ready to prove our desired result on the variance.

Proof of Proposition 5.6.1. By Lemmas 5.6.2 through 5.6.7 we only have to consider quadruples $(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})$ that are good, have no nested arcs, and which have $d\log n \leq k_1,k_2,\ell_1,\ell_2 \leq cn$. In this case, given k_1,k_2,ℓ_1,ℓ_2 there are $g(k_1,\ell_1)$ ways to pick x_1,y_1 and after that at most $g(k_2,\ell_2)$ ways to pick x_2,y_2.

Therefore, it suffices to show that for d large enough and c small enough we have

$$\mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})] \leq (1+\epsilon)\cdot\frac{1}{16\pi}k_1^{-3/2}\ell_1^{-3/2}\cdot\frac{1}{16\pi}k_2^{-3/2}\cdot\ell_2^{-3/2}, \tag{5.26}$$

as this implies that the desired contribution is at most

$$\sum_{k_1,k_2,\ell_1,\ell_2} g(k_1,\ell_1)\cdot g(k_2,\ell_2)\cdot(1+\epsilon)\cdot\frac{1}{16\pi}k_1^{-3/2}\ell_1^{-3/2}\cdot\frac{1}{16\pi}k_2^{-3/2}\cdot\ell_2^{-3/2},$$

which factors as

$$(1+\epsilon)\left(\sum_{k_1,\ell_1} g(k_1,\ell_1)\frac{1}{16\pi}k_1^{-3/2}\ell_1^{-3/2}\right)\cdot\left(\sum_{k_2,\ell_2} g(k_2,\ell_2)\frac{1}{16\pi}k_2^{-3/2}\ell_2^{-3/2}\right),$$

which by Proposition 5.4.6 is at most $(1+\epsilon)^3\cdot(\frac{1}{\pi}n\log n)^2$ for d large enough and c small enough.

In order to show (5.26) we follow the same approach as the proof of part 1 of Proposition 5.4.6. First, with probability 2^{-8} all of $x_i, x_i+k_i, y_i, y_i+\ell_i$ receive the correct color and with probability 2^{-4} the number of red points between x_i and x_i+k_i and the number of blue points between y_j and $y_j+\ell_j$ are all even. This follows immediately from the aforementioned proof when neither (x_1,x_1+k_1) and $(y_2,y_2+\ell_2)$ nor (x_2,x_2+k_2)

and $(y_1, y_1 + \ell_1)$ intersect. Otherwise, we may without loss of generality assume that $x_1 < y_1 < x_1 + k_1 < x_2 < y_1 + \ell_1 < y_2 < x_2 + k_2 < y_2 + \ell_2$. In this case, color all the remaining points between x_1 and $y_2 + \ell_2$ except for $x_1 + 1, y_1 + 1, x_2 + 1, y_2 + 1$. Then, given any such coloring there is a unique choice for the remaining four colors that makes the number of red/blue in the desired regions even, as first $y_2 + 1$ is uniquely determined, then $x_2 + 1$, then $y_1 + 1$ and lastly $x_1 + 1$.

Now suppose that r_i is half the number of red points between x_i and $x_i + k_i$ for $i = 1, 2$, r_0 is half the number of red points outside of the arcs, and b_0, b_1, b_2 are defined similarly. Conditioned on the values of r_i and b_j we can write the desired probability as

$$\frac{C_{r_0} C_{r_1} C_{r_2}}{C_{r_0+r_1+r_2+2}} \cdot \frac{C_{b_0} C_{b_1} C_{b_2}}{C_{b_0+b_1+b_2+2}}.$$

Again by Lemma , with high enough probability we can approximate r_i with $k_i/4$ ($i = 1, 2$) and r_0 with $n/4 - k_1/4 - k_2/4$, and similarly for the b_i, and the same asymptotic considerations as in Proposition 5.4.6 will now yield the desired result. □

5.6.1 Proofs of the technical lemmas

We are now ready to prove all the lemmas from above. First, we prove the lemma that concerns all pairs of edges coming from at most three arcs.

Proof of Lemma 5.6.2. Since it is clear that at least two arcs must be involved, there are two cases to consider. First, suppose that the total number of arcs involved equals two. Then both edges in the pair are the same edge, so the number of such pairs equals

$e(CP_n) \leq n^2$. On the other hand, if there are a total of three arcs involved, there are at most $n \cdot e(CP_n)$ pairs of such edges. Indeed, there are $e(CP_n)$ ways to choose the first edge in the pair, which yields two arcs, and then there are at most n ways to choose a third arc that interlaces with either of the two arcs used already. Therefore, in expectation there are at most

$$\mathbb{E}[n \cdot e(CP_n)] = \frac{1}{\pi} n^2 \log n = o(n^2 \log^2 n)$$

such pairs. □

The next two lemmas are used to show that we may assume that each of the gap sizes is of order at least $\log n$. We define $e'(CP_n)$ as the number of edges in CP_n, at least one of whose arcs has size at most $d \log n$ or at least cn. We refer to such edges as *exceptional edges*. In Proposition 5.4.4 we showed that $\mathbb{E}[e'(CP_n)] = o(n \log n)$.

Proof of Lemma 5.6.3. First we consider the number of pairs with $2n - k_1 < d \log n$. We claim that there are at most $(d \log n)^2 \cdot e(CP_n)$ such pairs. Indeed, we can pick the edge (x_2, k_2, y_2, ℓ_2) in at most $e(CP_n)$ ways, and the edge (x_1, k_1, y_1, ℓ_1) in at most $(d \log n)^2$ ways: we can pick k_1 in $d \log n$ ways, after which there is at most one x_1 such that x_1 and $x_1 + k_1$ are connected (since $k_1 > n$) and the vertex corresponding to this arc has degree at most $d \log n$ (as each interlacing arc must have an endpoint less than x_1 or larger than $x_1 + k_1$). By taking expectations we see that we have at most $(d \log n)^2 \mathbb{E}[e(CP_n)] = o(n^2 \log^2 n)$ such pairs.

Now suppose that $k_1 - k_2 < d \log n$ or $k_2 < d \log n$. First consider the pairs with

(x_1, k_1, y_1, ℓ_1) an exceptional edge. We claim that the number of such pairs is at most $n \cdot d \log n \cdot e'(CP_n)$, from which taking expectations will suffice. In order to prove this, note that there are at most $e'(CP_n)$ ways to pick an exceptional edge. Then, in the case $k_1 - k_2 < d \log n$, there are at most $d \log n$ ways to pick x_2, and the corresponding arc has degree at most n. Similarly, if $k_2 < d \log n$, there are at most n ways to pick x_2, and the corresponding arc has degree at most $d \log n$.

Therefore, we may assume that (x_1, k_1, y_1, ℓ_1) is not an exceptional edge. Assume that k_1 and ℓ_1 are given. By the same logic as the proof of Proposition 5.4.6, we know that there are at most $4n \min\{k_1, \ell_1\}$ options for x_1 and y_1, and by Lemma 5.4.2 the probability of having arcs connecting (x_1, x_1+k_1) and $(y_1, y_1+\ell_1)$ is $O(k_1^{-3/2}\ell_1^{-3/2})$. Furthermore, given (x_1, k_1, y_1, ℓ_1) there are at most $k_1 \cdot d \log n$ possible second edges by a similar argument as above, where we now use k_1 instead of n since we have fixed the size of the outer arc. Hence, the expected number of such pairs of edges is given by

$$O\left(2n \log n \cdot \sum_{k_1, \ell_1} \min\{k_1, \ell_1\} k_1^{-1/2} \ell_1^{-3/2}\right)$$

so it suffices to show that $\sum_{k_1,\ell_1} \min\{k_1, \ell_1\} k_1^{-1/2} \ell_1^{-3/2} = o(n \log n)$. The contribution from $k_1 \le \ell_1$ is at most

$$\sum_{\ell_1 \le cn} \ell_1^{-3/2} \sum_{k_1 \le \ell_1} k_1^{1/2} \le \sum_{\ell_1 \le cn} \ell_1^{-3/2} O(\ell_1^{3/2}) = O(n),$$

and the contribution from $\ell_1 \le k_1$ is at most

$$\sum_{k_1 \le cn} k_1^{-1/2} \sum_{\ell_1 \le k_1} \ell_1^{-1/2} = \sum_{k_1 \le cn} k_1^{-1/2} O(k_1^{1/2}) = O(n)$$

completing the proof. □

Proof of Lemma 5.6.4. We first consider the case that one of k_1, k_2 is less than $d \log n$. By symmetry we can assume that $k_1 < d \log n$. As in the previous lemma, the number of pairs of edges with (x_2, k_2, y_2, ℓ_2) an exceptional edge is at most $n \cdot d \log n \cdot e'(CP_n)$ as there are at most $n \cdot d \log n$ edges where one vertex has degree at most $d \log n$, and there are at most $e'(CP_n)$ ways to pick the second edge. Therefore, in expectation, there are at most $O(n \log n) \cdot \mathbb{E}[e'(CP_n)] = o(n^2 \log^2 n)$ such pairs.

Thus we may assume that (x_2, k_2, y_2, ℓ_2) is not an exceptional edge. Consider all pairs of edges where $k_1 < \sqrt{\log n}$. The number of such pairs is at most $n \cdot \sqrt{\log n} \cdot e(CP_n)$, as one can pick the arc $(x_1, x_1 + k_1)$ in at most n ways, this vertex has degree at most $\sqrt{\log n}$, and there are at most $e(CP_n)$ ways to pick the second edge. In particular, the expected number of such pairs is at most $n \cdot \sqrt{\log n} \cdot \mathbb{E}[e(CP_n)] = O(n^2 (\log n)^{3/2}) = o(n^2 \log^2 n)$.

Lastly we handle the case where $\sqrt{\log n} \leq k_1 \leq d \log n$. We consider the expected number of pairs of an arc and an edge $((x_1, k_1), (x_2, k_2, y_2, \ell_2))$ such that k_1 is in the given range, and the arcs $(x_1, x_1 + k_1)$ and $(x_2, x_2 + k_2)$ are not nested. If we can show that the expected number of such pairs is $o(n^2 \log n)$ the result follows. Indeed, any pair of edges of interest comes from such an arc-edge pair together with an arc that interlaces with (x_1, k_1), and there are at most $O(\log n)$ such arcs. Thus in total we will get at most $o(n^2 \log n) \cdot O(\log n) = o(n^2 \log^2 n)$ pairs of edges.

To accomplish this, consider any valid quadruple $q = ((x_1, x_2), (k_1, k_2), (y_2), (\ell_2))$ giving an arc-edge pair as described above. We show that we have

$$\mathbb{P}[A(q)] = O\left(k_2^{-3/2} \ell_2^{-3/2} \cdot \left(k_1^{-3/2} + e^{-\sqrt{\log n}/16}\right)\right). \tag{5.27}$$

Showing the above bound on the probability suffices because then the number of arc-edge pairs is at most

$$\sum_q \mathbb{P}[A(q)] = O\left(\left(\sum_{x_1,k_1} k_1^{-3/2} + e^{-\sqrt{\log n}/16}\right) \cdot \left(\sum_{x_2,k_2,y_2,\ell_2} k_2^{-3/2}\ell_2^{-3/2}\right)\right)$$

where we note that some combinations of some (x_1, k_1) used in the first sum and some (x_2, k_2, y_2, ℓ_2) used in the second sum will not give a desired quadruple q, but this is no issue since we are only interested in an upper bound. By Lemma 5.4.3 and Proposition 5.4.6 the first sum is $o(n)$ and the second sum is $O(n \log n)$, showing the desired result. We will deviate slightly and assume that ℓ_2 is at least $2d \log n$, but we note that this change will not affect our previous arguments.

To prove (5.27) we note that $\mathbb{P}[A(q)]$ can be written as

$$\mathbb{P}[A(q)] = 2^{-2n} \sum_c \frac{C_{n_0}C_{n_1}C_{n_2}}{C_{n_0+n_1+n_2+2}} \cdot \frac{C_{m_0}C_{m_2}}{C_{m_0+m_2+1}}, \tag{5.28}$$

where the sum is over all colorings c of the points such that all the points coming from q receive the correct color and the number of points of the desired color in each region is even. Here n_0 and m_0 are half the number of red and blue points outside of the desired arcs, n_1 is half the number of red points within arc $(x_1, x_1 + k_1)$ and n_2 and m_2 are half the number of red an blue points respectively in the arcs $(x_2, x_2 + k_2)$ and $(y_2, y_2 + \ell_2)$. Note that $\ell_2 > 2d \log n$, hence the number of points between y_2 and $y_2 + \ell_2$ that do not lie between x_1 and $x_1 + k_1$ is at least $d \log n$.

Consider all the possible colorings of all the points except for the points in the interval $[x_1, x_1+k_1]$. By using Lemma , for d large enough, we can say that with probability at least $1-O(n^{-10})$ we have $n_0, m_0 = \Omega(n)$, $n_2 = \Omega(k_2)$ and $m_2 = \Omega(\ell_2)$, where the bound on m_2 follows by the above remark that there are still at least $d \log n$ points that we are considering. Since

$$n^{-10} = o\left(\left(k_1^{-3/2} + e^{-\sqrt{\log n}/16}\right) \cdot \left(k_2^{-3/2} \ell_2^{-3/2}\right)\right)$$

we can restrict our attention to all colorings where the above bounds are satisfied. Now, for any such coloring, using the asymptotic formula for the Catalan numbers, we have

$$\frac{C_{m_0} C_{m_2}}{C_{m_0+m_2+1}} = O(\ell_2^{-3/2}).$$

Furthermore, we can rewrite

$$\frac{C_{n_0} C_{n_1} C_{n_2}}{C_{n_0+n_1+n_2+2}} = \frac{C_{n_0+n_2+1} C_{n_1}}{C_{n_0+n_1+n_2+2}} \cdot \frac{C_{n_0} C_{n_2}}{C_{n_0+n_2+1}},$$

then as in Lemma 5.4.3 we can show that $\frac{C_{n_0+n_2+1} C_{n_1}}{C_{n_0+n_1+n_2+2}}$, which is the probability of an arc connecting x_1 and $x_1 + k_1$, is given by $O\left(k_1^{-3/2} + e^{-\sqrt{\log n}/16}\right)$, where this case is even a bit easier since we already specified the number of red points outside the arc. Furthermore, plugging in $n_0 = \Omega(n)$ and $n_2 = \Omega(k_2)$ we find $\frac{C_{n_0} C_{n_2}}{C_{n_0+n_2+1}} = O(k_2^{-3/2})$, and plugging all these results into (5.28) yields

$$\mathbb{P}[A(q)] = 2^{-2n} \sum_c O\left(\left(k_1^{-3/2} + e^{-\sqrt{\log n}/16}\right) \cdot \left(k_2^{-3/2} \ell_2^{-3/2}\right)\right),$$

which is $O\left(\left(k_1^{-3/2} + e^{-\sqrt{\log n}/16}\right) \cdot \left(k_2^{-3/2} \ell_2^{-3/2}\right)\right)$ since there are at most 2^{2n} valid colorings c. The finishes the case that one of k_1, k_2 is less than $d \log n$.

Secondly, consider the case that $2n - (k_1 + k_2) < d \log n$. By the above we may assume that $k_1, k_2 > d \log n$. For any $d \log n < k_1 < 2n - d \log n$ there are at most $O(\log n)$

values of k_2 for which $2n - (k_1 + k_2)$ is satisfied. Furthermore, given k_1 and k_2 there are at most $O((\log n)^2)$ ways to pick x_1 and x_2, as there are at most $d \log n$ dots outside of the arcs $(x_1, x_1 + k_1)$ and $(x_2, x_2 + k_2)$. A variant of the proof of Lemma 5.4.2 shows that with probability $O(n^{3/2} k_1^{-3/2} k_2^{-3/2})$ we have arcs connecting x_1 and $x_1 + k_1$, and x_2 and $x_2 + k_2$.

Given k_1, k_2, x_1 and x_2, and assuming that $(x_1, x_1 + k_1)$ and $(x_2, x_2 + k_2)$ match there are at most $k_1 \cdot k_2$ edges involving these two arcs. Therefore, the expected number of pairs of edges is at most

$$\sum_{k_1,k_2} O((\log n)^2) \cdot O(n^{3/2} k_1^{-3/2} k_2^{-3/2}) \cdot k_1 k_2 = O(n^{3/2} (\log n)^2) \sum_{k_1,k_2} k_1^{-1/2} k_2^{-1/2}.$$

We now claim that $k_2 \geq \frac{1}{2}(2n - k_1)$. Indeed, if $k_1 \geq 2n - 2d \log n$ we have $\frac{1}{2}(2n - k_1) \leq d \log n$, whereas $k_2 \geq d \log n$. Otherwise, we have $k_2 \geq 2n - k_1 - d \log n \geq \frac{1}{2}(2n - k_1)$ since the last inequality is equivalent to $k_1 \leq 2n - 2d \log n$. Using this, together with the earlier observation that there are at most $O(\log n)$ choices for k_2 given k_1, we find

$$\begin{aligned} O(n^{3/2} (\log n)^2) \sum_{k_1,k_2} k_1^{-1/2} k_2^{-1/2} &= O(n^{3/2} (\log n)^2) \sum_{k_1,k_2} k_1^{-1/2} (2n - k_1)^{-1/2} \\ &= O(n^{3/2} (\log n)^3) \sum_{k_1} k_1^{-1/2} (2n - k_1)^{-1/2}. \end{aligned}$$

Using that $x \mapsto (x(2n - x))^{-1/2}$ is decreasing on $(0, n)$ and increasing on $(n, 2n)$ we can compare the last sum with an integral to find that

$$\sum_{k_1} k_1^{-1/2}(2n-k_1)^{-1/2} \leq \int_1^{2n-1} (x(2n-x))^{-1/2} \, \mathrm{d}x$$
$$= 2\arctan\left(\sqrt{\frac{x}{2n-x}}\right)\Bigg|_1^{2n-1}$$
$$= 2\arctan(\sqrt{2n-1}) - 2\arctan\left(\sqrt{\frac{1}{2n-1}}\right) \leq \pi.$$

Therefore, the expected number of pairs of these edges is $O(n^{3/2}(\log n)^3) = o(n^2 \log^2 n)$.

□

The next lemma takes care of the cases where the arcs on at least one side are nested.

Proof of Lemma 5.6.5. There are three cases to consider, based on the relative position of the arcs coming from the bottom:

1. These arcs are unnested.
2. We have $y_2 < y_1 < y_1 + \ell_1 < y_2 + \ell_2$.
3. We have $y_1 < y_2 < y_2 + \ell_2 < y_1 + \ell_1$.

We will prove that in each case we have

$$\mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] = O\left(n^3 k_1^{-3/2}(2n-k_1)^{-3/2}\ell_1^{-3/2}(2n-\ell_1)^{-3/2}k_m^{-3/2}\ell_m^{-3/2}\right), \tag{5.29}$$

where $k_m = \min\{k_2 - k_1, k_2\}$ and ℓ_m is defined based on which of the three cases we are working in. Furthermore, in all cases we will show an upper bound of $O(g(k_1, \ell_1) \cdot k_1 \cdot$

$\min\{k_m, \ell_m\}$) on the number of choices for x_1, x_2, y_1, y_2 given k_1, k_2, ℓ_1, ℓ_2. Here $g(k, \ell)$ is the number of pairs (x, y) such that (x, k, y, ℓ) is a good quadruple, as defined in Proposition 5.4.6 . We note that given k_m and k_1 there are only two possibilities for k_2 and we will define ℓ_m in such a way that the same thing holds for ℓ_2 given ℓ_m and ℓ_1. Therefore, the desired contribution will be of the order

$$\sum_{k_1,\ell_1,k_m,\ell_m} g(k_1, \ell_1) \cdot k_1 \cdot \min\{k_m, \ell_m\} \cdot n^3 k_1^{-3/2}(2n - k_1)^{-3/2} \cdot \ell_1^{-3/2}(2n - \ell_1)^{-3/2} k_m^{-3/2} \ell_m^{-3/2}.$$

Simply allowing all the variables in this sum to run between $d \log n$ and $2n - d \log n$ we can factor this as

$$\left(\sum_{k_m,\ell_m} \min\{k_m, \ell_m\} k_m^{-3/2} \ell_m^{-3/2}\right) \cdot \left(\sum_{k_1,\ell_1} g(k_1, \ell - 1) \cdot k_1 \cdot n^3 k_1^{-3/2}(2n - k_1)^{-3/2} \ell_1^{-3/2}(2n - \ell_1)^{-3/2}\right).$$

Note that the first sum is of order $O(\log n)$. Now, if $\max\{k_1, \ell_1\} > cn$ we can use the estimate $k_1 = O(n)$, to show that the total contribution is given by

$$O(n \log n) \cdot \left(\sum_{k_1,\ell_1} g(k_1, \ell - 1) \cdot n^3 k_1^{-3/2}(2n - k_1)^{-3/2} \ell_1^{-3/2}(2n - \ell_1)^{-3/2}\right) = o(n^2 \log^2 n),$$

as the last sum is of order $o(n \log n)$ by Proposition 5.4.4. Else, we can use $n^3(2n - k_1)^{-3/2}(2n - \ell_1)^{-3/2} = O(1)$ and the estimate $g(k_1, \ell_1) \leq 4n \min\{k_1, \ell_1\} \leq 4n\ell_1$ to see that the total contribution is of the order

$$O(n \log n) \cdot \left(\sum_{k_1,\ell_1} k_1^{-1/2} \ell_1^{-1/2}\right) = O(n^2 \log n) = o(n^2 \log^2 n),$$

where we used that

$$\sum_{k_1,\ell_1} k_1^{-1/2}\ell_1^{-1/2} = \left(\sum_{d\log n\le k_1\le cn} k_1^{-1/2}\right)\cdot\left(\sum_{d\log n\le k_1\le cn} k_1^{-1/2}\right) = O(\sqrt{n})\cdot O(\sqrt{n}).$$

We now show (5.29) and the desired bounds on the number of quadruples for each of the cases. We handle the first case in full detail, the other two cases are very similar so we only highlight the details.

1. We know from Lemma 5.4.2 that $\mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})]$ equals

$$O\left(n^3(2n-k_1)^{-3/2}(k_1-k_2)^{-3/2}k_2^{-3/2}(2n-\ell_1-\ell_2)^{-3/2}\ell_1^{-3/2}\ell_2^{-3/2}\right).$$

   In this case, we define $\ell_m = \min\{2n-\ell_1-\ell_2, \ell_2\}$. Now, since $(k_1-k_2)+k_2 = k_1$ we have $\max\{k_1-k_2, k_2\} \ge k_1/2$, so $(k_1-k_2)^{-3/2}k_2^{-3/2} = O(k_1^{-3/2}k_m^{-3/2})$ and similarly we find $(2n-\ell_1-\ell_2)^{-3/2}\ell_2^{-3/2} = O((2n-\ell_1)^{-3/2}\ell_m^{-3/2})$.

   Furthermore, given k_1, k_2, ℓ_1, ℓ_2 there are at most $g(k_1,\ell_1) + O(n) = O(g(k_1,\ell_1))$ ways to pick (x_1, y_1), where we have to add $O(n)$ to account for the option that (x_1, k_1, y_1, ℓ_1) is not a good quadruple. Now suppose that (x_1, y_1) has been chosen.

   If $k_m < \ell_m$ there are at most (k_1-k_2) ways to pick x_2 and after that at most $2k_2$ ways to pick y_2, so there are at most $O((k_1-k_2)k_2) = O(k_1k_m)$ ways to pick (x_2, y_2) (where we used $k_1-k_2, k_2 \le k_1$).

   Similarly, if $\ell_m < k_m$ there are at most k_1 ways to pick x_2 and we claim that there are at most $O(\ell_m)$ ways to pick y_2. Indeed, if $\ell_m = 2n-\ell_1-\ell_2$ then there are at most two ways to pick the relative order of the arcs, after which y_1 is determined by how many of the $2n-\ell_1-\ell_2=\ell_m$ outside points are to the left of y_2, whereas if $\ell_m=\ell_2$ the value of y_2 is determined by the relative order of x_2 and y_2 and by how many points the arcs (x_2, x_2+k_2) and $(y_2, y_2+\ell_2)$ have in common. For the first option we have two choices and for the last one we have $\ell_2=\ell_m$ choices.

2. In this case we see that $\mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})]$ equals

$$O\left(n^3(2n-k_1)^{-3/2}(k_1-k_2)^{-3/2}k_2^{-3/2}(2n-\ell_2)^{-3/2}\ell_1^{-3/2}(\ell_2-\ell_1)^{-3/2}\right),$$

so defining $\ell_m=\min\{2n-\ell_2, \ell_2-\ell_1\}$ gives the desired bound on the probability.

For the count of the number of options for (x_1, y_1, x_2, y_2) the only thing that changes is the number of ways to pick (x_2, y_2) given (x_1, y_1) and given $\ell_m \leq k_m$. Again, there are at most k_1 ways to pick x_2. If $\ell_m=\ell_2-\ell_1$ then y_2 is determined by the number of dots between y_1 and y_2, whereas if $\ell_m=2n-\ell_2$ the value of y_2 is determined by choosing how many of the outside points should be to the left of y_2.

3. Here we can bound $\mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})]$ by

$$O\left(n^3(2n-k_1)^{-3/2}(k_1-k_2)^{-3/2}k_2^{-3/2}(2n-\ell_1)^{-3/2}\ell_2^{-3/2}(\ell_1-\ell_2)^{-3/2}\right),$$

so we define $\ell_m=\min\{\ell_1-\ell_2, \ell_2\}$.

Again, the only thing that remains is to bound the number of ways to pick y_2 given (x_1, y_1, x_2) in the case $\ell_m \leq k_m$. If $\ell_m=\ell_1-\ell_2$ then y_2 is determined by picking the distance between y_1 and y_2, whereas if $\ell_m=\ell_2$ the value of y_2 is determined by picking the relative order of x_2 and y_2 and choosing the number of points that the two arcs $(x_2, x_2 + k_2)$ and $(y_2, y_2 + \ell_2)$ have in common. $\square$

Next we handle the case where at least one of the arcs has size linear in n.

Proof of Lemma 5.6.6. We assume $k_1 = \max\{k_1, k_2, \ell_1, \ell_2\}$ without loss of generality. Let $k_0 = 2n - k_1 - k_2$ and $\ell_0 = 2n - \ell_1 - \ell_2$ and set $m_i = \min\{k_i, \ell_i\}$ for $i = 0, 1, 2$. First assume that $\ell_1 \neq \max\{\ell_0, \ell_1, \ell_2\}$.

We claim that given k_1, k_2, ℓ_1, ℓ_2, the number of quadruples is at most $O(m_0^2 m_1 m_2) = O(m_0^2 \ell_1 m_2)$. Since there are only finitely many options for the orderings of the endpoints of the arcs, it suffices to show the bounds for each specific ordering. But, given the ordering of the arcs, we claim that there are at most $m_0 m_i$ ways to pick (x_i, y_i). Indeed, consider the case that $k_0 = \min\{k_0, \ell_0\}$. Then we can pick x_i in at most k_0 ways, as it is determined by the number of points to the left of x_i (if the arc $(x_i, x_i + k_i)$ is the leftmost arc) or to the number of points to the right of $x_i + k_i$ (if the arc is the rightmost arc), so x_i can be picked in at most $k_0 = m_0$ ways. After that, y_i is determined by the number of points that the arcs $(x_i, x_i + k_i)$ and $(y_i, y_i + \ell_i)$ have in common and this is at most m_i. The case $\ell_0 = \min\{k_0, \ell_0\}$ is similar.

Now given a quadruple, by Lemma 5.4.2 the probability that that all the desired arcs match is $O(n^{3/2} k_0^{-3/2} \ell_0^{-3/2} \ell_1^{-3/2} k_2^{-3/2} \ell_2^{-3/2})$ where we used that $k_1 \geq cn$. Therefore, the desired contribution is at most

$$O(n^{3/2}) \cdot \sum m_0^2 \ell_0^{-3/2} k_0^{-3/2} \cdot \ell_1^{-1/2} \cdot m_2 k_2^{-3/2} \ell_2^{-3/2}. \tag{5.30}$$

Since $\ell_0 + \ell_1 + \ell_2 = 2n$ we have $\max\{\ell_0, \ell_1, \ell_2\} \geq 2n/3$. Since we assumed that ℓ_1 is not the maximum we have two cases.

- ℓ_0 is the maximum. In this case $n^{3/2}\ell_0^{-3/2} = O(1)$. As $\ell_0 + \ell_1 + \ell_2 = 2n$, ℓ_0 is determined by ℓ_1 and ℓ_2 and similary k_1 is determined by k_0 and k_2, so the contribution to (5.30) is

$$O(1) \cdot \sum_{k_0,k_2,\ell_1,\ell_2} m_0^2 k_0^{-3/2} \ell_1^{-1/2} m_2 k_2^{-3/2} \ell_2^{-3/2},$$

where the sum is over some appropriate range. To find an upper bound we can split this sum as

$$O(1) \cdot \left(\sum_{k_0} m_0^2 k_0^{-3/2} \right) \cdot \left(\sum_{\ell_1} \ell_1^{-1/2} \right) \cdot \left(\sum_{k_2,\ell_2} m_2 k_2^{-3/2} \ell_2^{-3/2} \right),$$

which after merging back involves more terms than before, but that is fine as we are only interested in an upper bound. We will now estimate each individual sum. For the first one, if $k_0 \leq \ell_0$ this contributes $\sum_{k_0} k_0^{1/2} = O(n^{3/2})$, whereas if $k_0 \geq \ell_0$ this sum is at most $O(n^2) \sum k_0^{-3/2} = O(n^2) \cdot O(n^{-1/2}) = O(n^{3/2})$ where we used that $k_0 \geq 2n/3$ in this case. For the second sum we get a bound of $O(n^{1/2})$. For the last sum we may assume $k_2 \leq \ell_2$ by symmetry and see that this sum is

$$O\left(\sum_{\ell_2} \ell_2^{-3/2} \sum_{k_2 \leq \ell_2} k_2^{-1/2} \right) = O\left(\sum_{\ell_2} \ell_2^{-1} \right) = O(\log n),$$

so altogether we get $O(n^2 \log n)$ in this case.

- Now assume that $\ell_2 = \max\{\ell_0, \ell_1, \ell_2\}$. Using the estimate $O(n^{3/2}) \cdot \ell_2^{-3/2} = O(1)$

the contribution to (5.30) is at most

$$O\left(\left(\sum_{k_0,\ell_0} m_0^2 k_0^{-3/2}\ell_0^{-3/2}\right)\cdot\left(\sum_{\ell_1}\ell_1^{-1/2}\right)\cdot\left(\sum_{k_2} m_2 k_2^{-3/2}\right)\right).$$

Similar arguments to above give that the first sum is $O(n)$, the second one is $O(n^{1/2})$ and the last one is $O(n^{1/2})$ where here one has to distinguish cases based on whether $k_2 \geq \ell_2$ or $k_2 \leq \ell_2$ just as for the first sum in the case above, so the total contribution will be $O(n^2) = o(n^2 \log^2 n)$, as desired.

It remains to handle the case $\ell_1 = \max\{\ell_0, \ell_1, \ell_2\}$. In this setting, we claim that (after being given an ordering of the endpoints of the arcs) we can choose x_1, x_2, y_1 and y_2 in $k_0 \cdot \ell_0 \cdot m_0 \cdot m_2$ ways. Indeed, we can still pick x_2, y_2 in $m_0 \cdot m_2$ ways, whereas we have at most k_0 ways to pick x_1 and ℓ_0 ways to pick y_1. In this case, we get a contribution of at most

$$O(n^{3/2}) \cdot \sum m_0 \ell_0^{-1/2} k_0^{-1/2} \cdot \ell_1^{-3/2} \cdot m_2 k_2^{-3/2} \ell_2^{-3/2}.$$

Using $O(n^{3/2}) \cdot \ell_1^{-3/2} = O(1)$ we have to evaluate

$$\left(\sum_{k_0,\ell_0} m_0 \ell_0^{-1/2} k_0^{-1/2}\right) \cdot \left(\sum_{k_2,\ell_2} m_2 k_2^{-3/2} \ell_2^{-3/2}\right),$$

where the second sum is $O(\log n)$ as before and by a similar argument we find that the first sum is $O(n^2)$, showing that this contribution is $O(n^2 \log n) = o(n^2 \log^2 n)$. □

Lastly, we handle all quadruples that are valid but not good.

Proof of Lemma 5.6.7. By Lemma 5.4.2 we know

$$\mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] = O(k_1^{-3/2} k_2^{-3/2} \ell_1^{-3/2} \ell_2^{-3/2}).$$

Also, we know by Proposition 5.4.6 that $\sum_{k_i,\ell_i} g(k_i,\ell_i)k_i^{-3/2}\ell_i^{-3/2} = O(n\log n)$ and by Proposition 5.4.4 that $\sum_{k_i,\ell_i} nk_i^{-3/2}\ell_i^{-3/2} = o(n\log n)$.

Our goal is to show given (k_1,k_2,ℓ_1,ℓ_2) there are at most $O(g(k_1,\ell_1)n+ng(k_2,\ell_2)+n^2)$ quadruples $q \in Q_4$, since then the desired contribution is at most

$$O\left(\sum_{k_1,\ell_1,k_2,\ell_2} (g(k_1,\ell_1)n + ng(k_2,\ell_2) + n^2)k_1^{-3/2}\ell_1^{-3/2}k_2^{-3/2}\ell_2^{-3/2}\right),$$

which is the sum of

$$O\left(\left(\sum_{k_1,\ell_1} g(k_1,\ell_1)k_1^{-3/2}\ell_1^{-3/2}\right)\cdot\left(\sum_{k_2,\ell_2} nk_2^{-3/2}\ell_2^{-3/2}\right)\right) = O(n\log n)\cdot o(n\log n)$$
$$O\left(\left(\sum_{k_1,\ell_1} nk_1^{-3/2}\ell_1^{-3/2}\right)\cdot\left(\sum_{k_2,\ell_2} g(k_2,\ell_2)k_2^{-3/2}\ell_2^{-3/2}\right)\right) = o(n\log n)\cdot O(n\log n)$$
$$O\left(\left(\sum_{k_1,\ell_1} nk_1^{-3/2}\ell_1^{-3/2}\right)\cdot\left(\sum_{k_2,\ell_2} nk_2^{-3/2}\ell_2^{-3/2}\right)\right) = o(n\log n)\cdot o(n\log n)$$

so the total contribution is $o(n^2\log^2 n)$ as well.

Now, given (k_1,k_2,ℓ_1,ℓ_2) there are only a few ways in which we can have a valid but not good quadruple.

- (x_1,k_1,y_1,ℓ_1) is good, but (x_2,k_2,y_2,ℓ_2) is not good. In this case we can pick (x_1,y_1) in at most $g(k_1,\ell_1)$ ways and (x_2,y_2) in $O(n)$ ways, so we are done.

- (x_2,k_2,y_2,ℓ_2) is good, but (x_1,k_1,y_1,ℓ_1) is not good. Similarly to the previous case this will give a bound of $O(ng(k_2,\ell_2))$.

- Neither of the (x_i,k_i,y_i,ℓ_i) are good. In this case we get a bound of $O(n^2)$ as there are $O(n)$ ways to pick any individual (x_i,k_i,y_i,ℓ_i).

- Both of the (x_i, k_i, y_i, ℓ_i) are good, but the endpoint of one arc of the first four-tuple is adjacent to the endpoint of an arc of the second four-tuple. Note that there are only finitely many possible orderings of the endpoints of the arcs. Given an ordering, there are now at most $g(k_1, \ell_1)$ ways to pick (x_1, y_1), which determines either x_2 or y_2 since one of $\{x_2, x_2 + k_2, y_2, y_2 + \ell_2\}$ is adjacent to a now known point, and after that there are at most $2n$ ways to pick the other of x_2, y_2, so there are $O(g(k_1, \ell_1) \cdot n)$ possible quadruples in this case. □

5.7 Induced subgraphs and connected components

In this section we prove results on the number of induced subgraphs of CP_n isomorphic to a given Catalan-pair graph H on at least 3 vertices, and we will use this to prove Theorem 5.1.4. At the end of the section we will also discuss a result about the connected components of CP_n.

5.7.1 A lower bound for the number of induced subgraphs

Recall that $N^*_H(G)$ denotes the number of induced subgraphs of G isomorphic to H, and that $A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ denotes the intersection of the following events.

1. The points x_i and $x_i + k_i$ are colored red and the points y_j and $y_j + \ell_j$ are colored blue for all i, j.

2. For all i and j the number of red points x with $x_i < x < x_i + k_i$ and the number of

blue points y with $y_j < y < y_j + \ell_j$ is even.

3. For all i and j we have that $(x_i, x_i + k_i)$ and $(y_j, y_j + \ell_j)$ match in CP_n.

The following lemma will be a key step to proving the general lower bound. Note that this lemma can be seen as a converse to Lemma 5.4.2.

Lemma 5.7.1. *There exists a positive real number* $\alpha_{s,t}$ *with*

$$\mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] \geq \alpha_{s,t} \prod_{i=1}^{s} k_i^{-3/2} \prod_{j=1}^{t} \ell_j^{-3/2} \tag{5.31}$$

for all good quadruples $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ *where* $\mathbf{x}$ *and* $\mathbf{y}$ *have length* s *and* t *respectively.*

Proof. We first show that with probability $2^{-3(s+t)}$ the first two conditions are satisfied. It is clear that with probability $1/2$ all of the points $x_i, x_i + k_i, y_j, y_j + \ell_j$ receive the correct color, so with probability $2^{-2(s+t)}$ all of these points have the correct color. Now conditioned on all of these points having the correct color, we show that with probability $2^{-(s+t)}$ the second condition is satisfied. Consider all the points of the form $x_i + 1$ and $y_j + 1$, and note that by assumption of $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ being a good quadruple all of these points are different and not equal to any of the $x_i, x_i + k_i, y_j$ and $y_j + \ell_j$. Consider the rightmost of these points, and suppose that it is equal to $x_i + 1$ for some i. Since all of the points to the right have been colored, we have that in particular all of the points x with $x_i < x < x_i + k_i$ except for this one have been colored. Therefore there is a unique choice for the color of $x_i + 1$ that makes the number of red points x with $x_i < x < x_i + k_i$ even. Inductively apply this argument for the remaining points, always taking the rightmost uncolored point.

Now suppose the first two conditions are satisfied. We apply Lemma 5.3.2 to determine a lower bound for the probability that the third condition is met. To this end, for each $1 \leq i \leq s$, let $2r_i$ be the number of red points x with $x_i < x < x_i + k_i$ that do not satisfy $x_j \leq x \leq x_j + k_j$ for any $j \neq i$. Let $2r_0$ be the number of red points that have not been counted for any of the r_i and that are not of the form x_i or $x_i + k_i$. Define $b_0, b_1, \ldots, b_t$ similarly. Let R_n and B_n denote the total number of red and blue points respectively. Note that for any $1 \leq i \leq s$ we have $2r_i \leq k_i$, hence in particular $r_i \leq k_i$. Applying the aforementioned lemma we find that

$$\begin{aligned}
\mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] &\geq 2^{-3(s+t)} \cdot \alpha_s \prod{}' \frac{R_n^{3/2}}{r_i^{3/2}} \cdot \alpha_t \prod{}' \frac{B_n^{3/2}}{b_j^{3/2}} \\
&\geq \alpha_{s,t} \prod_{i=0}^{s} \frac{R_n^{3/2}}{\max(r_i, 1)^{3/2}} \cdot \prod_{j=0}^{t} \frac{B_n^{3/2}}{\max(b_j, 1)^{3/2}} \\
&\geq \alpha_{s,t} \prod_{i=1}^{s} k_i^{-3/2} \prod_{j=1}^{t} \ell_j^{-3/2},
\end{aligned}$$

where we used that $R_n \geq \max(r_0, 1)$, $B_n \geq \max(b_0, 1)$, $\max(r_i, 1) \leq k_i$ and $\max(b_j, 1) \leq \ell_j$. □

We are now ready to prove the lower bound of Theorem 5.1.4. In fact, we will give a lower bound for any Catlan-pair graph regardless of whether it is connected or not.

Proposition 5.7.2. *Let H be a Catalan-pair graph on v vertices with i isolated vertices and m isolated edges. Then*

$$\mathbb{E}[N_H^*(CP_n)] = \Omega(n^{\frac{v+i}{2}} (\log n)^m). \tag{5.32}$$

Proof. We will prove this by first showing that the result holds for $m = i = 0$, then for $i = 0$, and finally for arbitrary m and i. We note that one can prove the most general case without first going through the other two cases, but this would decrease the readability of the proof.

First assume $m = i = 0$, and let q_H be any quadruple representing H. Our goal will be to find a large number of "blowups" of q_H. Let $c \geq 4v$ be a fixed constant, and let

$$P_j := \{1 + (j-1)\lfloor n/c \rfloor, 2 + (j-1)\lfloor n/c \rfloor, \ldots, -1 + j\lfloor n/c \rfloor\},$$

$$P := P_1 \times \cdots \times P_{2v}.$$

Given $p = (p_1, \ldots, p_{2v}) \in P$, we will define a quadruple $q_c(p)$ as follows. If in q_H we have $x_j = a$ and $x_j + k_j = b$, then in $q_c(p)$ we let $x_j = p_a$ and $x_j + k_j = p_b$, and we similarly define y_j and $y_j + \ell_j$ to correspond to the bottom jth arc of q_H. We note that the reason we force all the points of the left of $2v\lfloor n/c \rfloor \leq n/2$ is to make sure that in the general case we have enough space left to place or find arcs yielding the isolated edges and vertices.

We claim that $q_c(p)$ is a good quadruple that represents H for any $p \in P$. First observe that the points of $q_c(p)$ have the same relative order as the points of q_H, which shows that $q_c(p)$ satisfies the third condition for being a good quadruple (since q_H satisfies this condition), and moreover that $q_c(p)$ represents H. The first condition for being a good quadruple follows since the largest point we could choose for $q_c(p)$ is $-1 + 2v\lfloor n/c \rfloor \leq n/2$ since $c \geq 4v$, and the second condition follows since $|\max P_j - \min P_k| \geq 2$ for all j, k by the way we defined these sets. This proves our claim.

Now let $Q_H(c)$ denote the set of all $q_c(p)$ with $p \in P$. Observe that for large n we have

$$|Q_H(c)| = (\lfloor n/c \rfloor - 1)^{2v} \geq (2c)^{-2v} n^{2v}.$$

Also observe since $k_j, \ell_j \leq 2n$ for all j, Lemma 5.7.1 gives that $\mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] \geq \alpha_v n^{-3v/2}$ for all $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l}) \in Q_H(c)$, where $\alpha_v := 2^{-3v/2} \max_{s+t=v} \alpha_{s,t}$. In particular, we have that

$$\mathbb{E}[N_*(H)] \geq \sum_{(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l}) \in Q_H(4v)} \mathbb{P}[A(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})] \geq (8v)^{-2v} n^{2v} \cdot \alpha_v n^{-3v/2} = \Omega(n^{v/2}).$$

Now assume that $i = 0$ and let $c = 4m + 4v$. We will say that two vectors $\mathbf{k}, \mathbf{l}$ each of length m are *nice* if we have $4 \leq k_j \leq \ell_j \leq \lfloor n/c \rfloor$ for all j. Let $Q_c(\mathbf{k}, \mathbf{l})$ denote the set of all quadruples $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ such that

$$1 + (2j - 2 + 2v)\lfloor n/c \rfloor \leq x_j \leq -1 + (2j - 1 + 2v)\lfloor n/c \rfloor,$$

$$x_j + 2 \leq y_j \leq x_j + k_j - 2,$$

We claim that each quadruple of $Q_c(\mathbf{k}, \mathbf{l})$ is good whenever $\mathbf{k}, \mathbf{l}$ is nice. The first condition follows since the largest point we pick is $y_m + \ell_m \leq -1 + (2m + 2v)\lfloor n/c \rfloor \leq \frac{n}{2}$ since $c = 4m + 4v$. Similarly one can verify that

$$x_j \leq y_j - 2 \leq x_j + k_j - 4 \leq y_j + \ell_j - 6 \leq x_{j+1} - 8,$$

where the first two inequalities follow from $x_j + 2 \leq y_j \leq x_j + k_j - 2$, the third inequality from $\ell_j \geq k_j$ and $y_j \geq x_j + 2$, and the last inequality from $y_j + \ell_j \leq -1 + (2j + 2v)\lfloor n/c \rfloor \leq x_{j+1} - 2$. This shows that the second and third conditions of being a good quadruple are satisfied, proving the claim. We also note that, for n sufficiently large,

$$|Q_c(\mathbf{k},\mathbf{l})| = (\lfloor n/c \rfloor - 1)^m \prod_{j=1}^{m} (k_j - 3) \geq (8c)^{-m} n^m \prod_{j=1}^{m} k_j,$$

where we have used that $k_j - 3 \geq \frac{1}{4} k_j$ for all j.

Now let H' denote H after deleting its m isolated edges. For $\mathbf{k},\mathbf{l}$ nice, let $Q(\mathbf{k},\mathbf{l})$ be the set of all quadruples q which are obtained by taking the union of the arcs of some $q_1 \in Q_{H'}(c)$ and some $q_2 \in Q_c(\mathbf{k},\mathbf{l})$. We claim that every such q is good. Indeed, the first condition holds since it holds for both q_1 and q_2. The second condition holds since it holds restricted to any two points of q_1 or q_2, and because the largest point of q_1 is at most $-1+2v\lfloor n/c \rfloor$ while the smallest point of q_2 is at least $1+2v\lfloor n/c \rfloor$. This also implies that the third condition is satisfied since it is satisfied for both q_1 and q_2, so the claim is proven.

Observe that each quadruple $(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l}) \in Q(\mathbf{k},\mathbf{l})$ represents H and that

$$\mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})] \geq \alpha_v n^{-3(v-2m)/2} \prod_{j=1}^{m} k_j^{-3/2} \ell_j^{-3/2}$$

by Lemma 5.7.1. Also observe that our previous work shows that

$$|Q(\mathbf{k},\mathbf{l})| = |Q_{H'}(c)| \cdot |Q_c(\mathbf{k},\mathbf{l})| \geq \beta_c n^{2v-3e} \prod_{j=1}^{m} k_j$$

for some absolute constant β_c. We conclude that

$$\begin{aligned}
\mathbb{E}[N_*(H)] &\geq \sum_{\mathbf{k},\mathbf{l}\text{ nice}} \; \sum_{(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l}) \in Q(\mathbf{k},\mathbf{l})} \mathbb{P}[A(\mathbf{x},\mathbf{k},\mathbf{y},\mathbf{l})] \geq \sum_{\mathbf{k},\mathbf{l}\text{ nice}} \alpha_v \beta_c n^{v/2} \prod_{j=1}^{m} k_j^{-1/2} \ell_j^{-3/2} \\
&= \alpha_v \beta_c n^{v/2} \left(\sum_{4 \leq k \leq \ell \leq \lfloor n/c \rfloor} k^{-1/2} \ell^{-3/2} \right)^m = \Omega(n^{v/2} (\log n)^m),
\end{aligned}$$

where we use the fact that the above sum is of order $\Omega(\log n)$.

Now let H be an arbitrary Catalan-pair graph. Let H'' be H with its isolated vertices removed, and let $N'_*(H)$ be the number of induced copies of H'' in CP_n which have all of its points in the interval $[1, n/2]$. Note that implicitly our above argument shows that $\mathbb{E}[N'_*(H)] = \Omega(n^{(v-i)/2}(\log n)^m)$.

We claim that, deterministically, $N_*(H) \geq N'_*(H) \cdot \binom{n/4}{i}$. Indeed, observe that there are at most $n/2$ arcs which have an endpoint in the interval $[1, n/2]$, and hence there exists at least $n/2$ arcs with both endpoints not in this interval. Let A_R denote the set of these arcs that are colored red, and similarly define A_B. One of these sets must have size at least $n/4$, so let C be such that $|A_C| \geq n/4$.

We claim that any induced copy of H'' contained in $[1, n/2]$ together with i arcs of A_C is an induced copy of H. Indeed, by definition no arc in A_C can interlace with any arc of the H'', and none of the A_C arcs interlace with one another since they are all colored the same way. Thus the graph that these arcs induce will be H'' together with i isolated vertices, which is precisely H. We conclude that

$$N_*(H) \geq \binom{|A_C|}{i} \cdot N'_*(H'') \geq \binom{n/4}{i} N'_*(H'').$$

The result now follows by taking expectations of the above inequality and using that $\mathbb{E}[N'_*(H)] = \Omega(n^{(v-i)/2}(\log n)^m)$. $\square$

5.7.2 An upper bound for the number of induced subgraphs

A key step in finding the expected number of edges was to bound the number of good quadruples (x, k, y, ℓ) for given k and ℓ. Therefore, for general H we would like to bound the number of valid quadruples $(\mathbf{x}, \mathbf{y}, \mathbf{k}, \mathbf{l})$ for given $\mathbf{k}$ and $\mathbf{l}$. One of the reasons this is more complicated in the general setting is that H might have several different representatives. However, since there are only finitely many representatives, it suffices to prove the desired bounds for each of them separately.

In order to do this we introduce some new notation. Let H be a Catalan-pair graph on v vertices and let $q = (\bar{\mathbf{x}}, \bar{\mathbf{k}}, \bar{\mathbf{y}}, \bar{\mathbf{l}})$ be a quadruple with $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ increasing such that the following conditions are satisfied.

- The lengths of $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ add to v.
- We have $\{\bar{x_i}\} \cup \{\bar{x_i} + \bar{k_i}\} \cup \{\bar{y_j}\} \cup \{\bar{y_j} + \bar{\ell_j}\} = \{1, 2, \ldots, 2v\}$.
- The quadruple q is valid and the resulting Catalan-pair graph is isomorphic to H.

We say that a valid quadruple $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ *represents H by q* if the relative order of the x_i, $x_i + k_i$, y_j and $y_j + \ell_j$ coincides with the relative order of $\bar{x_i}$, $\bar{x_i} + \bar{k_i}$, $\bar{y_j}$ and $\bar{y_j} + \bar{\ell_j}$. Note that the f_i and g_j as defined in the beginning of Section 5.4 depend solely on $\mathbf{k}$, $\mathbf{l}$, and q, and are independent of the exact values of $\mathbf{x}$ and $\mathbf{y}$.

We wish to prove a lemma that upper bounds the number of valid quadruples for given $\mathbf{k}$, $\mathbf{l}$, and representing quadruple q. From now on we assume that H is a connected Catalan-pair graph on $v \geq 3$ vertices that has s and t vertices in its bipartite components respectively. Additionally, let q be a quadruple as above where $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ have length s and t respectively.

When $\mathbf{k}$ and $\mathbf{l}$ are known we denote by (x_i) the arc $(x_i, x_i + k_i)$. For a valid quadruple $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ we say that (x_i) is a *maximal arc* if there is no j with $x_j < x_i < x_i + k_i < x_j + k_j$. We say that arc (x_i) covers arc (x_j) if we have $x_i < x_j < x_j + k_j < x_i + k_i$ and there is no i' with $x_i < x_{i'} < x_j < x_j + k_j < x_{i'} + k_{i'} < x_i + k_i$. Note that each arc is either maximal, or has a unique arc that covers it. However, a single arc can cover multiple arcs.

Lemma 5.7.3. *Let* $\mathbf{k}$ *and* $\mathbf{l}$ *be* s *and* t*-tuples of positive integers for which there exists a valid quadruple* $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ *representing* H *by* q*. The number of such quadruples is at most*

$$(\min\{f_0, g_0\} + 2v + 1) \cdot \prod_{\substack{i \geq 1 \\ i \neq i_0}} (f_i + 2v + 1) \cdot \prod_{j \geq 1} (g_j + 2v + 1) \tag{5.33}$$

for any $i_0 \neq 0$*, and it also at most*

$$(f_0 + 2v + 1)(g_0 + 2v + 1) \cdot \prod_{\substack{i \geq 1 \\ i \neq i_0}} (f_i + 2v + 1) \cdot \prod_{\substack{j \geq 1 \\ j \neq j_0}} (g_j + 2v + 1) \tag{5.34}$$

for any $i_0, j_0 \neq 0$.

Proof. In order to prove the first bound we first consider the case that $f_0 = \min\{f_0, g_0\}$. Let $i_0, i_1, \ldots, i_d$ be such that (x_{i_d}) is maximal and such that (x_{i_p}) covers $(x_{i_{p-1}})$ for all $1 \leq p \leq d$. We claim that there are at most

$$(f_0 + 2v + 1) \cdot \prod_{p=1}^{d} (f_{i_p} + 2v + 1)$$

ways to choose x_{i_d}, $x_{i_{d-1}}$, …, x_{i_1}, x_{i_0}. Indeed, since we specified q, $\mathbf{k}$, and $\mathbf{l}$ (and hence the f_i and g_j), we know how many points $m < x_{i_d}$ are of the form $m = y_j$, $m = y_j + \ell_j$, or which satisfy $x_{i'} \leq m \leq x_{i'} + k_{i'}$ for some i'. By definition of f_0, we know that there are at most $f_0 \leq f_0 + 2v$ points outside of arc x_{i_d} that are not of this form. We can choose amongst these at most $f_0 + 2v$ points how many lie to the left of x_{i_d}, and such a choice uniquely determines x_{i_d} (since we now know the total number of points which lie to the left of x_{i_d}). We conclude that we can place x_{i_d} in at most $f_0 + 2v + 1$ ways. A similar argument shows that there are at most $f_{i_j} + 2v + 1$ ways to place each $x_{i_j - 1}$ given that x_{i_j} has already been placed, where now f_{i_j} plays the role of f_0 by restricting our attention to points of the form $x_{i_j} < m < x_{i_j} + k_{i_j}$. This completes the proof of the claim.

Now suppose that we have inductively placed some (proper) subset of the arcs. Let Z denote the set of arcs z which have not been placed and whose endpoints alternate with some arc that has already been placed. Since H is connected, $Z \neq \emptyset$. Since Z is finite, let $z \in Z$ be such that z covers no other $z' \in Z$. Without loss of generality, assume that z is of the form (y_j). Then, we are in one of the following situations.

1. The arc (y_j) is minimal.

2. The arc (y_j) is not minimal and all the arcs covered by (y_j) have been placed already.

3. The arc (y_j) is not minimal, at least one arc covered by (y_j) has not been placed and any such arc does not alternate endpoints with any of the arcs placed so far.

We claim that in all cases there are at most $g_j + 2v + 1$ ways to choose y_j.

1. Note that in this case there are at most $g_j + 2v$ points between y_j and $y_j + \ell_j$. Indeed, there are g_j points that are not of the form x_i or $x_i + k_i$ and there are at most $2v$ points that are of this form. By assumption, the endpoints of (y_j) alternate with the endpoints of some (x_i). Consider the case where $y_j < x_i < y_j + \ell_j$. Then the number of points between y_j and x_i is at most $g_j + 2v$, else there would be too many points between y_j and $y_j + \ell_j$. Note that this number of intermediate points uniquely determines y_j since x_i is known. Therefore we have at most $g_j + 2v + 1$ ways to choose y_j.

2. In this case we can follow a similar argument as used when choosing x_{i_d}. Note that since the y_j are increasing, y_{j+1} is the leftmost arc that is covered by (y_j). By definition of g_j, there are at most $g_j + 2v$ points between y_j and y_{j+1} and the value of y_j is known, so we again have at most $g_j + 2v + 1$ ways to choose y_j.

3. In this case, suppose that (y_j) intersects (x_i) and that we have $y_j < x_i < y_j + \ell_j$. We again count the possible number of points between y_j and x_i. As before, there are between 0 and $g_j + 2v$ such points that do not lie below an arc covered by (y_j). We claim that we know how many of the other points lie between y_j and x_i, which again yields that there are at most $g_j + 2v + 1$ options for y_j.

   Indeed, consider an arc $(y_{j'})$ that is covered by (y_j). If $(y_{j'})$ has not been placed, then it does not alternate endpoints with (x_i) by assumption. Thus this arc either lies completely between y_j and x_i or completely between x_i and $y_j + \ell_j$, and since we specified the quadruple q representing H, we know which of these two cases happens. Thus we know exactly how many such points lie between y_j and x_i. Now if $(y_{j'})$ has been placed, we know all of $y_{j'}$, $y_{j'} + \ell_{j'}$ and x_i, so clearly we also know how many of the points between $y_{j'}$ and $y_{j'} + \ell_{j'}$ lie to the left of x_i.

Inductively, we can place the arcs one by one (in the order described above) and note that in this process we get the product of all of the numbers of the form $f_i + 2v + 1$ and $g_j + 2v + 1$ except for the numbers $f_{i_0} + 2v + 1$ and $g_0 + 2m + 1$, establishing the first bound when $f_0 = \min\{f_0, g_0\}$.

Now assume that $g_0 = \min\{f_0, g_0\}$. Since H is connected, there exists some $j_0 \neq 0$ such that (y_{j_0}) and (x_{i_0}) interlace, and moreover we can choose j_0 such that it does not cover any $(y_{j'})$ that also interlaces with (x_{i_0}). Let $j_0, j_1, \ldots, j_e$ be such that (y_{j_e}) is maximal and such that (y_{j_p}) covers $(x_{j_{p-1}})$ for all $1 \leq p \leq e$. By the same reasoning as above, there are at most $(g_0 + 2v + 1) \cdot \prod_{p=1}^{e}(g_{i_p} + 2v + 1)$ ways to choose y_{j_e}, $y_{j_{e-1}}$, …, y_{j_1}, y_{j_0}. We now place the remaining arcs Z as we did before. We use almost all of the same bounds as before, except we now use the bound $g_{j_0} + 2v + 1$ instead of $f_{i_0} + 2v + 1$ when we place (x_{i_0}). We are justified in using this bound since, by assumption of (y_{j_0}) not covering any arc that interlaces with (x_{i_0}), one of the endpoints of (x_{i_0}) must be one of the points counted by g_{j_0}. Ultimately this gives us the product of all of the numbers of the form $f_i + 2v + 1$ and $g_j + 2v + 1$ except for the numbers $f_{i_0} + 2v + 1$ and $f_0 + 2m + 1$ as desired.

To prove the final bound, let $i_0, i_1, \ldots, i_d$ be such that (x_{i_d}) is maximal and such that (x_{i_p}) covers $(x_{i_{p-1}})$ for all $1 \leq p \leq d$, and similarly define $j_0, j_1, \ldots, j_e$. By reasoning similar to that above, the number of ways we can place all of these arcs down in at most

$$(f_0 + 2v + 1)(g_0 + 2v + 1) \cdot \prod_{p=1}^{d} (f_{i_p} + 2v + 1) \cdot \prod_{p=1}^{e} (g_{i_p} + 2v + 1).$$

We then place the remaining arcs and use the same bounds as we did before, and this ultimately gives us a product of all of the terms except for $f_{i_0} + 2v + 1$ and $g_{j_0} + 2v + 1$. $\square$

Proposition 5.7.4. *Let* $\mathbf{k}$ *and* $\mathbf{l}$ *be* s *and* t*-tuples of positive integers for which there exists a valid quadruple* $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ *representing* H *by* q*. Then the number of such quadruples is at most*

$$(h_1 + 2v + 1) \cdot (h_2 + 2v + 1) \cdot (h_4 + 2v + 1) \cdot \prod_{i=6}^{v+2} (h_i + 2v + 1), \tag{5.35}$$

where $h_1 \leq h_2 \leq \ldots \leq h_{v+1} \leq h_{v+2}$ *are* $f_0, f_1, \ldots, f_s$ *and* $g_0, g_1, \ldots, g_t$ *written in increasing order.*

Proof. Without loss of generality we may assume that $f_{i_0} = \max_{i,j \neq 0}\{f_i, g_j\}$. Observe that $f_{i_0} \geq h_3$ since we assume $v \geq 3$, and further that $f_{i_0} \geq h_5$ if $\max\{f_0, g_0\} \leq h_4$. First assume that $\{f_0, g_0\} \neq \{h_1, h_2\}$. In this case we apply the first bound of Lemma 5.7.3 with our choice of i_0. This bound consists of the product of all the values $h_i + 2v + 1$ except for the terms $f_{i_0} + 2v + 1$ and $\max\{f_0, g_0\} + 2v + 1$, and in this case we say that our bound "omits" the values $f_{i_0} + 2v + 1$ and $\max\{f_0, g_0\} + 2v + 1$. If $\max\{f_0, g_0\} \geq h_5$ then these two terms are at least $h_3 + 2v + 1$ and $h_5 + 2v + 1$. If $\max\{f_0, g_0\} \leq h_4$, then we again omit at least $h_3 + 2v + 1$ and $h_5 + 2v + 1$ since $\{f_0, g_0\} \neq \{h_1, h_2\}$ implies that $\max\{f_0, g_0\} \geq h_3$. Thus in this case we achieve our desired result.

Now assume that $\{f_0, g_0\} = \{h_1, h_2\}$. In this case we apply the second bound of Lemma 5.7.3 to i_0 and $j_0 = 1$. Now we omit only $f_{i_0} + 2v + 1$ (which is at least $h_5 + 2v + 1$) and $g_1 + 2v + 1$ (which is at least $h_3 + 2v + 1$). We conclude the result. □

With this proposition we can prove an upper bound on the expected number of induced subgraphs.

Proposition 5.7.5. *Let H be a connected Catalan-pair graph on $v \geq 3$ vertices. Then*

$$\mathbb{E}[N_H^*(CP_n)] = O(n^{v/2}). \tag{5.36}$$

Proof. First notice that there are only finitely many valid quadruples $q = (\bar{\mathbf{x}}, \bar{\mathbf{k}}, \bar{\mathbf{y}}, \bar{\mathbf{l}})$ for which $\{\bar{x}_i\} \cup \{\bar{x}_i + \bar{k}_i\} \cup \{\bar{y}_j\} \cup \{\bar{y}_j + \bar{\ell}_j\} = \{1, 2, \ldots, 2v\}$ and such that the resulting Catalan-pair graph is isomorphic to H. Therefore, it suffices to show for each such q that the expected number of induced Catalan-pair graphs of CP_n that is represented by q is $O(n^{v/2})$.

Consider $1 \leq h_1 \leq h_2 \leq \ldots \leq h_{v+1} \leq h_{v+2} \leq 2n$. We claim that the number of pairs $(\mathbf{k}, \mathbf{l})$ such that there exist a valid quadruple $(\mathbf{x}, \mathbf{k}, \mathbf{y}, \mathbf{l})$ representing H by q and for which $\{h_i\} = \{f_i\} \cup \{g_j\}$ is at most $(v+2)!$. Indeed, note that since q defines the relative order of all the points, knowing the values of f_i and g_j uniquely determines $\mathbf{k}$ and $\mathbf{l}$. Since there are $(v+2)!$ ways to distribute the h_i over the f_i and g_j, there are at most $(v+2)!$ possible pairs $(\mathbf{k}, \mathbf{l})$.

Therefore, using Lemma 5.4.2 and Proposition 5.7.4 we find that the expected number of induced subgraphs isomorphic to H and represented by q is at most

$$(v+2)! \cdot \sum_h \Bigg((h_1+2v+1)\cdot(h_2+2v+1)\cdot(h_4+2v+1) \cdot \prod_{i=6}^{v+2}(h_i+2v+1)\cdot \beta_{s,t} n^3 \cdot \widetilde{\prod_i} h_i^{-3/2} \Bigg)$$

where the sum is over all possible sequences $h = (h_1, h_2, \ldots, h_{v+1}, h_{v+2})$ and $\widetilde{\prod}$ indicates the product over all i with $h_i \geq 16v \log n$. Note that implicitly this sum is over all possible $(\mathbf{k}, \mathbf{l})$, and we will break up this sum into the cases where $\{\max f_i, \max g_j\} = \{h_a, h_{v+2}\}$ for all possible a. We will show the desired upper bound of $O(n^{v/2})$ in each of these cases. Note that $\sum f_i = 2n - 2v$, so $\max f_i$ is at least linear and is uniquely determined by the other f_i. We first consider $v \geq 5$.

First, assume that $a \geq 6$. In this case, we can take out the factors $(h_a + 2v + 1) \cdot (h_{v+2} + 2v + 1) \cdot h_a^{-3/2} \cdot h_{v+2}^{-3/2}$ and note that this is $O(n^{-1})$, by virtue of h_a, h_{v+2} being linear in n. Therefore, the remaining part can (up to some large constant) be estimated by

$$n^2 \cdot \sum_{h_{v+1}=1}^{2n} \cdots \sum_{h_{a+1}=1}^{h_{a+2}} \sum_{h_{a-1}=1}^{h_{a+1}} \cdots \sum_{h_1=1}^{h_2} (h_1+2v+1)\cdot(h_2+2v+1) \cdot (h_4+2v+1)\cdot \prod_{\substack{i=6\\ i\neq a}}^{v+1}(h_i+2v+1)\cdot \widetilde{\prod_i} h_i^{-3/2}, \tag{5.37}$$

where the last product no longer involves h_a nor h_{v+2}. Note that this expression is actually independent of a, so for simplicity we assume that $a = v+1$. Let b be the number of h_i for which $h_i \leq 16v \log n$. First consider the case where $b = 0$. In this case, (5.37) is of the order

$$n^2 \cdot \sum_{h_v=1}^{2n} h_v^{-1/2} \sum_{h_{v-1}=1}^{h_v} h_{v-1}^{-1/2} \cdots \sum_{h_6=1}^{h_7} h_6^{-3/2} \sum_{h_5=1}^{h_6} h_5^{-3/2} \sum_{h_4=1}^{h_5} h_4^{-1/2} \sum_{h_3=1}^{h_4} h_3^{-3/2} \sum_{h_2=1}^{h_3} h_2^{-1/2} \sum_{h_1=1}^{h_2} h_1^{-1/2}.$$

Once again estimating these sums by integrals we find that

$$\begin{aligned}
&\sum_{h_5=1}^{h_6} h_5^{-3/2} \sum_{h_4=1}^{h_5} h_4^{-1/2} \sum_{h_3=1}^{h_4} h_3^{-3/2} \sum_{h_2=1}^{h_3} h_2^{-1/2} \sum_{h_1=1}^{h_2} h_1^{-1/2} \\
&\quad = O\left(\sum_{h_5=1}^{h_6} h_5^{-3/2} \sum_{h_4=1}^{h_5} h_4^{-1/2} \sum_{h_3=1}^{h_4} h_3^{-3/2} \sum_{h_2=1}^{h_3} 1 \right) \\
&\quad = O\left(\sum_{h_5=1}^{h_6} h_5^{-3/2} \sum_{h_4=1}^{h_5} h_4^{-1/2} \sum_{h_3=1}^{h_4} h_3^{-1/2} \right) \\
&\quad = O\left(\sum_{h_5=1}^{h_6} h_5^{-3/2} \sum_{h_4=1}^{h_5} 1 \right) \\
&\quad = O\left(\sum_{h_5=1}^{h_6} h_5^{-1/2} \right) = O(h_6^{1/2}) = O(n^{1/2}).
\end{aligned}$$

Furthermore, each of the remaining sums is at most $\sum_{x=1}^{2n} x^{-1/2} = O(n^{1/2})$, so the total sum is $O(n^2 \cdot (n^{1/2})^{v-5} \cdot n^{1/2}) = O(n^{v/2})$.

All of the cases $b = 0$ and $2 \leq a \leq 5$ have essentially the same proof as one another, so we will only explicitly go through one of these cases, namely $a = 3$. In this case we take out the factors $(h_{v+2} + 2v + 1)h_3^{-3/2} h_{v+2}^{-3/2} = O(n^{-2})$ from (5.37), and we use the fact that $h_i \geq h_3$ is linear for all $i \geq 3$ to conclude (5.37) is of the order of magnitude at most

$$\begin{aligned}
&n \cdot \sum_{h_{v+1}=1}^{2n} n^{-1/2} \cdots \sum_{h_6=1}^{2n} n^{-1/2} \sum_{h_5=1}^{2n} n^{-3/2} \sum_{h_4=1}^{2n} n^{-1/2} \sum_{h_2=1}^{2n} h_2^{-1/2} \sum_{h_1=1}^{h_2} h_1^{-1/2} \\
&\quad = O\left(n^{v/2-1} \sum_{h_2=1}^{2n} h_2^{-1/2} \sum_{h_1=1}^{h_2} h_1^{-1/2} \right) = O(n^{v/2}).
\end{aligned}$$

Now consider the case that $a > b \geq 5$, and again we can assume for simplicity that $a = v + 1$. Then (5.37) is at most of the order of

$$n^2 \cdot \sum_{h_v=1}^{2n} h_v^{-1/2} \sum_{h_{v-1}=1}^{h_v} h_{v-1}^{-1/2} \cdots \sum_{h_{b+1}=1}^{h_{b+2}} h_{b+1}^{-1/2} \cdot \sum_{h_b=1}^{16v\log n} (h_b + 2v + 1) \cdots$$
$$\cdots \sum_{h_6=1}^{16v\log n} (h_6 + 2v + 1) \sum_{h_5=1}^{16v\log n} \sum_{h_4=1}^{16v\log n} (h_4 + 2v + 1)$$
$$\sum_{h_3=1}^{16v\log n} \sum_{h_2=1}^{16m\log n} (h_2 + 2v + 1) \sum_{h_1=1}^{16v\log n} (h_1 + 2v + 1).$$

Note that each of the rightmost b sums will contribute at most $O((\log n)^2)$ each, and the remaining sums will contribute $O(n^{(v-b)/2})$ by an argument similar to the one above. Thus the total contribution will be of the order $O(n^2 \cdot n^{(v-b)/2} \cdot (\log n)^{2b}) = o(n^{v/2})$.

Similar arguments give a bound of $o(n^{v/2})$ when $b \in \{1, 2, 3, 4\}$ and for any $a > b$. Note that since h_a is linear in n, we always have $b < a$ for n large enough, so these finitely many cases are all that need to be checked for $v = 5$. The proofs for $v = 3, 4$ are essentially the same, and we note that we did not deal with these cases earlier because we could not write, for example, h_6. We omit the details. □

We note that the above proof shows the somewhat stronger result that the only quadruples that contribute to the order of magnitude of $n^{v/2}$ are those which have all of their gap sizes at least $16v \log n$. With this we can now prove Theorem 5.1.4.

Proof of Theorem 5.1.4. The result for induced subgraphs follows from Proposition 5.7.2 and 5.7.5. For any H we claim that

$$N_H^*(CP_n) \leq N_H(CP_n) \leq v! \cdot \sum_{H'} N_{H'}^*(CP_n),$$

where the sum is over all Catalan-pair graphs H' on v vertices that contain H as a subgraph. The lower bound is obvious. For the upper bound, note that for any given subgraph of CP_n isomorphic to H, the induced subgraph on these vertices is isomorphic to some H' appearing in this sum, and for given H' there are at most $v!$ subgraphs of H' isomorphic to H. Taking the expectation of both sides of this inequality and using the result for induced subgraphs gives the desired conclusion. □

5.7.3 The sizes of the connected components

Computational evidence suggest that a typical random Catalan-pair graph on n vertices will have one large component with roughly $n/2$ vertices and many smaller components. As we proved in Section 5.5, many of these components will be isolated vertices, but a significant amount will have larger size. In fact, we show that for *any* fixed Catalan-pair graph the number of connected components of CP_n isomorphic to this graph is linear in n.

Proposition 5.7.6. *Let H be a connected Catalan-pair graph on v vertices and let $n \geq v+2$. There exists a constant C, independent of H, such that the expected number of connected components of CP_n isomorphic to H is at least $C \cdot (n-v+1/2) \cdot 16^{-v}$.*

Proof. Let a and A be as in (5.6) and take $C = \left(\frac{a}{A}\right)^2$. Assume that H has bipartite components of sizes s and t. We show that for any $1 \leq x \leq 2n-2v+1$, we have probability at least $1/2 \cdot (a/A)^2 \cdot 16^{-v}$ that there are v arcs connecting $\{x, x+1, \ldots, x+2v-1\}$ and that the resulting Catalan-pair graph on these $2v$ points is isomorphic to H, which in particular yields a connected component of CP_n isomorphic to H.

Consider a fixed representative for H. With probability $(1/2)^{2v}$ the points $x, x+1, \ldots, x+2v-1$ are colored in the exact same order as the points in the representative. Furthermore, since there are at least four other points, with probability at least $1/2$ the other points do not all have the same color. Therefore, we have $r > s$ and $b > t$ red and blue points in total. Given r and s, the probability that we the arcs on the points $x, x+1, \ldots, x+2v-1$ exactly match those in the representative for H is given by

$$\frac{1}{2} \cdot \frac{C_{r-s}}{C_r} \cdot \frac{C_{b-t}}{C_b} \geq \frac{1}{2} \cdot \frac{a \cdot r^{3/2}}{4^s \cdot A \cdot (r-s)^{3/2}} \cdot \frac{a \cdot b^{3/2}}{4^t \cdot A \cdot (b-t)^{3/2}} \geq \frac{1}{2} \cdot \left(\frac{a}{A}\right)^2 \cdot \frac{1}{4^{s+t}}.$$

Since $s+t=v$ this implies that with probability at least $4^{-v} \cdot \frac{1}{2} \cdot (a/A)^2 \cdot 4^{-v}$ we get such a connected component isomorphic to H starting at point x. By linearity of expectation, the expected number of connected components isomorphic to H is at least

$$(2n-2v+1) \cdot 4^{-v} \cdot \frac{1}{2} \cdot \left(\frac{a}{A}\right)^2 \cdot 4^{-v} = (n-v+1/2) \cdot \left(\frac{a}{A}\right)^2 \cdot 16^{-v}. \qquad \square$$

In particular, we expect a typical Catalan-pair graph on n vertices to have connected components of size at least logarithmic in n.

This chapter contains material from: D. Kroes and S. Spiro, *"Random Graphs Induced by Catalan Pairs"*, Journal of Combinatorics, Accepted (2020). The dissertation author was one of the primary investigators and authors of this paper.

Bibliography

[AS08] Noga Alon and Joel H. Spencer, *The probabilistic method*, 3 ed., John Wiley & Sons, 2008.

[BDD+] Steve Butler, Erik D. Demaine, Martin L. Demaine, Ron Graham, Adam Hesterberg, Jason Ku, Jayson Lynch, and Tadashi Tokieda, *Paperclip graphs*, submitted.

[BHM+21] Jonah Blasiak, Mark Haiman, Jennifer Morse, Anna Pun, and George H. Seelinger, *A proof of the extended delta conjecture*, https://arxiv.org/abs/2102.08815.

[BO18] Benjamin Braun and McCabe Olsen, *Euler-mahonian statistics and descent bases for semigroup algebras.*, European Journal of Combinatorics **69** (2018), 237–254.

[Bol84] Béla Bollobás, *The evolution of random graphs*, Transactions of the American Mathematical Society **286** (1984), no. 1, 257–274.

[Che55] Claude Chevalley, *Invariants of finite groups generated by reflections*, American Journal of Mathematics **77** (1955), no. 4, 778–782.

[CL20] Federico Castillo and Fu Liu, *Deformation cones of nested braid fans*, International Mathematics Research Notices **rnaa090** (2020), https://doi.org/10.1093/imrn/rnaa090.

[CLO15] David A Cox, John Little, and Donal O'Shea, *Ideals, varieties, and algorithms*, 4 ed., Springer International Publishing, 2015.

[CR20] Jonathan Chan and Brendon Rhoades, *Generalized coinvariant algebras for wreath products*, Advances in Applied Mathematics **120** (2020), Article 102060.

[DFH$^+$99] Luc Devroye, Philippe Flajolet, Ferran Hurtado, Marc Noy, and William Steiger, *Properties of random triangulations and trees*, Discrete and Computational Geometry **1** (1999), no. 22, 105–117.

[DM20] Michele D'Adderio and Anton Mellit, *A proof of the compositional delta conjecture*, https://arxiv.org/abs/2011.11467.

[ER60] Paul Erdős and Alfréd Rényi, *On the evolution of random graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960), 17–61.

[FS09] Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge University Press, 2009.

[Gil59] Edgar N. Gilbert, *Random graphs*, The Annals of Mathematical Statistics **30** (1959), no. 4, 1141–1144.

[GS84] Adriano M. Garsia and Dennis Stanton, *Group actions on stanley-reisner rings and invariants of permutation groups*, Advances in Mathematics **51** (1984), no. 2, 107–201.

[HR18] Jia Huang and Brendon Rhoades, *Ordered set partitions and the* 0*-hecke algebra.*, Algebraic Combinatorics **1** (2018), no. 1, 47–80.

[HRS18] James Haglund, Brendon Rhoades, and Mark Shimozono, *Ordered set partitions, generalized coinvariant algebras, and the delta conjecture*, Advances in Mathematics **329** (2018), 851–915.

[HRW18] James Haglund, Jeffrey B. Remmel, and Andrew T. Wilson, *The delta conjecture*, Transactions of the American Mathematical Society **370** (2018), 4029–4057.

[Mac15] Percy A. MacMahon, *Combinatory analysis, volume 1*, Cambridge University Press, 1915.

[MR15] Anthony Mendes and Jeffrey Remmel, *Counting with symmetric functions*, Springer International Publishing, 2015.

[Net01] Eugen Netto, *Lehrbuch der combinatorik*, Chelsea, 1901.

[NG10] Nicholas Nash and David Gregg, *An output sensitive algorithm for computing a maximum independent set of a circle graph*, Information Processing Letters **110** (2010), no. 16, 630–634.

[NT06] Jean-Christophe Novelli and Jean-Yves Thibon, *Polynomial realizations of some trialgebras*, Proceedings of Formal Power Series and Algebraic Combinatorics (FPSAC) (2006), 243–255.

[Pak14] Igor Pak, *History of catalan numbers*, https://arxiv.org/abs/1408.5711.

[Rho18] Brendon Rhoades, *Ordered set partition statistics and the delta conjecture*, Journal of Combinatorial Theory, Series A **154** (2018), 172–217.

[RW15] Jeffrey B. Remmel and Andrew T. Wilson, *An extension of macmahon's equidistribution theorem to ordered set partitions*, Journal of Combinatorial Theory, Series A **134** (2015), 242–277.

[RW19] Brendon Rhoades and Andrew T. Wilson, *Line configurations and r-stirling partitions*, Journal of Combinatorics **10** (2019), no. 3, 411–431.

[RW20] ———, *Vandermondes in superspace*, Transactions of the American Mathematical Society **373** (2020), no. 4483–4516.

[Sag01] Bruce E. Sagan, *The symmetric group*, 2 ed., Springer-Verlag, 2001.

[Spi94] Jeremy P. Spinrad, *Recognition of circle graphs*, Journal of Algorithms **16** (1994), no. 2, 264–282.

[Sta15] Richard Stanley, *Catalan numbers*, Cambridge University Press, 2015.

[Ste19] Steingrímsson, *Statistics on ordered partitions of sets*, https://arxiv.org/abs/math/0605670.

[Tis15] Alexander Tiskin, *Fast distance multiplication of unit-monge matrices*, Algorithmica **71** (2015), no. 4, 859–888.

[Wil16] Andrew T. Wilson, *An extension of macmahon's equidistribution theorem to ordered multiset partitions*, The Electronic Journal of Combinatorics **23** (2016), no. 1.5.

[Zab19] Mike Zabrocki, *A module for the delta conjecture*, https://arxiv.org/abs/1902.08966.